



**HAL**  
open science

# Problèmes inverses et analyse en ondelettes adaptées

Thanh Mai Pham Ngoc

► **To cite this version:**

Thanh Mai Pham Ngoc. Problèmes inverses et analyse en ondelettes adaptées. Mathématiques [math].  
Université Pierre et Marie Curie - Paris VI, 2009. Français. NNT: . tel-00439112

**HAL Id: tel-00439112**

**<https://theses.hal.science/tel-00439112v1>**

Submitted on 6 Dec 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie - Paris-VI

**THÈSE DE DOCTORAT**

présentée pour obtenir le grade de

**DOCTEUR EN SCIENCES DE L'UNIVERSITÉ PARIS-VI**

Spécialité : **Mathématiques**

soutenue par

**Thanh Mai Pham Ngoc**

---

**PROBLÈMES INVERSES  
ET ANALYSE EN ONDELETTES ADAPTÉES**

---

**Rapporteurs :** M. Peter T. **KIM** University of Guelph  
M. Jean-Michel **LOUBES** Université Paul Sabatier

Soutenue publiquement le 27 novembre 2009 devant le jury composé  
de

M. Lucien	<b>BIRGÉ</b>	Université Paris-VI	Examineur
Mme Fabienne	<b>COMTE</b>	Université Paris-V	Examinatrice
M. Gilles	<b>FAÏ</b>	Université Lille-I	Examineur
M. Jean-Michel	<b>LOUBES</b>	Université Paul Sabatier	Rapporteur
Mme Dominique	<b>PICARD</b>	Université Paris-VII	Directrice
Mme Judith	<b>ROUSSEAU</b>	Université Paris Dauphine	Examinatrice



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Problèmes inverses . . . . .	5
1.1.1	Présentation générale . . . . .	5
1.1.2	Modélisation des problèmes inverses . . . . .	7
1.2	Méthodes de résolution : de la SVD aux Needlets . . . . .	9
1.2.1	La décomposition en valeurs singulières (méthode SVD) . . . . .	9
1.2.2	Les Ondelettes-Vaguelettes . . . . .	14
1.2.3	Needlets sur $\mathbb{R}$ . . . . .	16
1.3	Un Problème inverse sur la sphère : La déconvolution . . . . .	17
1.4	Quelques compléments de statistique non paramétrique . . . . .	20
1.4.1	Point de vue minimax . . . . .	20
1.4.2	Espaces de Besov forts . . . . .	21
1.4.3	Approche bayésienne non paramétrique . . . . .	23
1.5	Principaux résultats . . . . .	24
1.5.1	Approche statistique minimax du problème des moments de Hausdorff . . . . .	24
1.5.2	Déconvolution localisée sur la sphère . . . . .	26
1.5.3	Régression en design aléatoire et estimateurs bayésiens à ondelettes déformées . . . . .	33
1.6	Perspectives . . . . .	37
<b>2</b>	<b>Statistical minimax approach to the Hausdorff moment problem</b>	<b>39</b>
2.1	Introduction . . . . .	39
2.2	Statement of the problem . . . . .	42
2.2.1	The model. . . . .	42
2.2.2	The estimation procedure . . . . .	45
2.3	Proofs . . . . .	46
2.3.1	Proof of Theorem 2.1. . . . .	46

2.3.2	Proof of Theorem 2.2. . . . .	48
2.4	Conclusion . . . . .	52
2.5	Appendix . . . . .	52
<b>3</b>	<b>Localized deconvolution on the sphere</b>	<b>55</b>
3.1	Introduction . . . . .	55
3.2	Some preliminaries about harmonic analysis on $SO(3)$ and $\mathbb{S}^2$	58
3.2.1	The SVD Method . . . . .	60
3.3	Needlet construction . . . . .	61
3.3.1	Littlewood-Paley decomposition . . . . .	61
3.3.2	Discretization and localization properties . . . . .	62
3.3.3	Besov spaces on the sphere . . . . .	64
3.4	Needlet algorithm : thresholding needlet coefficients . . . . .	66
3.4.1	Performances of the procedure . . . . .	67
3.4.2	Condition (3.36) and the smoothness of $f_\varepsilon$ . . . . .	68
3.5	Practical performances . . . . .	69
3.6	Proof of Theorem 3.2 . . . . .	76
<b>4</b>	<b>Regression in random design and Bayesian warped wavelets estimators</b>	<b>83</b>
4.1	Introduction . . . . .	83
4.2	Methodology . . . . .	85
4.2.1	Warped bases . . . . .	85
4.2.2	Priors and estimators . . . . .	87
4.3	Functional spaces . . . . .	89
4.4	Minimax performances of the procedures . . . . .	90
4.4.1	Bayesian estimators based on Gaussian priors with small variances . . . . .	90
4.4.2	Bayesian estimators based on Gaussian priors with large variance . . . . .	91
4.5	Simulations and discussion . . . . .	91
4.6	Proofs . . . . .	98
	<b>Bibliographie</b>	<b>113</b>

# Chapitre 1

## Introduction

### 1.1 Problèmes inverses

#### 1.1.1 Présentation générale

Les problèmes inverses apparaissent dans de nombreux domaines scientifiques. Sans pour autant être exhaustif, citons entre autres les mathématiques, la sismologie, l'imagerie médicale, le traitement du signal, la cosmologie, l'acoustique architecturale... Ce caractère répandu, et ce dans des domaines aussi divers s'explique par la nature très générale des problèmes inverses. En effet, dans les grandes lignes l'idée directrice de ces derniers peut être décrite de façon assez simple : un modèle physique peut être vu comme une boîte qui, à partir de paramètres d'entrée, renvoie des données de sortie. Nous possédons des informations (une mesure) sur les données de sortie, et nous voulons en déduire une évaluation des paramètres d'entrée, d'où la dénomination de problèmes inverses. Les problèmes inverses sont en fait répandus, dès lors qu'il y a de la modélisation physique. En effet, aucune mesure n'est vraiment directe. Elle se fait toujours par l'intermédiaire d'un dispositif physique, qui nous donne une information indirecte sur le paramètre que l'on veut observer.

Nous allons à présent donner de façon plus formelle la forme d'un problème inverse linéaire. Celui-ci se présente de la façon suivante : Soit  $\mathbb{H}$  et  $\mathbb{K}$  deux espaces de Hilbert et soit  $A : \mathbb{H} \rightarrow \mathbb{K}$  un opérateur linéaire. On cherche à trouver une bonne approximation de la solution  $f$  de l'équation

$$g = Af,$$

lorsque l'on dispose d'une mesure de  $g$ .

Les problèmes inverses se divisent en deux catégories, celle des problèmes mal posés, par ailleurs les plus rencontrés et celle des problèmes bien posés. Ces derniers dont l'expression a été consacrée par Jacques Hadamard regroupent les problèmes où la solution  $f$  existe, est unique et dépend continûment des données  $g$  (i.e stabilité). Le problème de Dirichlet pour l'équation de Laplace en est une illustration. Revenons à présent aux problèmes mal posés qui vont occuper l'essentiel de notre attention dans cette thèse. Selon la nature et les propriétés de l'opérateur  $A$ , l'obtention de la solution  $f$  s'avère plus ou moins ardue. En effet, trois cas majeurs se distinguent :

- Si l'opérateur  $A$  n'est pas surjectif, la solution n'existe pas nécessairement.
- Si l'opérateur  $A$  n'est pas injectif, la solution peut ne pas être unique, auquel cas il nous incombe de réfléchir sur quelle solution privilégier.
- Si  $A^{-1}$  n'est pas continu, de petites perturbations sur les données  $g$  peuvent engendrer de forts écarts sur la solution  $f$ .

Aussi, il apparaît que la résolution des problèmes mal posés est sujette à des difficultés intimement liées à la nature de l'opérateur  $A$ . Pour illustrer notre propos, nous allons donner quelques exemples célèbres de problèmes inverses mal posés.

**Exemple 1.1.** *L'équation intégrale du premier ordre de Fredholm*

On définit l'opérateur de Fredholm comme suit

$$g(t) = \int_0^1 K(t, s)f(s)ds.$$

Si  $K(t, s) = K(t - s)$ , nous sommes dans le cadre du fameux problème de déconvolution sur  $\mathbb{R}$ .

**Exemple 1.2.** *Opérateur intégral d'Abel*

Ici  $\mathbb{H} = \mathbb{K} = L_2(0, 1)$

$$Ax(t) = \frac{1}{\pi} \int_t^1 \frac{x(\tau)}{\sqrt{\tau - t}} d\tau, \quad t \in (0, 1).$$

**Exemple 1.3.** *Equation de la chaleur 1D*

Si  $f : (x, t) \in [0, 1] \times \mathbb{R}^+ \rightarrow f(x, t) \in \mathbb{R}$  est une fonction continue donnée, et si  $k$  est une constante positive, le problème parabolique que nous considérons consiste à chercher  $u : (x, t) \in [0, 1] \times \mathbb{R}^+ \rightarrow u(x, t) \in \mathbb{R}$  qui satisfait :

$$\frac{\partial u}{\partial t}(x, t) - k \frac{\partial^2 u}{\partial x^2}(x, t) = f(x, t), \quad \forall x \in ]0, 1[, \quad \forall t > 0, \quad (1.1)$$

$$\frac{\partial u}{\partial x}(0, t) - k \frac{\partial^2 u}{\partial x^2}(1, t) = 0, \quad \forall t > 0, \quad (1.2)$$

$$u(x, 0) = w(x) \quad \forall x \in ]0, 1[, \quad (1.3)$$

où  $w : (x) \rightarrow w(x) \in \mathbb{R}$  est une condition initiale donnée. Le problème (1.1)-(1.3) est appelé problème de la chaleur ; si  $u(x, t)$  est la température d'un barreau métallique de longueur unité, alors l'équation (1.1) est l'équation de diffusion de la chaleur lorsque  $k$  et  $f$  sont respectivement la conductivité thermique et la puissance par unité de longueur fournie au barreau, toutes deux divisées par la densité volumique et la chaleur spécifique massique. Dans ce contexte, la condition aux limites (1.2) signifie que le barreau est thermiquement isolé en  $x = 0$  et  $x = 1$ .

**Exemple 1.4.** *Transformée de Radon*

La tomographie consiste à recueillir des mesures d'atténuation (projections) d'un rayon  $X$  à travers un tissu et à partir de ces projections de reconstruire la carte de densité de ce tissu.

Soit  $f$  la fonction d'atténuation au point  $x$  du tissu. L'opérateur de tomographie  $R$ , autre dénomination de la transformée de Radon, est défini pour  $f \in L_2([0, 1]^2)$  par :

$$(Rf)(t, \theta) = \int f(x) \delta(x_1 \cos \theta + x_2 \sin \theta - t) dx,$$

où  $\delta$  est l'argument de dirac. La valeur  $(Rf)(t, \theta)$  somme les contributions de la fonction originale  $f$  le long d'une droite paramétrée par sa pente  $\theta$  et son abscisse  $t$ .

### 1.1.2 Modélisation des problèmes inverses

#### Les observations

Bien évidemment, dans la réalité, il est difficile, voire impossible, d'obtenir des observations  $g$  qui ne soient pas entachées d'erreur. Aussi, est-il plus réaliste de considérer le modèle suivant :

$$Y = Af + \varepsilon\xi.$$



Deux cas de figures se présentent : le bruit  $\xi$  peut être déterministe ou aléatoire. Le cas du bruit déterministe qui constitue l'approche classique des problèmes inverses avec notamment l'hypothèse  $\|\xi\|_{\ell_2} < 1$  est l'objet d'une très vaste littérature. Le livre pionnier de Tikhonov et Arsenin [143] (voir chapitre 6), les travaux de Morozov [113], Traub et *al.* [146], Engl et *al.* [52], Wershhulz [154], Liskovets [100], Ivanov [75], Vainikko et Veretennikov [149], Aliev [6] en sont quelques éléments.

Le cas du modèle statistique qui nous importe comporte un bruit aléatoire. Son étude a été initiée par Sudakov et Gal'fin [134] et Bakushinskii [15]. Il se présente comme suit

$$Y = Af + \varepsilon\xi, \quad (1.4)$$

où  $\varepsilon$  avec  $0 < \varepsilon < 1$  désigne le niveau de bruit et  $\xi$  est un bruit blanc gaussien indexé par  $\mathbb{K}$  sur un espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ . Le modèle (1.4) peut être vu au sens où pour tout  $u \in \mathbb{K}$  la variable aléatoire  $Y(u)$  définie par

$$Y(u) = (Af, u) + \varepsilon\xi(u), \quad u \in \mathbb{K} \quad (1.5)$$

est observable, et où  $\xi(u) = (\xi, u)$  est une variable aléatoire gaussienne, centrée et de variance  $\|u\|^2$ .

Pour tout  $(u_1, u_2) \in \mathbb{K}^2$ , la covariance entre  $\xi(u_1)$  et  $\xi(u_2)$  est donnée par

$$\mathbb{E}((\xi(u_1)\xi(u_2))) = (u_1, u_2).$$

Le modèle (1.4) a beaucoup été étudié dans divers cadres statistiques. Dans le cadre minimax, des vitesses optimales de convergence pour le risque quadratique ont été mises en lumière entre autres par Korostelev et Tsybakov [94], [95], Johnstone et Silverman [81], Koo [93], Donoho [46], Mair et Ruymgaart [103]. Plus récemment, ce problème a été l'objet d'études dans un cadre minimax adaptatif. Nous pouvons citer les travaux de Goldenshluger [63] pour le risque ponctuel dans le problème de déconvolution, Johnstone [79] à l'aide d'estimateurs construits à partir des Ondelettes-Vaguelettes, Cavalier et Tsybakov [31] qui ont exhibé des estimateurs super-optimaux. Kerkycharian et *al.* [89] pour les problèmes de déconvolution et de Wicksell à l'aide de nouvelles bases appelées Needlets ont fourni de nouvelles vitesses de convergence tout à fait originales.

De prime abord, une tentative naturelle pour résoudre le problème (1.4) serait de prendre l'estimée  $\hat{f} = A^{-1}Y$ . Malheureusement dans le cas d'un opérateur  $A$  compact, et en fait de façon plus générale dans de nombreux

contextes scientifiques comme nous le verrons avec le chapitre 2, l'opérateur  $A^{-1}$  n'est pas un opérateur linéaire borné et de fait son inverse n'est pas continu. La solution  $f$  ne dépend donc pas continûment des données  $Y$ . Nous nous trouvons ainsi dans le cadre de problèmes inverses mal posés et un appareillage plus sophistiqué pour la résolution de ces problèmes s'impose, ce sur quoi nous allons nous pencher dans la section suivante.

## 1.2 Méthodes de résolution : de la SVD aux Needlets

Les méthodes pour résoudre les problèmes inverses mal posés en statistique sont légion. Nombre d'entre elles nous viennent de l'analyse numérique. Nous pouvons citer entre autres les méthodes de régularisation de Tikhonov-Phillips, les techniques d'estimation récursive dans les espaces de Hilbert, les méthodes de projection (Galerkin). Nous expliquerons dans ce qui suit la décomposition en valeurs singulières qui étaye de nombreuses procédures d'estimation statistique. Nous en verrons également les limites.

### 1.2.1 La décomposition en valeurs singulières (méthode SVD)

Pour reprendre les termes employés par Donoho dans [46], la méthode SVD de par la place fondamentale qu'elle tient dans les problèmes inverses linéaires constitue un véritable paradigme pour leur analyse et leur résolution. C'est donc tout naturellement que nous allons débiter cette partie par un théorème crucial à la base de la méthode SVD.

**Théorème 1.1.** *Si  $A$  est un opérateur compact, alors il existe deux bases orthonormées  $(e_k)_{k \in \mathbb{N}}$  de  $\mathbb{H}$  et  $(g_k)_{k \in \mathbb{N}}$  de  $\mathbb{K}$  et une suite strictement positive  $b_k \rightarrow 0$  quand  $k$  tend vers l'infini, telles que :*

$$Ae_k = b_k g_k, \quad A^* g_k = b_k e_k, \quad A^* A e_k = b_k^2 e_k, \quad A A^* g_k = b_k^2 g_k, \quad (1.6)$$

où  $A^*$  désigne l'opérateur adjoint de  $A$ .

C'est ce que l'on appelle la *décomposition en valeurs singulières* de l'opérateur  $A^*A$  et le triplet  $(e_k, g_k, b_k)$  est appelé *système singulier*. Dans la base  $e_k$  la matrice de l'opérateur  $A^*A$  est diagonale. La décomposition en valeurs singulières de  $A^*A$  va nous permettre de donner un estimateur simple de  $f$ . Pour cela, on décompose tout d'abord la fonction sur la base  $e_k$  :

$$f = \sum_{l=1}^{\infty} (f, e_l) e_l,$$

puis en notant  $y_k = (Y, g_k)$  et  $\xi_k = (\xi, g_k)$ , il vient

$$\begin{aligned} y_k &= (Af, g_k) + \varepsilon\xi_k \\ y_k &= \left(\sum_l b_l(f, e_l)g_l, g_k\right) + \varepsilon\xi_k \\ y_k &= b_k\theta_k + \varepsilon\xi_k, \quad k \in \mathbb{N}^* \end{aligned} \tag{1.7}$$

ou encore de façon équivalente :

$$y'_k = \theta_k + \varepsilon\sigma_k\xi_k, \quad k \in \mathbb{N}^* \tag{1.8}$$

avec  $\theta_k = (f, e_k)$  et  $\sigma_k = b_k^{-1}$ . Précisons que puisque le système  $g_k$  forme une base orthonormale, les  $\xi_k$  sont des variables gaussiennes i.i.d centrées et réduites. Il vient alors un estimateur naturel des coefficients de Fourier  $\theta_k$

$$\hat{\theta}_k = y_k/b_k.$$

Cette projection sur la base SVD nous mène à un modèle séquentiel hétéroscédastique de bruit blanc gaussien (1.7). Le comportement asymptotique de la suite des  $b_k$  et plus précisément la vitesse à laquelle les  $b_k$  vont tendre vers 0 va jouer un rôle primordial dans l'étude du modèle (1.7). En effet, puisque les  $b_k$  tendent vers 0, la suite des signaux  $b_k\theta_k$  va elle même tendre vers 0 et l'on va perdre ainsi de plus en plus d'information au profit du bruit qui va devenir de plus en plus prépondérant. Le signal est donc lissé par l'opérateur  $A^{-1}$ .

Il va donc falloir d'une part tronquer la série de Fourier de la fonction  $f$  jusqu'à un niveau convenable  $N$  et d'autre part poser une suite de filtre  $\lambda_k$  sur les coefficients afin de ne garder que les plus significatifs et se délester d'une multitude de petits coefficients bruités qui corrompent le signal. C'est en sens qu'il faut comprendre la régularisation du problème dans la méthode SVD. On obtient au final l'estimation suivante de  $f$  :

$$\hat{f} = \sum_{k=1}^N \lambda_k \frac{y_k}{b_k} e_k.$$

Plusieurs types de filtre existent, mentionnons en quelques uns :

- Filtre de Tikhonov-Phillips

$$\lambda_k = \frac{1}{1 + Ck^\tau},$$

où  $C$  et  $\tau$  sont des constantes strictements positives.

- Filtre de Pinsky

$$\lambda_k = (1 - (k/C)^\tau)_+.$$

- Méthode de Féjer

$$\lambda_k = \left(1 - \frac{k-1}{N}\right), \quad k \leq N.$$

- Méthode de Bernstein Rogozinski pour des bases trigonométriques. En posant  $N = 2m + 1$ ,

$$\lambda_{2k} = \lambda_{2k+1} = \cos \frac{\pi k}{2m} \quad k = 0, \dots, m.$$

Par ailleurs, une terminologie a été établie pour classer les problèmes inverses selon leur difficulté, difficulté inhérente au comportement asymptotique de la suite des  $b_k$ . C'est Wahba [153] qui la première a parlé de cette notion de "ill-posedness degree" ou encore degré du problème inverse. Cette notion n'est guère aisée à quantifier en général et se trouve fortement liée à la structure des espaces de Hilbert en question. Néanmoins, une "classification" semble faire référence.

- Problèmes moyennement mal posés.

Si la suite des  $b_k$  tend vers 0 polynômialement et les coefficients de Fourier  $\theta_k$  tendent vers 0 exponentiellement ou polynômialement, on parlera de problèmes moyennement mal posés. Dans ce cas bien connu et même pour un cadre élargi par rapport au modèle (1.8), Cavalier et Tsybakov [31], Cavalier et al. [29] ont fourni des estimateurs de  $f$  adaptatifs super optimaux.

- Problèmes sévèrement mal posés.

Si les  $b_k$  tendent vers 0 exponentiellement et les  $\theta_k$  polynômialement, on parlera de problèmes sévèrement mal posés. Ce type de problème sera abordé dans le chapitre 2.

- Problèmes 2-exp sévèrement mal posés.

Cavalier et al. [30], Tsybakov [147] et Golubev et Khasminskii [65] ont abordé le cas d'un problème sévèrement mal posés où et les  $b_k$  et les  $\theta_k$  tendaient vers 0 exponentiellement vite. Les auteurs le qualifient de 2-exp problème sévèrement mal posé et exhibent des estimateurs adaptatifs super optimaux. Ce genre de problème se distingue par des vitesses polynomiales et ce en dépit du caractère très mal posé.

D'autres auteurs ont également mis en lumière ce comportement dans un cadre minimax non adaptatif parmi lesquels Ermakov [53], Pensky et Vidakovic [120] dans le cadre de la déconvolution et Efromovich et Koltchinskii [51].

Nous renvoyons également aux articles de Landau et Pollack [97], Shure et al [131], Davison [37], Louis [102], Gori et Guattari [66], et les travaux de Bertero et al. [19], [21], [20], [22], qui donnent à voir la variété de bases SVD possibles selon la nature très spécifique de l'opérateur en question. Enfin les articles de Bertero et al [23], Johnstone et Silverman [82] [81] étudient en détails la base SVD dans de nombreux exemples et contextes différents. Du reste, nous allons à présent nous attarder sur quelques exemples de problèmes où la base SVD est explicite.

### Exemples de bases SVD

#### Exemple 1.5. *La déconvolution*

Cet exemple fameux a déjà été évoqué plus haut. Reprenons-le un peu plus en détails.

Ici les espaces de Hilbert  $\mathbb{H}$  et  $\mathbb{K}$  sont l'ensemble des fonctions 1-périodiques de carré intégrable pour la norme  $\mathbb{L}^2([0, 1], dx)$ . Et l'opérateur de convolution est donné par

$$Af(u) = \int_0^1 \gamma(u-t)f(t)dt, \quad f \in \mathbb{H},$$

où  $\gamma : [0, 1]^2 \rightarrow \mathbb{R}$  est appelée noyau de convolution. La base SVD se trouve être la base trigonométrique réelle sur  $[0, 1]$ . En effet, considérons

$$e_1(x) = 1, \quad e_{2k}(x) = \cos(2\pi kt) \text{ et } e_{2k+1}(x) = \sin(2\pi kt), \quad \forall k \in \mathbb{N}^*.$$

On peut alors montrer que

$$Ae_k(x) = b_k e_k(x), \quad \text{avec } b_k = \int_0^1 \gamma(u) e_k(u) du.$$

La relation (1.6) est obtenue en posant  $e_k = g_k$ .

Ce problème a été l'objet d'un nombre important de travaux. Plusieurs approches ont été mises en oeuvre dont celle par la méthode des noyaux, voir

les contributions de Carroll et Hall [28], Devroye [40], Diggle et Hall [41], Stefanski [133], Stefanski et Carroll [132], Efromovich [50], Fan [54], [55], Liu et Taylor [101], Masry [105], [106], Taylor et Zhang [140]. Des auteurs ont aussi abordé ce problème sous l'angle des ondelettes, voir Pensky et Vidakovic [120], Fan et Koo [56], Kalifa et Mallat [83], Johnstone et al. [80].

**Exemple 1.6.** *Equation de la chaleur*

Nous reprenons l'exemple (1.3). Pour ce problème la base SVD est de nouveau la base de Fourier et  $u(x, T)$  peut être écrit  $u(x, T) = \sqrt{2} \sum_{k=1}^{+\infty} \theta_k e^{-\pi^2 k^2 T} \sin(k\pi x)$ . Les valeurs singulières sont  $b_k = e^{-\pi^2 k^2 T/2}$ . Le problème inverse de l'équation de la chaleur est donc sévèrement mal posé.

**Exemple 1.7.** *Problème de Wicksell*

Le problème de Wicksell peut se résumer comme un problème d'estimation de la densité de probabilité des rayons de sphères dans un milieu donné à partir d'observations de coupes bi-dimensionnelles. On retrouve ce type de problème dans l'étude de l'état des métastases dans le foie.

Le milieu est découpé en tranches de manière indépendante de la taille et de la position des sphères. Soit  $G(x)$  la fonction de répartition des rayons des cercles (bi-dimensionnels) observés sur des coupes transversales effectuées au hasard sur le milieu. Le problème posé est alors, étant donné un échantillon issu de la loi dont la fonction de répartition est  $G$ , de pouvoir obtenir une bonne estimation de la densité  $f(r)$ . D'après Johnstone et Silverman [82], le problème de Wicksell est défini par un opérateur du type Abel :

$$\mathbb{H} = \mathbb{L}_2([0, 1], d\mu), \quad d\mu(x) = (4x)^{-1} dx, \quad \mathbb{K} = \mathbb{L}_2([0, 1], d\lambda), \quad d\lambda(x) = 4\pi^{-1}(1-y^2)^{1/2} dy,$$

et

$$Af(y) = \frac{\pi}{4} y(1-y^2)^{-1/2} \int_y^1 (x^2 - y^2)^{-1/2} f(x) d\mu.$$

Les fonctions singulières pour cet opérateur sont

$$\begin{aligned} e_k(x) &= 4(k+1)^{1/2} x^2 P_k^{0,1}(2x^2 - 1) \\ g_k(y) &= \mathcal{U}_{2k+1}(y), \end{aligned}$$

où  $P_k^{0,1}$  est le polynôme de Jacobi de type  $(0, 1)$  de degré  $k$  et  $\mathcal{U}_k$  est le polynôme de Chebishev de second espèce de degré  $k$ . Les valeurs singulières pour ce problème sont

$$b_k = \frac{\pi}{16} (1+k)^{-1/2}.$$

La décomposition en valeurs singulières pour la résolution des problèmes inverses mal posés, bien que très populaire et séduisante de par sa simplicité théorique, sa concision et son efficacité numérique, pâtit néanmoins de certains défauts au premier chef duquel la difficulté plus ou moins grande de déterminer le système singulier. Par ailleurs, il apparaît que les fonctions de base singulières  $(e_k)_k$  et  $(g_k)_k$  sont entièrement déterminées par l'opérateur  $A$  et non par la fonction  $f$  et de fait ne sont pas forcément adaptées à cette dernière. Plus précisément, les bases singulières mises en évidence par l'opérateur  $A$  ne vont pas nécessairement convenir à la fonction  $f$ , celle-ci pouvant avoir une représentation avec beaucoup voire trop de coefficients dans la base des  $(e_k)_k$ . Toutes ces raisons ont motivé la mise en oeuvre de la méthode des ondelettes-vaguelettes de Donoho, expliquée en détails dans l'article de référence Donoho [46] et dont nous donnons un aperçu dans la section suivante.

### 1.2.2 Les Ondelettes-Vaguelettes

Cette méthode permet à la fois de concilier les avantages incontestables de la méthode SVD qui représente parfaitement l'opérateur  $A$  et ceux de l'analyse multirésolution qui se trouve au coeur des ondelettes. La décomposition ondelettes-vaguelettes s'articule autour de l'idée de pouvoir donner une représentation qui convienne à la fois à l'opérateur et à la fonction cible. L'analyse multirésolution permet de disposer de bases orthonormées dont les atomes sont à la fois localisés en temps et en fréquence ce que ne permet pas la base SVD qui ne localise qu'en fréquence. Ainsi, les ondelettes de par leur capacité très forte de compression permettent une représentation "sparse" ou clairsemée du signal en question et qui constitue en quelque sorte une représentation optimale de celui-ci.

L'utilisation et l'efficacité des ondelettes dans le domaine de l'estimation non paramétrique n'est plus à démontrer comme en attestent les travaux de Donoho et al. [47], [48], [45], Kerkyacharien et Picard [90], [86], Hall et al. [69], Hall et Patil [70]. Pour de plus amples détails sur les ondelettes et leur construction on se référera aux ouvrages de Meyer [111], Mallat [104], Daubechies [36], Hardle et al. [71].

La décomposition ondelettes-vaguelettes (ou en anglais Wavelet Vaguelette Decomposition (WVD)) se fonde sur l'idée suivante :

on suppose l'existence d'une base d'ondelettes  $(\psi_\lambda)$  et de deux ensembles orthonormaux  $(u_\lambda)$  et  $(v_\lambda)$  avec  $\lambda = (j, k)$  qui doivent satisfaire les relations

ci dessous

$$A\psi_\lambda = \kappa_j v_\lambda \quad (1.9)$$

$$A^*u_\lambda = \kappa_j \psi_\lambda. \quad (1.10)$$

Ce premier jeu de relation est appelé quasi relations singulières et induit comme son nom l'indique que nous sommes "proches" de la décomposition SVD de l'opérateur  $A$ . Notons que si  $u = v$  nous retombons sur la décomposition SVD. La WVD suppose aussi que

$$(u_\lambda, u_\nu) = \delta_{\lambda\nu},$$

où  $\delta_{\lambda\nu} = 0$  si  $\lambda$  est différent de  $\nu$  et 1 sinon, et que :

$$\left\| \sum_\lambda a_\lambda u_\lambda \right\|_2 \simeq \|a_\lambda\|_{\ell^2} \simeq \left\| \sum_\lambda a_\lambda v_\lambda \right\|_2.$$

Cette décomposition nous amène à la représentation suivante de la fonction  $f$  :

$$f = \sum_\lambda (Af, u_\lambda) \kappa_j^{-1} \psi_\lambda.$$

Fort de cette décomposition, il nous faut trouver un estimateur de  $f$ . A l'instar de la méthodologie qui a prévalu pour la SVD, on estime  $(Af, u_\lambda)$  par  $(Y, u_\lambda)$  et on filtre en imposant un seuillage doux ou dur. On obtient donc l'estimateur suivant

$$\hat{f} = \sum_\lambda \delta_{t_j} [(Y, u_\lambda) \kappa_j^{-1}] \psi_\lambda. \quad (1.11)$$

Plus précisément, avec  $f = \sum_\lambda \beta_\lambda \psi_\lambda$  où  $\lambda = (j, k)$ , dans le cas du modèle (1.5) on obtient les observations qui justifient l'écriture précédente de  $f$  (1.11),

$$\begin{aligned} y_{\lambda'} &= (A \sum_\lambda \beta_\lambda \psi_\lambda, u_{\lambda'}) + \varepsilon \xi(u_{\lambda'}) = \sum_\lambda \beta_\lambda (A\psi_\lambda, u_{\lambda'}) + \varepsilon \xi(u_{\lambda'}) \\ &= \sum_\lambda \beta_\lambda (\psi_\lambda, A^* u_{\lambda'}) + \varepsilon \xi(u_{\lambda'}) = \sum_\lambda \beta_\lambda (\psi_\lambda, \kappa_j \psi_{\lambda'}) + \varepsilon \xi(u_{\lambda'}) \\ &= \kappa_j \beta_{\lambda'} + \varepsilon \xi(u_{\lambda'}). \end{aligned}$$

On en conclut

$$\hat{f} = \sum_\lambda \delta_{t_j} \left( \frac{y_\lambda}{\kappa_j} \right) \psi_\lambda,$$



où la procédure de seuillage  $\delta_{t_j}$  dépendant d'une suite de seuil  $t_j$  à fixer vaut dans le cas du seuillage doux :

$$\delta_{t_j}\left(\frac{y_\lambda}{\kappa_j}\right) = \operatorname{sgn}\left(\frac{y_\lambda}{\kappa_j}\right)\left(\left|\frac{y_\lambda}{\kappa_j}\right| - t_j\right)_+$$

et dans le cas du seuillage dur :

$$\delta_{t_j}\left(\frac{y_\lambda}{\kappa_j}\right) = \frac{y_\lambda}{\kappa_j} I\left(\left|\frac{y_\lambda}{\kappa_j}\right| > t_j\right).$$

Donoho [46] donne explicitement la construction de la décomposition WVD pour des problèmes inverses induits par l'opérateur d'intégration fractionnaire dont l'opérateur d'intégration d'Abel est un cas particulier ou encore la transformée de Radon fort utilisée en imagerie médicale. Néanmoins la WVD bien que très puissante et efficace dans les problèmes que nous venons de citer, ne s'applique pas dans tous les problèmes inverses, c'est notamment le cas du problème de Wicksell.

### 1.2.3 Needlets sur $\mathbb{R}$

Kerkyacharian et al. [89] ont jeté les bases de l'utilisation des Needlets en statistique pour les problèmes inverses. Le problème de Wicksell figure parmi l'une de leurs motivations puisque l'algorithme WVD ne s'y applique pas. Au demeurant, les needlets ont été développées par Petrushev et ses co-auteurs [122], [121], [114]. Les needlets dont la dénomination vient de "needle", "aiguille" en anglais, appellation justifiée pour leur très bonne localisation, permettent de s'affranchir des cas où la base SVD n'est pas la base de Fourier. Les Needlets contrairement à la base SVD ne forment pas nécessairement un système orthonormal mais constitue ce que l'on appelle en anglais un *frame*.

**Definition 1.1.** *Soit  $\mathbb{H}$  un espace de Hilbert. Une suite  $\psi_n$  est appelé *frame* si*

$$\|f\|^2 = \sum_n |(f, \psi_n)|^2.$$

Nous donnerons dans la suite de cette introduction de plus amples détails sur la construction des needlets sur  $\mathbb{S}^2$ , la sphère unité de  $\mathbb{R}^3$  comme nous traiterons dans cette thèse le problème de la déconvolution sur  $\mathbb{S}^2$ . Dans les grandes lignes, la construction des needlets est structurée par deux étapes fondamentales. La première étape est la décomposition de Calderón qui permet une analyse en fréquence de l'espace  $\mathbb{H}$ , où une base orthonormée entre

en jeu, en l'occurrence la base SVD ainsi qu'une fonction  $b$  régulière. Cette décomposition de Calderón est suivie par une seconde étape de discrétisation de chaque sous espace de fréquence. Cette discrétisation fait intervenir une formule de quadrature pour chacun d'entre eux. Cette formule de quadrature, bien sûr change selon les problèmes. Dans [89] pour le problème de Wicksell, la formule de quadrature fait intervenir les racines des polynômes de Jacobi.

### 1.3 Un Problème inverse sur la sphère : La déconvolution

Les problèmes inverses mal posés ont eu jusqu'ici dans leur très grande majorité trait à des problèmes sur  $\mathbb{R}$ , comme en témoignent les références susmentionnées. Ceci a notamment été le cas pour le problème de la déconvolution. Ce dernier fort populaire et étudié sur la droite réelle a été en fait très peu abordé sur la sphère unité  $\mathbb{S}^{d-1}$  en dimension  $d$ . Par souci de clarté, nous allons revenir sur la modélisation du problème de la déconvolution sur la sphère en introduisant les concepts essentiels de ce problème. Les premiers à en avoir fait état sont Van Rooij et Ruymgaart [150]. Ces auteurs considéraient le problème où des erreurs viennent corrompre des données initiales de  $\mathbb{S}^{d-1}$ . Cette perturbation dans le contexte de la sphère se traduit par une multiplication d'un processus aléatoire par une matrice aléatoire du groupe orthogonal  $O(d)$ . Van Rooij et Ruymgaart [150] considéraient alors le modèle suivant :

on observe le processus  $Z$  défini par

$$Z_i = \varepsilon_i X_i, \quad (1.12)$$

où  $Z_i \in \mathbb{S}^{d-1}$ ,  $\varepsilon_i \in O(d)$ ,  $X_i \in \mathbb{S}^{d-1}$ . On désigne par  $\mu$  la probabilité invariante sur  $O(d)$  et  $\bar{\mu}$  la mesure uniforme sur  $\mathbb{S}^{d-1}$ . On peut alors définir le produit de convolution par

$$f * h(\omega) = \int_{O(d)} f(O)h(O^t\omega)d\mu(O), \quad \omega \in \mathbb{S}^{d-1},$$

où  $h \in L_2(\mathbb{S}^{d-1}, \bar{\mu})$  et  $f \in L_1(O(d), \mu)$ . Les auteurs ont simplement posé le problème et soulevé des questions sans toutefois apporter de réponses à l'estimation de la densité  $f$  du processus  $X$ .

Healy et al. [73], puis Kim et Koo [91], [92] ont abordé le problème de la déconvolution sur la sphère  $\mathbb{S}^2$ , les dimensions supérieures étant enclines

à des difficultés et une complexification telles, qu'il convenait d'abord de considérer l'étude sur  $\mathbb{S}^2$  et de poser les jalons d'une procédure d'estimation sur  $\mathbb{S}^2$ . Le cadre reste peu ou prou le même que celui défini par Van Rooij et Ruymgaart [150], à savoir que l'on observe  $Z_1, \dots, Z_n$ ,  $n$  observations i.i.d

$$Z_i = \varepsilon_i X_i, \quad (1.13)$$

où  $Z_i \in \mathbb{S}^2$ ,  $\varepsilon_i \in SO(3)$ ,  $X_i \in \mathbb{S}^2$ , les processus  $\varepsilon$  et  $X$  sont supposés indépendants. Notons que la convolution est cette fois ci exprimée par la multiplication du processus  $X$  par une matrice orthogonale  $3 \times 3$  réelle de déterminant 1. Si l'on considère les densités des processus, on obtient le produit de convolution suivant

$$f_Z = f_\varepsilon * f, \quad (1.14)$$

avec,

$$g * h(\omega) = \int_{SO(3)} g(u)h(u^{-1}\omega)d\mu, \quad \omega \in \mathbb{S}^2,$$

où  $h \in L_2(\mathbb{S}^2, \bar{\mu})$  et  $g \in L_2(SO(3), \mu)$ .

Les outils techniques en analyse pour aborder le problème de la déconvolution sur  $\mathbb{S}^2$  et construire les estimateurs de la densité  $f$  viennent de l'analyse harmonique sur  $SO(3)$  et sur  $\mathbb{S}^2$ . Dans cette introduction, nous n'entrerons pas dans les détails de ces outils fort complexes et les laissons pour le chapitre 3. Nous préférons nous attacher à donner les grandes lignes de la procédure d'estimation mis en avant par Healy et al. [73], Kim et Koo [91] qui ont traité ce problème sous l'angle de la méthode SVD.

Il s'agit donc de l'approche "classique" de résolution des problèmes inverses mal posés. Il est possible d'étendre les concepts de la dimension 1 à la dimension 3 et de faire une analogie entre ce qui se passe sur la droite réelle et  $\mathbb{S}^2$ . La base de Fourier pour  $L_2(\mathbb{S}^2)$  est la base des harmoniques sphériques  $Y_m^l$  et les harmoniques rotationnelles  $D_{m,n}^l$  construites à partir des fonctions de Legendre généralisées constituent une base orthonormée de  $L_2(SO(3))$ . Nous allons donner les étapes majeures de la méthode SVD pour ce problème (les détails se trouvent dans le chapitre 3).

On réécrit le modèle (1.14) en passant aux transformées de Fourier (définies dans le chapitre 3 par les formules (3.6) et (3.11))

$$\hat{f}^l = (\hat{f}_\varepsilon^l)^{-1} \hat{f}_Z^l. \quad (1.15)$$

On suppose que  $f_\varepsilon$  est connue et que la matrice  $(\hat{f}_\varepsilon^l)^{-1}$  existe. La version empirique de (1.15) donne

$$\begin{aligned}\hat{f}_m^{l,N} &:= \frac{1}{N} \sum_{j=1}^N \sum_{n=-l}^l \hat{f}_{\varepsilon^{-1},mn}^l \bar{Y}_n^l(Z_j) \\ \hat{f}_{\varepsilon^{-1}}^l &:= (\hat{f}_\varepsilon^l)^{-1},\end{aligned}\tag{1.16}$$

où  $\hat{f}_{\varepsilon^{-1},mn}^l$  est le terme  $(m, n)$  de la matrice  $\hat{f}_{\varepsilon^{-1}}^l$ .  
On obtient donc l'estimateur suivant de  $f$

$$f^N(\omega) = \sum_{l=0}^{\tilde{N}} \sum_{m=-l}^l \hat{f}_{X,m}^{l,N} Y_m^l(\omega),\tag{1.17}$$

où  $\tilde{N}$  est à fixer judicieusement.

Il faut encore définir la régularité de la distribution du processus  $\varepsilon$ . La matrice  $(\hat{f}_\varepsilon^l)^{-1}$  joue le même rôle que les valeurs singulières  $b_k$  pour la droite réelle. La distribution de  $\varepsilon$  est dite *super régulière* si la norme d'opérateur de la transformée de fourier de  $f_\varepsilon$  satisfait

$$\|(\hat{f}_\varepsilon^l)^{-1}\|_{op} \leq d_0 l^{\beta_1} \exp(-l^\beta/\gamma), \quad l \rightarrow \infty,$$

où  $d_1, \beta_1, \beta$  et  $\gamma$  sont des constantes positives et cette distribution sera dite *régulière* si

$$\|(\hat{f}_\varepsilon^l)^{-1}\|_{op} \leq d_1 l^{-\beta}.$$

Nous rappelons que

$$\|\hat{f}_\varepsilon^l\|_{op} = \sup_{h \neq 0, h \in \mathcal{E}_l} \frac{\|\hat{f}_\varepsilon^l h\|_2}{\|h\|_2},$$

où  $\mathcal{E}_l$  est le sous espace vectoriel de dimension  $(2l+1)$  engendré par les harmoniques sphériques  $\{Y_m^l : -l \leq m \leq l\}$ .

Kim et Koo [91] fournissent pour leur estimateur par la méthode SVD des vitesses de convergence minimax dans le cas des erreurs régulières ou très régulières pour des densités appartenant à des classes de Sobolev étendues à  $\mathbb{S}^2$ .

Cet aperçu de la méthode SVD pour le problème de la déconvolution sur  $\mathbb{S}^2$  permet de clarifier les outils et concepts qui serviront en partie à l'approche que nous proposons à ce problème, à savoir une procédure d'estimation basée sur les needlets sphériques.

Avant de passer au paragraphe consacré aux principaux résultats de cette thèse, nous allons nous attarder sur quelques notions de statistique non paramétrique qui seront utiles à la compréhension des travaux que nous présentons.

## 1.4 Quelques compléments de statistique non paramétrique

### 1.4.1 Point de vue minimax

Il s'agit du point de vue théorique classique pour mesurer la performance d'une procédure d'estimation. On considère le problème statistique suivant : On dispose de  $n$  observations  $X_1, \dots, X_n$  issues d'un modèle statistique noté  $(\mathcal{X}_n, \mathcal{A}_n, \mathbb{P}_f^n, f \in \mathcal{F}(\mathbb{R}^d, \mathbb{R}^{d'}))$ , où  $\mathcal{F}(\mathbb{R}^d, \mathbb{R}^{d'})$  désigne l'ensemble des applications de  $\mathbb{R}^d$  dans  $\mathbb{R}^{d'}$ . Soit  $\hat{f}_n$  un estimateur de  $f$  construit à partir de ces  $n$  observations. Le risque de  $\hat{f}_n$  associé à une perte  $\rho$  est défini par :

$$R_\rho^n(\hat{f}_n, f) = \mathbb{E}_f \rho(\hat{f}_n, f).$$

Les fonctions de perte usuelles décrivent des normes  $L_p$  ou des normes associées à des espaces de Sobolev, de Hölder ou de Besov. Sans hypothèses de régularité sur  $f$ , on ne peut pas en général obtenir des résultats de convergence pour

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{F}(\mathbb{R}^d, \mathbb{R}^{d'})} R_\rho^n(\hat{f}_n, f).$$

Aussi, il nous faut choisir un espace fonctionnel  $V$  auquel appartiendra la fonction à estimer  $f$ . Le risque minimax pour  $V$  est alors défini par

$$\mathcal{R}_\rho^n(V) = \inf_{\hat{f}_n} \sup_{f \in V} \mathbb{E}_f \rho(\hat{f}_n, f),$$

où l'infimum est pris sur l'ensemble de tous les estimateurs. Si  $r_n$  est une suite qui tend vers 0 telle que

$$C_1 r_n \leq \mathcal{R}_\rho^n(V) \leq C_2 r_n,$$

avec  $C_1$  et  $C_2$  deux constantes, alors on dit que  $r_n$  est la vitesse minimax pour l'espace  $V$  associée à la perte  $\rho$ . Un estimateur  $\hat{f}_n$  sera dit optimal au sens minimax s'il atteint la vitesse  $r_n$ , c'est-à-dire si

$$\sup_{f \in V} \mathbb{E}_f \rho(\hat{f}_n, f) \leq C_3 r_n,$$

où  $C_3$  est une constante.

### 1.4.2 Espaces de Besov forts

L'objet de cette section est de présenter succinctement quelques propriétés des espaces de Besov forts notamment dans le cadre de l'analyse par ondelettes. Pour plus de détails, on se référera aux travaux de Peetre [118], Meyer [111] ou DeVore et Lorentz [39].

Commençons par donner la définition de la classe des espaces de Besov forts en termes de module de continuité. Soit  $0 < s < 1$ ,  $1 \leq p \leq \infty$ ,  $1 \leq q < \infty$ , et posons  $\forall (x, h) \in \mathbb{R}^2$ ,  $\tau_h f(x) = f(x - h)$ . On définit

$$\gamma_{pq}^s(f) = \left( \int_{\mathbb{R}} \left( \frac{\|\tau_h f - f\|_{L_p}}{|h|^s} \right)^q \frac{dh}{|h|} \right)^{1/q},$$

et

$$\gamma_{p\infty}^s(f) = \sup_{h \in \mathbb{R}} \frac{\|\tau_h f - f\|_{L_p}}{|h|^s}.$$

Si  $s = 1$ , on pose,

$$\gamma_{pq}^1(f) = \left( \int_{\mathbb{R}} \left( \frac{\|\tau_h f + \tau_{-h} f - 2f\|_{L_p}}{|h|} \right)^q \frac{dh}{|h|} \right)^{1/q},$$

et

$$\gamma_{p\infty}^1(f) = \sup_{h \in \mathbb{R}} \frac{\|\tau_h f + \tau_{-h} f - 2f\|_{L_p}}{|h|}.$$

En s'aidant de ces définitions, on définit  $\forall 0 < s \leq 1$  et  $\forall 1 \leq p, q \leq \infty$ , l'espace de Besov fort de paramètres  $s$ ,  $p$ , et  $q$

$$B_{p,q}^s = \{f \in L_p(\mathbb{R}) : \gamma_{pq}^s(f) < \infty\},$$

muni de la norme

$$\|f\|_{B_{p,q}^s} = \|f\|_{L_p} + \gamma_{pq}^s(f).$$

Si  $s = N + \alpha$ , avec  $N \in \mathbb{N}$  et  $0 < \alpha \leq 1$ , on dit que  $f \in B_{p,q}^s$  si et seulement si  $f^{(m)} \in B_{p,q}^\alpha$ , pour tout  $m \leq N$ . Cet espace est muni de la norme :

$$\|f\|_{B_{p,q}^s} = \|f\|_{L_p} + \sum_{m \leq N} \gamma_{pq}^\alpha(f^{(m)}).$$

Rappelons que l'espace de Sobolev  $H^s$  correspond à l'espace de Besov  $B_{2,2}^s$ , l'espace de Hölder  $C^s$  (avec  $0 < s \notin \mathbb{N}$ ) à  $B_{\infty,\infty}^s$ .

Une caractérisation essentielle des espaces de Besov forts repose sur la notion de vitesse d'approximation. On a le résultat suivant :

**Théorème 1.2.** *Supposons donné un système "fonction d'échelle/ondelette"  $(\phi, \psi)$  de régularité  $r$ . Considérons les opérateurs de projection  $P_j$ ,  $j \geq 0$ , sur les espaces  $V_j$  de l'analyse multirésolution engendré par  $(\phi, \psi)$ . Alors, pour tous  $0 < s < r$ ,  $1 \leq p, q \leq \infty$ ,  $f$  appartient à l'espace de Besov fort  $B_{p,q}^s$  si et seulement si  $f \in L_p(\mathbb{R})$  et il existe une suite de nombres positifs  $(\mathcal{E}_j)_{j \in \mathbb{N}} \in l_q(\mathbb{N})$  telle que*

$$\forall j \in \mathbb{N}, \quad \|f - P_j f\|_{L_p} \leq 2^{-js} \mathcal{E}_j.$$

Enfin, rappelons que les espaces de Besov forts admettent une caractérisation en termes de coefficients d'ondelette. Sous les hypothèses du théorème précédent, pour toute fonction  $f \in L_p(\mathbb{R})$ , on peut définir les coefficients :

$$\alpha_{0k} = \int f(x) \phi_{0k}(x) dx, \quad \beta_{jk} = \int f(x) \psi_{jk}(x) dx,$$

et  $f$  appartient à l'espace de Besov fort  $B_{p,q}^s$  si et seulement si

$$J_{p,q}^s(f) = \|\alpha_0\|_{l_p} + \left( \sum_{j \geq 0} 2^{jq(s+1/2-1/p)} \|\beta_j\|_{l_p}^q \right)^{1/q} < \infty, \quad \text{si } q < \infty,$$

et

$$J_{p,\infty}^s(f) = \|\alpha_0\|_{l_p} + \sup_{j \geq 0} 2^{j(s+1/2-1/p)} \|\beta_j\|_{l_p} < \infty \quad \text{si } q = \infty.$$

De plus, la norme  $\|\cdot\|_{B_{p,q}^s}$  est équivalente à la norme  $J_{p,q}^s$ . On en déduit aisément les inclusions suivantes :

$$B_{p,q'}^{s'} \subset B_{p,q}^s, \quad \text{pour } s' > s \text{ ou pour } s' = s \text{ et } q' \leq q,$$

$$B_{p,q}^s \subset B_{p',q}^{s'}, \quad \text{pour } p' > p, \quad s' = s - (1/p - 1/p').$$

En particulier, pour  $s > 1/p$ ,  $q > 1$ ,  $B_{p,q}^s \subset B_{\infty,\infty}^{s'}$  est inclus dans l'espace des fonctions continues et bornées. On a aussi :

$$B_{p',\min(p',2)}^0 \subset L_{p'}, \quad p' \geq 1$$

où  $B_{p',\min(p',2)}^0$  est défini à travers la norme  $J_{p',q}^s$  avec  $s = 0$ .

### 1.4.3 Approche bayésienne non paramétrique

A la fin du paragraphe 1.2.2 consacré aux ondelettes-vaguelettes, nous avons énoncé les deux procédures de Donoho et Johnstone [43], [42] pour seuiller les estimateurs des coefficients de la fonction d'intérêt dans la base des ondelettes-vaguelettes. Ces procédures de seuillage doux et dur peuvent bien sûr aussi s'appliquer pour les estimateurs des coefficients de la fonction exprimés dans une base d'ondelettes classiques et ce dans des problèmes autres que les problèmes inverses comme celui de la régression non paramétrique. Si les bonnes performances en statistique non paramétrique tant sur le plan théorique que pratique des procédures qui combinent à la fois "décomposition dans une base d'ondelettes" et "seuillage" ne sont plus à démontrer, il est à propos de mettre en exergue le seuillage au travers de l'approche bayésienne non paramétrique.

Les procédures bayésiennes par ondelettes sont devenues populaires car fort efficaces d'un point de vue pratique. Elles se montrent d'ailleurs supérieures sur ce plan là aux estimateurs par seuillage doux ou dur, comme se sont appliqués à le montrer entre autres, Abramovich et al. [3], Abramovich et Sapatinas [4], et Antoniadis et al. [10].

L'approche bayésienne va consister à placer sur les coefficients de la fonction dans la base d'ondelettes un modèle a priori sensé exprimer la "sparsité" de la fonction dans ladite base.

Soit la décomposition de  $f$  dans une base d'ondelettes  $(\psi_{jk})_{j \geq -1, k \in \mathbb{Z}}$  :

$$f = \sum_{j \geq -1} \sum_k \beta_{jk} \psi_{jk}.$$

On met sur les coefficients  $\beta_{jk}$  le modèle a priori suivant :

$$\beta_{jk} = w_{jk} \gamma_{jk}^l + (1 - w_{jk}) \gamma_{jk}^s,$$

avec  $w_{jk}$  qui suit une loi de Bernoulli,  $w_{jk}$  représentant la proportion de coefficients non négligeables. La loi de  $\gamma_{jk}^l$  est général une loi normale et celle de  $\gamma_{jk}^s$  une loi normale de faible variance (voir Chipman et al. [32]) ou la plupart du temps la mesure de dirac en 0. Les  $\beta_{jk}$  sont supposés indépendants.

Chipman et al. [32], Clyde et al. [34] et Vidakovic [151] ont étudié l'estimateur de la moyenne a posteriori qui est un estimateur par contraction alors que Abramovich et al. [3] considèrent l'estimateur de la médiane a posteriori qui est un estimateur par seuillage. Ce dernier sera d'ailleurs mis en valeur dans le chapitre 4 sur une base particulière, la base des ondelettes déformées.



## 1.5 Principaux résultats

Dans cette partie, nous allons présenter les principaux résultats de cette thèse que l'on retrouve en intégralité ainsi que leur preuves détaillées dans les chapitres suivants.

### 1.5.1 Approche statistique minimax du problème des moments de Hausdorff

L'objet du chapitre 2 est l'étude du problème des moments de Hausdorff sous l'angle des problèmes inverses et dans un cadre statistique minimax. Le problème classique des moments est un problème qui apparaît dans de nombreux domaines mathématiques tels que les probabilités, l'analyse numérique, la théorie des opérateurs etc... Il s'agit de déterminer une mesure à partir de la connaissance de ses moments. Le problème des moments de Hausdorff traite le cas d'une mesure à support compact. L'approche par l'angle des problèmes inverses mal posés a été initiée par Ang et al. [9]. En effet si on définit l'opérateur des moments par :  $A : L_2(0, 1) \rightarrow l^2$

$$(Au)_k = \int_0^1 u(x)x^k dx \quad k = 0, 1, \dots,$$

alors Ang et al. [9] ont démontré que cet opérateur n'est pas continûment inversible.

Ce problème a été assez peu étudié en statistique. En effet, à première vue, il semble pas tout à fait évident d'observer dans la réalité une suite de moments de la densité que l'on cherche à estimer. En général les observations portent sur d'autres quantités. Et pourtant, dans certains domaines tels que la physique quantique ou la gravité quantique (voir Ash et Mac Donald [12], Gaspard et Cyrot-Lackman [61], Mead et Papanicolaou [110] pour quelques exemples), les moments de certains états quantiques sont mesurables et les physiciens s'interrogent sur la densité sous-jacente. Ceci constitue une première motivation à l'étude de ce problème. Une seconde motivation nous vient de l'étude des processus AR(1). Dans Leipus et al. [99], les auteurs s'intéressaient à des processus à longue mémoire obtenus en agrégeant des processus AR(1) indépendants. On rappelle qu'un processus AR(1) est défini comme suit

$$Y_{t-1}^{(j)} = a^{(j)}Y_{t-1}^{(j)} + \varepsilon_t^{(j)}.$$

Ils cherchaient à estimer la densité sous-jacente du processus  $a$  à support compact, à partir de l'observation des processus agrégés. Il s'avérait que

la covariance des processus agrégés n'était autre que les moments de la densité de  $a$ . Ils aboutissaient alors à un problème des moments. Les vitesses de convergence mises en lumière en perte  $L_2$  étaient logarithmiques donc très lentes. Les auteurs n'avaient pas prouvé de borne inférieure pour cette vitesse. Aussi, nous sommes-nous interrogés sur le type de vitesse que l'on peut obtenir dans un problème inverse statistique "classique" faisant intervenir des moments.

Nous considérons donc le modèle suivant. On observe une suite de moments bruités :

$$y_k = \mu_k + \varepsilon \xi_k \quad k = 0, 1, \dots \quad (1.18)$$

où  $\xi_k$  sont des variables gaussiennes i.i.d centrées et réduites et  $\varepsilon$  tend vers 0 et

$$\mu_k = \int_{-1}^1 x^k f(x) dx \quad k = 0, 1, \dots,$$

désignent les moments de la densité que l'on cherche estimer.

Nous construisons un estimateur  $\hat{f}_N$  par la méthode des séries orthogonales en projetant la densité sur l'espace des polynômes de Legendre (pour le détail de la construction de l'estimateur par projection voir Chapitre 2, section 2.2.2). On se place dans un cadre d'estimation minimax et sur une classe de régularité de Sobolev  $W_r^2$  (définie en (2.5)) avec décroissance polynomiale des coefficients de Fourier sur la base des polynômes de Legendre.

Nous obtenons les théorèmes suivants de borne supérieure et de borne inférieure qui montrent que notre estimateur  $\hat{f}_N$  converge à la vitesse optimale.

**Théorème 1.3.** *Soit  $\alpha > 0$ , on définit l'entier  $N = \lfloor \alpha \log(1/\varepsilon) \rfloor$ . On a alors*

$$\sup_{f \in \mathcal{W}_2^r} \mathbb{E}_f \|\hat{f}_N - f\|^2 \leq C [\log(1/\varepsilon)]^{-2r},$$

où  $C$  est une constante strictement positive.

**Théorème 1.4.** *On a*

$$\inf_{\hat{f}} \sup_{f \in \mathcal{W}_2^r} \mathbb{E}_f \|\hat{f} - f\|^2 \geq c [\log(1/\varepsilon)]^{-2r},$$

où  $c$  est une constante strictement positive.

Ces deux théorèmes montrent que la vitesse minimax dans le problème des moments de Hausdorff pour le modèle (1.18) est logarithmique et que l'estimateur  $\hat{f}_N$  est optimal au sens minimax. Ce dernier est par ailleurs

adaptatif.

La vitesse minimax étant très lente, la démonstration de la borne inférieure s'en trouve plus difficile. Celle-ci s'appuie sur l'application du lemme de Fano qui demande la spécification d'un sous-problème. Ce dernier amène à la construction d'un sous ensemble de densités suffisamment distantes les unes des autres en norme  $L_2$  et en même temps telles que les lois correspondantes du processus  $Y$  des observations soient suffisamment proches pour la distance de Kullback-Leibler.

Cette vitesse logarithmique peut également s'expliquer de façon qualitative. En effet, le problème des moments tel qu'il a été énoncé dans le modèle (1.18) fait partie des problèmes inverses sévèrement mal posés pour suivre la terminologie de classification des problèmes inverses définie dans la section 1.2.1 de cette introduction. Pour cela il faut réécrire le modèle (1.18) en le projetant sur la base des polynômes de Legendre. On obtient ainsi un modèle hétéroscédastique de bruit blanc gaussien (1.8) caractéristique des problèmes inverses (pour la réécriture du modèle, voir (2.4)). Ce modèle équivalent a d'ailleurs prévalu pour la démonstration de la borne inférieure. On montre alors que la suite des  $\sigma_k$  tend exponentiellement vers l'infini (voir le lemme 2.5), si bien que nous nous trouvons en présence d'un problème statistique inverse sévèrement mal posé.

### 1.5.2 Déconvolution localisée sur la sphère

Dans le chapitre 3 nous nous intéressons à un autre problème inverse, celui de la déconvolution sur la sphère. Comme nous l'avons précisé plus haut ce problème qui a été beaucoup étudié sur  $\mathbb{R}$  a été très peu abordé sur la sphère. Les quelques articles qui s'y sont attelés n'ont en tout cas pas fourni de performances pratiques des estimateurs proposés. On considère donc le modèle suivant de déconvolution comme l'ont fait Healy et al. [73] : on observe  $Z_1, \dots, Z_N$  des variables aléatoires i.i.d avec

$$Z_i = \varepsilon_i X_i, \quad (1.19)$$

où les  $\varepsilon_i$  sont des variables aléatoires i.i.d de  $SO(3)$ , les  $X_i$  et  $Z_i$  sont des variables aléatoires i.i.d de  $\mathbb{S}^2$ . Les distributions des variables aléatoires  $X_i$ ,  $Z_i$  et  $\varepsilon_i$  sont absolument continues par rapport à la mesure de Haar (mesure invariante) respectivement sur  $\mathbb{S}^2$ ,  $\mathbb{S}^2$  et  $SO(3)$  avec des densités  $f$ ,  $f_Z$  et  $f_\varepsilon$ . On a alors

$$f_Z = f_\varepsilon * f, \quad (1.20)$$

où  $*$  désigne le produit de convolution défini comme suit :

$$g * h(\omega) = \int_{SO(3)} g(u)h(u^{-1}\omega)du,$$

pour  $\omega \in \mathbb{S}^2$ .

A l'instar de ce qui est usité sur  $\mathbb{R}$ , afin de construire un estimateur, on passe dans le domaine de Fourier où l'on obtient une relation simple liant la transformée de Fourier de  $f_\varepsilon$  et  $f$ . Notre approche pour estimer  $f$  sera celle d'un estimateur par seuillage construit sur le frame des needlets sphériques.

### Les needlets sphériques

La construction des needlets sur la sphère s'apparente à celle sur  $\mathbb{R}$  et se divise en deux temps : une décomposition de Caldéron et une discrétisation des espaces de fréquence. Tous les détails de cette construction ainsi que les caractéristiques et propriétés des needlets se trouvent dans le chapitre 3. Les changements notables dans le cas sphérique se situent au niveau de la quadrature qui apparaît dans l'étape de discrétisation, nous avons besoin ici d'une quadrature de la sphère.

Une needlet sphérique est définie par :

$$\psi_{j\eta}(\omega) = \sqrt{\lambda_{j\eta}} \sum_{\ell \in \mathbb{N}} b(\ell/2^j) \sum_{m=-\ell}^{\ell} \bar{Y}_m^\ell(\omega) Y_m^\ell(\xi_{j\eta}),$$

où  $\omega \in \mathbb{S}^2$ ,  $j \geq 0$ ,  $\eta \in \mathcal{Z}_j$  avec  $\mathcal{Z}_j$  l'ensemble indexant la quadrature et  $\xi_{j\eta}$  sont les points de quadrature. On peut choisir  $\mathcal{Z}_j$  de façon à ce que son cardinal soit de l'ordre de  $C2^{2j}$  où  $C$  est une constante positive. La fonction  $b$  est une fonction très régulière à support sur  $[1/2, 2]$ .

Ces needlets sont très bien localisées autour des points de la quadrature. En effet pour tout  $k \in \mathbb{N}$  il existe une constante  $c_k$  telle que pour  $\xi \in \mathbb{S}^2$  :

$$|\psi_{j,\eta}(\xi)| \leq \frac{c_k 2^j}{(1 + 2^j d(\eta, \xi))^k}. \quad (1.21)$$

où  $d$  est la distance géodésique sur la sphère ( $d(\xi, \eta) = \arccos\langle \eta, \xi \rangle$ ). En d'autres mots, les needlets sont localisées de façon quasi exponentielle autour de leur point de quadrature associé, comme en témoignent les figures 1.1 en coordonnées sphériques. Notons que plus le niveau  $j$  augmente, meilleure se trouve la localisation autour des points de quadrature.

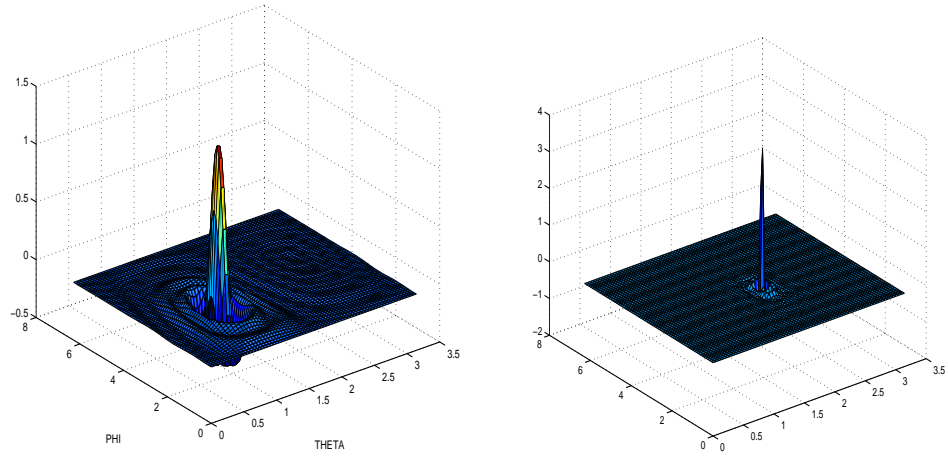


FIGURE 1.1 – Needlets pour  $j = 3$ ,  $\eta = 250$  et  $j = 5$ ,  $\eta = 5000$ .

Les harmoniques sphériques  $Y_m^l$  en revanche ne sont pas localisées.

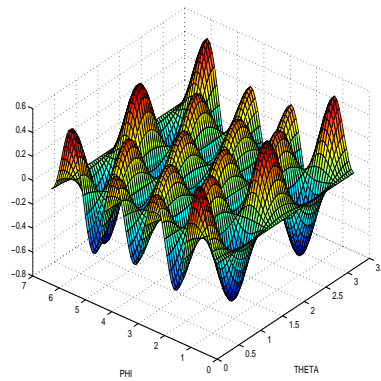


FIGURE 1.2 – Harmonique sphérique  $l = 8$   $m = 2$ .

Enfin, les needlets possèdent des propriétés de concentration qui facilitent l'étude du risque en norme  $L_p$  : pour tout  $1 \leq p < \infty$ , il existe des constantes  $c_p$ ,  $C_p$ ,  $c$ ,  $C$  and  $D_p$  telles que

$$\begin{aligned}
c_p 2^{2j(\frac{p}{2}-1)} &\leq \|\psi_{j\eta}\|_p^p \leq C_p 2^{2j(\frac{p}{2}-1)}, \\
c 2^j &\leq \|\psi_{j\eta}\|_\infty \leq C 2^j, \\
\left\| \sum_{\eta \in \mathcal{Z}_j} \lambda_\eta \psi_{j,\eta} \right\|_\pi &\leq c 2^{2j(\frac{1}{2}-\frac{1}{\pi})} \left( \sum_{\eta \in \mathcal{Z}_j} |\lambda_\eta|^\pi \right)^{1/\pi}, \\
\left\| \sum_{\eta \in \mathcal{Z}_j} \lambda_\eta \psi_{j\eta} \right\|_p^p &\leq D_p \sum_{\eta \in \mathcal{Z}_j} |\lambda_\eta|^p \|\psi_{j\eta}\|_p^p, \\
\left\| \sum_{\eta \in \mathcal{Z}_j} u_\eta \psi_{j\eta} \right\|_\infty &\leq C \sup_{\eta \in \mathcal{Z}_j} |u_\eta| 2^j.
\end{aligned}$$

### Procédure d'estimation

La procédure d'estimation par les needlets va combiner à la fois la simplicité d'inversion de la méthode SVD et une procédure de seuillage efficace. La décomposition de  $f$  sur le frame des needlets est la suivante :

$$f = \sum_j \sum_{\eta \in \mathcal{Z}_j} (f, \psi_{j\eta})_{\mathbb{L}_2(\mathbb{S}^2)} \psi_{j\eta} := \sum_j \sum_{\eta \in \mathcal{Z}_j} \beta_{jk} \psi_{j\eta}.$$

Par la relation de Parseval on a l'égalité

$$\beta_{j\eta} = \sum_{l,m} (f, Y_m^l) (\psi_{j\eta}, Y_m^l) := \sum_{l,m} \hat{f}_m^l \psi_{j\eta}^{lm},$$

que l'on estime par

$$\hat{\beta}_{j\eta} = \sum_{l,m} \hat{f}_m^{l,N} \psi_{j\eta}^{lm}.$$

Enfin, on applique une procédure de seuillage dur aux coefficients estimés  $\hat{\beta}_{j\eta}$

$$t(\hat{\beta}_{j\eta}) = \hat{\beta}_{j\eta} I\{|\hat{\beta}_{j\eta}| \geq \kappa \sqrt{\frac{\log N}{N}} |\sigma_j|\},$$

avec

$$\sigma_j^2 := M^2 \sum_{ln} \left| \sum_m \psi_{j\eta}^{lm} \hat{f}_{\varepsilon^{-1}mn}^l \right|^2,$$

où  $M$  tel que  $\|f\|_\infty \leq M$ .

Un point essentiel est le contrôle de la variance de l'estimateur  $\hat{\beta}_{j\eta}$ . Si la distribution de  $\varepsilon$  est régulière i.e  $\|\hat{f}_\varepsilon^l\|_{op} \leq d_0 l^{-\nu}$ , nous imposons la condition fondamentale suivante sur les coefficients de Fourier des needlets :

$$\sigma_j^2 \leq C 2^{2j\nu},$$

avec  $C$  une constante strictement positive.

Les needlets que nous construisons satisfont bien cette condition du fait du support de la fonction  $b$ .

### Performances de la procédure d'estimation

On suppose que  $f$  appartient à la boule de Besov sur la sphère de rayon  $M$ ,  $B_{\pi,r}^s(M)$ . L'espace  $B_{\pi,r}^s$  est une généralisation à  $\mathbb{S}^2$  des espaces de Besov.

On a les vitesses de convergence pour notre estimateur  $\hat{f}$  :

**Théorème 1.5.** *Soit  $1 \leq p < \infty$ ,  $\nu > 0$ , et supposons que*

$$\sigma_j^2 := M^2 \sum_{ln} \left| \sum_m \psi_{jn}^{lm} f_{\varepsilon^{-1}mn}^l \right|^2 \leq C 2^{2j\nu}, \quad \forall j \geq 0. \quad (1.22)$$

*Maintenant soit  $\kappa^2 \geq 16p$  et  $2^J = d[t_N]^{\frac{-1}{(\nu+1)}}$  avec  $t_N = \sqrt{\frac{\log N}{N}}$  et  $d$  une constante positive. Alors si  $\pi \geq 1$ ,  $s \geq 2/\pi$ ,  $r \geq 1$  (avec la restriction  $r \leq \pi$  si  $s = (\nu + 1)(\frac{p}{\pi} - 1)$ ), alors il existe une constante  $C$  telle que :*

$$\sup_{f \in B_{\pi,r}^s(M)} \mathbb{E} \|\hat{f} - f\|_p^p \leq C (\log(N))^{p-1} [N^{-1/2} \sqrt{\log(N)}]^{\mu p}, \quad (1.23)$$

où

$$\begin{aligned} \mu &= \frac{s}{s + \nu + 1}, \quad \text{si } s \geq (\nu + 1)\left(\frac{p}{\pi} - 1\right) \\ \mu &= \frac{s - 2/\pi + 2/p}{s + \nu - 2/\pi + 1}, \quad \text{si } \frac{2}{\pi} \leq s < (\nu + 1)\left(\frac{p}{\pi} - 1\right). \end{aligned}$$

### Performances pratiques de la procédure d'estimation

Nous donnons à présent les résultats des performances numériques de cette procédure. Nous cherchons à estimer deux sortes de densités, l'uniforme sur la sphère et une densité unimodale qui sont deux types de densités importantes dans le domaine de l'astrophysique. Le bruit va correspondre à une rotation selon l'axe Oz des observations avec pour chaque observation un angle  $\Phi$  qui suit une loi uniforme. Plusieurs lois uniformes avec différents supports sont testées. Plus le support de la loi uniforme sera grand et plus le niveau de bruit sera important.

Pour la densité uniforme sur la sphère,  $f = 1/(4\pi)$ . Un calcul très simple montre que  $(f, \psi_{j,\eta}) = 0$  pour tous les  $j$  et les  $\eta$ . De fait une façon de mesurer

les performances de l'estimateur dans ce cas est de compter le nombre de coefficients qui ne sont pas mis à 0.

	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$\kappa = 0.2$	2	1	75	337
$\kappa = 0.7$	0	0	2	8
$\kappa = 0.8$	0	0	0	2

TABLE 1.1 – nombre de coefficients non nuls après seuillage pour différentes valeurs de  $\kappa$ ,  $\phi \sim U[0, \pi/8]$ .

	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$\kappa = 0.1$	4	35	135	556
$\kappa = 0.4$	0	4	24	103
$\kappa = 0.7$	0	0	1	5
$\kappa = 0.8$	0	0	0	3

TABLE 1.2 – nombre de coefficients non nuls après seuillage pour différentes valeurs de  $\kappa$ ,  $\phi \sim U[0, \pi]$ .

Pour la densité unimodale,  $f(\omega) = ce^{-4|\omega-\omega_1|^2}$  avec  $\omega_1 = (0, 1, 0)$  et  $c = 1/0.7854$ . Les observations non bruitées sont concentrées autour du point  $(0, 1, 0)$ . On présente pour différentes perturbations les données bruitées sur la sphère ainsi que sur la sphère aplatie pour une meilleure visibilité et les reconstructions graphiques de la densité estimée en coordonnées sphériques.

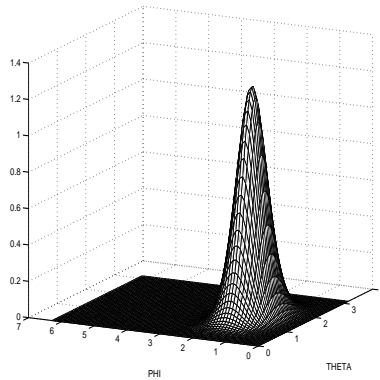


FIGURE 1.3 – Densité cible.



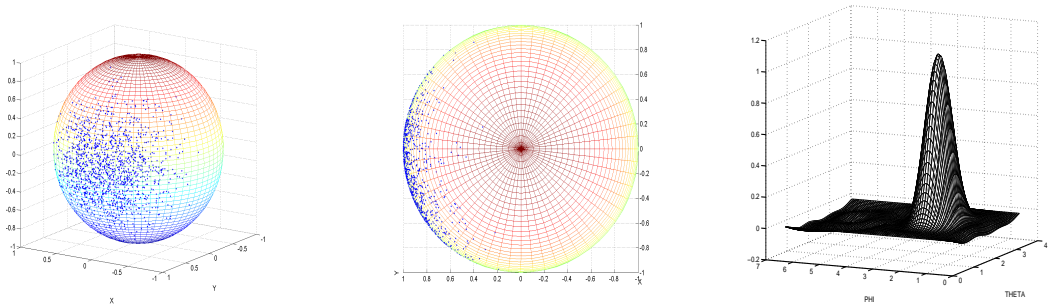


FIGURE 1.4 – Observations, reconstruction de la densité estimée,  $\phi \sim U[0, \pi/8]$ .

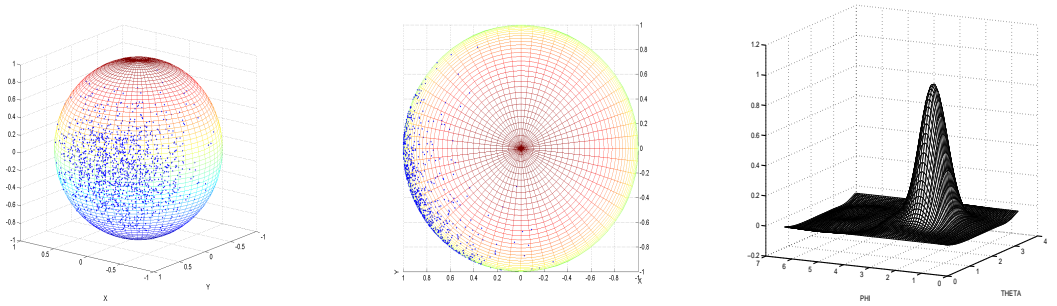


FIGURE 1.5 – Observations, reconstruction de la densité estimée,  $\phi \sim U[0, \pi/4]$ .

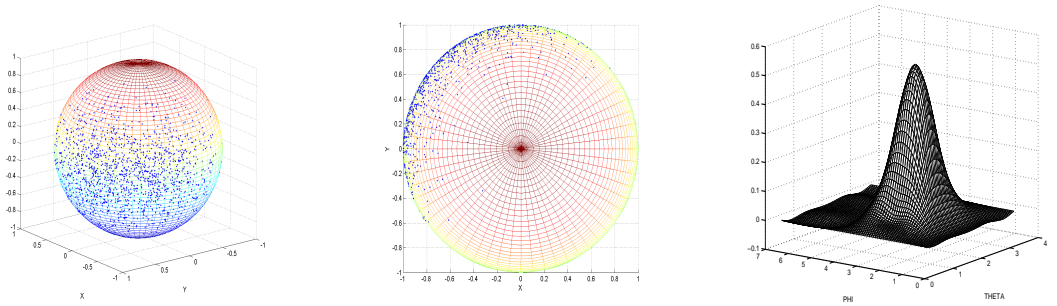


FIGURE 1.6 – Observations, reconstruction de la densité estimée,  $\phi \sim U[0, \pi/2]$ .

On remarque que la position du pic quelque soit l'intensité du bruit est bien localisée et que la forme de la densité est conservée. Il y a un léger

élargissement de la base, cela est dû aux observations, qui du fait de la perturbation, vont être plus dispersées sur la sphère comme le montrent les différents graphiques de cette dernière.

### 1.5.3 Régression en design aléatoire et estimateurs bayésiens à ondelettes déformées

Dans le chapitre 4, nous nous intéressons au problème de la régression en design aléatoire à l'aide d'une procédure bayésienne basée sur des ondelettes déformées. Bien que nous ne soyons plus dans un cadre de problème inverse et traitons d'un problème statistique différent, ce chapitre marque une continuité avec le précédent sur la déconvolution sur la sphère à l'aide des needlets puisqu'il fait la part belle à des estimateurs bayésiens exprimés dans une base particulière et intéressante qui permet une expression simple et naturel du problème statistique en question. Comme pour les needlets où nous perdions quelques propriétés des bases classiques usuelles au profit d'une grande flexibilité et d'une simplicité dans la procédure d'estimation, la base mise en exergue dans le chapitre 4 n'est certes plus orthogonale mais permet également une grande simplicité et efficacité quant à l'estimation.

On considère le modèle de régression en design aléatoire. Celui-ci s'exprime de la façon suivante : on observe des couples  $(X_i, Y_i)$ , pour  $i = 1, \dots, n$  donnés par :

$$Y_i = f(X_i) + \varepsilon_i, \quad 1 \leq i \leq n, \quad (1.24)$$

où  $f$  est la fonction à estimer,  $\varepsilon_i$  sont des erreurs i.i.d gaussiennes centrées et de variance égale à  $\sigma^2$ . Les points  $X_i$  du design sont des variables i.i.d à support sur  $[0, 1]$  de densité  $g$ . Nous supposons que  $g$  est bornée inférieurement.

On s'applique à construire un estimateur qui combine à la fois les ondelettes déformées et les techniques bayésiennes. Celles-ci fort populaires, ont été la plupart du temps utilisées dans un cadre de régression en design fixe  $X_i = \frac{i}{n}$ , voir par exemple les travaux de Abramovich et al. [2], [3]. Explicitons maintenant l'estimateur bayésien à ondelettes déformées.

#### L'estimateur bayésien à ondelettes déformées.

Les ondelettes déformées ont été étudiées de façon très approfondie par Kerkycharian et Picard [87] qui ont développé une procédure pour estimer  $f$  dans le modèle (1.24) qui utilisait son développement dans ladite base et seuillait les coefficients estimés. Cette base est obtenue en déformant les

bases d'ondelettes usuelles par la fonction de répartition  $G(x) = \int_0^x g(t)dt$ . Le développement de  $f$  dans cette base qui n'est plus orthogonale est

$$f(x) = \sum_{J \geq -1} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}(G(x)).$$

On estime les coefficients  $\beta_{jk}$  par

$$\hat{\beta}_{jk} = (1/n) \sum_{i=1}^n \psi_{jk}(G(X_i)) Y_i.$$

Nous plaçons sur les coefficients  $\beta_{jk}$  un modèle a priori comme l'ont fait précédemment Abramovich et al. [2], [3] et Rivoirard [127].

$$\beta_{jk} \sim \pi_j N(0, \tau_j^2) + (1 - \pi_j) \delta(0),$$

où  $0 \leq \pi_j \leq 1$  et  $\delta(0)$  représente la masse de dirac en 0. Le paramètre  $\pi_j$  représente la proportion de coefficients non négligeables de la fonction  $f$ .

On s'intéresse à l'estimateur de la médiane a posteriori :

$$\tilde{f}(x) = \sum_{j \leq J} \sum_{k=0}^{2^j-1} \tilde{\beta}_{jk} \psi_{jk}(G(x)), \quad \text{avec } \tilde{\beta}_{jk} = \text{Med}(\beta_{jk} | \hat{\beta}_{jk}).$$

L'estimateur de la médiane est de type seuillage comme l'ont montré Abramovich et al. [2], [3], c'est à dire qu'il s'annule dès que  $\hat{\beta}_{jk}$  est inférieure à un certain seuil. Nous envisageons deux types de scénarios pour les paramètres  $\pi_j$  et  $\tau_j$ .

Dans un premier temps, nous considérons comme l'ont fait Abramovich et al. [2], [3] les valeurs suivantes pour les hyperparamètres :

$$\tau_j^2 = c_1 2^{-j\alpha} \quad \pi_j = \min(1, c_2 2^{-j\beta}), \quad (\text{MS})$$

où  $c_1, c_2, \alpha$  et  $\beta$  sont des constantes positives.

Dans un second temps, comme l'ont suggéré Autin et al. [14], nous nous attacherons à des paramètres qui ne dépendent que du nombre d'observations  $n$  et qui se présentent sous la forme :

$$\tau_j^2 := \tau(n)^2 = 1/\sqrt{n \log n} \quad n^{-q_1/2} \leq w(n) \leq n^{-q_2/2}. \quad (\text{ML})$$

On étudie les performances théoriques dans un cadre minimax ainsi que les performances pratiques de cet estimateur dans les deux scénarios qui viennent d'être évoqués.

### Performances théoriques.

Nous nous plaçons sur les boules de Besov de rayon  $R$ ,  $B_{2,\infty}^s(R)$  et obtenons les vitesses de convergence suivantes :

**Théorème 1.6.** *On observe le modèle (1.24) et on considère les paramètres de la loi a priori (MS). On pose  $J := J_\alpha$  de façon à ce que  $2^{J_\alpha} = (3/(2n))^{-1/\alpha}$  avec  $\alpha > 1$  et  $\alpha > s$ . Alors on a :*

$$\sup_{f(G^{-1}) \in B_{2,\infty}^s(R)} \mathbb{E} \|\hat{f} - f\|_2^2 = O((1/n)^{1-1/\alpha} \log^2(n)) + O((1/n)^{2s/\alpha}).$$

**Théorème 1.7.** *On observe le modèle (1.24) et on considère les paramètres de la loi a priori (ML). On pose  $J := J_n$  de façon à ce que  $2^{J_n} = n/\log(n)$ . Soit  $\alpha > 1$  et  $\alpha > s$ , alors on a :*

$$\sup_{f(G^{-1}) \in B_{2,\infty}^s(R)} \mathbb{E} \|\hat{f} - f\|_2^2 = O\left(\left(\frac{\log n}{n}\right)^{2s/(2s+1)}\right).$$

Les résultats sont exprimés pour des régularités qui concernent la fonction  $f(G^{-1})$ . On remarque que pour le choix spécifique  $\alpha = 2s + 1$  dans la cas des choix de paramètres (MS), on obtient les vitesses de convergence minimax obtenues dans le cas d'un design uniforme. La vitesse obtenue pour les choix (ML) est la vitesse minimax pour la régression en design uniforme. Notons aussi que l'estimateur bayésien construit avec les paramètres (ML) est adaptatif.

### Performances pratiques.

On présente les performances des procédures bayésiennes basées sur les choix de paramètres (ML) et (MS) ainsi que celles de l'estimateur de seuillage dur basé sur les ondelettes déformées proposé par Kerkycharian et Picard [87]. L'estimateur de Kerkycharian et Picard [87] a la forme suivante :

$$f^\dagger(x) = \sum_{j=-1}^{J(n)} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} I(|\hat{\beta}_{j,k}| \geq \kappa \sigma \sqrt{2 \log(n)/n}) \psi_{jk}(G(x)),$$

avec  $J(n)$  tel que  $2^{J(n)}$  soit de l'ordre de  $\sqrt{n/\log(n)}$ .

Les estimateurs bayésiens construits à partir des hyperparamètres (ML) et (MS) sont notés respectivement E1 et E2, l'estimateur de Kerkycharian et Picard sera noté E3. Nous testons ces procédures pour les quatre fonctions

"test" de Donoho et Johnstone, *Blocks*, *Bumps*, *HeaviSine*, *Doppler* et pour chacune d'entre elles pour deux types de densité de design, *Sine* et *Hole2*.

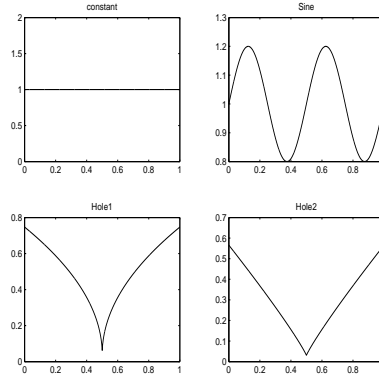


FIGURE 1.7 – Densité du design.

		RSNR=4			RSNR=7		
design density		E1	E2	E3	E1	E2	E3
<i>Blocks</i>	Sine	0.0194	0.0219	0.0227	0.0113	0.0161	0.0129
	Hole2	0.0196	0.0220	0.0226	0.0114	0.0163	0.0130
<i>Bumps</i>	Sine	0.0243	0.240	0.259	0.0156	0.0167	0.0172
	Hole2	0.0241	0.0237	0.0253	0.0155	0.0167	0.0169
<i>HeaviSine</i>	Sine	0.0164	0.0141	0.0133	0.0103	0.0092	0.0093
	Hole2	0.0169	0.0146	0.0138	0.0107	0.0097	0.0096
<i>Doppler</i>	Sine	0.0236	0.0231	0.0236	0.0157	0.0238	0.0248
	Hole2	0.0244	0.0238	0.0248	0.0166	0.0172	0.0176

TABLE 1.3 – RMSE pour les estimateurs E1, E2 et E3, pour différentes fonctions test, différentes densités de design et différents niveaux de bruit.

Nos choix de paramètres a priori ou de réglages ont été pour les procédures :

- E1 (MS) :  $c_1 = 1$ ,  $c_2 = 2$ ,  $\alpha = 0.5$ ,  $\beta = 1$ .
- E2 (ML) :  $q_1 = q_2 = q = 1$ ,  $w(n) = 20n^{-q/2}$ ,  $\tau(n) = 20\sigma^2/(n \log(n))$ .

- E3 (Kerkyacharian Picard) :  $\kappa = 1$ .

Les résultats figurant dans le tableau 1.3 indiquent que l'on obtient de meilleurs résultats pour les procédures bayésiennes, en particulier pour des signaux irréguliers tels que Bumps, Blocks et Doppler. Plus précisément, c'est la procédure E1 qui se montre la plus performante car ses RMSE sont les plus faibles dans huit cas sur seize, on rappelle qu'elle est par ailleurs adaptative. Ces meilleurs résultats pour les procédures bayésiennes peuvent s'expliquer en partie par le fait que les estimateurs bayésiens appliquent un seuil local et ainsi se montrent plus à même de déceler les irrégularités des signaux que les estimateurs à seuillage universel.

## 1.6 Perspectives

Le chapitre portant sur la déconvolution sur la sphère a mis en jeu un estimateur basé sur le *frame* des needlets et une procédure de seuillage dur. Le choix du seuil occupe bien évidemment une place fondamentale tant sur le plan théorique que pratique dans toute procédure de seuillage. En effet, les simulations ont fait apparaître que ce choix n'était pas des plus aisés et qu'il fallait éprouver plusieurs valeurs du paramètre  $\kappa$  appelé dans la littérature "tuning parameter", avant d'obtenir de bonnes reconstructions de la densité en question. Aussi, l'étude théorique d'une calibration du seuil nous semble-t-elle un axe de recherche possible et nécessaire qui s'inscrit dans la ligne droite de ce thème de la déconvolution et qui pourrait en pratique donner de bons résultats.

Les simulations menées dans le chapitre sur la déconvolution sphérique sont une amorce dans le domaine qui jusqu'à présent a été très peu étudié voire pas en pratique. Ces premiers résultats se sont avérés concluants et nous invitent vivement à poursuivre d'autres études numériques qui cette fois mettraient en jeu des lois plus complexes pour le processus de perturbations  $\varepsilon$  et d'autres fonctions cibles plus irrégulières. Des lois possibles pour le processus  $\varepsilon$  seraient celles de Rosenthal venant des marches aléatoires sur les groupes, ou celle de Laplace Rotationnelle évoquées dans Healy et al. [73] et Kim et Koo [91], ou encore le mouvement brownien sur la sphère. Un champ d'application où les perspectives sont fort nombreuses se trouve en cosmologie avec l'étude statistique de champs aléatoires sphériques, dont le Fond diffus cosmologique (CMB) est un exemple. Il serait donc possible de tester notre procédure sur des données réelles. Ce domaine de l'astrophysique a d'ailleurs été abordé sous l'angle également des needlets par Guilloux et al. [67], [38], [58].

Toujours sur ce thème de la déconvolution sphérique, notre étude a mis en lumière deux types de bases, les needlets et celle des harmoniques sphériques. Au lieu de considérer comme ce qui a été entrepris dans cette thèse une décomposition sur l'unique *frame* des needlets, nous pourrions considérer une décomposition du signal sur un dictionnaire de bases constitué de needlets et des harmoniques sphériques. Une telle décomposition n'est pas envisageable en vue de la construction d'un estimateur par seuillage qui est un estimateur par pénalisation au sens  $\ell_0$ , aussi une alternative réside-t-elle dans les estimateurs par pénalisation au sens  $\ell_1$  (Dantzig, LASSO) que nous pourrions considérer.

Le chapitre 4 mettait en exergue une procédure bayésienne construite à partir des ondelettes déformées pour traiter le problème de la régression en random design. La loi de mélange a priori mise sur les coefficients de la fonction de régression sur la base des ondelettes déformées faisait intervenir des hyperparamètres, celui de la variance de la loi gaussienne et celui de proportion des coefficients non négligeables de la fonction. Comme l'attestent les simulations, ces hyperparamètres ont été calibrés en pratique à la main, reflétant une fois encore la difficulté du calibrage qui des hyperparamètres, qui des seuils. Aussi, une voie intéressante à étudier serait de mener cette même étude sous le prisme du bayésien empirique. En effet, Johnstone et Silverman [77], [78] ont privilégié cette approche en plaçant sur les coefficients d'ondelettes de la fonction de régression en design uniforme une loi a priori dont les hyperparamètres sont calculés à partir des observations. Au demeurant, il est connu qu'en pratique le bayésien empirique donne d'excellents résultats numériques et se montre plus performant que les procédures de seuillage comme nous l'avons souligné dans le paragraphe 1.4.3. Nous pourrions d'ailleurs mettre à profit une approche bayésienne pour le seuillage dans le problème de la déconvolution sphérique.

Le chapitre 2 de cette thèse fait l'objet d'une publication, voir Pham Ngoc [123]. Le chapitre 4 va apparaître dans la revue *Electronic Journal of Statistics*, voir Pham Ngoc [124]. Enfin, le chapitre 3 écrit en collaboration avec G. Kerkycharian et D. Picard fait l'objet d'un article soumis dans une revue à comité de lecture.

## Chapitre 2

# Statistical minimax approach to the Hausdorff moment problem<sup>1</sup>

The purpose of this chapter is to study the problem of estimating a compactly supported function from noisy observations of its moments. We provide a statistical approach to the famous Hausdorff classical moment problem. We prove a lower bound on the rate of convergence of the mean integrated squared error and provide an estimator which attains minimax rate over the corresponding smoothness classes.

### 2.1 Introduction

The classical moment problem can be stated as follows : it consists in getting some information about a distribution  $\mu$  from the knowledge of its moments  $\int x^k d\mu(x)$ . This problem has been largely investigated in many mathematical topics, among others, in operator theory, mathematical physics, inverse spectral theory, probability theory, inverse problems, numerical analysis but also in a wide variety of settings in physics and engineering such as quantum spin models, speech processing, radio astronomy (see for instance Kay and Marple [84], Lang and McClellan [98] ...). We may cite the classical and pioneer books in the field (see Akhiezer [5], Shohat and Tamarkin [130]) which put emphasis on the existence aspect of the solution and its uniqueness. According to the support of the distribution of interest,

---

1. The bulk of this chapter has appeared in *Inverse Problems*, see Pham Ngoc [123].



one may refer to one of the three types of classical moment problems : the Hamburger moment problem whose support of  $\mu$  is the whole real line, the Stieljes problem on  $[0, +\infty)$  and finally the Hausdorff problem on a bounded interval. In this chapter, we shall focus on the last issue under the inverse problem angle.

The Hausdorff moment problem which dates back to 1921 (see Hausdorff [72]) occupies a central place in the field of inverse problems and has been an object of great interest in literature since then. For instance, the particular case when only a finite number of moments are known which is called the Truncated Hausdorff moment problem, has recently aroused much attention (see Talenti [138], Fasino and Inglese [57], Inglese [74], Tagliani [135]). Another interest and aspect of the Hausdorff moment problem lies in its very close link to the inversion of the Laplace transform when this latter is given at equidistant points ( see for instance Brianzi and Frontini [25], Dung, Huy, Quan and Trong [49], Tagliani [136], Tagliani and Velazquez [137]). In fact, by a simple change of variable, the problem of Laplace transform inversion is equivalent to the Hausdorff moment problem. More recently, Ang, Gorenflo, Le and Trong [9] have presented the Hausdorff moment problem under the angle of ill posed problems, in a sense that solutions do not depend continuously on the given data. Nonetheless, until now, as far as we know, the statistical approach which consists in assuming that the noise is stochastic has been very little put forward and rarely raised.

We consider in this chapter a statistical point of view of the Hausdorff moment problem. We aim at estimating an unknown probability density from noisy measurements of its moments on a symmetric bounded interval. In reality, it is barely impossible to measure moments with having any corruption. Without loss of generality we may and will suppose that  $[-a, a] = [-1, 1]$ . In practise, observations of moments appear especially in engineering mechanics (see Athanassoulis and Gavriliadis [13]) and all the references therein), in quantum physics (see for instance Ash and Mc Donald [12] and Mead and Papanicolaou [110]) and image analysis (see Teague [141]). We shall give now some concrete examples where we can have access to moments.

Examples. 1. In quantum physics, a big issue aims at reconstructing a positive density of states in harmonic solids from its moments on a finite interval. It is explained how to measure these moments (p 2410 and 2411 in Mead and Papanicolaou [110]).

2. In the field of quantum gravity, the transition probabilities for a Markov chain related to the causal set approach to modeling discrete theories of quantum gravity satisfy a moment problem (see Ash and McDonald [12]). One have to measure those probabilities.

3. In the context of non classical moment problem, we may cite the following example (see Ang et al. [8]). It deals with the determination of the shape of an object in the interior of the Earth by gravimetric methods. The density of that object differs from the density of the surrounding medium. Assuming a flat earth model, the problem consists in finding a curve  $x \rightarrow \sigma(x)$  in the half plane  $0 \leq \sigma(x) < H, 0 \leq x \leq 1$ ,  $\sigma(x)$  satisfying a non linear integral equation of the first kind of the form

$$\frac{1}{2\pi} \int_0^1 \frac{H - \sigma(\xi)}{(x - \xi)^2 + (H - \sigma(\xi))^2} d\xi = f(x),$$

where  $f(x)$  is a given function. The nonlinear integral equation can be approximated by the following linear integral equation in  $\varphi$  :

$$\int_0^1 \frac{\varphi(\xi)}{(M + x + \xi)^2} d\xi = 2\pi f(-M - x),$$

with  $\varphi(x) = H - \sigma(x)$ ,  $M$  is large enough and  $x \geq 0$ . By taking  $x = 1, 2, \dots, n, \dots$ , we get the following equivalent moment problem :

$$\int_0^1 \frac{\varphi(\xi)}{(M + n + \xi)^2} d\xi = \mu_n,$$

where  $\mu_n = 2\pi f(-M - n)$ ,  $n = 1, 2, \dots$

In the first examples cited above, the authors aim at estimating an unknown density from measurements of its moments. The following chapter presents the case of the reconstruction of an unknown function which could be in particular a density of probability, when one has its moments corrupted with some white noise. This approach constitutes one among other ill-posed inverse problems point of view.

Estimation in statistics using moments has already been put forward (see Mnatsakanov and Ryumgaart [112]), but the approach there was based on empirical moments and empirical processes, the results were expressed in terms of weak convergence whereas our chapter is built upon an ill-posed inverse approach with minimax results.

The estimation procedure we use is based on the expansion of the unknown function through the basis of Legendre polynomials and an orthogonal series method. We establish an upper bound and a lower bound on the estimation accuracy of the procedure showing that it is optimal in a minimax sense. We show that the achieved rate is only of logarithmic order. This fact has

already been underlined by Goldenshluger and Spokoiny [62]. In their paper, Goldenshluger and Spokoiny [62] tackled the problem of reconstructing a planar convex set from noisy geometric moments observations. They pointed out that in view of reconstructing a planar region from noisy measurements of moments, the upper bound was only in the order of logarithmic rate. The lower bound has not been proved. In a second part, they consider reconstruction from Legendre moments to get faster rates of convergence. Legendre moments can be observed in the context of shapes reconstruction. In our present work, instead of considering a planar region, we deal with functions belonging to a Sobolev scale and we stay focused on the classic moments with respect to the monomials  $x^k$ . Moreover, recently in the context of long-memory processes obtained by aggregation of independent parameter AR(1) processes and in view of estimating the density of the underlying random parameter, Leipus, Oppenheim, Philippe and Viano [99] had to deal with a problem of moments. They obtained very slow logarithmic rate but without showing that this could be the best possible. In a certain way, our minimax results provide a piece of answer.

One might question this chronic slow rate which seems inherent to moment problems. In fact, the underlying problem lies in the non orthogonal nature of the monomials  $x^k$ . They actually hamper the convergence rate to be improved for bringing a small amount of information. This remark is highlighted in our proof of the upper bound.

This chapter is organized as follows : in section 2 we introduce the model and the estimator of the unknown function and we finally state the two theorems. Section 3 contains the proofs. The last section is an appendix in which we prove some useful inequalities about binomial coefficients.

## 2.2 Statement of the problem

### 2.2.1 The model.

First of all, let us recall the usual statistical framework of ill-posed inverse problems (see Mathe and Pereverzev [108]) :

Let  $A : H \rightarrow H$  be a known linear operator on a Hilbert space  $H$ . The problem is to estimate an unknown function  $f \in H$  from indirect observations

$$Y = Af + \varepsilon\xi, \tag{2.1}$$

where  $\varepsilon$  is the amplitude of the noise. It is supposed to be a small positive parameter which tends to 0,  $\xi$  is assumed to be a zero-mean Gaussian random process indexed by  $H$  on a probability space  $(\Omega, \mathcal{A}, \mathbf{P})$ .

In the Hausdorff moment problem, the operator  $A_k$  which determines the way the observations are indirect is defined by :

$$A_k(f) = \int_{-1}^1 x^k f(x) dx, \quad k \in \mathbb{N}. \quad (2.2)$$

Let us now state the Hausdorff moment problem model. From the equation (4.1) and (2.2), we derive the following sequence of moments observations pertubated by a stochastic noise :

$$y_k = \mu_k + \varepsilon \xi_k \quad k = 0, 1, \dots \quad (2.3)$$

where  $\varepsilon$  is the noise level, it is supposed to tend to 0,  $\xi_k$  are assumed to be i.i.d standard Gaussian random variables and  $\mu_k$  are the moments of the unknown probability density  $f$  given by :

$$\mu_k = \int_{-1}^1 x^k f(x) dx \quad k = 0, 1, \dots$$

The assumption that the noise is modelled as i.i.d standard Gaussian random variables is a natural modeling and has already been used (see Rodriguez and Seatzu [128], Brianzi and Frontini ([25], Goldenshluger and Spokoiny [62]). It is a standard assumption in statistical inverse problems. Moreover, we may consider (2.3) as a first approach of the problem at hand.

The objective is to estimate the unknown function  $f$  supported on the bounded interval  $[-1, 1]$  in the model (2.3). The solution of this problem is unique in the case  $k = \infty$  (see Bertero, De Mol and Pike [23]), so that the statistical problem of recovering  $f$  from (2.3) is relevant. In the particular case of estimating a density of probability, the measure having density  $f$  is unique (see Feller [59], Chap.7)).

The use of the Legendre polynomials in the Hausdorff classical moment problem in order to approximate the unknown measure is a completely standard and quite natural procedure (see Ang, Gorenflo, Le and Trong [9], Bertero, De Mol and Pike [23], Papoulis [117], Teague [141], Goldenshluger and Spokoiny [62] as those polynomials directly result from the Gram-Schmidt orthonormalization of the family  $\{x^k\}$ ,  $k = 0, 1, \dots$ . Consequently, expanding the density  $f$  to be estimated in the basis of Legendre polynomials falls naturally and fits the problem's nature.

Denote  $\beta_{n,j}$  the coefficients of the normalized Legendre polynomial of degree  $n$  :

$$P_n(x) = \sum_{j=0}^n \beta_{n,j} x^j$$

By considering the Legendre polynomials, if we multiply both sides of the model (2.3) by the coefficients  $\beta_{k,j}$  we get the following model (for the proof see Lemma 2.3. in Appendix) :

$$\tilde{y}_k = \theta_k + \varepsilon \sigma_k \xi_k \tag{2.4}$$

where  $\sigma_k^2 = \sum_{j=0}^k \beta_{k,j}^2$ ,  $\tilde{y}_k = \sum_{j=0}^k \beta_{k,j} y_j$ ,  $\theta_k = \sum_{j=0}^k \beta_{k,j} \mu_j = \int_{-1}^1 f(x) P_k(x) dx$  and  $\xi_k$  are i.i.d standard Gaussian random variables. So the model (2.3) is equivalent to the model (2.4). The model (2.4) will be used for the proof of the lower bound.

Before going any further, we can make a remark at this stage concerning the model (2.4) which is an heteroscedastic gaussian sequence space model. Depending on the asymptotic behavior of the intensity noise  $\sigma_n^2$  one may characterize the nature of the problem's ill-posedness (see Cavalier, Golubev, Lepski and Tsybakov [30]). Here, in our case,  $\sigma_n^2 \geq \frac{1}{4} 4^n$  (see Lemma 2.5. from Appendix) and hence tends to infinity exponentially. We may say that we are dealing with a severely ill-posed problem with log-rates .

We assume that  $f$  belongs to the Sobolev space  $W_2^r$  defined by :

$$W_2^r = \{f \in L^2[-1, 1] : \sum_k k^{2r} |\theta_k|^2 < \infty\} \tag{2.5}$$

where  $\theta_k = \int_{-1}^1 f(x) P_k(x) dx$  is the Legendre Fourier coefficient and  $P_k$  denotes the normalized Legendre polynomial of degree  $k$ . Note that we consider more general functions than densities of probability which are included in this smoothness class.

Sobolev spaces associated with various kinds of underlying orthonormal basis constitute quite standard smoothness assumption classes in classical ill-posed problems (see for instance Mair and Ruymgaart [103], Mathe and Pereverzev [109], Goldenshluger and Pereverzev [64]). In the Hausdorff moment problem, the underlying basis is the Legendre polynomials.

Let us now give some highlights of the Sobolev space  $W_2^r$  regarding Legendre polynomials. Rafal'son [126] and Tomin [144] have shown in the more general case of Jacobi polynomials (and thus in the particular case of Legendre polynomials we are considering here) that Sobolev space  $W_2^r$  consists of all

functions  $f$  which have their derivatives  $f', f'', \dots, f^{(r-1)}$  being absolutely continuous on each interval  $[a, b] \subset (-1, 1)$  (see also Mathe and Pereverzev [109]).

### 2.2.2 The estimation procedure

Let us define now the estimator of  $f$ . This latter is induced by an orthogonal series method through the Legendre polynomials. Any function in  $L^2[-1, 1]$  has an expansion :

$$f(x) = \sum_{k=0}^{\infty} \theta_k P_k(x) \quad \text{with} \quad \theta_k = \int_{-1}^1 f(x) P_k(x) dx.$$

The problem of estimating  $f$  reduces to estimation of the sequence  $\{\theta_k\}_{k=1}^{+\infty}$  for Legendre polynomials form a complete orthogonal function system in  $L^2[-1, 1]$ .

We have

$$\theta_k = \int_{-1}^1 f(x) P_k(x) dx = \sum_{j=0}^k \beta_{k,j} \int_{-1}^1 f(x) x^j dx = \sum_{j=0}^k \beta_{k,j} \mu_j.$$

This leads us to consider the following estimator of  $\theta$  :

$$\hat{\theta}_k = \sum_{j=0}^k \beta_{k,j} y_j$$

and hence the estimator  $\hat{f}_N$  of  $f$  :

$$\hat{f}_N(x) = \sum_{k=0}^N \hat{\theta}_k P_k(x) = \sum_{k=0}^N \sum_{j=0}^k \beta_{k,j} y_j P_k(x)$$

where  $y_j$  is given by (2.3) and  $N$  is an integer to be properly selected later.

The mean integrated square error of the estimator  $\hat{f}_N$  is :

$$\mathbb{E}_f \|\hat{f}_N - f\|^2,$$

where  $\mathbb{E}_f$  denotes the expectation w.r.t the distribution of the data in the model (2.3) and for a function  $g \in L^2[-1, 1]$ ,

$$\|g\| = \left( \int_{-1}^1 g^2(x) dx \right)^{1/2}.$$

In this chapter we shall consider the problem of estimating  $f$  using the mean integrated square risk in the model (2.3).

We state now the two results of the chapter. The first theorem establishes an upper bound.

**Théorème 2.1.** *For  $\alpha > 0$ , define the integer  $N = \lfloor \alpha \log(1/\varepsilon) \rfloor$ . Then we have*

$$\sup_{f \in \mathcal{W}_2^r} \mathbb{E}_f \|\hat{f}_N - f\|^2 \leq C [\log(1/\varepsilon)]^{-2r},$$

where  $C$  is an absolute positive constant and  $\lfloor \cdot \rfloor$  denotes the floor function.

We recall that the floor function is defined by :  $\lfloor x \rfloor = \max\{n \in \mathbb{Z} \mid n \leq x\}$ . The second theorem provides a lower bound.

**Théorème 2.2.** *We have*

$$\inf_{\hat{f}} \sup_{f \in \mathcal{W}_2^r} \mathbb{E}_f \|\hat{f} - f\|^2 \geq c [\log(1/\varepsilon)]^{-2r},$$

where  $c$  is a positive constant which depends only on  $r$  and the infimum is taken over all estimators  $\hat{f}$ .

These two theorems show that our estimator  $\hat{f}_N$  converges at optimal rate in minimax sense while being adaptive.

## 2.3 Proofs

### 2.3.1 Proof of Theorem 2.1.

For the following proof, we consider the genuine model (2.3). By the usual MISE decomposition which involves a variance term and a bias term, we get

$$\mathbb{E}_f \|\hat{f}_N - f\|^2 = \mathbb{E}_f \sum_{k=0}^N (\hat{\theta}_k - \theta_k)^2 + \sum_{k \geq N+1} \theta_k^2$$

but

$$\begin{aligned} \mathbb{E}_f \sum_{k=0}^N (\hat{\theta}_k - \theta_k)^2 &= \mathbb{E}_f \sum_{k=0}^N \left( \sum_{j=0}^k \beta_{k,j} (y_j - \mu_j) \right)^2 \\ &= \varepsilon^2 \mathbb{E}_f \sum_{k=0}^N \left( \sum_{j=0}^k \beta_{k,j} \xi_j \right)^2 \end{aligned}$$

and since  $\xi_j \stackrel{iid}{\sim} N(0, 1)$ , it follows that

$$\begin{aligned} \mathbb{E}_f \|\hat{f}_N - f\|^2 &= \varepsilon^2 \sum_{k=0}^N \sum_{j=0}^k \beta_{k,j}^2 + \sum_{k \geq N+1} \theta_k^2 \\ &= V_N + B_N^2 \end{aligned}$$

We first deal with the variance term  $V_N$ . To this end, we have to upper bound the sum of the squared coefficients of the normalized Legendre polynomial of degree  $k$ . Set  $\sigma_k^2 = \sum_{j=0}^k \beta_{k,j}^2$ . An explicit form of  $P_k(x)$  is given by (see Abramowitz and Stegun [1]) :

$$P_k(x) = \left( \frac{2k+1}{2} \right)^{1/2} \frac{1}{2^k} \sum_{j=0}^{[k/2]} (-1)^j \binom{k}{j} \binom{2k-2j}{k} x^{k-2j},$$

where  $[\cdot]$  denotes the integer part and  $\binom{k}{j}$  denotes the binomial coefficient,  $\binom{k}{j} = \frac{k!}{(k-j)!j!}$ . This involves

$$\begin{aligned} \sigma_k^2 &= \frac{2k+1}{2} \frac{1}{4^k} \sum_{j=0}^{[k/2]} \left\{ \binom{k}{j} \binom{2k-2j}{k} \right\}^2 \\ &\leq \frac{2k+1}{2} \frac{1}{4^k} \left\{ \binom{2k}{k} \right\}^2 \sum_{j=0}^{[k/2]} \left\{ \binom{k}{j} \right\}^2 \end{aligned} \quad (2.6)$$

$$\leq \frac{2k+1}{2} \frac{1}{4^k} \left\{ \binom{2k}{k} \right\}^2 (2^k)^2, \quad (2.7)$$

the inequality (2.6) is due to the fact that for  $0 \leq j \leq [k/2]$ , we have  $\binom{2k-2j}{k} \leq \binom{2k}{k}$ . As for (2.7), we have  $\sum_{j=0}^{[k/2]} \left\{ \binom{k}{j} \right\}^2 \leq \left\{ \sum_{j=0}^k \binom{k}{j} \right\}^2$  and it is well known that  $\left\{ \sum_{j=0}^k \binom{k}{j} \right\}^2 = (2^k)^2$ .

By using now that  $\left\{ \binom{2k}{k} \right\}^2 \leq \frac{4^{2k}}{\sqrt{k}}$  (see Lemma 2.4. from Appendix) we have

$$\sigma_k^2 \leq \frac{2k+1}{2} \frac{4^{2k}}{\sqrt{k}}$$

which yields

$$V_N \leq C \varepsilon^2 N^{3/2} 4^{2N},$$



where  $C > 0$  denotes an absolute positive constant.

Now, it remains to upper bound the bias term  $B_N^2$ .

$$\begin{aligned} B_N^2 &= \sum_{k \geq N+1} \theta_k^2 \\ &= \sum_{k \geq N+1} \frac{k^{2r}}{k^{2r}} \theta_k^2 \\ &\leq N^{-2r} \sum_{k=1}^{\infty} k^{2r} \theta_k^2 \end{aligned}$$

Since the density  $f$  belongs to the space  $W_2^r$ ,  $\sum_k k^{2r} |\theta_k|^2 < \infty$ , we get

$$B_N^2 = \mathcal{O}(N^{-2r})$$

Finally we have the upper bound for the MISE :

$$\mathbb{E}_f \|\hat{f}_N - f\|^2 \leq C\varepsilon^2 N^{3/2} 4^{2N} + C' N^{-2r} \quad (2.8)$$

At last, it remains to choose the optimal  $N$  which will minimize the expression (2.8). This  $N$  is obtained by equalizing the upper bounds of the bias and the variance term, namely :

$$C\varepsilon^2 N^{3/2} 4^{2N} = C' N^{-2r},$$

as  $4^{2N} \gg N^{2r+3/2}$ , consequently  $N \asymp \log(\frac{1}{\varepsilon^2})$ . Once one plugs  $N \asymp \log(\frac{1}{\varepsilon^2})$  in (2.8), the desired result of the Theorem 1. follows.

### 2.3.2 Proof of Theorem 2.2.

From now on, to prove the lower bound and for practical reasons, we shall consider the model (2.4) which constitutes a heteroscedastic gaussian sequence space model. We recall that the equivalence between the models (2.3) and (2.4) is proved in Lemma 2.3. in Appendix.

A successful approach and standard tool to obtain lower bounds for minimax risk consists in specifying a subproblem namely constructing a subset of densities based on the observations (2.4). Then we lean on the application of the following particular version of Fano's lemma (see Birgé and Massart [24]) which will allow us to evaluate the difficulty of the specified subproblem and will give us a lower bound for the MISE associated to this subproblem. One crucial point in the Fano's lemma is the use of the Kullback-Leibler

divergence  $K(p_1, p_0)$  between two probability densities  $p_1$  and  $p_0$  defined by :

$$K(p_1, p_0) = \begin{cases} \int_{\mathbb{R}} \log\left(\frac{p_1(x)}{p_0(x)}\right) p_1(x) dx & \text{if } \mathbb{P}_1 \ll \mathbb{P}_0 \\ +\infty & \text{otherwise} \end{cases}$$

Here's the version of Fano's lemma, we are going to exploit :

**Lemma 2.1.** *Let  $\eta$  be a strictly positive real number and  $\mathcal{C}$  be a finite set of probability densities  $\{f_0, \dots, f_M\}$  on  $\mathbb{R}$  with  $|\mathcal{C}| \geq 6$  such that :*

(i)  $\|f_i - f_j\| \geq \eta > 0, \quad \forall 0 \leq i < j \leq M.$

(ii)  $\mathbb{P}_j \ll \mathbb{P}_0, \quad \forall j = 1, \dots, M,$  and

$$K(\mathbb{P}_j, \mathbb{P}_0) = K(f_j, f_0) \leq H < \log M$$

then for any estimator  $\hat{f}$  and any nondecreasing function  $\ell$

$$\sup_{f \in \mathcal{C}} \mathbb{E}_f [\ell(\|\hat{f} - f\|)] \geq \ell\left(\frac{\eta}{2}\right) \left[1 - \left(\frac{2}{3} \vee \frac{H}{\log M}\right)\right].$$

First of all, we have to construct an appropriate set of functions  $\mathcal{E}$ . We are going to define  $\mathcal{E}$  as a set of functions of the following type

$$\mathcal{E} = \left\{ f_\delta \in W_2^r : f_\delta = \mathbf{1}_{[-1,1]} \frac{c_0}{m^{(4r+3)/2}} \sum_{k=m}^{2m-1} \delta_k k^{(2r+2)/2} P_k, \right. \\ \left. \delta = (\delta_m, \dots, \delta_{2m-1}) \in \Delta = \{0, 1\}^m \right\}$$

We verify that  $f_\delta$  belongs to  $W_2^r$ . In this aim, we have to calculate the Legendre-Fourier coefficients associated with the density  $f_\delta$  :

$$\begin{aligned} \theta_{\delta l} &= \int_{-1}^1 f_\delta(x) P_l(x) dx \\ &= \begin{cases} \frac{c_0}{m^{(4r+3)/2}} \cdot l^{(2r+2)/2} \cdot \delta_l & \text{if } l \in [m, 2m-1] \\ 0 & \text{else} \end{cases} \end{aligned} \quad (2.9)$$

hence

$$\begin{aligned} \sum_{k=0}^{+\infty} k^{2r} \theta_{\delta k}^2 &= \frac{c_0^2}{m^{4r+3}} \sum_{k=m}^{2m-1} k^{2r} k^{2r+2} \delta_k^2 \\ &\leq \frac{c_0^2}{m^{4r+3}} \sum_{k=m}^{2m-1} k^{4r+2} \delta_k^2 \\ &\leq \frac{c_0^2 (2m)^{4r+2}}{m^{4r+3}} \sum_{k=m}^{2m-1} \delta_k^2 \leq c_0^2 2^{4r+2} < \infty, \end{aligned}$$

since  $\delta_k \in \{0, 1\}$ .

We set  $\delta^{(0)} = (0, \dots, 0)$  and  $f_{\delta^{(0)}} \equiv f_0$ . The Legendre-Fourier coefficients of  $f_0$  are null :

$$\theta_{0l} = 0 \quad \forall l \in \mathbb{N}. \quad (2.10)$$

We are now going to exhibit the suitable subset of densities  $\mathcal{C}$  of the Lemma 1. To this purpose, we only take into consideration a subset of  $M+1$  densities of  $\mathcal{E}$  :

$$\mathcal{C} = \{f_{\delta^{(0)}}, \dots, f_{\delta^{(M)}}\}$$

where  $\{\delta^{(1)}, \dots, \delta^{(M)}\}$  is a subset of  $\{0, 1\}^m$ .

The vector of observations in the model (2.4),  $\tilde{Y} = (\tilde{y}_1, \dots, \tilde{y}_\infty)$  follows the law  $\mathbb{P}_{f_\delta}$ , for  $\delta \in \mathcal{C}$ .

We are now going to apply Lemma 1. We first check the condition (i), accordingly, we have to assess the distance  $\|f_{\delta^{(i)}} - f_{\delta^{(j)}}\|^2$ . By the orthogonality of the system  $\{P_k\}_k$  and thanks to Parseval equality we get, for  $0 \leq i < j \leq M$ ,

$$\begin{aligned} \|f_{\delta^{(i)}} - f_{\delta^{(j)}}\|^2 &= \frac{c_0^2}{m^{4r+3}} \sum_{k=m}^{2m-1} k^{2r+2} (\delta_k^{(i)} - \delta_k^{(j)})^2 \\ &\geq \frac{c_0^2}{m^{4r+3}} \cdot m^{2r+2} \sum_{k=m}^{2m-1} (\delta_k^{(i)} - \delta_k^{(j)})^2 \\ &\geq \frac{c_0^2}{m^{2r+1}} \sum_{k=m}^{2m-1} (\delta_k^{(i)} - \delta_k^{(j)})^2 \\ &= \frac{c_0^2}{m^{2r+1}} \rho(\delta^{(i)}, \delta^{(j)}), \end{aligned}$$

where  $\rho(\cdot, \cdot)$  is the *Hamming distance*. We are going to resort to the Varshamov-Gilbert bound which is stated in the following lemma to find a lower bound of the quantity  $\rho(\delta^{(i)}, \delta^{(j)})$  :

**Lemma 2.2.** (Varshamov-Gilbert bound, 1962). *Fix  $m \geq 8$ . Then there exists a subset  $\{\delta^{(0)}, \dots, \delta^{(M)}\}$  of  $\Delta$  such that  $M \geq 2^{m/8}$  and*

$$\rho(\delta^{(j)}, \delta^{(k)}) \geq \frac{m}{8}, \quad \forall 0 \leq j < k \leq M.$$

Moreover we can always take  $\delta^{(0)} = (0, \dots, 0)$ .

For a proof of this lemma see for instance Tsybakov [148], p 89.

Hence

$$\|f_{\delta^{(i)}} - f_{\delta^{(j)}}\|^2 \geq (c_0^2)/(8m^{2r}) \equiv \eta^2$$

We are now going to check the condition (ii) in Lemma 1. and evaluate the Kullback-Leibler divergence. It is well known (see for instance Birgé and Massart [24] p 62) that for the Kullback-Leibler divergence in the case of a gaussian sequence space model we have

$$K(\mathbb{P}_{f_\delta}, \mathbb{P}_{f_0}) = K(f_\delta, f_0) = \frac{1}{\varepsilon^2} \sum_{l=1}^{\infty} \frac{|\theta_{\delta l} - \theta_{0l}|^2}{\sigma_l^2}. \quad (2.11)$$

Hence, by virtue of (2.9), (2.10) and (2.11), the Kullback-Leibler divergence between the two probability densities  $f_0$  and  $f_\delta$  corresponding to the observations in the model (2.4) for all  $\delta \in \mathcal{C}$  satisfies

$$\begin{aligned} K(f_\delta, f_0) &= \frac{1}{\varepsilon^2} \frac{c_0^2}{m^{4r+3}} \sum_{l=m}^{2m-1} \frac{l^{2r+2} \delta_l^2}{\sigma_l^2} \\ &\leq \frac{c_0^2 2^{2r+2}}{\varepsilon^2} \frac{m^{2r+2}}{m^{4r+3}} \sum_{l=m}^{2m-1} \frac{\delta_l^2}{\sigma_l^2} \end{aligned}$$

but thanks to Lemma 2.5. (see Appendix) we have

$$\frac{1}{\sigma_l^2} \leq \frac{1}{4^{l-1}}$$

which implies

$$K(f_\delta, f_0) \leq \frac{c_0^2 2^{2r+4}}{\varepsilon^2} \frac{1}{m^{2r+1} 4^m} \sum_{l=m}^{2m-1} \delta_l^2 \leq \frac{c_0^2 2^{2r+4}}{\varepsilon^2} \frac{1}{m^{2r+1} 4^m} \cdot m \leq \frac{c_0^2 2^{2r+4} m}{\varepsilon^2 4^m}$$

One chooses  $m = \frac{1}{\log 4} \log(\frac{1}{\varepsilon^2})$  so that  $1/4^m = \varepsilon^2$ , hence :

$$K(f_\delta, f_0) \leq c_0^2 2^{2r+4} m$$

and since  $m \leq 8 \log M / \log 2$  (see Lemma 2)

$$K(f_\delta, f_0) \leq \frac{c_0^2 2^{2r+7}}{\log 2} \log M$$

Eventually one can choose  $c_0$  small enough to have  $c \equiv \frac{c_0^2 2^{2r+7}}{\log 2} < 1$ .

Since now all the conditions of the Lemma 1. are fulfilled, we are in position to apply its result with the loss function  $\ell(x) = x^2$  and  $\eta = (c_0)/(2\sqrt{2}m^r)$ . Therefore, we derive that whatever the estimator  $\hat{f}$ ,

$$\sup_{f \in \mathcal{C}} \mathbb{E}_f [\|\hat{f} - f\|^2] \geq \frac{c_0^2}{32m^{2r}} \left[ 1 - \left( \frac{2}{3} \vee \frac{c}{\log M} \right) \right] \geq \frac{c_0^2}{96m^{2r}}.$$

but from above we had  $m = \frac{1}{\log 4} \log(\frac{1}{\varepsilon^2})$  which gives the desired lower bound.

## 2.4 Conclusion

The two theorems of this chapter show that in the problem of estimating a density of probability on a compact interval from noisy moments observations, the best rate of convergence one can achieve, supposing Sobolev scale smoothness and considering the mean integrated squared error, is only of logarithmic order. In a future work, one could try to generalise this result for  $L^p$  loss and may obtain faster rate of convergence if one assumes a more restricted smoothness class involving super smooth densities. Besides, one may consider an heteroscedastic gaussian noise instead of a white noise model.

## 2.5 Appendix

**Lemma 2.3.** *The models (2.3) and (2.4) are equivalent.*

**Proof.** We recall the model (2.3) :

$$y_j = \mu_j + \varepsilon \xi_j = \int_{-1}^1 f(x) x^j dx + \varepsilon \xi_j.$$

We are going now to multiply both sides of (2.3) by the coefficient  $\beta_{jk}$  of the

Legendre polynomial :

$$\begin{aligned}
 (2.3) \quad &\Leftrightarrow \beta_{kj}y_j = \beta_{kj} \int_{-1}^1 f(x)x^j dx + \varepsilon\beta_{kj}\xi_j \\
 &\Leftrightarrow \sum_{j=0}^k \beta_{kj}y_j = \sum_{j=0}^k \beta_{kj} \int_{-1}^1 f(x)x^j dx + \sum_{j=0}^k \varepsilon\beta_{kj}\xi_j \\
 &\Leftrightarrow \sum_{j=0}^k \beta_{kj}y_j = \int_{-1}^1 f(x) \sum_{j=0}^k \beta_{kj}x^j dx + \varepsilon \sum_{j=0}^k \beta_{kj}\xi_j \\
 &\Leftrightarrow \sum_{j=0}^k \beta_{kj}y_j = \int_{-1}^1 f(x)P_k dx + \varepsilon \sum_{j=0}^k \beta_{kj}\xi_j
 \end{aligned}$$

Let us set  $\tilde{\xi}_k = \sum_{j=0}^k \beta_{kj}\xi_j$ . Since  $\xi_j$  are i.i.d standard Gaussian random variables, the random variable  $\tilde{\xi}_k$  follows a normal law with zero mean and variance equal to  $\sum_{j=0}^k \beta_{kj}^2$ . Hence (2.3) is equivalent to :

$$\tilde{y}_k = \theta_k + \varepsilon\sigma_k\xi_k$$

where  $\sigma_k^2 = \sum_{j=0}^k \beta_{kj}^2$ ,  $\tilde{y}_k = \sum_{j=0}^k \beta_{kj}y_j$ ,  $\theta_k = \sum_{j=0}^k \beta_{kj}\mu_j = \int_{-1}^1 f(x)P_k(x)dx$  and  $\xi_k$  are i.i.d standard Gaussian random variables.

**Lemma 2.4.** *For all  $n \geq 1$  we have :*

$$\binom{2n}{n} \leq \frac{4^n}{n^{1/4}} \quad (2.12)$$

**Proof.** Let us prove (2.12) by recursion on  $n$ . The inequality is clearly true for  $n = 1$ .

Suppose (2.12) true for a certain  $n \geq 1$ .

$$\binom{2(n+1)}{n+1} = \binom{2n}{n} \frac{2(2n+1)}{n+1} \leq \frac{4^n}{n^{1/4}} \frac{2(2n+1)}{n+1},$$

by recursion hypothesis. It remains to prove that

$$\frac{4^n}{n^{1/4}} \frac{2(2n+1)}{n+1} \leq \frac{4^{n+1}}{(n+1)^{1/4}}. \quad (2.13)$$

$$\begin{aligned}
 (2.13) \quad &\Leftrightarrow \frac{2(2n+1)}{n^{1/4}(n+1)} \leq \frac{4}{(n+1)^{1/4}} \\
 &\Leftrightarrow \frac{n+1}{n} \left(\frac{2n+1}{n+1}\right)^4 \leq 2^4 \\
 &\Leftrightarrow \left(n + \frac{1}{2}\right)^{1/4} \leq n(n+1)^3,
 \end{aligned}$$

which is true because we have  $(n + \frac{1}{2})^{1/4} \leq (n + \frac{1}{2})^3(n + 1)$  and  $(n + \frac{1}{2})^3 \leq n(n + 1)^2$  since  $\frac{1}{8} \leq n^2/2 + n/4$ . This completes the proof.

**Lemma 2.5.** *For all  $n \geq 1$ , we have :*

$$\sigma_n^2 \geq 4^{n-1} \tag{2.14}$$

where  $\sigma_n$  is defined in (2.4).

**Proof.** Firstly, let us recall the value of the noise intensity  $\sigma_n^2$  :

$$\begin{aligned} \sigma_n^2 &= \frac{2n+1}{2} \frac{1}{4^n} \sum_{j=0}^{[n/2]} \left\{ \binom{n}{j} \binom{2n-2j}{n} \right\}^2 \\ &\geq \frac{n}{4^n} \binom{2n}{n}^2. \end{aligned}$$

And so, in order to prove (2.14) it remains to prove that

$$\binom{2n}{n} \geq \frac{4^n}{2\sqrt{n}} \quad n \geq 1.$$

We again use a recursion on  $n$ .

The inequality (2.14) is clear for  $n = 1$ . We suppose the property true for a certain  $n \geq 1$  and we shall prove it at the rank  $(n + 1)$ .

$$\begin{aligned} \binom{2(n+1)}{n+1} &= \binom{2n}{n} \frac{2(2n+1)}{n+1} \\ &\geq \frac{4^n}{2\sqrt{n}} \frac{2(2n+1)}{n+1} \\ &> \frac{4^{n+1}}{2\sqrt{n+1}} \end{aligned} \tag{2.15}$$

the inequality (2.15) is true because it is equivalent to  $4n^2 + 4n + 1 > 4n^2 + 4n$  what we always have.

## Chapitre 3

# Localized deconvolution on the sphere<sup>1</sup>

We provide a new algorithm for the treatment of deconvolution on the sphere which combines the traditional SVD inversion with an appropriate thresholding technique in a well chosen new basis. We establish upper bounds for the behaviour of our procedure for any  $\mathbb{L}_p$  loss. It is important to emphasize the adaptation properties of our procedures with respect to the regularity (sparsity) of the object to recover as well as to inhomogeneous smoothness. We also perform a numerical study which proves that the procedure shows very promising properties in practice as well.

### 3.1 Introduction

The spherical deconvolution problem was first proposed by Rooij and Ruymgaart (1991) [150] and subsequently solved in Healy et al. (1998) [73]. Kim and Koo (2002) [91] established minimaxity for the  $\mathbb{L}_2$ -rate of convergence. The optimal procedures obtained there are using orthogonal series methods associated with spherical harmonics. One important problem arising with these procedures is their poor local performances due to the fact that spherical harmonics are spread all over the sphere. This explains for instance the fact that although they are optimal in the  $\mathbb{L}_2$  sense, they cease to be optimal for other losses, such as  $\mathbb{L}_p$  losses for instance.

In our approach we focus on two important points. We aim at a procedure of estimation which is efficient from a  $\mathbb{L}_2$  point of view, as well as it performs satisfactorily from a local point of view (for other  $\mathbb{L}_p$  losses for instance).

---

1. This is a joint work with Gérard Kerkycharian and Dominique Picard.



Deconvolution is an inverse problem and in such a problem there is a notable conflict between the inversion part which in presence of noise creates an instability reasonably handled by a Singular Value Decomposition (SVD) approach and the fact that the SVD basis very rarely is localized and capable of representing local features of images, which are especially important to recover. Our strategy is to follow the approach started in Kerkyacharian et al. (2007) [89] for the Wicksell case, Kerkyacharian et al. (2009) [85] for the Radon transform, which utilizes the construction borrowed from Narkowich Petrushev and Wald (2006) [116], [115] of a tight frame (i.e. a redundant family) staying sufficiently close to the SVD decomposition but which enjoys at the same time enough localisation properties to be successfully used for statistical estimation (see for instance Baldi et al. (2009) Pietrobon et al. (2008) [16, 17, 125] for other types of applications). The construction [116] produces a family of functions which very much resemble to wavelets, the needlets.

To achieve the goals presented above, and especially adaptation to different regularities and local inhomogeneous smoothness, we essentially use a projection method on the needlets (which enables a stable inversion of the deconvolution, due to the closeness to the SVD basis) with a fine tuning subsequent thresholding process.

This provides a reasonably simple algorithm with very good performances, both from a theoretical point of view and a numerical point of view. In effect, this new algorithm provides a much better spatial adaptation, as well as adaptation to wider classes of regularity. We give here upper bounds obtained by the procedure over a large class of Besov spaces and any  $\mathbb{L}_p$  losses.

It is important to notice that especially because we consider different  $\mathbb{L}_p$  losses, we provide rates of convergence of new types attained by our procedure, which, of course, coincide with the usual ones for  $\mathbb{L}_2$  losses.

Again, the problem of choosing appropriated spaces of regularity on the sphere is a serious question, and we decided to consider the spaces which may be the closest to our natural intuition : those which generalize to the sphere case the classical approximation properties of usual regularity spaces such as Hölder spaces and include at the same time the Sobolev regularity spaces used in Kim and Koo (2002) [91].

Sphere deconvolution has a vast domain of application ; among others, medical imaging where this technique has become very important (see [145]) and astrophysics. Indeed, our results are especially motivated by many recent developments in the area of observational astrophysics.

It is a common problem in astrophysics to analyse data sets consisting of a number of objects (such as galaxies of a particular type) or of events (such

as cosmic rays or gamma ray bursts) distributed on the celestial sphere. In many cases, such objects trace an underlying probability distribution  $f$  on the sphere, which itself depends on the physics which governs the production of the objects and events.

The case for instance of ultra high energy cosmic rays (UHECR) illustrates well the type of applications of our results. Ultra high energy cosmic rays are particles of unknown nature which arrive at the earth from apparently random directions of the sky. They could originate from long-lived relic particles from the Big Bang, about 13 billion years old. Alternatively, they could be generated by the acceleration of standard particles, such as protons, in extremely violent astrophysical phenomena, such as cluster shocks. They could also originate from Active Galactic Nuclei (AGN), or from neutron stars surrounded by extremely high magnetic fields.

Hence, in some hypotheses, the underlying probability distribution for the directions of incidences of observed UHECRs would be a finite sum of point-like sources – or near point like, taken into account the deflection of the cosmic rays by magnetic fields. In other hypotheses, the distribution could be uniform, or smooth and correlated with the local distribution of matter in the universe. The distribution could also be a superposition of the above. Identifying between these hypotheses is of primordial importance for understanding the origin and mechanism of production of UHECRs.

Of course, the observations of these events ( $X_i$ 's in the sequel) are always most often perturbed by a secondary noise ( $\varepsilon_i$ ) which leads to the deconvolution problem described below. Following Healy et al. (1998) [73], Kim and Koo (2002) [91], the spherical deconvolution problem can be described as follows. Consider the situation where we observe  $Z_1, \dots, Z_N$   $N$  i.i.d. observation with

$$Z_i = \varepsilon_i X_i, \tag{3.1}$$

where the  $\varepsilon_i$ 's are i.i.d. random elements in  $SO(3)$  (the group of  $3 \times 3$  rotation matrices), and the  $Z_i$ 's and  $X_i$ 's are i.i.d. random elements of  $\mathbb{S}^2$  (two-dimensional unit sphere of  $\mathbb{R}^3$ ) random elements, with  $\varepsilon_i$  and  $X_i$  assumed to be independent. We suppose that the distribution of resp.  $X$ ,  $Z$  and  $\varepsilon$  are absolutely continuous with the Haar measure of resp.  $\mathbb{S}^2$ ,  $\mathbb{S}^2$  and  $SO(3)$  with the densities  $f_Z$ ,  $f_\varepsilon$ ,  $f_X$ .

Then,

$$f_Z = f_\varepsilon * f_X, \tag{3.2}$$

where  $*$  denotes convolution and is defined below. In the sequel,  $f_X$  will be denoted by  $f$  to emphasize the fact that it is the object to recover.

The following paragraph recall the necessary definitions. It is largely inspired by Kim and Koo (2002) [91] and Healy et al. (1998) [73].

### 3.2 Some preliminaries about harmonic analysis on $SO(3)$ and $\mathbb{S}^2$

We will provide a brief overview of Fourier analysis on  $SO(3)$  and  $\mathbb{S}^2$ . Most of the material can be found in expanded form in Vilenkin (1969) [152], Talman (1968) [139], Terras (1985) [142], Kim and Koo (2002) [91], and Healy et al. (1998) [73]. Let

$$u(\phi) = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad a(\theta) = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix},$$

where  $\phi \in [0, 2\pi)$ ,  $\theta \in [0, \pi)$ . It is well known that any rotation matrix can be decomposed as a product of three elemental rotations, one around the  $z$ -axis first with angle  $\psi$ , followed by a rotation around the  $y$ -axis with angle  $\theta$ , and finally another rotation again around the  $z$ -axis with angle  $\phi$ . Indeed, the well known Euler angle decomposition says that any  $g \in SO(3)$  can almost surely be uniquely represented by three angles  $(\phi, \theta, \psi)$ , with the following formula (see Healy et al. (1998) [73] for details) :

$$g = u(\phi)a(\theta)u(\psi), \quad (3.3)$$

where  $\phi \in [0, 2\pi)$ ,  $\theta \in [0, \pi)$ ,  $\psi \in [0, 2\pi)$ . Consider the functions, known as the rotational harmonics,

$$D_{mn}^l(\phi, \theta, \psi) = e^{-i(m\phi+n\psi)} P_{mn}^l(\cos \theta), \quad (3.4)$$

where the generalized Legendre associated functions  $P_{mn}^l$  for  $-l \leq m, n \leq l$ ,  $l = 0, 1, \dots$  are fully described in Vilenkin (1969) [152]. The functions  $D_{mn}^l$  for  $-l \leq m, n \leq l$ ,  $l = 0, 1, \dots$  are the eigenfunctions of the Laplace Beltrami operator on  $SO(3)$ , hence,  $\sqrt{2l+1}D_{mn}^l$ ,  $-l \leq m, n \leq l$ ,  $l = 0, 1, \dots$  is a complete orthonormal basis for  $\mathbb{L}_2(SO(3))$  with respect to the probability Haar measure. In addition, if we define the  $(2l+1) \times (2l+1)$  matrices by

$$D^l(g) = [D_{mn}^l(g)], \quad (3.5)$$

where for  $-l \leq m, n \leq l$ ,  $l = 0, 1, \dots$  and  $g \in SO(3)$ , they constitute the collection of inequivalent irreducible representations of  $SO(3)$  (for further details see Vilenkin (1969) [152]).

Hence, for  $f \in \mathbb{L}_2(SO(3))$ , we define the rotational Fourier transform on  $SO(3)$  by

$$\hat{f}_{mn}^l = \int_{SO(3)} f(g) D_{mn}^l(g) dg, \quad (3.6)$$

where again we think of (3.6) as the matrix entries of the  $(2l+1) \times (2l+1)$  matrix

$$\hat{f}^l = [\hat{f}_{mn}^l]_{-l \leq m, n \leq l}, \quad l = 0, 1, \dots$$

and  $dg$  is the probability Haar measure on  $SO(3)$ . The rotational inversion can be obtained by

$$\begin{aligned} f(g) &= \sum_l \sum_{-l \leq m, n \leq l} \hat{f}_{mn}^l \overline{D_{mn}^l(g)} \\ &= \sum_l \sum_{-l \leq m, n \leq l} \hat{f}_{mn}^l D_{mn}^l(g^{-1}) \end{aligned} \quad (3.7)$$

(3.7) is to be understood in  $\mathbb{L}_2$ -sense although with additional smoothness conditions, it can hold pointwise.

A parallel spherical Fourier analysis is available on  $\mathbb{S}^2$ . Any point on  $\mathbb{S}^2$  can be represented by

$$\omega = (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)^t,$$

with  $\phi \in [0, 2\pi)$ ,  $\theta \in [0, \pi)$ . We also define the functions :

$$Y_m^l(\omega) = Y_m^l(\theta, \phi) = (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_m^l(\cos \theta) e^{im\phi}, \quad (3.8)$$

for  $-l \leq m \leq l$ ,  $l = 0, 1, \dots$ ,  $\phi \in [0, 2\pi)$ ,  $\theta \in [0, \pi)$  and where  $P_m^l(\cos \theta)$  are the associated Legendre functions. The functions  $Y_m^l$  obey

$$Y_{-m}^l(\theta, \phi) = (-1)^m \bar{Y}_m^l(\theta, \phi). \quad (3.9)$$

The set  $\{Y_m^l, -l \leq m \leq l, l = 0, 1, \dots\}$  is forming an orthonormal basis of  $\mathbb{L}_2(\mathbb{S}^2)$ , generally referred to as the spherical harmonic basis.

Again, as above, for  $f \in \mathbb{L}_2(\mathbb{S}^2)$ , we define the spherical Fourier transform on  $\mathbb{S}^2$  by

$$\hat{f}_m^l = \int_{\mathbb{S}^2} f(\omega) \overline{Y_m^l(\omega)} d\omega, \quad (3.10)$$

where  $d\omega$  is the probability Haar measure on the sphere  $\mathbb{S}^2$ . The spherical inversion can be obtained by

$$f(\omega) = \sum_l \sum_{-l \leq m \leq l} \hat{f}_m^l Y_m^l(\omega). \quad (3.11)$$

The bases detailed above are important because they realize a singular value decomposition of the convolution operator created by our model. In effect, we define for  $f_\varepsilon \in \mathbb{L}_2(SO(3))$ ,  $f \in \mathbb{L}_2(\mathbb{S}^2)$  the convolution by the following formula :

$$f_\varepsilon * f(\omega) = \int_{SO(3)} f_\varepsilon(u) f(u^{-1}\omega) du$$

and we have for all  $-l \leq m \leq l$ ,  $l = 0, 1, \dots$ ,

$$(\widehat{f_\varepsilon * f})_m^l = \sum_{n=-l}^l \hat{f}_{\varepsilon, mn}^l \hat{f}_n^l := (\hat{f}_\varepsilon^l \hat{f}^l)_m. \quad (3.12)$$

### 3.2.1 The SVD Method

The singular value method (see Healy et al. (1998) [73] and Kim Koo (2002) [91]) consists in expanding  $f$  in the spherical harmonics basis  $Y_m^l$  and estimating the spherical Fourier coefficients using the formula above (3.12). We get the following estimator of the spherical Fourier transform of  $f$  :

$$\begin{aligned} \hat{f}_m^{l,N} &:= \frac{1}{N} \sum_{j=1}^N \sum_{n=-l}^l \hat{f}_{\varepsilon^{-1}, mn}^l \bar{Y}_n^l(Z_j) \\ \hat{f}_{\varepsilon^{-1}}^l &:= (\hat{f}_\varepsilon^l)^{-1}, \end{aligned} \quad (3.13)$$

provided, of course, that these inverse matrices exist, and then the estimator of the distribution  $f$  is

$$f^N(\omega) = \sum_{l=0}^{\tilde{N}} \sum_{m=-l}^l \hat{f}_m^{l,N} Y_m^l(\omega), \quad (3.14)$$

where  $\tilde{N}$  depending on the number of observations has to be properly selected.

### 3.3 Needlet construction

This construction is due to Narcowich et al. (2006) [115]. Its aim is essentially to build a very well localized tight frame constructed using spherical harmonics, as discussed below. It was recently extended to more general euclidean settings with fruitful statistical applications (see Kerkycharian et al. (2007) [89]), Baldi et al. (2009) Pietrobon et al. (2008) [16, 17, 125]. As described above, we have the following decomposition :

$$\mathbb{L}_2(dx) = \bigoplus_{l=0}^{\infty} \mathbb{H}_l, \quad (3.15)$$

where  $\mathbb{H}_l$  is the space of spherical harmonics of  $\mathbb{S}^2$ , of degree  $l$  (which dimension is  $2l + 1$ ).

The orthogonal projector on  $\mathbb{H}_l$  can be written using the following the kernel operator

$$\forall f \in \mathbb{L}_2(dx), P_{\mathbb{H}_l} f(x) = \int_{\mathbb{S}^d} L_l(\langle x, y \rangle) f(y) dy \quad (3.16)$$

where  $\langle x, y \rangle$  is the standard scalar product of  $\mathbb{R}^3$ , and  $L_l$  is the Gegenbauer polynomial with parameter  $\frac{1}{2}$  of degree  $l$ , defined on  $[-1, +1]$  and normalized so that

$$\int_{-1}^1 L_l(t) L_k(t) dt = \frac{2l + 1}{8\pi^2} \quad (3.17)$$

Let us point out the following reproducing property of the projection operators :

$$\int_{\mathbb{S}^d} L_l(\langle x, y \rangle) L_k(\langle y, z \rangle) dy = \delta_{l,k} L_l(\langle x, z \rangle). \quad (3.18)$$

The following construction is based on two fundamental steps : Littlewood-Paley decomposition and discretization, which are summarized in the two following subsections.

#### 3.3.1 Littlewood-Paley decomposition

Let  $\phi$  be a  $C^\infty$  function on  $\mathbb{R}$ , symmetric and decreasing on  $\mathbb{R}^+$  supported in  $|\xi| \leq 1$ , such that  $1 \geq \phi(\xi) \geq 0$  and  $\phi(\xi) = 1$  if  $|\xi| \leq \frac{1}{2}$ .

$$b^2(\xi) = \phi\left(\frac{\xi}{2}\right) - \phi(\xi) \geq 0$$

so that

$$\forall |\xi| \geq 1, \quad \sum_{j \geq 0} b^2\left(\frac{\xi}{2^j}\right) = 1. \quad (3.19)$$

Remark that  $b(\xi) \neq 0$  only if  $\frac{1}{2} \leq |\xi| \leq 2$ . Let us now define the operator  $\Lambda_j = \sum_{l \geq 0} b^2\left(\frac{l}{2^j}\right) L_l$  and the associated kernel

$$\Lambda_j(x, y) = \sum_{l \geq 0} b^2\left(\frac{l}{2^j}\right) L_l(\langle x, y \rangle) = \sum_{2^{j-1} < l < 2^{j+1}} b^2\left(\frac{l}{2^j}\right) L_l(\langle x, y \rangle).$$

We obviously have :

$$\forall f \in \mathbb{L}_2(\mathbb{S}^2), \quad f = \lim_{J \rightarrow \infty} L_0(f) + \sum_{j=0}^J \Lambda_j(f). \quad (3.20)$$

and if  $M_j(x, y) = \sum_{l \geq 0} b\left(\frac{l}{2^j}\right) L_l(\langle x, y \rangle)$ , then

$$\Lambda_j(x, y) = \int M_j(x, z) M_j(z, y) dz. \quad (3.21)$$

### 3.3.2 Discretization and localization properties

Let us define

$$\mathcal{P}_l = \bigoplus_{m=0}^l \mathbb{H}_m,$$

the space of the restrictions to  $S^2$  of the polynomials of degree less than  $l$ .

The following quadrature formula is true : for all  $l \in \mathbb{N}$  there exists a finite subset  $\mathcal{X}_l$  of  $S^2$  and positive real numbers  $\lambda_\eta > 0$ , indexed by the elements  $\eta$  of  $\mathcal{X}_l$ , such that

$$\forall f \in \mathcal{P}_l, \quad \int_{\mathbb{S}^2} f(x) dx = \sum_{\eta \in \mathcal{X}_l} \lambda_\eta f(\eta). \quad (3.22)$$

Then the operator  $M_j$  defined in the subsection above is such that :

$$z \mapsto M_j(x, z) \in \mathcal{P}_{[2^{j+1}]},$$

so that

$$z \mapsto M_j(x, z) M_j(z, y) \in \mathcal{P}_{[2^{j+2}]},$$

and we can write :

$$\Lambda_j(x, y) = \int M_j(x, z)M_j(z, y)dz = \sum_{\eta \in \mathcal{X}_{[2^j+2]}} \lambda_\eta M_j(x, \eta)M_j(\eta, y) .$$

This implies :

$$\begin{aligned} \Lambda_j f(x) &= \int \Lambda_j(x, y)f(y)dy = \int \sum_{\eta \in \mathcal{X}_{[2^j+2]}} \lambda_\eta M_j(x, \eta)M_j(\eta, y)f(y)dy \\ &= \sum_{\eta \in \mathcal{X}_{[2^j+2]}} \sqrt{\lambda_\eta} M_j(x, \eta) \int \sqrt{\lambda_\eta} M_j(y, \eta)f(y)dy . \end{aligned}$$

We denote

$$\mathcal{X}_{[2^j+2]} = \mathcal{Z}_j, \quad \psi_{j,\eta}(x) := \sqrt{\lambda_\eta} M_j(x, \eta) \text{ for } \eta \in \mathcal{Z}_j ,$$

It can also be proved that the set of cubature points  $\mathcal{X}_l$  can be chosen so that :

$$\frac{1}{c} 2^{2j} \leq \#\mathcal{Z}_j \leq c 2^{2j} \tag{3.23}$$

for some  $c > 0$ . Actually in the simulations of §3.5 we make use of some sets of cubature points such that  $\#\mathcal{Z}_j = 12 \cdot 2^{2j}$  exactly. It holds, using (3.20)

$$f = L_0(f) + \sum_j \sum_{\eta \in \mathcal{Z}_j} \langle f, \psi_{j,\eta} \rangle_{\mathbb{L}_2(\mathbb{S}^d)} \psi_{j,\eta} .$$

The main result of Narcowich et al. (2006) [115] is the following localization property of the  $\psi_{j,\eta}$ , called needlets : for any  $k \in \mathbb{N}$  there exists a constant  $c_k$  such that, for every  $\xi \in \mathbb{S}^2$  :

$$|\psi_{j,\eta}(\xi)| \leq \frac{c_k 2^j}{(1 + 2^j d(\eta, \xi))^k} . \tag{3.24}$$

where  $d$  is the natural geodesic distance on the sphere ( $d(\xi, \eta) = \arccos \langle \eta, \xi \rangle$ ). In other words needlets are almost exponentially localized around their associated cubature point, which motivates their name.

A major consequence of this localization property can be summarized in the following properties which will play an essential role in the sequel.

For any  $1 \leq p < \infty$ , there exist positive constants  $c_p, C_p, c, C$  and  $D_p$  such that



$$c_p 2^{2j(\frac{p}{2}-1)} \leq \|\psi_{j\eta}\|_p^p \leq C_p 2^{2j(\frac{p}{2}-1)}, \quad (3.25)$$

$$c2^j \leq \|\psi_{j\eta}\|_\infty \leq C2^j, \quad (3.26)$$

$$\left\| \sum_{\eta \in \mathcal{Z}_j} \lambda_\eta \psi_{j,\eta} \right\|_\pi \leq c2^{2j(\frac{1}{2}-\frac{1}{\pi})} \left( \sum_{\eta \in \mathcal{Z}_j} |\lambda_\eta|^\pi \right)^{1/\pi} \quad (3.27)$$

$$\left\| \sum_{\eta \in \mathcal{Z}_j} \lambda_\eta \psi_{j\eta} \right\|_p^p \leq D_p \sum_{\eta \in \mathcal{Z}_j} |\lambda_\eta|^p \|\psi_{j\eta}\|_p^p \quad (3.28)$$

$$\left\| \sum_{\eta \in \mathcal{Z}_j} u_\eta \psi_{j\eta} \right\|_\infty \leq C \sup_{\eta \in \mathcal{Z}_j} |u_\eta| 2^j \quad (3.29)$$

To conclude this section, let us give a graphic representation of a spherical needlet in the spherical coordinates in order to illustrate the above theory. In the following graphic, we chose  $j = 3$  and  $\eta = 250$ .

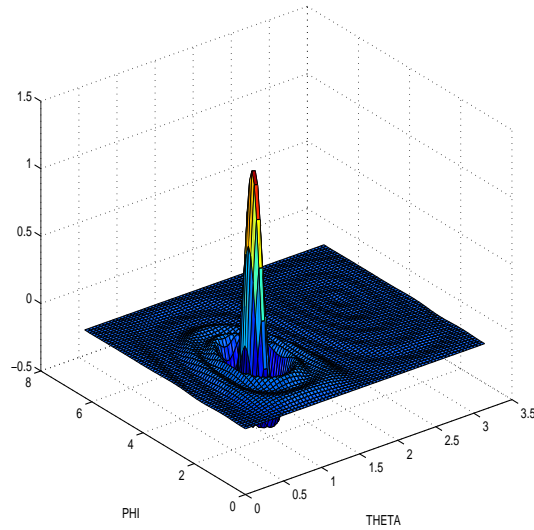


FIGURE 3.1 – A spherical needlet.

### 3.3.3 Besov spaces on the sphere

The problem of choosing appropriated spaces of regularity on the sphere is a serious question, and we decided to consider the spaces which may be the closest to our natural intuition : those which generalize to the sphere case the classical approximation properties used to define for instance Sobolev

spaces. In this section we summarize the main properties of Besov spaces which will be used in the sequel, as established in [115].

Let  $f : \mathbb{S}^2 \rightarrow \mathbb{R}$  be a measurable function. We define

$$E_k(f, \pi) = \inf_{P \in \mathcal{P}_k} \|f - P\|_\pi,$$

the  $\mathbb{L}_\pi$  distance between  $f$  and the space of polynomials of degree  $k$ . The Besov space  $B_{\pi,r}^s$  is defined as the space of functions such that

$$f \in \mathbb{L}_\pi \text{ and } \left( \sum_{k=0}^{\infty} (k^s E_k(f, \pi))^r \frac{1}{k} \right)^{1/r} < +\infty.$$

Remarking that  $k \rightarrow E_k(f, \pi)$  is decreasing, by a standard condensation argument this is equivalent to

$$f \in \mathbb{L}_\pi \text{ and } \left( \sum_{j=0}^{\infty} (2^{js} E_{2^j}(f, \pi))^r \right)^{1/r} < +\infty.$$

and the following theorem states that as it is the case for Besov spaces in  $\mathbb{R}^d$ , the needlet coefficients are good indicators of the regularity and in fact Besov spaces of  $\mathbb{S}^2$  are Besov bodies, when expressed using the needlet expansion.

**Théorème 3.1.** *Let  $1 \leq \pi \leq +\infty$ ,  $s > 0$ ,  $0 \leq r \leq +\infty$ . Let  $f$  a measurable function and define*

$$\langle f, \psi_{j,\eta} \rangle = \int_{\mathbb{S}^d} f(x) \psi_{j,\eta}(x) dx \stackrel{\text{def}}{=} \beta_{j,\eta},$$

*provided the integrals exist. Then  $f$  belongs to  $B_{\pi,r}^s$  if and only if, for every  $j = 1, 2, \dots$ ,*

$$\left( \sum_{\eta \in \mathcal{X}_j} (|\beta_{j,\eta}| \|\psi_{j,\eta}\|_\pi)^\pi \right)^{1/\pi} = 2^{-js} \delta_j,$$

*where  $(\delta_j)_j \in \ell_r$ .*

As has been seen above,

$$c2^{2j(\frac{1}{2}-\frac{1}{\pi})} \leq \|\psi_{j,\eta}\|_\pi \leq C2^{2j(\frac{1}{2}-\frac{1}{\pi})},$$

for some positive constants  $c, C$ , the Besov space  $B_{\pi,r}^s$  turns out to be a Banach space associate to the norm

$$\|f\|_{B_{\pi,r}^s} := \|(2^{j[s+2(\frac{1}{2}-\frac{1}{\pi})]}\|(\beta_{j\eta})_{\eta \in \mathcal{X}_j}\|_{\ell_r})_{j \geq 0}\|_{\ell_r} < \infty, \quad \text{and} \quad (3.30)$$

Using standard arguments (reducing to comparisons of  $l_q$  norms), it is easy to prove the following embeddings :

$$\begin{aligned} B_{\pi,r}^s &\subset B_{p,r}^s \quad \text{for } p \leq \pi \\ B_{\pi,r}^s &\subset B_{p,r}^{s-2(\frac{1}{\pi}-\frac{1}{p})} \quad \text{for } \pi \leq p \text{ and } s > 2(\frac{1}{\pi}-\frac{1}{p}). \end{aligned} \quad (3.31)$$

Moreover, it is also true that for  $s > \frac{2}{\pi}$ , if  $f$  belongs to  $B_{\pi,r}^s$ , then it is continuous, and as a consequence bounded.

In the sequel we shall denote by  $B_{\pi,r}^s(M)$  the ball of radius  $M$  of the Besov space  $B_{\pi,r}^s$ .

### 3.4 Needlet algorithm : thresholding needlet coefficients

The first step is to construct a needlet system (frame)  $\{\psi_{j\eta} : \eta \in \mathcal{Z}_j, j \geq -1\}$  as described in section 3. The needlet decomposition of any  $f \in \mathbb{L}_2(\mathbb{S}^2)$  takes the form

$$f = \sum_j \sum_{\eta \in \mathcal{Z}_j} (f, \psi_{j\eta})_{\mathbb{L}_2(\mathbb{S}^2)} \psi_{j\eta}.$$

Using Parseval's identity, we have  $\beta_{j\eta} = (f, \psi_{j\eta})_{\mathbb{L}_2(\mathbb{S}^2)} = \sum_{lm} \hat{f}_m^l \psi_{j\eta}^{lm}$  with  $\hat{f}_m^l = (f, Y_m^l)$  and  $\psi_{j\eta}^{lm} = (\psi_{j\eta}, Y_m^l)$ .

Thus

$$\hat{\beta}_{j\eta} = \sum_{lm} \hat{f}_m^{l,N} \psi_{j\eta}^{lm}, \quad (3.32)$$

is an unbiased estimate of  $\beta_{j\eta}$ . We recall that  $\hat{f}_m^{l,N}$  has been defined in (3.13). Notice that from the needlet construction (see the previous section) it follows that the sum above is finite. More precisely,  $\psi_{j\eta}^{lm} \neq 0$  only for  $2^{j-1} < l < 2^{j+1}$ .

Let us consider the following estimate of  $f$  :

$$\hat{f} = \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} t(\hat{\beta}_{j\eta}) \psi_{j\eta},$$

where  $t$  is a thresholding operator defined by

$$t(\hat{\beta}_{j\eta}) = \hat{\beta}_{j\eta} I\{|\hat{\beta}_{j\eta}| \geq \kappa t_N |\sigma_j|\} \quad \text{with} \quad (3.33)$$

$$t_N = \sqrt{\frac{\log N}{N}}, \quad (3.34)$$

$$\sigma_j^2 = M^2 \sum_{ln} \left| \sum_m \psi_{j\eta}^{lm} \hat{f}_{\varepsilon^{-1}mn}^l \right|^2. \quad (3.35)$$

Here  $\kappa$  is a tuning parameter of the method which will be properly selected later on.  $M$  is such that  $\|f\|_\infty \leq M$ . Notice that the thresholding depends on the resolution level  $j$  through the constant  $\sigma_j$  which will also be specified later on, and the same with regard to the upper level of details  $J$ .

### 3.4.1 Performances of the procedure

**Théorème 3.2.** *Let  $1 \leq p < \infty$ ,  $\nu > 0$ , and let us assume that*

$$\sigma_j^2 := M^2 \sum_{ln} \left| \sum_m \psi_{jn}^{lm} \hat{f}_{\varepsilon^{-1}mn}^l \right|^2 \leq C 2^{2j\nu}, \quad \forall j \geq 0. \quad (3.36)$$

*Let us take  $\kappa^2 \geq 16p$  and  $2^J = d[t_N]^{\frac{-1}{(\nu+1)}}$  with  $t_N$  as in (3.34) and  $d$  is a positive constant. Then if  $\pi \geq 1$ ,  $s > 2/\pi$ ,  $r \geq 1$  (with the restriction  $r \leq \pi$  if  $s = (\nu + 1)(\frac{p}{\pi} - 1)$ ), there exists a constant  $C$  such that :*

$$\sup_{f \in B_{\pi,r}^s(M)} \mathbb{E} \|\hat{f} - f\|_p^p \leq C (\log(N))^{p-1} [N^{-1/2} \sqrt{\log(N)}]^{\mu p}, \quad (3.37)$$

where

$$\begin{aligned} \mu &= \frac{s}{s + \nu + 1}, \quad \text{if } s \geq (\nu + 1)\left(\frac{p}{\pi} - 1\right) \\ \mu &= \frac{s - 2/\pi + 2/p}{s + \nu - 2/\pi + 1}, \quad \text{if } \frac{2}{\pi} < s < (\nu + 1)\left(\frac{p}{\pi} - 1\right). \end{aligned}$$

The proof of this theorem is given in section 3.6.

#### Remarks

1. The rates of convergence found here are standard in inverse problems. They can be related to rates found in Kim and Koo (2002) in the same deconvolution problem, with a  $\mathbb{L}_2$  loss and constraints on the spaces comparable to  $B_{s_2}^2(M)$ . In the deconvolution problem on the interval, similar rates are found even for  $\mathbb{L}_p$  losses (with standard modifications since the dimension here is 2 instead of 1) : see for instance Johnstone et al. (2004) [76]. These results are proved to be minimax (see Kim and Koo (2002) [91]) up to logarithmic factors, for the case  $p = 2$  with a  $B_{s_2}^2(M)$  constraint on the object to estimate.

2. It is worthwhile to notice that the procedure is adaptive, meaning that it does not require a priori knowledge on the regularity (or sparsity) of the function. It also adapt to non homogeneous smoothness of the function. The logarithmic factor is a standard price to pay for adaptation.
3. The parameter  $\nu$  appearing here is often called degree of ill-posedness of the problem (DIP). It appears here through condition (3.36) which is essential in this problem. In [91] for instance, and very often in diverse inverse problems, this DIP parameter is introduced with the help of the eigenvalues of the operator (i.e. here the discrepancy of the coefficients of  $f_\varepsilon$  in its expansion along the spherical harmonics). In the following subsection, we prove that (3.36) is in fact a consequence of the standard 'ordinary smooth' condition.

### 3.4.2 Condition (3.36) and the smoothness of $f_\varepsilon$

Following Kim and Koo (2002) [91] (condition 3.6), we can define the smoothness of  $f_\varepsilon$  spectrally. We place ourselves in the 'ordinary smooth' case

$$\|(\hat{f}_\varepsilon^l)^{-1}\|_{op} \leq d_0 l^\nu \quad \text{and} \quad \|\hat{f}_\varepsilon^l\|_{op} \leq d_1 l^{-\nu} \quad \text{as } l \rightarrow \infty, \quad (3.38)$$

for some positive constants  $d_0, d_1$  and nonnegative constant  $\nu$ , and where the operator norm of the rotational Fourier transform  $\hat{f}_\varepsilon^l$  is defined as

$$\|\hat{f}_\varepsilon^l\|_{op} = \sup_{h \neq 0, h \in \mathcal{E}_l} \frac{\|\hat{f}_\varepsilon^l h\|_2}{\|h\|_2},$$

$\mathcal{E}_l$  being the  $(2l+1)$ -dimensional vector space spanned by  $\{Y_m^l : -l \leq m \leq l\}$ . The following proposition states that condition (3.35) is satisfied in the ordinary smooth case by the needlets system.

**Proposition 3.1.** *If  $\|\hat{f}_\varepsilon^l\|_{op} \leq d_0 l^{-\nu}$ , then*

$$|\sigma_j|^2 := M^2 \sum_{ln} \left| \sum_m \psi_{j\eta}^{lm} \hat{f}_{\varepsilon^{-1}mn}^l \right|^2 \leq C 2^{2j\nu}, \quad \forall j \geq 0.$$

*Proof.* Since  $\psi_{j\eta}^{lm} \neq 0$  only for  $2^{j-1} < l < 2^{j+1}$ ,

$$\begin{aligned} \sum_{l=2^{j-1}}^{2^{j+1}} \sum_{n=-l}^l \left| \sum_m \psi_{j\eta}^{lm} \hat{f}_{\varepsilon^{-1}mn}^l \right|^2 &= \sum_{l=2^{j-1}}^{2^{j+1}} \|(\hat{f}_\varepsilon^l)^{-1} \psi_{j\eta}\|_2^2 \\ &\leq \sum_{l=2^{j-1}}^{2^{j+1}} \|(\hat{f}_\varepsilon^l)^{-1}\|_{op}^2 \|\psi_{j\eta}\|_2^2 \end{aligned}$$

which proves the result using inequality (3.25).

We now give a brief review of some examples of smooth distributions which are discussed in depth in Healy et al (1998) [73] and Kim and Koo (2002) [91].

### Rotational Laplace distribution

This distribution can be viewed as an exact analogy on  $SO(3)$  of the Laplace distribution on  $\mathbb{R}$ . Spectrally, for some  $\rho^2 > 0$ , this distribution is characterized by

$$\hat{f}_{\varepsilon, mn}^l = (1 + \rho^2 l(l+1))^{-1} \delta_{mn}, \quad (3.39)$$

for  $-l \leq m, n \leq l$  and  $l = 0, 1, \dots$ , and where  $\delta_{mn} = 1$  if  $m = n$  and 0 otherwise.

### The Rosenthal distribution

This distribution has its origin in random walks in groups (for details see Rosenthal (1994) [129]).

If one considers the situation where  $f_\varepsilon$  is a  $p$ -fold convolution product of conjugate invariant random for a fixed axis, then, Rosenthal (1994) [129] (p.407) showed that

$$\hat{f}_{\varepsilon, mn}^l = \left( \frac{\sin(l+1/2)\theta}{(2l+1)\sin\theta/2} \right)^p \delta_{mn},$$

for  $-l \leq m, n \leq l$  and  $l = 0, 1, \dots$  and where  $0 < \theta \leq \pi$  and  $p > 0$ .

## 3.5 Practical performances

In this section we produce the results of numerical experiments on the sphere  $\mathbb{S}^2$ . The sets of cubature points  $\xi_{j\eta}$  in the simulations that follow have been generated with the HEALPix pixelisation. For each resolution level  $j$ , the HEALPix pixelisation gives  $12 \cdot 2^{2j}$  cubature points.

In the two examples below we considered samples of cardinality  $N = 1500$ . The maximal resolution level  $J$  is taken such that  $J = (1/2) \log_2 \left( \frac{N}{\log N} \right)$ . In order not to have more cubature points than observations we set  $J = 3$  for  $N = 1500$ . We recall the expression of the estimate of the needlets coefficients

of the density of interest :

$$\hat{\beta}_{j\eta} = \frac{1}{N} \sqrt{\lambda_{j\eta}} \sum_{l=2^{j-1}}^{2^{j+1}} b(l/2^j) \sum_{m=-l}^l Y_m^l(\xi_{j\eta}) \sum_{n=-l}^l \hat{f}_{\varepsilon^{-1},mn}^l \sum_{u=1}^N \bar{Y}_n^l(Z_u), \quad (3.40)$$

where the cubature weight is equal to  $\lambda_{j\eta} = 4\pi/(12.2^{2j})$ .

We replace the rotational Fourier transform  $(\hat{f}_\varepsilon^l)_{mn} := \hat{f}_{\varepsilon,mn}^l$  (defined in (3.6)) by its empirical version. Indeed, we suppose that the noise is unknown, this is a situation which is very likely to occur for instance in the context of astrophysics.

We precise again that  $\hat{f}_{\varepsilon^{-1},mn}^l$  denotes the  $(m,n)$  element of the matrix  $(\hat{f}_\varepsilon^l)^{-1} := \hat{f}_{\varepsilon^{-1}}^l$  which is the inverse of the  $(2l+1) \times (2l+1)$  matrix  $(\hat{f}_\varepsilon^l)$ . In order to get the empirical version  $\hat{f}_{\varepsilon^{-1},mn}^{l,N}$  of  $\hat{f}_{\varepsilon^{-1},mn}^l$ , we have first to compute the empirical matrix  $(\hat{f}_\varepsilon^{l,N})$  then to inverse it to get the matrix  $(\hat{f}_\varepsilon^{l,N})^{-1} := \hat{f}_{\varepsilon^{-1}}^{l,N}$ . The  $(m,n)$  entry of the matrix  $(\hat{f}_\varepsilon^{l,N})$  is given by the formula :

$$\hat{f}_{\varepsilon,mn}^{l,N} = 1/N \sum_{j=1}^N D_{m,n}^l(\varepsilon_j),$$

where the rotational harmonics  $D_{m,n}^l$  have been defined in (3.4). The  $\varepsilon_j$ 's are i.i.d realizations of the variable  $\varepsilon \in SO(3)$ .

For the generation of the random variable  $\varepsilon \in SO(3)$ , we chose the Oz as the rotation axis and an angle  $\phi$  following a uniform law. Various cases are investigated for the angle  $\phi$ , that is to say a uniform law with different supports such as  $[0, \pi/8]$ ,  $[0, \pi/4]$ ,  $[0, \pi/2]$ . The more the support of the uniform law is large, the more the level noise will be intense.

This particular choice of rotation matrix entails in the decomposition of an element of  $SO(3)$  (see formula (3.3)) the angles  $\psi$  and  $\theta$  to be both equal to zero. For this specific setting of perturbation, we deduce the following form for the rotational harmonics :

$$D_{m,n}^l(\varepsilon_j) = P_{mn}^l(1)e^{-i\phi_j} = \delta_{mn}e^{-i\phi_j},$$

where  $\phi_j \sim U[0, a]$  and  $a$  is a positive constant which will be specified later. For the reconstruction of the density  $f$ , we recall that

$$\hat{f} := \frac{1}{|\mathbb{S}^2|} + \sum_{j=0}^J \sum_{\eta=1}^{12.2^{2j}} \hat{\beta}_{j\eta} \psi_{j\eta} I_{\{|\hat{\beta}_{j\eta}| \geq \kappa t_n \sigma_j\}},$$

the expression of  $\sigma_j$  is given in (3.35), and  $t_N = \sqrt{\log(N)/N}$ . The constant  $M$  which appears in the expression of  $\sigma_j$  is such that  $\|f\|_\infty \leq M$ .

**Example 1.** In this first example, we consider the case of the uniform density  $f = \frac{1}{4\pi}$ . It is easy to verify that  $\beta_{j\eta} = \langle f, \psi_{j\eta} \rangle_{L^2} = 0$  for every  $j$  and every  $\eta$ . Following Baldi et al. (2009) [16], a simple way of assessing the performance of the procedure is to count the number of coefficients surviving thresholding.

	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$\kappa = 0.08$	2	1	75	337
$\kappa = 0.29$	0	0	2	8
$\kappa = 0.34$	0	0	0	2

TABLE 3.1 – number of coefficients surviving thresholding for various values of  $\kappa$ ,  $\phi \sim U[0, \pi/8]$ .

	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$\kappa = 0.05$	4	35	135	556
$\kappa = 0.17$	0	4	24	103
$\kappa = 0.30$	0	0	1	5
$\kappa = 0.34$	0	0	0	3

TABLE 3.2 – number of coefficients surviving thresholding for various values of  $\kappa$ ,  $\phi \sim U[0, \pi]$ .

We precise that both in the cases of an angle following a law  $U[0, \pi/8]$  or  $U[0, \pi]$ , all the coefficients are killed for  $\kappa = 0.38$ . Accordingly, we can conclude that the thresholding procedure based on spherical needlets is very efficient.

**Example 2.** We will now deal with the example of a density of the form  $f(\omega) = ce^{-4|\omega-\omega_1|^2}$ , with  $\omega_1 = (0, 1, 0)$  and  $c = 1/0.7854$ . With this choice of density, it turns out that  $\|f\|_\infty = 1/0.7854 = 1.2732$ . Hence we can set  $M$  such that  $M = 1.2732$ . The graph of  $f$  in the spherical coordinates  $(\Phi, \Theta)$  ( $\Phi = \text{longitude}$ ,  $0 \leq \Phi \leq 2\pi$ ,  $\Theta = \text{colatitude}$ ,  $0 \leq \Theta \leq \pi$ ) is given in Fig 2. We also plot the noisy observations for different cases of perturbations.



We remark that for rather big rotation angles such that  $\phi \sim U[0, \pi/4]$  or  $\phi \sim U[0, \pi/2]$ , the observations tend to be spread over a large region on the sphere and not being concentrated in a specific region any more. Consequently, denoising might prove to be difficult. In the context of the deconvolution on the sphere, a large amount of noise corresponds to a rotation around the Oz axis with a large angle.

A fundamental issue in the field of astrophysics is to detect the place of the peak of the bell which in the present density case is localized in  $(\Theta = \pi/2, \Phi = \pi/2)$ . For each case of noise, we plot the observations both on the sphere and on the flattened sphere and give the reconstructed density in the spherical coordinates.

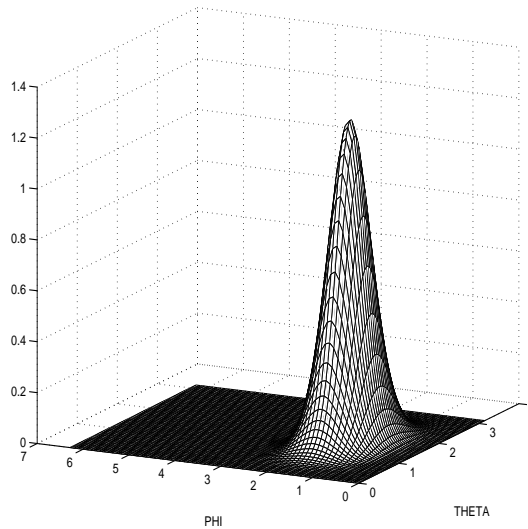


FIGURE 3.2 – The target density

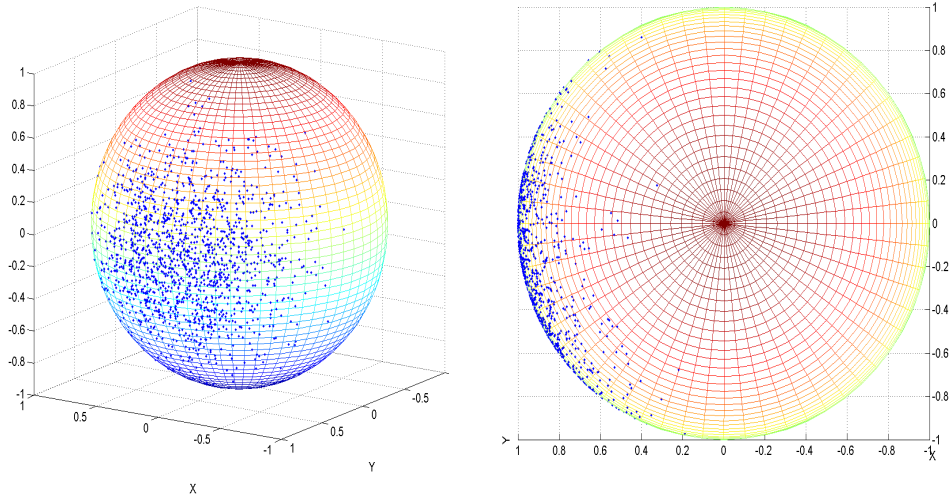


FIGURE 3.3 – Observations  $\phi \sim U[0, \pi/8]$ .

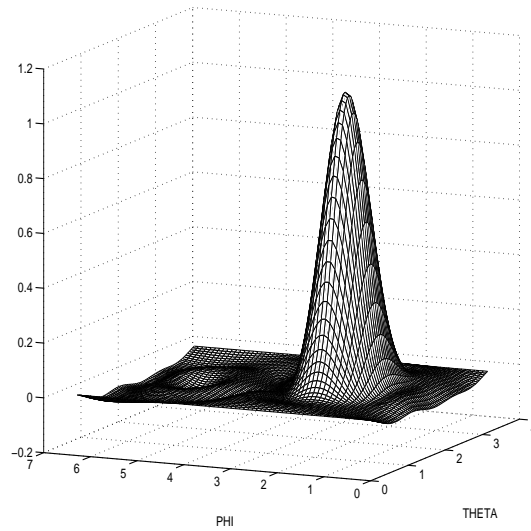


FIGURE 3.4 – The estimated density, bandwidth  $\kappa = 0.43$ .

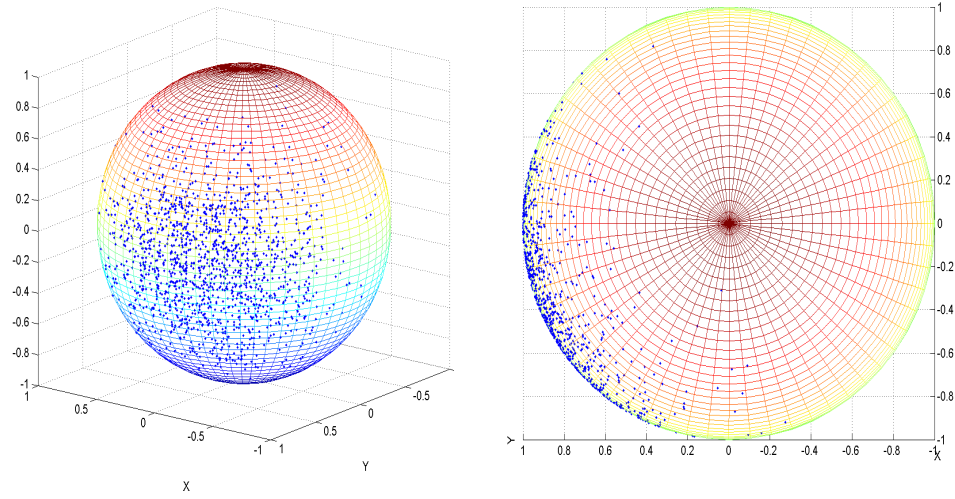


FIGURE 3.5 – Observations  $\phi \sim U[0, \pi/4]$ .

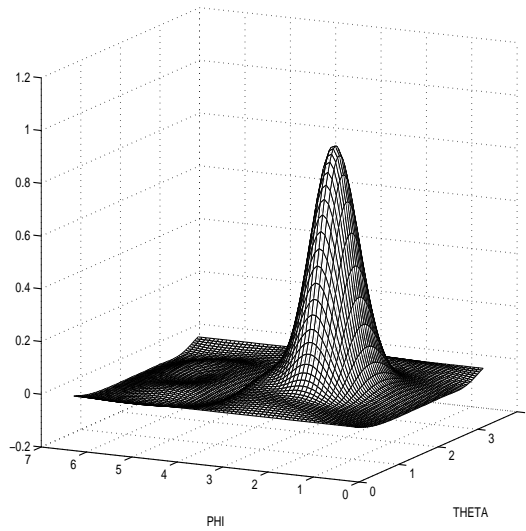


FIGURE 3.6 – The estimated density, bandwidth  $\kappa = 0.46$ .

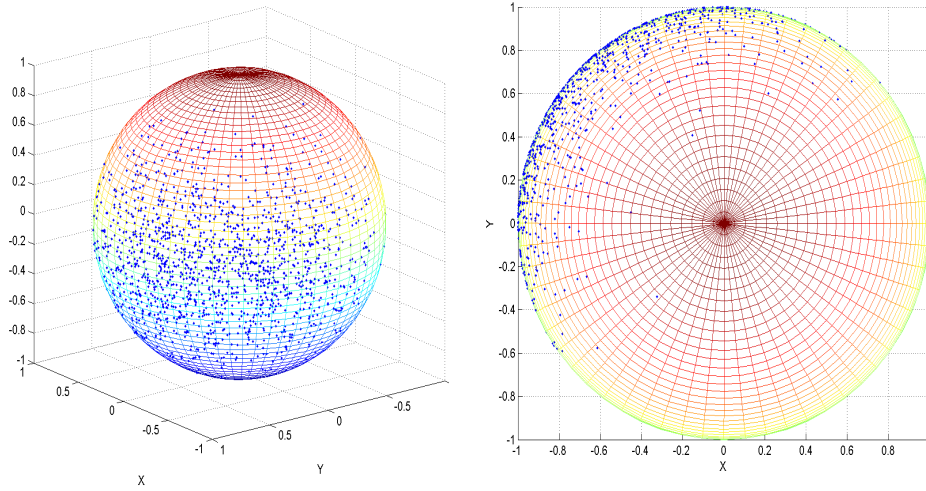


FIGURE 3.7 – Observations  $\phi \sim U[0, \pi/2]$ .

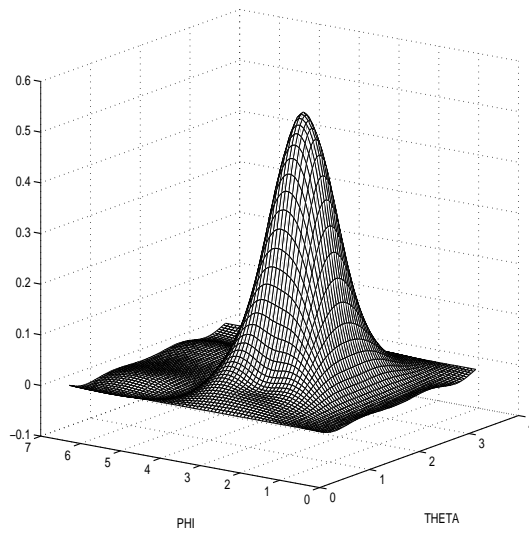


FIGURE 3.8 – The estimated density, bandwidth  $\kappa = 0.56$ .

At a closer inspection, we notice that the position of the peak of the estimated bell is pretty well localized whatever the amount of noise. Nevertheless in the case of the law  $U[0, \pi/2]$ , the longitude coordinate of the peak tends to slightly move, its colatitude coordinate remaining well localized.

Therefore even if in the case of rather big rotations such that  $\phi \sim U[0, \pi/4]$  and  $\phi \sim U[0, \pi/2]$ , our estimation procedure allows us to detect the position of the peak. Of course, one remarks that the base of the bell tends to become a bit larger when the noise increases, this is due to the fact that the observations are not concentrated in a specific region any longer, but the genuine form of the density is well preserved.

### 3.6 Proof of Theorem 3.2

In this proof,  $C$  will denote an absolute constant which may change from one line to the other.

We begin with the following proposition

**Proposition 3.2.** *For any  $p \geq 1$ ,*

$$\mathbb{E}|\hat{\beta}_{j\eta} - \beta_{j\eta}|^p \leq c\left[\frac{\sigma_j^2}{N}\right]^{\frac{p}{2}} \quad (3.41)$$

$$\mathbb{P}[|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \sigma_j \kappa t_N] \leq cN^{-\gamma(\kappa)} \quad (3.42)$$

$$\gamma(\kappa) = -\kappa^2/4 \quad (3.43)$$

To prove (3.41), we will use Rosenthal inequality :

$$\begin{aligned} \hat{\beta}_{j\eta} - \beta_{j\eta} &= \frac{1}{N} \sum_{i=1}^N G(Z_i) - \mathbb{E}G(Z_i) \\ G(x) &= \sum_{lm} \psi_{j\eta}^{lm} \sum_n \hat{f}_{\varepsilon^{-1}mn}^l Y_n^l(x). \end{aligned}$$

Hence,

$$\text{Var}G(Z) \leq \int_{\mathbb{S}^2} |G(\omega)|^2 f_Z(\omega) d\omega \leq M \int_{\mathbb{S}^2} |G(\omega)|^2 d\omega,$$

since  $\|f_Z\|_\infty \leq M$  (this is a consequence of the fact that  $s > 2/\pi$  which in turns implies that  $f$  is continuous, and so is  $f_Z$  on the sphere). Then using Parseval :

$$\int_{\mathbb{S}^2} |G(\omega)|^2 d\omega = \sum_{ln} \left| \sum_m \psi_{j\eta}^{lm} \hat{f}_{\varepsilon^{-1}mn}^l \right|^2 = \sigma_j^2.$$

Moreover, using Cauchy Schwarz inequality, and noticing again that  $\psi_{j\eta}^{lq} \neq 0$  only for  $2^{j-1} \leq l \leq 2^{j+1}$  :

$$|G(\omega)| \leq \left[ \sum_{ln} \left| \sum_m \psi_{j\eta}^{lm} \hat{f}_{\varepsilon^{-1}mn}^l \right|^2 \right]^{1/2} \left[ \sum_{lm} I\{2^{j-1} \leq l \leq 2^{j+1}\} |Y_m^l(\omega)|^2 \right]^{1/2} \leq C\sigma_j 2^{j/2},$$

since  $\sum_m |Y_m^l(\omega)|^2 = \frac{2l+1}{4\pi}$  because of the addition formula ( see Kim and Koo (2002) [91]) inequality (6.1)). Then using Rosenthal inequality, we get (we recall that  $2^j \leq N^{1/2}$  for  $j \leq J$ ) :

$$\mathbb{E} \left| \frac{1}{N} \sum_{i=1}^N G(Z_i) - \mathbb{E}G(Z_i) \right|^p \leq C \left[ \frac{1}{N^p} \sigma_j^2 (\sigma_j 2^{j/2})^{p-2} + \frac{1}{N^p} (N \sigma_j^2)^{\frac{p}{2}} \right] \leq C \sigma_j^p N^{-p/2}.$$

To prove (3.42), we use Bernstein inequality :

$$\mathbb{P}[|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \sigma_j \kappa t_N] \leq 2 \exp - \frac{\kappa^2 \sigma_j^2 \log N}{2(\sigma_j^2 + \frac{1}{3} \kappa \sigma_j t_N \sigma_j 2^{j/2} \pi^{-1/2})} \leq 2N^{-\kappa^2/4}.$$

This ends up the proof of the proposition.

Now, to get the result of Theorem 3.2, we begin by the following decomposition :

$$\begin{aligned} \mathbb{E} \|\hat{f} - f\|_p^p &\leq 2^{p-1} \{ \mathbb{E} \left\| \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} (t(\hat{\beta}_{j\eta}) - \beta_{j\eta}) \psi_{j\eta} \right\|_p^p + \left\| \sum_{j>J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_p^p \} \\ &=: I + II \end{aligned}$$

The term  $II$  is easy to analyse, as follows. We observe first that since  $B_{\pi,r}^s(M) \subset B_{p,r}^s(M')$  for  $\pi \geq p$ , this case will be assimilated to the case  $\pi = p$  and from now on, we will only consider  $\pi \leq p$ . Since  $f$  belongs to  $B_{\pi,r}^s(M)$ , using the embedding results recalled above in (3.31), we have that  $f$  also belongs to  $B_{p,r}^{s-(\frac{2}{\pi}-\frac{2}{p})}(M')$ , for some constant  $M'$  and for  $\pi \leq p$ . Hence

$$\left\| \sum_{j>J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_p \leq C 2^{-J[s-2(\frac{1}{\pi}-\frac{1}{p})]}.$$

Then we only need to verify that  $\frac{s-2(\frac{1}{\pi}-\frac{1}{p})}{\nu+1}$  is always larger than  $\mu$ , which is not difficult.

Indeed, on the first zone  $s \geq (\nu+1)(p/\pi - 1)$ . So,  $s + \nu + 1 \geq (\nu+1)\frac{p}{\pi}$  which entails that  $\frac{s}{(s+\nu+1)} \leq \frac{s}{(\nu+1)\frac{p}{\pi}}$ . We need to check that  $s - 2(\frac{1}{\pi} - \frac{1}{p}) \geq \frac{s\pi}{p}$ . We have that  $s - \frac{2}{\pi} + \frac{2}{p} - \frac{s\pi}{p} = 2(\frac{s\pi}{2} - 1)(\frac{1}{\pi} - \frac{1}{p}) \geq 0$  since  $s \geq \frac{2}{\pi}$  and  $p \geq \pi$ .

On the second zone, we obviously have that  $\frac{s-2(\frac{1}{\pi}-\frac{1}{p})}{\nu+1}$  is always larger than  $\mu = \frac{s-2/\pi+2/p}{s+\nu-2/\pi+1}$ .

Bounding the term I is more involved. Using the triangular inequality together with Hölder inequality, and property (3.28) for the second line, we get

$$\begin{aligned} I &\leq 2^{p-1} J^{p-1} \sum_{j=-1}^J \mathbb{E} \left\| \sum_{\eta \in \mathcal{Z}_j} (t(\hat{\beta}_{j\eta}) - \beta_{j\eta}) \psi_{j\eta} \right\|_p^p \\ &\leq 2^{p-1} J^{p-1} C \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} \mathbb{E} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^p \|\psi_{j\eta}\|_p^p. \end{aligned}$$

Now, we separate four cases :

$$\begin{aligned} \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} \mathbb{E} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^p \|\psi_{j\eta}\|_p^p &= \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} \mathbb{E} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^p \|\psi_{j\eta}\|_p^p \left\{ I\{|\hat{\beta}_{j\eta}| \geq \kappa t_N \sigma_j\} \right. \\ &\quad \left. + I\{|\hat{\beta}_{j\eta}| < \kappa t_N \sigma_j\} \right\} \\ &\leq \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} \left[ \mathbb{E} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^p \|\psi_{j\eta}\|_p^p I\{|\hat{\beta}_{j\eta}| \geq \kappa t_N \sigma_j\} \right. \\ &\quad \left\{ I\{|\beta_{j\eta}| \geq \frac{\kappa}{2} t_N \sigma_j\} + I\{|\beta_{j\eta}| < \frac{\kappa}{2} t_N \sigma_j\} \right\} \\ &\quad \left. + |\beta_{j\eta}|^p \|\psi_{j\eta}\|_p^p I\{|\hat{\beta}_{j\eta}| < \kappa t_N \sigma_j\} \left\{ I\{|\beta_{j\eta}| \geq 2\kappa t_N \sigma_j\} \right. \right. \\ &\quad \left. \left. + I\{|\beta_{j\eta}| < 2\kappa t_N \sigma_j\} \right\} \right] \\ &= : Bb + Bs + Sb + Ss. \end{aligned}$$

Using Proposition 3.2, we have

$$\begin{aligned} Bb &\leq C \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} \sigma_j^p N^{-p/2} \|\psi_{j\eta}\|_p^p I\{|\beta_{j\eta}| \geq \frac{\kappa}{2} t_N \sigma_j\} \\ Ss &\leq \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^p \|\psi_{j\eta}\|_p^p I\{|\beta_{j\eta}| < 2\kappa t_N \sigma_j\}. \end{aligned}$$

And, using in addition (3.25) and (3.36)

$$\begin{aligned}
 Bs &\leq \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} [\mathbb{E}|\hat{\beta}_{j\eta} - \beta_{j\eta}|^{2p}]^{1/2} [\mathbb{P}\{|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \frac{\kappa}{2} t_N \sigma_j\}]^{1/2} \|\psi_{j\eta}\|_p^p I\{|\beta_{j\eta}| < \frac{\kappa}{2} t_N \sigma_j\} \\
 &\leq C \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} \sigma_j^p N^{-p/2} 2^{1/2} N^{-\kappa^2/32} \|\psi_{j\eta}\|_p^p I\{|\beta_{j\eta}| < \frac{\kappa}{2} t_N \sigma_j\} \\
 &\leq C \sum_{j=-1}^J 2^{jp(\nu+1)} N^{-p/2} N^{-\kappa^2/32} \leq CN^{-\kappa^2/32}.
 \end{aligned}$$

Now, since  $f$  belongs to  $B_{\pi,r}^s(M)$ , using again Proposition 3.2 and (3.31), we have

$$\begin{aligned}
 Sb &\leq \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^p \|\psi_{j\eta}\|_p^p \mathbb{P}\{|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq 2\kappa t_N \sigma_j\} I\{|\beta_{j\eta}| \geq 2\kappa t_N \sigma_j\} \\
 &\leq \sum_{j=-1}^J \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^p \|\psi_{j\eta}\|_p^p 2N^{-\kappa^2} I\{|\beta_{j\eta}| \geq 2\kappa t_N \sigma_j\} \\
 &\leq C \sum_{j=-1}^J 2^{-jp(s-2(\frac{1}{\pi}-\frac{1}{p})_+)} N^{-\kappa^2} \leq CN^{-\kappa^2} 2^{2J} \leq CN^{-\kappa^2 + \frac{1}{(\nu+1)}}.
 \end{aligned}$$

It is easy to check that in any cases if  $\kappa \geq 1 - p$  the terms  $Bs$  and  $Sb$  are smaller than the rates announced in the theorem.

We have using Proposition 3.2, (3.25) and condition (3.36) for any  $p \geq z \geq 0$  :

$$\begin{aligned}
 Bb &\leq CN^{-p/2} \sum_{j=-1}^J \sigma_j^p 2^{j(p-2)} \sum_{\eta \in \mathcal{Z}_j} I\{|\beta_{j\eta}| \geq \frac{\kappa}{2} t_N \sigma_j\} \\
 &\leq CN^{-p/2} \sum_{j=-1}^J \sigma_j^p 2^{j(p-2)} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z [t_N \sigma_j]^{-z} \\
 &\leq Ct_N^{p-z} \sum_{j=-1}^J 2^{j[\nu(p-z)+p-2]} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z.
 \end{aligned}$$



Also, for any  $p \geq z \geq 0$

$$\begin{aligned} Ss &\leq C \sum_{j=-1}^J 2^{j(p-2)} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z \sigma_j^{p-z} [t_N]^{p-z} \\ &\leq C [t_N]^{p-z} \sum_{j=-1}^J 2^{j(\nu(p-z)+p-2)} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z. \end{aligned}$$

So in both cases we have the same bound to investigate. We will write this bound on the following form (forgetting the constant) :

$$A+B := t_N^{p-z_1} \left[ \sum_{j=-1}^{j_0} 2^{j[\nu(p-z_1)+p-2]} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^{z_1} \right] + t_N^{p-z_2} \left[ \sum_{j=j_0+1}^J 2^{j[\nu(p-z_2)+p-2]} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^{z_2} \right].$$

The constants  $z_i$  and  $j_0$  will be chosen depending on the cases, with the only constraint  $p \geq z_i \geq 0$ .

We recall that we only need to investigate the case  $p \geq \pi$ , since when  $p \leq \pi$ ,  $B_{\pi r}^s(M) \subset B_{pr}^s(M')$ .

Let us first consider the case where  $s \geq (\nu+1)(\frac{p}{\pi}-1)$ , put

$$q = \frac{p(\nu+1)}{s+\nu+1},$$

and observe that on the considered domain,  $q \leq \pi$  and  $p > q$ . In the sequel it will be useful to observe that we have  $s = (\nu+1)(\frac{p}{q}-1)$ . Now, taking  $z_2 = \pi$ , we get :

$$B \leq t_N^{p-\pi} \left[ \sum_{j=j_0+1}^J 2^{j[\nu(p-\pi)+p-2]} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^\pi \right].$$

Now, as

$$\frac{p}{q} - \frac{2}{\pi} + \nu \left( \frac{p}{q} - 1 \right) = s + 1 - \frac{2}{\pi}$$

and

$$\sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^\pi = 2^{-j\pi(s+1-\frac{2}{\pi})} \tau_j^\pi,$$

with  $(\tau_j)_j \in l_r$  (this last thing is a consequence of the fact that  $f \in B_{\pi,r}^s(M)$ ), we can write :

$$\begin{aligned} B &\leq t_N^{p-\pi} \sum_{j=j_0+1}^J 2^{jp(1-\frac{\pi}{q})(\nu+1)} \tau_j^\pi \\ &\leq C t_N^{p-\pi} 2^{j_0 p(1-\frac{\pi}{q})(\nu+1)}. \end{aligned}$$

The last inequality is true for any  $r \geq 1$  if  $\pi > q$  and for  $r \leq \pi$  if  $\pi = q$ . Notice that  $\pi = q$  is equivalent to  $s = (\nu + 1)(\frac{p}{\pi} - 1)$ . Now if we choose  $j_0$  such that  $2^{j_0 \frac{p}{q}(\nu+1)} \sim t_N^{-1}$  we get the bound

$$t_N^{p-q},$$

which exactly gives the rate announced in the theorem for this case. As for the first part of the sum (before  $j_0$ ), we have, taking now  $z_1 = \tilde{q}$ , with  $\tilde{q} \leq \pi$  (and also  $\tilde{q} \leq p$  since we investigate the case  $p \geq \pi$ ), so that  $[\frac{1}{2^{2j}} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^{\tilde{q}}]^{\frac{1}{\tilde{q}}} \leq [\frac{1}{2^{2j}} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^\pi]^{\frac{1}{\pi}}$ , we get

$$\begin{aligned} A &\leq t_N^{p-\tilde{q}} \left[ \sum_{-1}^{j_0} 2^{j[\nu(p-\tilde{q})+p-2]} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^{\tilde{q}} \right] \\ &\leq t_N^{p-\tilde{q}} \left[ \sum_{-1}^{j_0} 2^{j[\nu(p-\tilde{q})+p-\frac{2\tilde{q}}{\pi}]} \left[ \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^\pi \right]^{\frac{\tilde{q}}{\pi}} \right] \\ &\leq t_N^{p-\tilde{q}} \sum_{-1}^{j_0} 2^{j[(\nu+1)p(1-\frac{\tilde{q}}{q})]} \tau_j^{\tilde{q}} \\ &\leq Ct_N^{p-\tilde{q}} 2^{j_0 p[(\nu+1)(1-\frac{\tilde{q}}{q})]} \\ &\leq Ct_N^{p-q}. \end{aligned}$$

The last two lines are valid if  $\tilde{q}$  is chosen strictly smaller than  $q$  (this is possible since  $\pi \geq q$ ).

Let us now consider the case where  $s < (\nu + 1)(\frac{p}{\pi} - 1)$ , and choose now

$$q = p \frac{\nu + 1 - \frac{2}{p}}{s + \nu - \frac{2}{\pi} + 1}.$$

In such a way that we easily verify that  $p - q = p \frac{s-2/\pi+2/p}{1+\nu+s-2/\pi}$ ,  $q - \pi = \frac{(p-\pi)(1+\nu)-\pi s}{s+\nu-\frac{2}{\pi}+1} > 0$ . Furthermore we also have  $s + 1 - \frac{2}{\pi} = \frac{p}{q} - \frac{2}{q} + \nu(\frac{p}{q} - 1)$ . Hence taking  $z_1 = \pi$  and using again the fact that  $f$  belongs to  $B_{\pi,r}^s(M)$ ,

$$\begin{aligned} A &\leq t_N^{p-\pi} \left[ \sum_{-1}^{j_0} 2^{j[\nu(p-\pi)+p-2]} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^\pi \right] \\ &\leq t_N^{p-\pi} \sum_{-1}^{j_0} 2^{j[(\nu+1-\frac{2}{p})\frac{p}{q}(q-\pi)]} \tau_j^\pi \\ &\leq Ct_N^{p-\pi} 2^{j_0 [(\nu+1-\frac{2}{p})\frac{p}{q}(q-\pi)]}. \end{aligned}$$

This is true since  $\nu + 1 - \frac{2}{p}$  is also strictly positive since  $\nu + 1 > \frac{s}{\pi-1} \geq \frac{2}{p-\pi} \geq \frac{2}{p}$ . If we now take  $2^{j_0 \frac{p}{q}(\nu+1-\frac{2}{p})} \sim t_N^{-1}$  we get the bound

$$t_N^{p-q},$$

which is the rate announced in the theorem for this case.

Again, for  $B$ , we have, taking now  $z_2 = \tilde{q} > q(> \pi)$

$$B \leq t_N^{p-\tilde{q}} \left[ \sum_{j=j_0+1}^J 2^{j[\nu(p-\tilde{q})+p-2]} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^{\tilde{q}} \right].$$

But

$$\sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^{\tilde{q}} \leq C 2^{-j\tilde{q}(s+1-\frac{2}{\pi})} \tau_j^{\tilde{q}},$$

and  $s + 1 - \frac{2}{\pi} = \frac{p}{q} - \frac{2}{q} + \nu(\frac{p}{q} - 1)$ , hence

$$\begin{aligned} B &\leq C t_N^{p-\tilde{q}} \sum_{j=j_0+1} 2^{j[(\nu+1-\frac{2}{p})\frac{p}{q}(q-\tilde{q})]} \tau_j^{\tilde{q}} \\ &\leq C t_N^{p-\tilde{q}} 2^{j_0[(\nu+1-\frac{2}{p})\frac{p}{q}(q-\tilde{q})]} \\ &\leq C t_N^{p-q}, \end{aligned}$$

which completes the proof of Theorem 2.

## Chapitre 4

# Regression in random design and Bayesian warped wavelets estimators<sup>1</sup>

In this paper we deal with the regression problem in a random design setting. We investigate asymptotic optimality under minimax point of view of various Bayesian rules based on warped wavelets. We show that they nearly attain optimal minimax rates of convergence over the Besov smoothness class considered. Warped wavelets have been introduced recently, they offer very good computable and easy-to-implement properties while being well adapted to the statistical problem at hand. We particularly put emphasis on Bayesian rules leaning on small and large variance Gaussian priors and discuss their simulation performances, comparing them with a hard thresholding procedure.

### 4.1 Introduction

We observe independent pairs of variables  $(X_i, Y_i)$ , for  $i = 1, \dots, n$ , under a random design regression model :

$$Y_i = f(X_i) + \varepsilon_i, \quad 1 \leq i \leq n, \quad (4.1)$$

where  $f$  is an unknown regression function that we aim at estimating, and  $\varepsilon_i$  are independent normal errors with  $\mathbb{E}(\varepsilon_i) = 0$ ,  $\text{Var}(\varepsilon_i) = \sigma^2 < \infty$ . The

---

1. The bulk of this chapter is to appear in *Electronic Journal of Statistics* see Pham Ngoc [124]

design points  $X_i$  are assumed to be supported in the interval  $[0, 1]$  and have a density  $g$  which will be supposed to be known. Furthermore we assume that the design density  $g$  is bounded from below, i.e.  $0 < m \leq g$ , where  $m$  is a constant. Many approaches have been proposed to tackle the problem of regression in random design, we mention among others the work of Hall and Turlach [68], Kovac and Silverman [96], Antoniadis et al. [11], Cai and Brown [27] and the model selection point of view adopted by Baraud [18]. The present paper provides a Bayesian approach to this problem based on *warped* wavelet basis. Warped wavelet basis  $\{\psi_{jk}(G) \ j \geq -1, k \in \mathbb{Z}\}$  in regression with random design were recently introduced by Kerkyacharian and Picard in [87]. The authors proposed an approach which would depart as little as possible from standard wavelet thresholding procedures which enjoy optimality and adaptivity properties. These procedures have been largely investigated in the case of equispaced samples (see a series of pioneering articles by Donoho et al. [43], [44], [42]). Kerkyacharian and Picard actually pointed out that expanding the unknown regression function  $f$  in the warped basis instead of the standard wavelet basis could be very interesting. Of course, this basis has no longer the orthonormality property nonetheless it behaves under some conditions as standard wavelets. Kerkyacharian and Picard investigated the properties of this new basis. They showed that not only is it well adapted to the statistical problem at hand by avoiding unnecessary calculations, but it also offers very good theoretical features while being easily implemented. More recently Brutti [26] highlighted their easy-to-implement computational properties.

The novelty of our contribution lies in the combination of Bayesian techniques and *warped* wavelets to treat regression in random design. We actually want to investigate whether this yields optimal theoretical results and promising practical performances, which will prove to be the case. We do not deal with the case of an unknown design density  $g$  which requires further machinery and will be the object of another paper.

Bayesian techniques for shrinking wavelet coefficients have become very popular in the last few years. The majority of them were devoted to fixed design regression scheme. Let us cite among others, papers of Abramovich et al. [2], [3], Clyde et al. [33], [34], [35], [14], Chipman et al. [32], Rivoirard [127], Pensky [119] in the case of i.i.d errors not necessarily Gaussian.

Most of those works are taking as distribution prior a mixture of Gaussian distributions. In particular, Abramovich et al. in [2] and [3] have explored optimality properties of Gaussian prior mixed with a point mass at zero and

which may be viewed as an extreme case of a Gaussian mixture :

$$\beta_{jk} \sim \pi_j N(0, \tau_j^2) + (1 - \pi_j) \delta(0),$$

where  $\beta_{jk}$  are the wavelet coefficients of the unknown regression function,  $\tau_j^2 = c_1 2^{-j\alpha}$  and  $\pi_j = \min(1, c_2 2^{-j\beta})$  are the hyperparameters. This particular form was devised to capture the sparsity of the expansion of the signal in the wavelet basis.

Our approach will consist in a first time in using the same prior but in the context of warped wavelets. In Theorem 1, we show that the Bayesian estimator built, using warped wavelets with this prior and this form of hyperparameters achieves the optimal minimax rate within logarithmic term, on the considered Besov functional space. Unfortunately, the Bayesian estimator turns out not to be adaptive. Indeed, the hyperparameters depend on the Besov smoothness class index. In order to compensate this drawback, Autin et al. in [14] suggested to consider Bayesian procedures based on Gaussian prior with large variance. Following this suggestion, we will consider priors still specified in terms of a normal density mixed with a point mass at zero but with large variance Gaussian densities. In Theorem 2, we prove again that the Bayesian estimator built with this latter form of prior, still combined with warped wavelets achieves nearly optimal minimax rate of convergence while being adaptive. Eventually, our simulation results highlight the very good performances and behaviour of these Bayesian procedures, whatever the regularity of the test functions, the noise level and the design density which can be far from the uniform case.

This paper is organized as follows. In section 2 some necessary methodology is given : we start with a short review of wavelets and warped wavelets, explain the prior model and discuss the two hyperparameters form we consider. We give in section 3 some definitions of the functional spaces we consider. In section 4, we investigate the performances of our Bayesian estimators in terms of minimax rates in two cases : the first one when the Gaussian prior has small variance, the second case focuses on Gaussian prior with large variance. Section 5 is devoted to simulation results and discussion. Finally, all proofs of main results are given in the Appendix.

## 4.2 Methodology

### 4.2.1 Warped bases

Wavelet series are generated by dilations and translations of a function  $\psi$  called the mother wavelet. Let  $\phi$  denote the orthogonal father wavelet

function. The function  $\phi$  and  $\psi$  are compactly supported. Assume  $\psi$  has  $r$  vanishing moments. Let :

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k), \quad j, k \in \mathbb{Z}$$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \quad j, k \in \mathbb{Z}.$$

For a given square-integrable function  $f$  in  $\mathbb{L}_2[0, 1]$ , let us denote

$$\zeta_{j,k} = \langle f, \psi_{j,k} \rangle .$$

In this paper, we use decompositions of 1- periodic functions on wavelet basis of  $\mathbb{L}_2[0, 1]$ . We consider periodic orthonormal wavelet bases on  $[0, 1]$  which allow to have the following series representation of a function  $f$  :

$$f(x) = \sum_{j \geq -1} \sum_{k=0}^{2^j-1} \zeta_{jk} \psi_{jk}(x) \quad (4.2)$$

where we have denoted  $\psi_{-1,k} = \phi_{0,k}$  the scaling function.

We are now going to give the essential background of *warped* wavelets which were introduced in details in [87]. First of all, let us define

$$G(x) = \int_0^x g(x)dx. \quad (4.3)$$

$G$  is assumed to be a known function, continuous and strictly monotone from  $[0, 1]$  to  $[0, 1]$ .

Let us expand the regression function  $f$  in the following sense :

$$f(G^{-1})(x) = \sum_{j \geq -1} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}(x)$$

or equivalently

$$f(x) = \sum_{j \geq -1} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}(G(x))$$

where

$$\beta_{jk} = \int f(G^{-1})(x) \psi_{jk}(x) dx = \int f(x) \psi_{jk}(G(x)) g(x) dx.$$

Hence, one immediately notices that expanding  $f(G^{-1})$  in the standard basis is equivalent to expand  $f$  in the new *warped* wavelet basis  $\{\psi_{jk}(G), j \geq$

$-1, k \in \mathbb{Z}$ . This may give a natural explanation that in the follow-on, regularity conditions will be expressed not for  $f$  but for  $f(G^{-1})$ .

We set  $\hat{\beta}_{jk} = (1/n) \sum_{i=1}^n \psi_{jk}(G(X_i))Y_i$ .  $\hat{\beta}_{jk}$  is an unbiased estimate of  $\beta_{jk}$  since

$$\begin{aligned} \mathbb{E}(\hat{\beta}_{jk}) &= (1/n) \sum_{i=1}^n \mathbb{E}(\psi_{j,k}(G(X_i))(f(X_i) + \epsilon_i)) = \mathbb{E}(\psi_{j,k}(G(X))f(X)) \\ &= \int f(x)\psi_{jk}(G(x))g(x)dx = \int f(G^{-1})(x)\psi_{jk}(x)dx = \beta_{jk}. \end{aligned}$$

#### 4.2.2 Priors and estimators

We set in the following

$$\gamma_{jk}^2 = \frac{\sigma^2}{n^2} \sum_{i=1}^n \psi_{jk}^2(G(X_i)). \quad (4.4)$$

As in Abramovich et al. (see [2], [3]), we use the following prior on the wavelet coefficients  $\beta_{jk}$  of the unknown function  $f$  with respect to the *warped* basis  $\{\psi_{jk}(G), j \geq -1, k \in \mathbb{Z}\}$  :

$$\beta_{jk} \sim \pi_j N(0, \tau_j^2) + (1 - \pi_j)\delta(0).$$

Considering the  $L_1$  loss, from this form of prior we derive the following Bayesian rule which is the posterior median :

$$\tilde{\beta}_{jk} = \text{Med}(\beta_{jk}|\hat{\beta}_{jk}) = \text{sign}(\hat{\beta}_{jk}) \max(0, \zeta_{jk}) \quad (4.5)$$

where

$$\zeta_{jk} = \frac{\tau_j^2}{\gamma_{jk}^2 + \tau_j^2} |\hat{\beta}_{jk}| - \frac{\tau_j \gamma_{jk}}{\sqrt{\gamma_{jk}^2 + \tau_j^2}} \Phi^{-1}\left(\frac{1 + \min(\eta_{jk}, 1)}{2}\right) \quad (4.6)$$

where  $\Phi$  is the normal cumulative distributive function and

$$\eta_{jk} = \frac{1 - \pi_j}{\pi_j} \frac{\sqrt{\tau_j^2 + \gamma_{jk}^2}}{\gamma_{jk}} \exp\left(-\frac{\tau_j^2 \hat{\beta}_{jk}^2}{2\gamma_{jk}^2(\tau_j^2 + \gamma_{jk}^2)}\right). \quad (4.7)$$

We set :

$$w_j(n) := \frac{\pi_j}{1 - \pi_j}. \quad (4.8)$$



We introduce now the estimator of the unknown regression  $f$

$$\tilde{f}(x) = \sum_{j \leq J} \sum_{k=0}^{2^j-1} \tilde{\beta}_{jk} \psi_{jk}(G(x)), \quad (4.9)$$

where  $J$  is a parameter which will be specified later.

Note that in our case, the estimator resembles the usual ones in [14], [2] and [3], except that the deterministic noise variance has been replaced by a stochastic noise level  $\gamma_{jk}^2$ . Its expression is given by (4.4). This change will have a marked impact both on the proofs of theorems by using now large deviation inequalities and on simulation results.

Futhermore, such  $\mathbb{L}_1$  rule is of thresholding type. Indeed, as underlined in [2] and [3],  $\tilde{\beta}_{jk}$  is null whenever  $\hat{\beta}_{jk}$  falls below a certain threshold  $\lambda_B$ . Some properties of the threshold  $\lambda_B$  that will be used in the sequel are given in lemma 1 in Appendix.

### Gaussian priors with small variance

In this paper, two cases of hyperparameters will be considered. The first one involves Gaussian priors with small variances. We will state as suggested in Abramovich et al. (see [2], [3]) :

$$\tau_j^2 = c_1 2^{-j\alpha} \quad \pi_j = \min(1, c_2 2^{-j\beta}), \quad (4.10)$$

where  $\alpha$  and  $\beta$  are non-negative constants,  $c_1, c_2 > 0$ .

This choice of hyperparameters is exhaustively discussed in Abramovich et al. [3]. The authors stressed that this form of hyperparameters was actually designed in order to capture the sparsity of wavelet expansion. They pointed out the connection between Besov spaces parameters and this particular form of hyperparameters. They investigate various practical choices.

For this case of hyperparameters (4.10), the estimator of  $f$  will be denoted  $\hat{f}$ .

### Gaussian priors with large variance

The second form of hyperparameters considered in the paper involves Gaussian priors with large variance as suggested in Autin et al. [14].

As a matter of fact, we suppose that the hyperparameters do not depend on  $j$  and we set :

$$\tau_j^2 := \tau(n)^2 = 1/\sqrt{n \log(n)}. \quad (4.11)$$

Besides,  $w_j(n) := w(n)$ . We suppose that there exist  $q_1$  and  $q_2$  such that for  $n$  large enough

$$n^{-q_1/2} \leq w(n) \leq n^{-q_2/2}. \quad (4.12)$$

This form of hyperparameters was emphasized in [14] in order to mimic heavy tailed priors such as Laplace or Cauchy distributions. Indeed, Johnstone and Silverman in [77], [78] showed that their empirical Bayes approach for regular regression setting, with a prior mixing a heavy-tailed density and a point mass at zero proved fruitful, both in theory and practice. Pensky in [119] also underlined the efficiency of this kind of hyperparameters.

We underscore that contrary to the first form of hyperparameters (4.10), these latter forms (4.11) and (4.12) lead to an adaptive Bayesian estimator. For this case of hyperparameters (4.11) and (4.12), the estimator of  $f$  will be denoted  $\tilde{f}$ .

### 4.3 Functional spaces

In this paper, the functional classes of interest are Besov bodies and weak Besov bodies. Let us define them. Using the decomposition (4.2), we characterize the Besov spaces by using the following norm

$$\|f\|_{spq} = \begin{cases} [\sum_{j \geq -1} 2^{jq(s+1/2-1/p)} \|(\beta_{j,k})_k\|_{\ell_p}^q]^{1/q} & \text{if } q < \infty \\ \sup_{j \geq -1} 2^{j(s+1/2-1/p)} \|(\beta_{j,k})_k\|_{\ell_p} & \text{if } q = \infty. \end{cases}$$

If  $\max(0, 1/p - 1/2) < s < r$  and  $p, q \geq 1$

$$f \in B_{p,q}^s \iff \|f\|_{spq} < \infty.$$

The Besov spaces have the following simple relationship

$$B_{p,q_1}^{s_1} \subset B_{p,q}^s, \quad \text{for } s_1 > s \text{ or } s_1 = s \text{ and } q_1 \leq q$$

and

$$B_{p,q}^s \subset B_{p_1,q}^{s_1}, \quad \text{for } p_1 > p \text{ and } s_1 \geq s - 1/p + 1/p_1.$$

The index  $s$  indicates the smoothness of the function. The Besov spaces capture a variety of smoothness features in a function including spatially inhomogeneous behavior when  $p < 2$ .

We recall and stress that in this paper as mentioned above, the regularity conditions will be expressed for the function  $f(G^{-1})$  due to the *warped* basis

context.

More precisely we shall focus on the space  $B_{2,\infty}^s$ . We have in particular

$$f \in B_{2,\infty}^s \iff \sup_{J \geq -1} 2^{2Js} \sum_{j \geq J} \sum_k \beta_{jk}^2 < \infty. \quad (4.13)$$

We define the Besov ball of some radius  $R$  as  $B_{2,\infty}^s(R) = \{f : \|f\|_{s,2,\infty} \leq R\}$ . Let us define now the weak Besov space  $W(r, 2)$

**Definition 4.1.** *Let  $0 < r < 2$ . We say that a function  $f$  belongs to the weak Besov body  $W(r, 2)$  if and only if :*

$$\|f\|_{W_r} := \left[ \sup_{\lambda > 0} \lambda^{r-2} \sum_{j \geq -1} \sum_k \beta_{jk}^2 I\{|\beta_{jk}| \leq \lambda\} \right]^{1/2} < \infty. \quad (4.14)$$

And we have the following proposition

**Proposition 4.1.** *Let  $0 < r < 2$  and  $f \in W(r, 2)$ . Then*

$$\sup_{\lambda > 0} \lambda^r \sum_{j \geq -1} \sum_k I\{|\beta_{jk}| > \lambda\} \leq \frac{2^{2-r} \|f\|_{W_r}^2}{1 - 2^{-r}}. \quad (4.15)$$

For the proof of this proposition see for instance [88]. To conclude this section, we have the following embedding

$$B_{2,\infty}^s \subset W_{2,2/(1+2s)},$$

which is not difficult to prove (see for instance [88]).

## 4.4 Minimax performances of the procedures

### 4.4.1 Bayesian estimators based on Gaussian priors with small variances

**Théorème 4.1.** *Assume that we observe model (4.1). We consider the hyperparameters defined by (4.10). Set  $J := J_\alpha$  such that  $2^{J\alpha} = (3/(2n))^{-1/\alpha}$ . Let  $\alpha > 1$  and  $\alpha \geq s$ , then we have the following upper bound :*

$$\sup_{f \in (G^{-1}) \in B_{2,\infty}^s(R)} \mathbb{E} \|\hat{f} - f\|_2^2 = \mathcal{O}((1/n)^{1-1/\alpha} \log^2(n)) + \mathcal{O}((1/n)^{2s/\alpha}). \quad (4.16)$$

The optimal choice of the hyperparameter  $\alpha$  in Theorem 4.1 should minimize the upper bound derived in (4.16). Consequently, let us choose now in (4.16)  $\alpha = 2s + 1$ , we immediately deduce the following corollary.

**Corollary 4.1.** *If one chooses for the prior parameter  $\alpha = 2s + 1$ , one gets*

$$\sup_{f(G^{-1}) \in B_{2,\infty}^s(R)} \mathbb{E} \|\hat{f} - f\|_2^2 = \mathcal{O}(\log^2(n) n^{-2s/(2s+1)}).$$

This corollary shows that with this specific choice of hyperparameter  $\alpha$ , one recovers the minimax rate of convergence up to a logarithmic factor that one achieves in a uniform design.

#### 4.4.2 Bayesian estimators based on Gaussian priors with large variance

**Théorème 4.2.** *We consider the model (4.1). We assume that the hyperparameters are defined by (4.11) and (4.12). Set  $J := J_n$  such that  $2^{J_n} = n/\log n$ , then we have :*

$$\sup_{f(G^{-1}) \in B_{2,\infty}^s(R)} \mathbb{E} \|\check{f} - f\|_2^2 \leq C \left( \frac{\log(n)}{n} \right)^{2s/(2s+1)}.$$

It is worthwhile to make some comments about the results of Theorem 2. Here, the estimator turns out to be adaptive and contrary to the similar results in Proposition 2 in [87], we no longer have the limitation on the regularity index  $s > 1/2$ . Moreover, Kerkyacharian and Picard [87] had to stop the highest level  $J$  such that  $2^J = (n/\log(n))^{1/2}$ , here we stop at the usual level  $J_n$  such that  $2^{J_n} = n/\log(n)$ , one gets in standard thresholding .

### 4.5 Simulations and discussion

A simulation study is conducted in order to compare the numerical performances of the two Bayesian estimators based on warped wavelets and on Gaussian prior with small or large variance, described respectively in section 2.2.1 and 2.2.2 and the hard thresholding procedure using the universal threshold  $\sigma\sqrt{2\log(n)}$ , based on warped basis and introduced by Kerkyacharian and Picard [87] for the nonparametric regression model in a random design setting. For more details on Kerkyacharian and Picard procedure, the readers are referred to Willer [155], see also [60]. In fact, we have decided

to concentrate on the procedure of Kerkyacharian and Picard. Indeed, it is interesting to point out differences and compare performances obtained by Bayesian procedures which apply local thresholds and a universal threshold procedure.

The main difficulties lie in implementing the Bayesian procedures with the stochastic variance (4.4). Note also the responses proposed by Amato et al. [7] and Kovac and Silverman [96].

All the simulations done in the present paper have been conducted with MATLAB and the Wavelet toolbox of MATLAB.

We consider here four test functions of Donoho and Johnstone [42] representing different level of spatial variability. The test functions are plotted in Fig. 1. For each of the four objects under study, we compare the three estimators at two noise levels, one with signal-to-noise ratio  $RSNR = 4$  and another with  $RSNR = 7$ . As in Willer [155], we also consider different cases of design density which are plotted in Fig. 2. The first two densities are uniform or slightly varying whereas the last two ones aim at depicting the case where a hole occurs in the density design. The sample size is equal to  $n = 1024$  and the wavelet we used is the Symmlet8.

In order to compare the behaviors of the estimators, the RMSE criterion was retained. More precisely, if  $\hat{f}(X_i)$  is the estimated function value at  $X_i$  and  $n$  is the sample size, then

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{f}(X_i) - f(X_i))^2}. \quad (4.17)$$

The RMSE displayed in Tab. 1 are computed as the average over 100 runs of expression (4.17). In each run, we hold all factors constant, except the design points (random design) and the noise process that were regenerated. E1 corresponds to the Bayesian estimator based on Gaussian prior with large variance, E2 to the Bayesian estimator based on Gaussian prior with small variance and E3 to the estimator built following the Kerkyacharian and Picard procedure in [87].

In order to implement E1, we made the following choices of hyperparameters described in section 2.2.2 : in (4.12),  $q_1 = q_2 = q = 1$  proved to be a good compromise whatever the function of interest to be estimated while leading to good graphic reconstructions. We set  $w(n) = 20 \times n^{-q/2}$  and  $\tau(n) = 20 \times \sigma^2 / (n \log(n))$ . To implement E2, we set  $c_1 = 1$ ,  $c_2 = 2$ ,  $\alpha = 0.5$  and  $\beta = 1$ , following the choices recommended in [3].

The following plots compare the visual quality of the reconstructions (see

Fig. 3. to Fig. 8). The solid line is the estimator and the dotted line is the true function.

		RSNR=4			RSNR=7		
design density		E1	E2	E3	E1	E2	E3
<i>Blocks</i>	Sine	0.0194	0.0219	0.0227	0.0113	0.0161	0.0129
	Hole2	0.0196	0.0220	0.0226	0.0114	0.0163	0.0130
<i>Bumps</i>	Sine	0.0243	0.240	0.259	0.0156	0.0167	0.0172
	Hole2	0.0241	0.0237	0.0253	0.0155	0.0167	0.0169
<i>HeaviSine</i>	Sine	0.0164	0.0141	0.0133	0.0103	0.0092	0.0093
	Hole2	0.0169	0.0146	0.0138	0.0107	0.0097	0.0096
<i>Doppler</i>	Sine	0.0236	0.0231	0.0236	0.0157	0.0238	0.0248
	Hole2	0.0244	0.0238	0.0248	0.0166	0.0172	0.0176

TABLE 4.1 – Values of RMSE over 100 runs

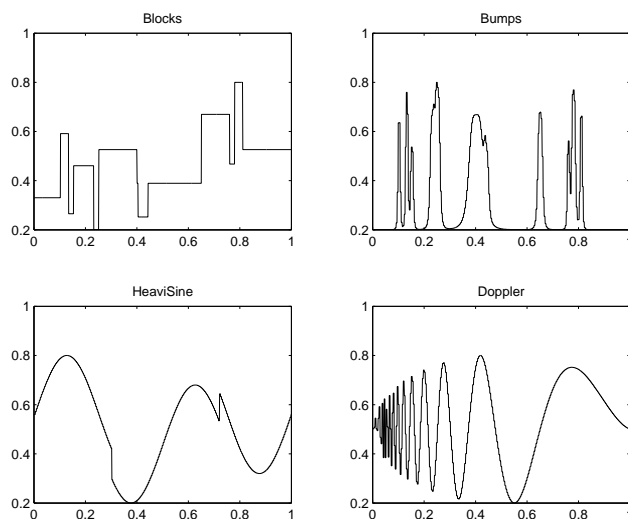


Fig. 1 Test functions

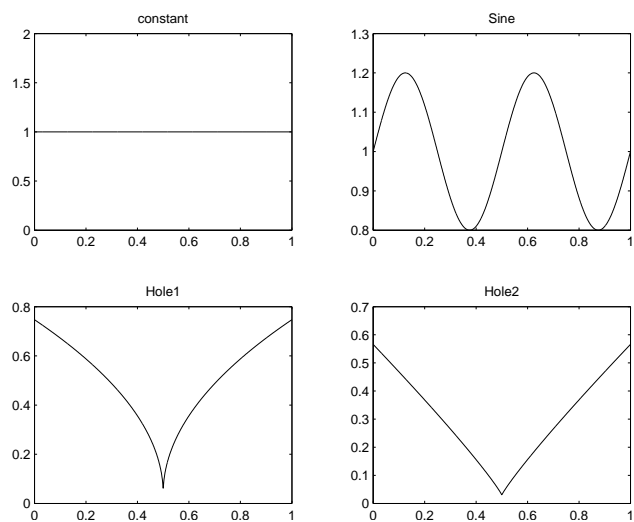


Fig. 2 Design density

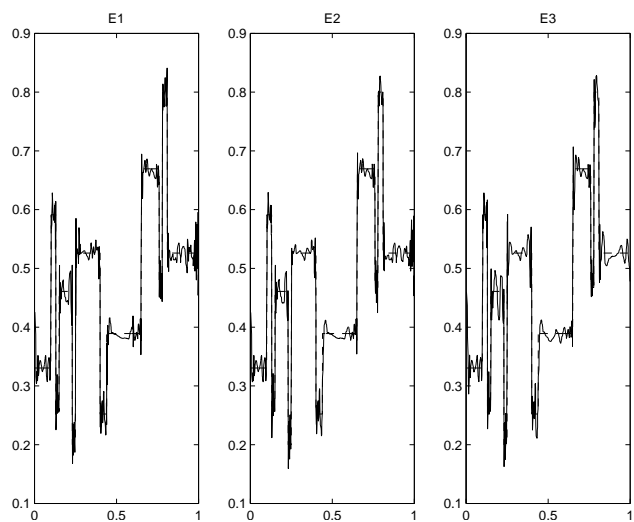


Fig. 3 Blocks target and Sine density, RSNR=4

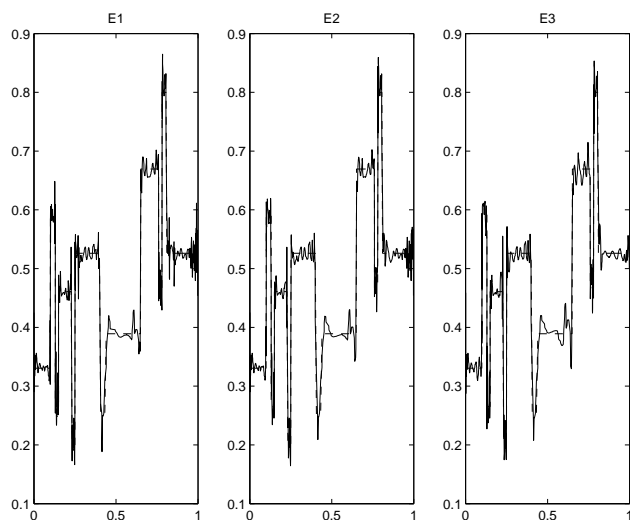


Fig. 4 Blocks target and Hole2 design density, RSNR=4

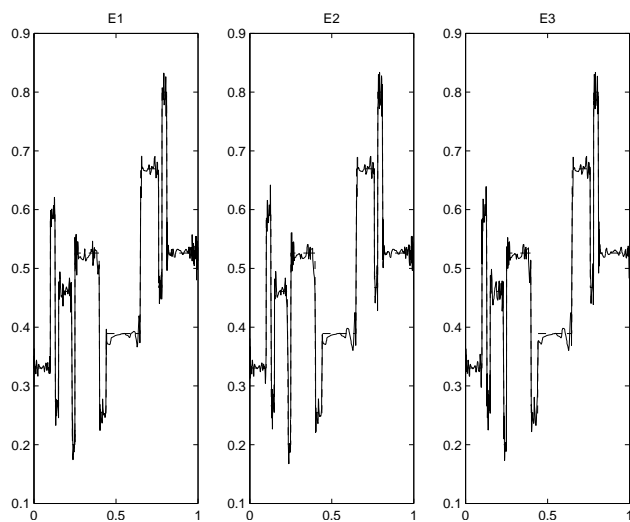


Fig. 5 Blocks target and Hole2 design density, RSNR=7



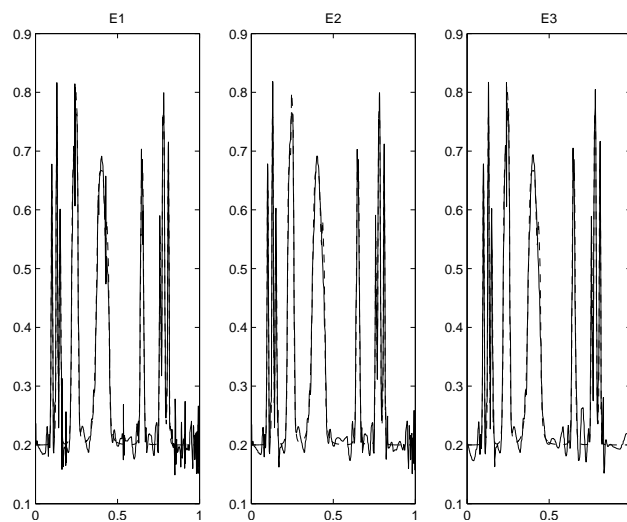


Fig. 6 Bumps target and Sine design density, RSNR=4

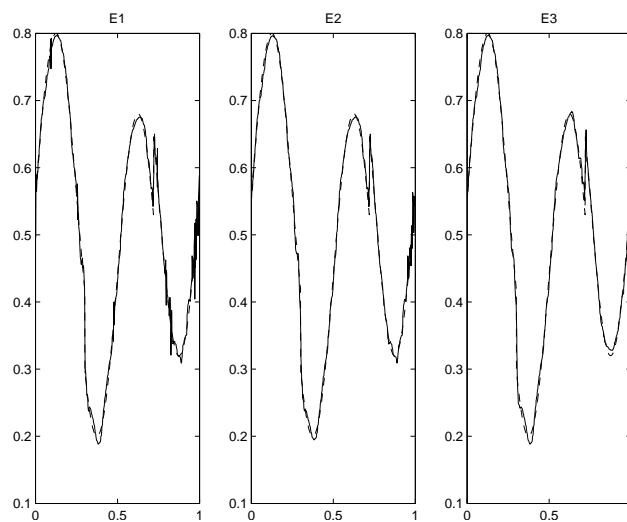


Fig. 7 HeaviSine target and Sine design density, RSNR=7

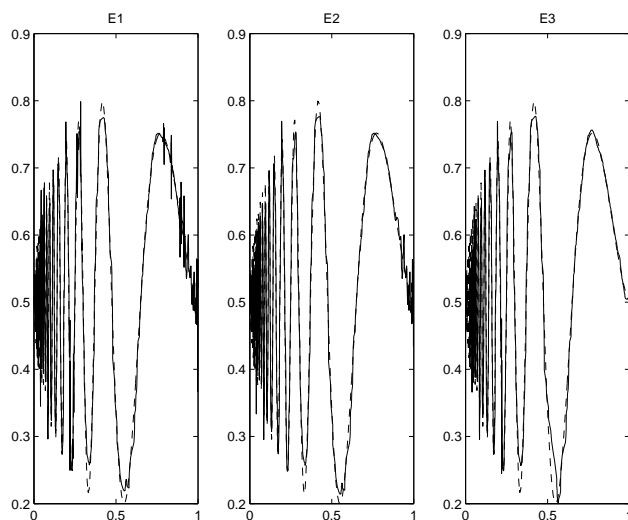


Fig. 8 Doppler target and Hole2 design density, SNR=4

We shall now comment and discuss the results displayed in Tab.1 as well as the various visual reconstructions.

The performances are always better for the Bayesian estimators except for the case of the HeaviSine test function. More precisely, the RMSE for Blocks whatever the noise level and design densities are smaller for Estimator 1. Moreover the RMSE are almost equal for Estimator 1 and 2 in the case of Bumps test function, whatever the design densities and for a noise level RSNR=4. This may be due to the irregularity of the Bumps, Blocks and Doppler test functions which are much rougher than the HeaviSine which is more regular. Indeed, Estimator 1 and 2 tend to detect better the corner of Blocks, the high peaks in Bumps, and the high frequency parts of Doppler as the graphics show it. We may explain this by the fact that Estimators 1 and 2 have level-dependent thresholds whereas Estimator 3 has a hard universal threshold.

As for the reconstructions, one can see that they are slightly better in the case of Sine density and small noise, whereas there are small deteriorations when a hole occurs in the design density, but this change does not affect the visual quality in too big proportions. This fact highlights the interest of "warping" the wavelet basis. Warping the basis allows the estimators to behave still correctly when the design densities are far from the uniform density such as in the case of Hole2.

## 4.6 Proofs

In the sequel  $C$  denotes some positive constant which may change from one line to another line. We also assume without loss of generality that  $\sigma = 1$  in model (4.1).

We have that

$$\mathbb{E}(\psi_{jk}^2(G(X))) = \int \psi_{jk}^2(G(x))g(x)dx = \int \psi_{jk}^2(y)dy = 1.$$

hence we get  $\mathbb{E}(\gamma_{jk}^2) = 1/n$ , the expression of  $\gamma_{jk}^2$  being given by (4.4). Let us define the following event :

$$\Omega_n^\delta = \{|\gamma_{jk}^2 - 1/n| \leq \delta\}. \quad (4.18)$$

To make proofs clearer, we recall the Bernstein inequality that we will use in the sequel. (see in [107] Proposition 2.8 and formula (2.16))

**Proposition 4.2.** *Let  $Z_1, \dots, Z_n$  be independent and square integrable random variables such that for some nonnegative constant  $b$ ,  $Z_i \leq b$  almost surely for all  $i \leq n$ . Let*

$$S = \sum_{i=1}^n (Z_i - \mathbb{E}[Z_i])$$

and  $v = \sum_{i=1}^n \mathbb{E}(Z_i^2)$ . Then for any positive  $x$ , we have

$$\mathbb{P}[S \geq x] \leq \exp\left(\frac{-v}{b^2} h\left(\frac{bx}{v}\right)\right)$$

where  $h(u) = (1+u) \log(1+u) - u$ .

It is easy to prove that

$$h(u) \leq \frac{u^2}{2(1+u/3)}$$

which immediately yields

$$\mathbb{P}[S \geq x] \leq \exp\left(\frac{-x^2}{2(v+bx/3)}\right).$$

**Lemma 4.1.** *Let  $\varsigma$  be some positive constant. We have*

$$\mathbb{P}(|\gamma_{jk}^2 - 1/n| > \varsigma/n) \leq 2e^{-n^{1-1/\alpha} \frac{\varsigma^2}{2C(1+\varsigma/3)}} \quad \forall j \leq J_\alpha \quad (4.19)$$

$$\mathbb{P}(|\gamma_{jk}^2 - 1/n| > \varsigma/n) \leq 2e^{-\varsigma^2 \log(n)/(C\|\psi\|_4^4 + \varsigma\|\psi\|_\infty^2)} \quad \forall j \leq J_n. \quad (4.20)$$

**Proof of Lemma 4.1**

Let us deal with the first case  $j \leq J_\alpha$ . To bound  $\mathbb{P}(|\gamma_{jk}^2 - 1/n| > \varsigma/n)$  we will use the Bernstein inequality and apply Proposition 4.2. In the present situation  $Z_i = (1/n^2)\psi_{jk}^2(G(X_i))$ .

First of all, in order to apply the Bernstein inequality, we need the value of the sum

$$v = \sum_{i=1}^n \mathbb{E}[(1/n^2)\psi_{jk}^2(G(X_i))]^2$$

we have

$$\begin{aligned} \mathbb{E}\psi_{j,k}^4(G(X)) &= \int_0^1 \psi_{j,k}^4(G(x))g(x)dx = \int_0^1 \psi_{j,k}^4(y)dy \\ &\leq \int_0^1 2^{2j}\psi^4(2^j y - k)dy \leq 2^j \int \psi^4(y)dy \leq C\|\psi\|_4^4 \end{aligned} \quad (4.21)$$

hence

$$(1/n^4) \sum_{i=1}^n \mathbb{E}\psi_{j,k}^4(G(X_i)) \leq (C/n^3)2^{J_\alpha} = \frac{C}{n^{3-1/\alpha}}$$

moreover

$$\psi_{jk}^2(G(X)) \leq \|\psi\|_\infty^2 2^j \leq Cn^{1/\alpha} \quad j \leq J_\alpha \quad a.s$$

so

$$\mathbb{P}(|\gamma_{jk}^2 - 1/n| > \varsigma/n) \leq 2 \exp\left(-\frac{\varsigma^2}{2C(1+\varsigma/3)} \frac{n^{-2}}{n^{-3+1/\alpha}}\right).$$

Let us now deal with the second case  $j \leq J_n$ . To bound  $\mathbb{P}(|\gamma_{jk}^2 - 1/n| > \varsigma/n)$  we will follow the lines of the proof of the first case. Here again

$$Z_i = 1/n^2\psi_{jk}^2(G(X_i)).$$

According to (4.21), we have

$$\mathbb{E}(1/n^4\psi_{jk}^4(G(X))) \leq C2^j/n^4 \leq C/(n^3 \log(n)), \quad j \leq J_n$$

and

$$v = \sum_{i=1}^n \mathbb{E}(1/n^4\psi_{jk}^4(G(X))) \leq C\|\psi\|_4^4/(n^2 \log(n))$$

and

$$1/n^2\psi_{jk}^2(G(X)) \leq \|\psi\|_\infty^2 2^j/(n^2) \leq \|\psi\|_\infty^2/(n \log(n)), \quad j \leq J_n \quad a.s$$

consequently

$$\mathbb{P}(|\gamma_{jk}^2 - 1/n| > \varsigma/n) \leq 2e^{-\varsigma^2 \log(n)/(C\|\psi\|_4^4 + \varsigma\|\psi\|_\infty^2)}.$$

The following lemma shows that the properties of the Bayesian estimators  $\check{f}$  and  $\hat{f}$  can be controlled on the event  $\Omega_n^\delta$ . To lighten the notations for the proof of this lemma, we will denote  $\Omega_n$  for  $\Omega_n^\delta$  and  $\Omega_n^c$  the complementary of  $\Omega_n$ .

**Lemma 4.2.** *We have*

$$\begin{aligned} \mathbb{E}[I(\Omega_n^c)\|\check{f} - f\|_2^2] &= o((\log(n)/n)^{2s/(2s+1)}) \\ \mathbb{E}[I(\Omega_n^c)\|\hat{f} - f\|_2^2] &= o((1/n)^{1-1/\alpha} \log(n)). \end{aligned}$$

**Proof of Lemma 4.2.**

We have

$$\begin{aligned} \mathbb{E}[I(\Omega_n^c)\|\check{f} - f\|_2^2] &\leq CJ_n \mathbb{E} \left[ \sum_{j \leq J_n} \sum_k (\tilde{\beta}_{jk} - \beta_{jk})^2 I(\Omega_n^c) \right] + \mathbb{P}(\Omega_n^c) \sum_{j > J_n} \sum_k \beta_{jk}^2 \\ &\leq V + B. \end{aligned}$$

Let us first deal with the variance term  $V$ . The estimator  $\tilde{\beta}_{jk}$  can be written as  $\tilde{\beta}_{jk} = w_{jk}\hat{\beta}_{jk}$  with  $0 \leq w_{jk} \leq 1$ . We have

$$\begin{aligned} V &\leq CJ_n \mathbb{E} \left[ \sum_{j \leq J_n, k} \left( w_{jk}(\hat{\beta}_{jk} - \beta_{jk}) - (1 - w_{jk})\beta_{jk} \right)^2 I(\Omega_n^c) \right] \\ &\leq 2CJ_n \mathbb{E} \left[ \sum_{j \leq J_n} \sum_k w_{jk}^2 (\hat{\beta}_{jk} - \beta_{jk})^2 I(\Omega_n^c) \right] + 2CJ_n \sum_{j \leq J_n} \sum_k \mathbb{E} [(1 - w_{jk})^2 \beta_{jk}^2 I(\Omega_n^c)] \\ &\leq 2CJ_n \mathbb{E} \left[ \sum_{j \leq J_n} \sum_k (\hat{\beta}_{jk} - \beta_{jk})^2 I(\Omega_n^c) \right] + 2CJ_n \sum_{j \leq J_n} \sum_k \mathbb{E} [\beta_{jk}^2 I(\Omega_n^c)] \end{aligned}$$

because  $0 \leq w_{jk} \leq 1$ . Then, using Cauchy Scharwz inequality we get

$$V \leq 2CJ_n \sum_{j \leq J_n} \sum_k \left[ \mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^4 \right]^{\frac{1}{2}} \mathbb{P}(\Omega_n^c)^{\frac{1}{2}} + 2CJ_n \sum_{j \leq J_n} \sum_k \beta_{jk}^2 \mathbb{P}(\Omega_n^c).$$

Using (4.20) and (4.40) we have

$$V \leq 2CJ_n 2^{J_n} e^{-\varsigma^2 \log(n)/(2C\|\psi\|_4^4/n + \varsigma\|\psi\|_\infty^2)} + 2CJ_n \|f(G^{-1})\|_2^2 e^{-\varsigma^2 \log(n)/(C\|\psi\|_4^4 + \varsigma\|\psi\|_\infty^2)}.$$

We recall that  $2^{J_n} = n/\log(n)$ , accordingly by choosing  $\varsigma$  large enough we have

$$V = o((\log(n)/n)^{2s/(2s+1)})$$

As for the term  $B$

$$B \leq C 2^{-2J_n s} e^{-\varsigma^2 \log(n)/(C\|\psi\|_4^4 + \varsigma\|\psi\|_\infty^2)}$$

which completes the proof for  $\tilde{f}$ .

The proof for  $\hat{f}$  is similar, all inequalities hold a fortiori since, in the case of the estimator  $\hat{f}$  we have  $\mathbb{P}(\Omega_n^c) \leq e^{-Cn^{1-1/\alpha}}$  (see (4.19)).

We consider the setting of Theorem 1. We recall that  $\tilde{\beta}_{jk}$  is zero whenever  $|\hat{\beta}_{jk}|$  falls below a threshold  $\lambda_B$  and we have the following lemma concerning the behavior of  $\lambda_B$ .

**Lemma 4.3.** *On the event  $\Omega_n^\delta$  defined by (4.18) with  $\delta = 1/(2n)$ , for  $\alpha > 1$  we have*

$$\lambda_B \approx \sqrt{\frac{\log(n)}{n}}, \quad j < J_\alpha \tag{4.22}$$

and  $J_\alpha$  is taken such that  $2^{J_\alpha} = (\frac{3}{2n})^{-1/\alpha}$ .

**Proof of Lemma 4.3.** We follow the lines of the proof of lemma 1. in [2].

On the one hand we have (see proof of lemma 1. in [2] page 228)

$$\lambda_B^2 \leq \frac{2\gamma_{jk}^2(\gamma_{jk}^2 + \tau_j^2)}{\tau_j^2} \log \left( \frac{1 - \pi_j}{\pi_j} \frac{\sqrt{\gamma_{jk}^2 + \tau_j^2}}{\gamma_{jk}} + c \right)$$

where  $c$  is some suitable positive constant. Besides, we have  $1/(2n) \leq \gamma_{jk}^2 \leq 3/(2n)$ , therefore

$$\lambda_B^2 \leq \frac{2(3/(2n))((3/(2n)) + c_1(3/(2n)))}{c_1(3/(2n))} \log \left( \frac{1 - c_2(3/(2n))^{\beta/\alpha}}{c_2(3/(2n))^{\beta/\alpha}} \frac{\sqrt{(1 + c_1)(3/(2n))}}{\sqrt{1/(2n)}} + c \right)$$

hence we get

$$\lambda_B^2 \leq \tilde{c}(1/n) \log(\tilde{c}(1/n)^{-\frac{\beta}{\alpha}} + c)$$

where  $\tilde{c}$  denotes a positive constant depending on  $c_1$  and  $c_2$  and which may be different at different places. Since

$$\tilde{c}(1/n) \log(\tilde{c}(1/n)^{(-\frac{\beta}{\alpha})} + c) \approx -\tilde{c}(\beta/\alpha)(1/n) \log(1/n)$$

we finally get

$$\lambda_B^2 \leq -\tilde{c}(\beta/\alpha)(1/n) \log(1/n).$$

On the other hand, for the reverse inequality, we have (see proof of lemma 1. in [2] page 228 and formula (14) in [2] page 221)

$$\lambda_B^2 \geq \frac{2\gamma_{jk}^2(\gamma_{jk}^2 + \tau_j^2)}{\tau_j^2} \log\left(\frac{1 - \pi_j}{\pi_j} \frac{\sqrt{\gamma_{jk}^2 + \tau_j^2}}{\gamma_{jk}}\right)$$

but  $|\gamma_{jk}^2 - 1/n| \leq 1/(2n)$  consequently one has

$$\lambda_B^2 \geq -\tilde{c}(\beta/\alpha)(1/n)(\log(1/n))$$

which completes the proof.

**Proof of Theorem 1.**

On the event  $\Omega_n^\delta$  defined by (4.18) with  $\delta = 1/(2n)$ , by the usual decomposition of the MISE into a variance and a bias term, we get

$$\begin{aligned} \mathbb{E}\|\hat{f} - f\|_2^2 &\leq 2[\mathbb{E}\|\sum_{j \leq J_\alpha} \sum_k (\tilde{\beta}_{jk} - \beta_{jk})\psi_{j,k}(G)\|_2^2 + \|\sum_{j > J_\alpha} \sum_k \beta_{jk}\psi_{j,k}(G)\|_2^2] \\ &\leq 2(V + B) \end{aligned}$$

with

$$\begin{aligned} V &= \mathbb{E}\|\sum_{j \leq J_\alpha} \sum_k (\tilde{\beta}_{jk} - \beta_{jk})\psi_{j,k}(G)\|_2^2 \\ B &= \|\sum_{j > J_\alpha} \sum_k \beta_{jk}\psi_{j,k}(G)\|_2^2. \end{aligned}$$

We first deal with the term  $V$ . We have

$$\|\sum_{j \leq J_\alpha} \sum_k (\tilde{\beta}_{jk} - \beta_{jk})\psi_{j,k}(G)\|_2^2 \leq J_\alpha \sum_{j \leq J_\alpha} \|\sum_k (\tilde{\beta}_{jk} - \beta_{jk})\psi_{j,k}(G)\|_2^2.$$

We want to show that

$$\|\sum_k (\tilde{\beta}_{jk} - \beta_{jk})\psi_{j,k}(G)\|_2^2 \leq C \sum_k (\tilde{\beta}_{jk} - \beta_{jk})^2.$$

For this purpose we have

$$\begin{aligned} \left\| \sum_k (\tilde{\beta}_{jk} - \beta_{jk}) \psi_{j,k}(G) \right\|_2^2 &= \int \left| \sum_k (\tilde{\beta}_{jk} - \beta_{jk}) \psi_{j,k}(G(x)) \right|^2 dx \\ &= \int \left| \sum_k (\tilde{\beta}_{jk} - \beta_{jk}) \psi_{j,k}(x) \right|^2 \frac{1}{g(G^{-1}(x))} dx \\ &= \left\| \sum_k (\tilde{\beta}_{jk} - \beta_{jk}) \psi_{j,k} \right\|_{\mathbb{L}_2(\varrho)}^2 \end{aligned}$$

where  $\varrho(x) = 1/(g(G^{-1}))(x)$ .

Now using inequality (44) p. 1075 in [87] we have

$$\left\| \sum_k (\tilde{\beta}_{jk} - \beta_{jk}) \psi_{j,k} \right\|_{\mathbb{L}_2(\varrho)}^2 \leq C 2^j \sum_k |\tilde{\beta}_{jk} - \beta_{jk}|^2 \varrho(I_{j,k})$$

where  $I_{j,k}$  denotes the interval  $[\frac{k}{2^j}, \frac{k+1}{2^j}]$  and  $\varrho(I_{j,k}) = \int_{I_{j,k}} \varrho(x) dx$ . But the design density  $g$  is bounded from below by  $m$ . Hence we get

$$\varrho(I_{j,k}) \leq 2^{-j}/m$$

and consequently

$$\left\| \sum_k (\tilde{\beta}_{jk} - \beta_{jk}) \psi_{j,k} \right\|_{\mathbb{L}_2(\varrho)}^2 \leq C \sum_k (\tilde{\beta}_{jk} - \beta_{jk})^2.$$

We decompose now  $V$  into three terms

$$V \leq C J_\alpha \mathbb{E} \sum_{j \leq J_\alpha} \sum_k [(\tilde{\beta}_{jk} - \beta'_{jk})^2 + (\beta'_{jk} - \beta''_{jk})^2 + (\beta''_{jk} - \beta_{jk})^2]$$

where

$$\beta'_{jk} = b_j \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq \kappa \lambda_B\}$$

with  $\kappa$  a positive constant and

$$b_j = \frac{\tau_j^2}{\tau_j^2 + \gamma_{jk}^2}$$

$$\beta''_{jk} = b_j \beta_{jk}.$$

As a consequence we have

$$V \leq C J_\alpha (A_1 + A_2 + A_3).$$



We are now going to upperbound each term  $A_1$ ,  $A_2$  and  $A_3$ . We start with  $A_1$

$$A_1 = \sum_{j \leq J_\alpha} \sum_k \mathbb{E}[(\tilde{\beta}_{jk} - \beta'_{jk})^2].$$

As precised in the beginning of section 2.2 p 6,  $\tilde{\beta}_{jk} = 0$  for  $|\hat{\beta}_{jk}| < \lambda_B$ . As well,  $\beta'_{jk} = 0$  for  $|\hat{\beta}_{jk}| < \kappa \lambda_B$  and  $\tilde{\beta}_{jk} - \beta'_{jk} \rightarrow 0$  monotonically as  $\hat{\beta}_{jk} \rightarrow \infty$ . Hence

$$\max_{\hat{\beta}_{jk}} |\tilde{\beta}_{jk} - \beta'_{jk}| = b_j \lambda_B$$

which implies

$$A_1 \leq C \sum_{j < J_\alpha} \sum_k \mathbb{E}(b_j^2 \lambda_B^2).$$

We have  $\lambda_B \approx \sqrt{\frac{\log n}{n}}$  and  $b_j \leq 1$  for  $j \leq J_\alpha$  hence we get

$$A_1 \leq C \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \frac{\log(n)}{n}$$

so

$$A_1 \leq C \frac{\log(n)}{n} \sum_{j \leq J_\alpha} 2^j \tag{4.23}$$

$$\leq C \frac{\log(n)}{n} \left(\frac{1}{n}\right)^{-1/\alpha} \tag{4.24}$$

finally

$$A_1 = \mathcal{O}(\log(n) \left(\frac{1}{n}\right)^{1-1/\alpha})$$

Let us now consider the second term  $A_2$

$$\begin{aligned} A_2 &= \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \mathbb{E}(\beta'_{jk} - \beta''_{jk})^2 \\ &= \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \mathbb{E}(b_j \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq \kappa \lambda_B\} - b_j \beta_{jk})^2 \end{aligned}$$

We have that  $b_j \leq 1$ , consequently it follows

$$\begin{aligned} A_2 &= \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \mathbb{E}((\hat{\beta}_{jk} - \beta_{jk})^2 I\{|\hat{\beta}_{jk}| \geq \kappa \lambda_B\}) + \mathbb{E} \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \beta_{jk}^2 I\{|\hat{\beta}_{jk}| < \kappa \lambda_B\} \\ &= A'_2 + A''_2 \end{aligned}$$

We have

$$A'_2 \leq \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^2.$$

Using inequality (64) in [87] p. 1086 we have

$$\mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^2 \leq C \frac{1 + \|f\|_\infty^2}{n} \quad (4.25)$$

hence

$$A'_2 = \mathcal{O}((1/n)^{1-1/\alpha}).$$

We now bound the term  $A''_2$  :

$$\begin{aligned} A''_2 &= \mathbb{E} \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \beta_{jk}^2 I\{|\hat{\beta}_{jk}| < \kappa \lambda_B\} (I\{|\beta_{jk}| < 2\kappa \lambda_B\} + I\{|\beta_{jk}| > 2\kappa \lambda_B\}) \\ &\leq \mathbb{E} \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \beta_{jk}^2 I\{|\beta_{jk}| < 2\kappa \lambda_B\} + \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \beta_{jk}^2 \mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > \kappa \lambda_B) \\ &= T_3 + T_4 \end{aligned} \quad (4.26)$$

We have

$$T_3 \leq C \sum_{j \leq J_\alpha} \lambda_B^2 2^j \leq C \frac{\log(n)}{n} n^{1/\alpha} = C \log(n) n^{-1+1/\alpha}.$$

Let us focus on  $T_4$ , we have

$$\hat{\beta}_{jk} - \beta_{jk} = 1/n \sum_{i=1}^n \psi_{j,k}(G(X_i))(f(X_i) + \varepsilon_i) - \mathbb{E} \psi_{j,k}(G(X))f(X)$$

Hence

$$\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > \kappa \sqrt{\log(n)/n}) \leq \mathbb{P}_1 + \mathbb{P}_2$$

where

$$\mathbb{P}_1 = \mathbb{P}\left(\left|1/n \sum_{i=1}^n \psi_{j,k}(G(X_i))(f(X_i)) - \mathbb{E}\psi_{j,k}(G(X))f(X)\right| > \kappa/2\sqrt{(\log(n)/n)}\right) \quad (4.27)$$

and

$$\mathbb{P}_2 = \mathbb{P}\left(\left|1/n \sum_{i=1}^n \psi_{j,k}(G(X_i))\varepsilon_i\right| > \kappa/2\sqrt{(\log(n)/n)}\right) \quad (4.28)$$

Kerkyacharian and Picard in [87] in order to prove inequality (65) in [87] showed p. 1088 that

$$\mathbb{P}_1 \leq 2 \exp\left(-\frac{3\kappa^2 \log(n)}{4\|f\|_\infty(3 + \kappa)}\right) \quad (4.29)$$

if  $2^j \leq n/\log(n)$ . As for  $\mathbb{P}_2$ , conditionally on  $(X_1, \dots, X_n)$  we have

$$1/n \sum_{i=1}^n \psi_{j,k}(G(X_i))\varepsilon_i \sim N(0, \gamma_{jk}^2),$$

where  $\gamma_{jk}^2$  has been defined in (4.4). Using exponential inequality for Gaussian random variable we have

$$\begin{aligned} \mathbb{P}_2 &\leq \mathbb{E}\left(\exp\left(-\frac{\kappa^2 \log(n)}{8n\gamma_{jk}^2}\right)\right) \\ &= \mathbb{E}e^{-\frac{\kappa^2 \log(n)}{8n\gamma_{jk}^2}} \left(I(|\gamma_{jk}^2 - 1/n| \leq 1/2n) + I(|\gamma_{jk}^2 - 1/n| > 1/(2n))\right) \\ &\leq e^{-\frac{\kappa^2 \log(n)}{12}} + \mathbb{P}(|\gamma_{jk}^2 - 1/n| > 1/(2n)). \end{aligned} \quad (4.30)$$

Using (4.19) with  $\varsigma = 1/2$ , we have for  $\alpha > 1$

$$\begin{aligned} T_4 &\leq (2e^{-Cn^{1-1/\alpha}} + e^{-\frac{\kappa^2 \log(n)}{12}} + 2 \exp\left(\frac{-3\kappa^2 \log(n)}{4\|f\|_\infty(3 + \kappa)}\right)) \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \beta_{jk}^2 \\ &\leq (2e^{-Cn^{1-1/\alpha}} + e^{-\frac{\kappa^2 \log(n)}{12}} + 2 \exp\left(\frac{-3\kappa^2 \log(n)}{4\|f\|_\infty(3 + \kappa)}\right)) \|f(G^{-1})\|_2^2 \end{aligned}$$

It remains to fix  $\kappa$  large enough so that we get

$$T_4 = \mathcal{O}(\log(n)n^{-1+1/\alpha}).$$

So we have for  $A_2''$ , with  $\alpha > 1$ ,

$$A_2'' = \mathcal{O}\left(\frac{\log(n)}{n^{1-1/\alpha}}\right)$$

Finally we get for  $A_2$

$$A_2 = \mathcal{O}\left(\log(n)\left(\frac{1}{n}\right)^{1-1/\alpha}\right).$$

Let us consider now the term  $A_3$

$$\begin{aligned} A_3 &\leq C \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \mathbb{E}(\beta_{jk}'' - \beta_{jk})^2 \\ &= C \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \beta_{jk}^2 (1 - b_j)^2 = \sum_{j \leq J_\alpha} \sum_{k=0}^{2^j-1} \left(\frac{\gamma_{jk}^2}{\tau_j^2 + \gamma_{jk}^2}\right)^2 \beta_{jk}^2. \end{aligned}$$

Since  $|\gamma_{jk}^2 - 1/n| \leq 1/(2n)$ , we get

$$A_3 \leq \sum_{j \leq J_\alpha} \left(\frac{3/(2n)}{c_1 2^{-j\alpha} + 1/(2n)}\right)^2 \sum_{k=0}^{2^j-1} \beta_{jk}^2$$

but  $f(G^{-1})$  belongs to the Besov ball  $B_{2,\infty}^s(R)$  which entails

$$\sum_{k=0}^{2^j-1} \beta_{jk}^2 \leq M 2^{-2js}$$

hence

$$A_3 \leq C/n^2 \sum_{j \leq J_\alpha} 2^{2j(-s+\alpha)}$$

We have

$$A_3 \leq C/n^2 (1/n)^{\frac{-2(-s+\alpha)}{\alpha}} = \mathcal{O}(1/n)^{2s/\alpha}.$$

We are now in position to give an upper bound for the variance term  $V$  namely

$$V \leq C J_\alpha (\log(n) (1/n)^{1-1/\alpha} + (1/n)^{2s/\alpha}).$$

It remains to bound the bias term  $B$ . In [87] p.1083 using inequality (44) the authors have proved that for any  $l$  we get

$$\left\| \sum_{j \geq l} \sum_k \beta_{jk} \psi_{j,k}(G) \right\|_2 \leq \sum_{j \geq l} \left\| \sum_k \beta_{jk} \psi_{j,k}(G) \right\|_2 \leq C \sum_{j \geq l} 2^{j/2} \left( \sum_k |\beta_{jk}|^2 \varrho(I_{j,k}) \right)^{1/2}. \quad (4.31)$$

Applying (4.31) with in our case of a bounded from below design density,  $\varrho(I_{j,k}) \leq 2^{-j}/m$  and  $l = J_\alpha$ , it follows

$$\begin{aligned} \left\| \sum_{j \geq J_\alpha} \sum_k \beta_{jk} \psi_{j,k}(G) \right\|_2 &\leq C \sum_{j \geq J_\alpha} \left( \sum_k |\beta_{jk}|^2 \right)^{1/2} \\ &\leq C \sum_{j \geq J_\alpha} 2^{-js} \leq C 2^{-J_\alpha s} \end{aligned}$$

hence

$$B = \left\| \sum_{j > J_\alpha} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{j,k}(G(x)) \right\|_2^2 \leq C 2^{-2J_\alpha s} = C(1/n)^{2s/\alpha}$$

which completes the proof of Theorem 1.

**Lemma 4.4.** *Let  $w_{jk}$  a sequence of random weights lying in  $[0, 1]$ . We assume that there exist positive constants  $c$ ,  $m$  and  $K$  such that for any  $\varepsilon > 0$ ,*

$$\check{\beta}_n = (w_{jk} \hat{\beta}_{jk})_{jk}$$

is a shrinkage rule verifying for any  $n$ ,

$$w_{jk}(n) = 0, \quad a.e. \quad \forall j \geq J_n \quad \text{with} \quad 2^{J_n} \sim n/\log(n) := t_n^2, \quad \forall k \quad (4.32)$$

$$|\hat{\beta}_{jk}| \leq m t_n \Rightarrow w_{jk} \leq c t_n, \quad \forall j \leq J_n, \quad \forall k, \quad (4.33)$$

$$(1 - w_{jk}(n)) \leq K \left( \frac{t_n}{|\hat{\beta}_{jk}|} + t_n \right) \quad a.e. \quad \forall j \leq J_n, \quad \forall k. \quad (4.34)$$

and let

$$\check{f} = \sum_{j < J_n} \sum_k w_{jk} \hat{\beta}_{jk} \psi_{jk}(G(x))$$

Then

$$\sup_{f(G^{-1}) \in B_{2,\infty}^s(R)} \mathbb{E} \|\check{f} - f\|_2^2 \leq (\log(n)/n)^{2s/(2s+1)}.$$

**Proof of Lemma 4.4.**

$$\begin{aligned} \mathbb{E}\|\check{f} - f\|_2^2 &\leq 2C(J_n \sum_{j \leq J_n} \sum_k \mathbb{E}(\check{\beta}_{jk} - \beta_{jk})^2 + \|\sum_{j > J_n} \sum_k \beta_{jk}^2 \psi_{jk}(G(x))\|_2^2) \\ &\leq V_1 + B_1. \end{aligned}$$

We first consider the term  $V_1$

$$\begin{aligned} V_1 &\leq 2J_n \mathbb{E} \sum_{j \leq J_n} \sum_k (w_{jk}^2 (\hat{\beta}_{jk} - \beta_{jk})^2 + (1 - w_{jk})^2 \beta_{jk}^2) I\{|\hat{\beta}_{jk}| \leq mt_n\} \\ &\quad + J_n \mathbb{E} \sum_{j \leq J_n} \sum_k (w_{jk}^2 (\hat{\beta}_{jk} - \beta_{jk})^2 + (1 - w_{jk})^2 \beta_{jk}^2) I\{|\hat{\beta}_{jk}| > mt_n\} \\ &= V_1' + V_1'' \end{aligned}$$

$$V_1' = J_n(T_5 + T_6)$$

$$T_5 = \mathbb{E} \sum_{j \leq J_n} \sum_k w_{jk}^2 (\hat{\beta}_{jk} - \beta_{jk})^2 I\{|\hat{\beta}_{jk}| \leq mt_n\}$$

but according to (4.25) we have for  $2^j \leq \log(n)/n$

$$\mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^2 \leq C \frac{1 + \|f\|_\infty^2}{n}$$

hence using (4.33) it follows

$$T_5 \leq Ct_n^2 2^{J_n} 1/n.$$

As for  $T_6$

$$\begin{aligned} T_6 &= \mathbb{E} \sum_{j \leq J_n} \sum_k (1 - w_{jk})^2 \beta_{jk}^2 I\{|\hat{\beta}_{jk}| \leq mt_n\} \\ &\leq \mathbb{E} \sum_{j \leq J_n} \sum_k (1 - w_{jk})^2 \beta_{jk}^2 I\{|\hat{\beta}_{jk}| \leq mt_n\} [I\{|\beta_{jk}| \leq 2mt_n\} + I\{|\beta_{jk}| > 2mt_n\}]. \end{aligned}$$

By (4.15) we get

$$T_6 \leq 2(mt_n)^{2s/(2s+1)} \|f\|_{W_{2/(1+2s)}}^2 + \sum_{j \leq J_n} \sum_k \beta_{jk}^2 \mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > mt_n).$$

We are going to bound  $\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > mt_n)$ . We have

$$\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| \geq m\sqrt{\log(n)/n}) \leq \mathbb{P}_3 + \mathbb{P}_4$$

where

$$\mathbb{P}_3 = \mathbb{P}\left(\left|1/n \sum_{i=1}^n \psi_{j,k}(G(X_i))(f(X_i) - \mathbb{E}\psi_{j,k}(G(X))f(X))\right| \geq m/2\sqrt{\log(n)/n}\right) \quad (4.35)$$

and

$$\mathbb{P}_4 = \mathbb{P}\left(\left|1/n \sum_{i=1}^n \psi_{j,k}(G(X_i))\varepsilon_i\right| > m/2\sqrt{\log(n)/n}\right). \quad (4.36)$$

Kerkyacharian and Picard in [87] in order to prove inequality (65) in [87] showed p. 1088 that

$$\mathbb{P}_3 \leq 2 \exp\left(-\frac{3m^2 \log(n)}{4\|f\|_\infty(3+m)}\right) \quad (4.37)$$

if  $2^j \leq n/\log(n)$ . As for  $\mathbb{P}_4$ , conditionally on  $(X_1, \dots, X_n)$  we have

$$1/n \sum_{i=1}^n \psi_{j,k}(G(X_i))\varepsilon_i \sim N(0, \gamma_{jk}^2)$$

where  $\gamma_{jk}^2$  has been defined in (4.4).

$$\begin{aligned} \mathbb{P}_4 &\leq \mathbb{E}\left(\exp\left(-\frac{m^2 \log(n)}{8n\gamma_{jk}^2}\right)\right) \\ &= \mathbb{E}e^{-\frac{m^2 \log(n)}{8n\gamma_{jk}^2}} \left(I(|\gamma_{jk}^2 - 1/n| \leq \varsigma/n) + I(|\gamma_{jk}^2 - 1/n| > \varsigma/n)\right) \\ &\leq e^{-\frac{m^2 \log(n)}{8(\varsigma+1)}} + \mathbb{P}(|\gamma_{jk}^2 - 1/n| > \varsigma/n). \end{aligned} \quad (4.38)$$

Using (4.20) to bound  $\mathbb{P}(|\gamma_{jk}^2 - 1/n| > \varsigma/n)$  we get

$$\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > mt_n) \leq 2e^{-\varsigma^2 \log(n)/(C\|\psi\|_4^4 + \varsigma\|\psi\|_\infty^2)} + e^{-\frac{m^2 \log(n)}{8(\varsigma+1)}} + 2e^{-\frac{3m^2 \log(n)}{4\|f\|_\infty(3+m)}}$$

thus

$$\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > mt_n) \leq 2n^{\frac{-\varsigma^2}{C\|\psi\|_4^4 + \varsigma\|\psi\|_\infty^2}} + n^{\frac{-m^2}{8(\varsigma+1)}} + 2n^{\frac{-3m^2}{4\|f\|_\infty(3+m)}} \quad (4.39)$$

which entails by fixing  $m$  and  $\varsigma$  large enough

$$\begin{aligned} T_6 &\leq 2(mt_n)^{4s/(2s+1)} \|f\|_{W_{2/(1+2s)}}^2 + t_n^2 \sum_{j \leq J_n} \sum_k \beta_{jk}^2 \\ &\leq 2(mt_n)^{4s/(2s+1)} \|f\|_{W_{2/(1+2s)}}^2 + \|f(G^{-1})\|_2^2 t_n^2. \end{aligned}$$

Let us look at the term  $V_1''$

$$V_1'' = \mathbb{E} \sum_{j \leq J_n} \sum_k (w_{jk}^2 (\hat{\beta}_{jk} - \beta_{jk})^2 + (1 - w_{jk})^2 \beta_{jk}^2) I\{|\hat{\beta}_{jk}| > mt_n\}$$

$$\begin{aligned} V_1'' &= \mathbb{E} \sum_{j \leq J_n} \sum_k (w_{jk}^2 (\hat{\beta}_{jk} - \beta_{jk})^2 + (1 - w_{jk})^2 \beta_{jk}^2) I\{|\hat{\beta}_{jk}| > mt_n\} [I\{|\beta_{jk}| \leq mt_n/2\} + I\{|\beta_{jk}| > mt_n/2\}] \\ &= T_7 + T_8 \end{aligned}$$

for the term  $T_7$ , we use the Cauchy Scharwz inequality

$$\begin{aligned} T_7 &\leq \sum_{j \leq J_n} \sum_k (\mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^4)^{1/2} (\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > mt_n/2))^{1/2} \\ &\quad + \sum_{j \leq J_n} \sum_k \beta_{jk}^2 I\{|\hat{\beta}_{jk}| > mt_n\} I\{|\beta_{jk}| \leq mt_n/2\}. \end{aligned}$$

Furthermore, using inequality (64) p. 1086 in [87] we get for  $2^j \leq n/\log(n)$

$$\mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^4 \leq C \frac{1 + \|f\|_\infty^4}{n^2} \quad (4.40)$$

and by (4.39)

$$\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > mt_n/2) \leq 2n^{\frac{-\varsigma^2}{C\|\psi\|_4^{4+\varsigma}\|\psi\|_\infty^2}} + n^{\frac{-m^2}{32(\varsigma+1)}} + 2n^{\frac{-3m^2}{16\|f\|_\infty(3+m)}}$$

from which follows by fixing again  $m$  and  $\varsigma$  large enough

$$\begin{aligned} T_7 &\leq C/n \cdot 2^{J_n} (n^{\frac{-\varsigma^2}{C\|\psi\|_4^{4+\varsigma}\|\psi\|_\infty^2}} + n^{\frac{-m^2}{32(\varsigma+1)}} + 2n^{\frac{-3m^2}{16\|f\|_\infty(3+m)}})^{1/2} + \sum_{j \leq J_n} \sum_k \beta_{jk}^2 I\{|\beta_{jk}| \leq mt_n/2\} \\ &\leq t_n^2 + ((m/2)t_n)^{4s/(1+2s)} \|f\|_{W_{s/(2s+1)}}^2. \end{aligned}$$

For the term  $T_8$

$$\begin{aligned} T_8 &= \mathbb{E} \sum_{j \leq J_n} \sum_k (w_{jk}^2 (\hat{\beta}_{jk} - \beta_{jk})^2 + (1 - w_{jk})^2 \beta_{jk}^2) I\{|\hat{\beta}_{jk}| > mt_n\} I\{|\beta_{jk}| > mt_n/2\} \\ &\leq \frac{4m^{-2/(2s+1)}}{(1 - 2^{-2/(1+2s)})} \|f\|_{W_{2/(1+2s)}}^2 (t_n)^{4s/(1+2s)} \\ &\quad + \mathbb{E} \sum_{j \leq J_n} \sum_k (1 - w_{jk})^2 \beta_{jk}^2 I\{|\hat{\beta}_{jk}| > mt_n\} I\{|\beta_{jk}| > mt_n/2\} [I\{|\hat{\beta}_{jk}| \geq |\beta_{jk}/2\} + I\{|\hat{\beta}_{jk}| < |\beta_{jk}/2\}]. \end{aligned}$$



Hereafter we decompose

$$\begin{aligned} \mathbb{E} \sum_{j \leq J_n} \sum_k (1 - w_{jk})^2 \beta_{jk}^2 I\{|\hat{\beta}_{jk}| > mt_n\} I\{|\beta_{jk}| > mt_n/2\} [I\{|\hat{\beta}_{jk}| \geq |\beta_{jk}|/2\} + I\{|\hat{\beta}_{jk}| < |\beta_{jk}|/2\}] \\ = T'_8 + T''_8 \end{aligned}$$

$$T''_8 \leq \sum_{j \leq J_n} \sum_k \beta_{jk}^2 \mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > mt_n/4)$$

using (4.39) we get for  $m$  and  $\varsigma$  large enough

$$T''_8 \leq (2n^{\frac{-\varsigma^2}{C\|\psi\|_4^4 + \varsigma\|\psi\|_\infty^2}} + n^{\frac{-m^2}{128(\varsigma+1)}} + 2n^{\frac{-3m^2}{64\|f\|_\infty(3+m)}}) \sum_{j \leq J_n} \sum_k \beta_{jk}^2 \leq t_n^2$$

as for  $T'_8$

$$T'_8 = \mathbb{E} \sum_{j \leq J_n} \sum_k (1 - w_{jk})^2 \beta_{jk}^2 I\{|\hat{\beta}_{jk}| > mt_n\} I\{|\beta_{jk}| > mt_n/2\} I\{|\hat{\beta}_{jk}| \geq |\beta_{jk}|/2\}$$

using (4.34) we get

$$\begin{aligned} T'_8 &\leq \mathbb{E} \sum_{j \leq J_n} \sum_k K^2 \beta_{jk}^2 \left( \frac{t_n}{|\hat{\beta}_{jk}|} + t_n \right)^2 I\{|\hat{\beta}_{jk}| \geq |\beta_{jk}|/2\} I\{|\beta_{jk}| > mt_n/2\} \\ &\leq K^2 \sum_{j \leq J_n} \sum_k \beta_{jk}^2 \left( \frac{2t_n}{|\beta_{jk}|} + t_n \right)^2 I\{|\beta_{jk}| > mt_n/2\} \\ &\leq 2K^2 \sum_{j \leq J_n} \sum_k \beta_{jk}^2 \left( \frac{4t_n^2}{|\beta_{jk}|^2} + t_n^2 \right) I\{|\beta_{jk}| > mt_n/2\} \\ &= 8K^2 t_n^2 \sum_{j \leq J_n} \sum_k I\{|\beta_{jk}| > mt_n/2\} + 2K^2 t_n^2 \|f(G^{-1})\|_2^2 \end{aligned}$$

using (4.15) it follows

$$\begin{aligned} T'_8 &\leq 8K^2 t_n^2 \left( \frac{mt_n}{2} \right)^{-2/(1+2s)} \frac{2^{2-2/(1+2s)}}{1 - 2^{-2/(1+2s)}} \|f\|_{W_{2/(1+2s)}}^2 + 2K^2 t_n^2 \|f(G^{-1})\|_2^2 \\ &\leq 32K^2 \frac{m^{-2/(1+2s)}}{1 - 2^{-2/(1+2s)}} t_n^{4s/(1+2s)} + 2K^2 t_n^2 \|f(G^{-1})\|_2^2. \end{aligned}$$

It remains to bound the bias term  $B_1$ . To this purpose we use the fact that  $f \in B_{2,\infty}^s$

$$B_{-1} = \left\| \sum_{j>J_n} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{j,k}(G(x)) \right\|_2^2 \leq C 2^{-2J_n s} = C t_n^{2s} \leq C t_n^{4s/(2s+1)}$$

which completes the proof.

**Proof of Theorem 2.**

In order to prove the Theorem 2., we have to prove that the Bayesian estimators (4.5) based on Gaussian priors with large variance (4.11) and (4.12) satisfy the conditions of Lemma 4.4.

We will not get into details of the proof because this latter is identical to the proof of Theorem 3. in [14], with the sole exception that here, the proof is carried over the event  $\Omega_n^\delta$  with  $\delta = \varsigma/n$ ,  $\varsigma$  some positive constant. Indeed, as precised above in section 2.2, a key observation is that instead of having a deterministic noise  $\varepsilon = 1/\sqrt{n}$  like in [14], here we have to deal with a stochastic noise  $\gamma_{jk}^2$  which expression is given by (4.4).



# Bibliographie

- [1] *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*. Edited by Milton Abramowitz and Irene A. Stegun. Dover Publications Inc., New York, 1966.
- [2] F. Abramovich, U. Amato, and C. Angelini. On optimality of Bayesian wavelet estimators. *Scand. J. Statist.*, 31(2) :217–234, 2004.
- [3] F. Abramovich, T. Sapatinas, and B. W. Silverman. Wavelet thresholding via a bayesian approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(4) :725–749, 1998.
- [4] Felix Abramovich and Theofanis Sapatinas. Bayesian approach to wavelet decomposition and shrinkage. In *Bayesian inference in wavelet-based models*, volume 141 of *Lecture Notes in Statist.*, pages 33–50. Springer, New York, 1999.
- [5] N. I. Akhiezer. *The classical moment problem and some related questions in analysis*. Translated by N. Kemmer. Hafner Publishing Co., New York, 1965.
- [6] B. Aliev. Estimation of the regularization method for the problem of the summability of the Fourier series. *Dokl. Akad. Nauk Tadzhik. SSR*, 21(4) :3–6, 1978.
- [7] Umberto Amato, Anestis Antoniadis, and Marianna Pensky. Wavelet kernel penalized estimation for non-equispaced design regression. *Stat. Comput.*, 16(1) :37–55, 2006.
- [8] D. D. Ang, V. N. Nhan, and D. N. Thanh. A nonlinear integral equation of gravimetry : Uniqueness and approximation by linear moments. *Vietnam J. Math.*, 27(1) :61–67, 1999.
- [9] Dang Dinh Ang, Rudolf Gorenflo, Vy Khoi Le, and Dang Duc Trong. *Moment theory and some inverse problems in potential theory and heat conduction*, volume 1792 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002.

- [10] A. Antoniadis, J. Bigot, and T. Sapatinas. Wavelet estimators in non-parametric regression : a comparative simulation study. *J. Statist. Software*, 6 :1–83, 1996.
- [11] A. Antoniadis, G. Grégoire, and P. Vial. Random design wavelet curve smoothing. 35 :225–232, 1997.
- [12] Avner Ash and Patrick McDonald. Moment problems and the causal set approach to quantum gravity. *J. Math. Phys.*, 44(4) :1666–1678, 2003.
- [13] G.A. Athanassoulis and P. N. Gavriiadis. The truncated hausdorff moment problem solved by using kernel density functions. *Probabilistic Engeneering Mechanics*, 17 :273–291, 2002.
- [14] F. Autin, D. Picard, and V. Rivoirard. Large variance Gaussian priors in Bayesian nonparametric estimation : a maxiset approach. *Math. Methods of Statist.*, 15(4) :349–373, 2006.
- [15] A. B. Bakursinskiĭ. The construction of regularizing algorithms in the case of random noise. *Dokl. Akad. Nauk SSSR*, 189 :231–233, 1969.
- [16] Paolo Baldi, Gérard Kerkyacharian, Domenico Marinucci, and Dominique Picard. Adaptive density estimation for directional data using needlets. *Ann. Statist. to appear*, 2009.
- [17] Paolo Baldi, Gérard Kerkyacharian, Domenico Marinucci, and Dominique Picard. Subsampling needlet coefficients on the sphere. *Bernoulli*, 15(2) :438–463, 2009.
- [18] Y. Baraud. Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6 :127–146, 2002.
- [19] M. Bertero and P. Boccacci. Computation of the singular system for a class of integral operators related to data inversion in confocal microscopy. *Inverse Problems*, 5(6) :935–957, 1989.
- [20] M. Bertero, P. Boccacci, and E. R. Pike. On the recovery and resolution of exponential relaxation rates from experimental data : a singular value analysis of the laplace transform inversion in the presence of noise. *Proc. Roy. Soc. Lond. A*, 383(2) :15–29, 1982.
- [21] M. Bertero, P. Boccacci, and E. R. Pike. Particle-size distributions from Fraunhofer diffraction : the singular-value spectrum. *Inverse Problems*, 1(2) :111–126, 1985.
- [22] M. Bertero, P. Boccacci, and E. R. Pike. Particle size distributions from spectral turbidity : a singular-system analysis. *Inverse Problems*, 2(2) :247–258, 1986.

- 
- [23] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data. I. General formulation and singular system analysis. *Inverse Problems*, 1(4) :301–330, 1985.
- [24] Lucien Birgé and Pascal Massart. Gaussian model selection. *J. Eur. Math. Soc. (JEMS)*, 3(3) :203–268, 2001.
- [25] P. Brianzi and M. Frontini. On the regularized inversion of the Laplace transform. *Inverse Problems*, 7(3) :355–368, 1991.
- [26] P. Brutti. Warped wavelets and vertical thresholding. Preprint. arxiv :0801.3319v1.
- [27] T.T. Cai and L.D. Brown. Wavelet shrinkage for nonequispaced samples. *Ann. Statist.*, 26 :1783–1799, 1998.
- [28] Raymond J. Carroll and Peter Hall. Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83(404) :1184–1186, 1988.
- [29] L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov. Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3) :843–874, 2002. Dedicated to the memory of Lucien Le Cam.
- [30] L. Cavalier, Y. Golubev, O. Lepski, and A. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Teor. Veroyatnost. i Primenen.*, 48(3) :534–556, 2003.
- [31] Laurent Cavalier and Alexandre Tsybakov. Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields*, 123(3) :323–354, 2002.
- [32] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch. Adaptive Bayesian Wavelet Shrinkage. *J. Amer. Statist. Assoc.*, 92 :1413–1421, 1997.
- [33] M. Clyde and E. I. George. Flexible Empirical Bayes Estimation for Wavelets. *J. Roy. Statist. Soc., Ser. B.*, 62(4) :681–698, 2000.
- [34] M. Clyde, G. Parmigiani, and B. Vidakovic. Multiple shrinkage and subset selection in wavelets. *Biometrika*, 85(2) :391–401, 1998.
- [35] M. A. Clyde and E. I. George. Empirical Bayes estimation in wavelet nonparametric regression. In *Bayesian inference in wavelet-based models*, volume 141 of *Lecture Notes in Statist.*, pages 309–322. Springer, New York, 1999.
- [36] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

- [37] M. E. Davison. A singular value decomposition for the Radon transform in  $n$ -dimensional Euclidean space. *Numer. Funct. Anal. Optim.*, 3(3) :321–340, 1981.
- [38] J. Delabrouille, J.-F. Cardoso, M. Le Jeune, M. Betoule, G. Faÿ, and F. Guilloux. A full sky, low foreground, high resolution CMB map from WMAP. *Astron. Astrophys.*, 493 :835–857, January 2009.
- [39] Ronald A. DeVore and George G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.
- [40] Luc Devroye. Consistent deconvolution in density estimation. *Canad. J. Statist.*, 17(2) :235–239, 1989.
- [41] Peter J. Diggle and Peter Hall. A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B*, 55(2) :523–531, 1993.
- [42] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3) :425–455, 1994.
- [43] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90(432) :1200–1224, 1995.
- [44] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage : asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57(2) :301–369, 1995. With discussion and a reply by the authors.
- [45] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam*, pages 183–218. Springer, New York, 1997.
- [46] David L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2(2) :101–126, 1995.
- [47] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Wavelet shrinkage : asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57(2) :301–369, 1995. With discussion and a reply by the authors.
- [48] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2) :508–539, 1996.

- [49] Nguyen Dung, Nguyen Vu Huy, Pham Hoang Quan, and Dang Duc Trong. A Hausdorff-like moment problem and the inversion of the Laplace transform. *Math. Nachr.*, 279(11) :1147–1158, 2006.
- [50] Sam Efromovich. Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.*, 92(438) :526–535, 1997.
- [51] Sam Efromovich and Vladimir Koltchinskii. On inverse problems with unknown operators. *IEEE Trans. Inform. Theory*, 47(7) :2876–2894, 2001.
- [52] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [53] M. S. Ermakov. Minimax estimation of the solution of an ill-posed problem of convolution type. *Problemy Peredachi Informatsii*, 25(3) :28–39, 1989.
- [54] Jianqing Fan. Global behavior of deconvolution kernel estimates. *Statist. Sinica*, 1(2) :541–551, 1991.
- [55] Jianqing Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3) :1257–1272, 1991.
- [56] Jianqing Fan and Ja-Yong Koo. Wavelet deconvolution. *IEEE Trans. Inform. Theory*, 48(3) :734–747, 2002.
- [57] Dario Fasino and Gabriele Inglese. Recovering a probability density from a finite number of moments and local a priori information. *Rend. Istit. Mat. Univ. Trieste*, 28(1-2) :183–200 (1997), 1996.
- [58] Gilles Faÿ, Frédéric Guilloux, Marc Betoule, Jean-François Cardoso, Jacques Delabrouille, and Maude Le Jeune. CMB power spectrum estimation using wavelets, October 2008.
- [59] William Feller. *An introduction to probability theory and its applications. Vol. II*. John Wiley & Sons Inc., New York, 1966.
- [60] I. Gannaz. *Estimation par ondelettes dans les modèles partiellement linéaires*. Ph.D. Thesis , Université Joseph Fourier, 2007.
- [61] J.P. Gaspard and F. Cyrot-Lackman. Density of states from moments. application to the impurity band. *J. Phys. C : Solid State Phys.*, 6 :3077–3096, 1973.
- [62] A. Goldenshluger and V. Spokoiny. On the shape-from-moments problem and recovering edges from noisy Radon data. *Probab. Theory Related Fields*, 128(1) :123–140, 2004.



- [63] Alexander Goldenshluger. On pointwise adaptive nonparametric deconvolution. *Bernoulli*, 5(5) :907–925, 1999.
- [64] Alexander Goldenshluger and Sergei V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Related Fields*, 118(2) :169–186, 2000.
- [65] G. K. Golubev and R. Z. Khasminskii. A statistical approach to some inverse problems for partial differential equations. *Problemy Peredachi Informatsii*, 35(2) :51–66, 1999.
- [66] F. Gori and G. Guattari. Signal restoration for linear systems with weighted inputs. Singular value analysis for two cases of low-pass filtering. *Inverse Problems*, 1(1) :67–85, 1985.
- [67] Frédéric Guilloux, Gilles Faÿ, and Jean-François Cardoso. Practical wavelet design on the sphere. *Appl. Comput. Harmon. Anal.*, 26(2) :143–160, 2009.
- [68] P. Hall and B. A. Turlach. Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Ann. Statist.*, 25 :1912–1925, 1997.
- [69] Peter Hall, Gérard Kerkycharian, and Dominique Picard. Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.*, 26(3) :922–942, 1998.
- [70] Peter Hall and Prakash Patil. Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.*, 23(3) :905–928, 1995.
- [71] Wolfgang Härdle, Gerard Kerkycharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.
- [72] Felix Hausdorff. Summationsmethoden und Momentfolgen. I. *Math. Z.*, 9(1-2) :74–109, 1921.
- [73] Dennis M. Healy, Jr., Harrie Hendriks, and Peter T. Kim. Spherical deconvolution. *J. Multivariate Anal.*, 67(1) :1–22, 1998.
- [74] Gabriele Inglese. Christoffel functions and finite moment problems. *Inverse Problems*, 11(4) :949–960, 1995.
- [75] V. K. Ivanov, V. V. Vasin, and V. P. Tanana. *Teoriya lineinykh nekorrektnykh zadach i ee prilozheniya*. “Nauka”, Moscow, 1978.
- [76] I. Johnstone, G. Kerkycharian, D. Picard, and M. Raimondo. Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3) :547–573, 2004.

- [77] I. M. Johnstone and B. W. Silverman. Needles and straw in haystacks : empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.*, 32(4) :1594–1649, 2004.
- [78] I. M. Johnstone and B. W. Silverman. Empirical Bayes selection of wavelet thresholds. *Ann. Statist.*, 33(4) :1700–1752, 2005.
- [79] Iain M. Johnstone. Wavelet shrinkage for correlated data and inverse problems : adaptivity results. *Statist. Sinica*, 9(1) :51–83, 1999.
- [80] Iain M. Johnstone, Gérard Kerkyacharian, Dominique Picard, and Marc Raimondo. Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3) :547–573, 2004.
- [81] Iain M. Johnstone and Bernard W. Silverman. Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.*, 18(1) :251–280, 1990.
- [82] Iain M. Johnstone and Bernard W. Silverman. Discretization effects in statistical inverse problems. *J. Complexity*, 7(1) :1–34, 1991.
- [83] Jérôme Kalifa and Stéphane Mallat. Thresholding estimators for linear inverse problems and deconvolutions. *Ann. Statist.*, 31(1) :58–109, 2003.
- [84] M. Kay and S. L. Marple. Spectrum analysis-a modern perspective. *Proc. IEEE*, 69 :1380–1419, 1981.
- [85] G. Kerkyacharian, G. Kyriazis, E. Le Pennec, P. Petrushev, and D. Picard. Inversion of noisy radon transform by svd based needlets. *ACHA*, 2009.
- [86] G. Kerkyacharian and D. Picard. Density estimation in Besov spaces. *Statist. Probab. Lett.*, 13(1) :15–24, 1992.
- [87] G. Kerkyacharian and D. Picard. Regression in random design and warped wavelets. *Bernoulli*, 10(6) :1053–1105, 2004.
- [88] G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and well concentrated basis. *Test*, 9(2) :283–344, 2004.
- [89] Gérard Kerkyacharian, Pencho Petrushev, Dominique Picard, and Thomas Willer. Needlet algorithms for estimation in inverse problems. *Electron. J. Stat.*, 1 :30–76 (electronic), 2007.
- [90] Gérard Kerkyacharian and Dominique Picard. Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.*, 24(2) :485–507, 1996.
- [91] Peter T. Kim and Ja-Yong Koo. Optimal spherical deconvolution. *J. Multivariate Anal.*, 80(1) :21–42, 2002.

- [92] Peter T. Kim, Ja-Yong Koo, and Heon Jin Park. Sharp minimaxity and spherical deconvolution for super-smooth error distributions. *J. Multivariate Anal.*, 90(2) :384–392, 2004.
- [93] Ja-Yong Koo. Optimal rates of convergence for nonparametric statistical inverse problems. *Ann. Statist.*, 21(2) :590–599, 1993.
- [94] A. P. Korostelëv and A. B. Tsybakov. Optimal rates of convergence of estimates in a probabilistic formulation of the tomography problem. *Problemy Peredachi Informatsii*, 27(1) :92–103, 1991.
- [95] A. P. Korostelëv and A. B. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993.
- [96] A. Kovac and B. W. Silverman. Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *J. Amer. Statist. Assoc.*, 95 :172–183, 2000.
- [97] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty. II. *Bell System Tech. J.*, 40 :65–84, 1961.
- [98] S. W. Lang and J. H. McClellan. Spectral estimation for sensor arrays. *IEEE Trans. Acoust. Speech Signal Processing*, 31 :349–358, 1983.
- [99] Remigijus Leipus, George Oppenheim, Anne Philippe, and Marie-Claude Viano. Orthogonal series density estimation in a disaggregation scheme. *J. Statist. Plann. Inference*, 136(8) :2547–2571, 2006.
- [100] O. A. Liskovets. Theory and methods of solving ill-posed problems. In *Mathematical analysis, Vol. 20*, Itogi Nauki i Tekhniki, pages 116–178, 264. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Informatsii, Moscow, 1982.
- [101] Ming Chung Liu and Robert L. Taylor. A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.*, 17(4) :427–438, 1989.
- [102] Alfred K. Louis. Incomplete data problems in x-ray computerized tomography. I. Singular value decomposition of the limited angle transform. *Numer. Math.*, 48(3) :251–262, 1986.
- [103] Bernard A. Mair and Frits H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5) :1424–1444, 1996.
- [104] Stéphane Mallat. *A wavelet tour of signal processing*. Academic Press Inc., San Diego, CA, 1998.

- [105] Elias Masry. Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory*, 37(4) :1105–1115, 1991.
- [106] Elias Masry. Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Process. Appl.*, 47(1) :53–74, 1993.
- [107] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003.
- [108] Peter Mathé and Sergei V. Pereverzev. Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods. *SIAM J. Numer. Anal.*, 38(6) :1999–2021 (electronic), 2001.
- [109] Peter Mathé and Sergei V. Pereverzev. Stable summation of orthogonal series with noisy coefficients. *J. Approx. Theory*, 118(1) :66–80, 2002.
- [110] Lawrence R. Mead and N. Papanicolaou. Maximum entropy in the problem of moments. *J. Math. Phys.*, 25(8) :2404–2417, 1984.
- [111] Yves Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1992. Translated from the 1990 French original by D. H. Salinger.
- [112] Robert Mnatsakanov and Frits H. Ruymgaart. Some results for moment-empirical cumulative distribution functions. *J. Nonparametr. Stat.*, 17(6) :733–744, 2005.
- [113] V. A. Morozov. *Methods for solving incorrectly posed problems*. Springer-Verlag, New York, 1984. Translated from the Russian by A. B. Aries, Translation edited by Z. Nashed.
- [114] F. Narcowich, P. Petrushev, and J. Ward. Decomposition of Besov and Triebel-Lizorkin spaces on the sphere. *J. Funct. Anal.*, 238(2) :530–564, 2006.
- [115] F. Narcowich, P. Petrushev, and J. Ward. Decomposition of Besov and Triebel-Lizorkin spaces on the sphere. *J. Funct. Anal.*, 238(2) :530–564, 2006.
- [116] F. Narcowich, P. Petrushev, and J. Ward. Local tight frames on spheres. *SIAM J. Math. Anal.*, 2006.
- [117] Athanasios Papoulis. A new method of inversion of the Laplace transform. *Quart. Appl. Math.*, 14 :405–414, 1957.

- [118] Jaak Peetre. *New thoughts on Besov spaces*. Mathematics Department, Duke University, Durham, N.C., 1976. Duke University Mathematics Series, No. 1.
- [119] M. Pensky. Frequentist optimality of Bayesian wavelet shrinkage rules for Gaussian and non-Gaussian noise. *Ann. Statist.*, 34(2) :769–807, 2006.
- [120] M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics*, 27 :2033–2053, 1999.
- [121] P. Petrushev and Y. Xu. Localized polynomials frames on the interval with jacobi weights. *J. Fourier Anal. Appl.*, 11(5)(5) :557–575, 2005.
- [122] P. Petrushev and Y. Xu. Localized polynomials kernels and frames (needlets) on the ball. *IMI*, 2005.
- [123] Thanh Mai Pham Ngoc. A statistical minimax approach to the Hausdorff moment problem. *Inverse Problems*, 24(4) :045018, 13, 2008.
- [124] Thanh Mai Pham Ngoc. Regression in random design and Bayesian warped wavelets estimators. *Electronic Journal of Statistics*, 3 :1084–1112, 2009.
- [125] D. Pietrobon, Paolo Baldi, Gérard Kerkyacharian, Domenico Marinucci, and Dominique Picard. Spherical needlets for cmb data analysis. 383 :539–545, 2008.
- [126] S. Z. Rafal'son. The approximation of functions by Fourier-Jacobi sums. *Izv. Vysrs. Urcebn. Zaved. Matematika*, 1968(4 (71)) :54–62, 1968.
- [127] V. Rivoirard. Bayesian modeling of sparse sequences and maxisets for Bayes rules. *Math. Methods Statist.*, 14(3) :346–376, 2005.
- [128] Giuseppe Rodriguez and Sebastiano Seatzu. Approximation methods for the finite moment problem. *Numer. Algorithms*, 5(1-4) :391–405, 1993. Algorithms for approximation, III (Oxford, 1992).
- [129] Jeffrey S. Rosenthal. Random rotations : characters and random walks on  $SO(N)$ . *Ann. Probab.*, 22(1) :398–423, 1994.
- [130] J. A. Shohat and J. D. Tamarkin. *The Problem of Moments*. American Mathematical Society Mathematical surveys, vol. II. American Mathematical Society, New York, 1943.
- [131] L. Shure, R. L. R.L. Parker, and G.E. Backus. Harmonic splines for geomagnetic modeling. *Phys. Earth Plan. Interiors*, 28 :215–229, 1982.

- 
- [132] Leonard Stefanski and Raymond J. Carroll. Deconvoluting kernel density estimators. *Statistics*, 21(2) :169–184, 1990.
- [133] Leonard A. Stefanski. Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.*, 9(3) :229–235, 1990.
- [134] V. N. Sudakov and L. A. Halfin. A statistical approach to the correctness of the problems of mathematical physics. *Dokl. Akad. Nauk SSSR*, 157 :1058–1060, 1964.
- [135] Aldo Tagliani. Entropy estimate of probability densities having assigned moments : Hausdorff case. *Appl. Math. Lett.*
- [136] Aldo Tagliani. Numerical inversion of Laplace transform on the real line of probability density functions. *Appl. Math. Comput.*, 123(3) :285–299, 2001.
- [137] Aldo Tagliani and Yurayh Velásquez. Numerical inversion of the Laplace transform via fractional moments. *Appl. Math. Comput.*, 143(1) :99–107, 2003.
- [138] Giorgio Talenti. Recovering a function from a finite number of moments. *Inverse Problems*, 3(3) :501–517, 1987.
- [139] James D. Talman. *Special functions : A group theoretic approach*. Based on lectures by Eugene P. Wigner. With an introduction by Eugene P. Wigner. W. A. Benjamin, Inc., New York-Amsterdam, 1968.
- [140] R. L. Taylor and H. M. Zhang. On a strongly consistent nonparametric density estimator for the deconvolution problem. *Comm. Statist. Theory Methods*, 19(9) :3325–3342, 1990.
- [141] M. R Teague. Image analysis via the general theory of moments. *J. Opt. Soc. Am.*, 70 :920–924, 1980.
- [142] Audrey Terras. *Harmonic analysis on symmetric spaces and applications. I*. Springer-Verlag, New York, 1985.
- [143] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of ill-posed problems*. V. H. Winston & Sons, Washington, D.C. : John Wiley & Sons, New York, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.
- [144] N. G. Tomin. An application of the interpolation of linear operators to questions of the convergence of series of Fourier coefficients with respect to classical orthogonal polynomials. *Dokl. Akad. Nauk SSSR*, 212 :1074–1077, 1973.

- [145] Gadian D. G. Tournier, J. D. and A. Connelly. Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution. *NeuroImage*, 23 :1176–1185, 2004.
- [146] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-based complexity*. Computer Science and Scientific Computing. Academic Press Inc., Boston, MA, 1988. With contributions by A. G. Werschulz and T. Boulton.
- [147] Alexandre Tsybakov. On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(9) :835–840, 2000.
- [148] Alexandre B. Tsybakov. *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004.
- [149] G. M. Vaňnikko and A. Yu. Veretennikov. *Iteratsionnye protsedury v nekorrektnykh zadachakh*. “Nauka”, Moscow, 1986.
- [150] Arnoud C. M. van Rooij and Frits H. Ruymgaart. Regularized deconvolution on the circle and the sphere. In *Nonparametric functional estimation and related topics (Spetses, 1990)*, volume 335 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 679–690. Kluwer Acad. Publ., Dordrecht, 1991.
- [151] B. Vidakovic. Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *Journ. Amer. Statist. Assoc.*, 93 :173–179, 1998.
- [152] N. Ja. Vilenkin. *Fonctions spéciales et théorie de la représentation des groupes*. Traduit du Russe par D. Hérault. Monographies Universitaires de Mathématiques, No. 33. Dunod, Paris, 1969.
- [153] Grace Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14(4) :651–667, 1977.
- [154] Arthur G. Werschulz. *The computational complexity of differential and integral equations*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 1991. An information-based approach, Oxford Science Publications.
- [155] T. Willer. *Estimation non paramétrique et problèmes inverses*. Ph.D. Thesis , Université Paris VII, 2006.