Point de vue maxiset en estimation non paramétrique

Florent Autin

UNIVERSITÉ PARIS 7 - DENIS DIDEROT
UFR de Mathématiques

THÈSE

pour l'obtention du Diplôme de

DOCTEUR DE L'UNIVERSITÉ PARIS 7

Spécialité : MATHÉMATIQUES APPLIQUÉES

présentée par

Florent AUTIN

Titre :

POINT DE VUE MAXISET EN ESTIMATION NON PARAMÉTRIQUE

Directrice de thèse : **Dominique PICARD**

Soutenue publiquement le **07 Décembre 2004**, devant le jury composé de

M. Lucien BIRGÉ, Université Paris 6
M. Stéphane BOUCHERON, Université Paris 7
M. Stéphane JAFFARD, Université de Paris 12
M. Oleg LEPSKI, Université d'Aix-Marseille 1
Mme Dominique PICARD, Université Paris 7
M. Alexandre TSYBAKOV, Université Paris 6

au vu des rapports de **M. Anestis ANTONIADIS** (Université de Grenoble 1) et de **Mme Sara van de GEER** (Université de Leiden).

Je tiens à remercier toutes les personnes sans qui cette thèse n'aurait pu aboutir :
En premier lieu, je tiens à exprimer toute ma gratitude à l'égard de Dominique Picard qui a su guider mes premiers pas dans la recherche et m'accorder toute sa confiance. Sa rigueur bienveillante, son enthousiasme constant et ses connaissances étendues sur l'ensemble des thèmes statistiques m'ont été d'une aide considérable pour la réalisation de ce travail.
Je suis très reconnaissant envers Anestis Antoniadis et Sara van de Geer pour avoir accepté de rapporter ma thèse et tiens à remercier très chaleureusement Lucien Birgé, Stéphane Boucheron, Stéphane Jaffard, Oleg Lepski et Alexandre Tsybakov qui me font l'honneur d'être présents aujourd'hui en tant que membres du jury.
Je profite de cette occasion pour remercier Laure Elie pour l'intérêt qu'elle porte à tous les anciens de son D.E.A et Michèle Wasse pour ses qualités humaines. Merci aussi à Fabrice Gamboa qui m'a donné goût aux statistiques.
Ce fut un réel plaisir d'effectuer mes services d'enseignements sous les responsabilités successives de Francis Comets, Gabrielle Viennet et Christian Léonard. Qu'ils en soient tous remerciés.
Un grand merci aux équipes des laboratoires PMA et MODAL'X pour leur accueil et à l'ensemble des thésards du bureau 5B1 (auxquels il me faut associer Agnès, Anne, Christian, Erwan, Karine, Tristan et Wadie) avec qui j'ai partagé de très bons moments dans une ambiance des plus chaleureuses. Il est important pour moi de remercier tout particulièrement Vincent pour son écoute, sa gentillesse et ses nombreux conseils.
Enfin j'adresse mes plus tendres remerciements à toute ma famille ainsi qu'à tous mes amis pour leur soutien moral. Un grand merci à Christelle, Fabien, Stéphane et Mab.

A mon grand-père

Table des matières

Chapitre 1

Introduction

1.1 Ondelettes et statistique

1.1.1 Motivations

Cette thèse a pour objet l'étude de certaines propriétés statistiques de diverses classes d'estimateurs. En effet, nous nous intéresserons à de grandes familles de procédures contenant la plupart des procédures déjà connues dans la littérature statistique. Plus précisément, nous chercherons à déterminer les espaces fonctionnels maximaux (ou maxisets) sur lesquels ces procédures atteignent une vitesse de convergence donnée, afin de pouvoir comparer ces procédures entre elles et d'établir dans la mesure du possible un estimateur *optimal* au sens maxiset pour chacune des familles considérées. En particulier, cette approche maxiset permettra d'apporter de nouvelles réponses théoriques à certains phénomènes observés en pratique.

Un des principaux enjeu des statistiques non paramétriques consiste à estimer une fonction à valeurs réelles inconnue f à partir d'observations émanant de celle-ci, aussi diverses soient elles. Dans notre travail, nous supposerons que le signal f admet une décomposition unique sur une base d'ondelettes de $\mathbb{L}_2(\mathbb{R})$ fixée :

$$f = \sum_{j \geq -1} \sum_{k \in \mathbb{Z}} \beta_{jk} \psi_{jk}. \tag{1.1}$$

Partir de l'idée que le signal f se décompose sur une famille de fonctions n'est pas nouvelle. En effet, Young (1977[119]), Wahba (1981[118]), Silverman (1985[106]) et Steinberg (1990[107]) considérèrent dans leurs travaux une décomposition sur respectivement des polynômes de Legendre, des polynômes trigonométriques, des B-spline et des polynômes d'Hermite. Cependant, même si les polynômes trigonométriques offraient l'avantage de constituer une base orthogonale de $\mathbb{L}_2(\mathbb{R})$, le choix idéal d'une telle famille restait discutable. Plutôt que de se restreindre à la décomposition sur une famille de polynômes, ce fut plutôt l'idée de décomposer le signal sur une base de $\mathbb{L}_2(\mathbb{R})$ qui fut retenue ensuite et l'apparition des ondelettes au début des années 90 permit alors d'apporter un nouvel élan à l'estimation fonctionnelle en proposant de nouvelles méthodes d'estimation rivalisant avec les méthodes d'estimation par noyaux introduites par Parzen (1962[97]) et beaucoup utilisées jusqu'alors, comme par exemple dans les travaux de Rejtö et Rèvész (1973[100]), Nadaraya (1992[91]), Mammen (1990[86], 1995[87], 1998[88]), Lepski (1991[78]), Lepski, Mammen et Spokoiny (1997[79]), Lepski et Spokoiny (1997[80]), Golubev et Levit (1996[54]) ou encore Tsybakov (2004[112]).

1.1.2 Intérêt des ondelettes

Les travaux de Yves Meyer et de son école (Daubechies, Mallat, Cohen,...) sont les premiers travaux relatifs aux ondelettes. La construction de bases d'ondelettes naquit de l'idée d'exhiber des bases orthogonales dont les atomes seraient à la fois localisés en fréquence et en temps. En effet, on utilisait jusqu'alors des bases orthogonales localisées seulement en temps, comme la base de Haar (voir section 2.1.1) qui offrait des reconstructions non lisses, ou seulement en fréquence, comme la base de Fourier dont un changement, même de faible amplitude, autour d'une fréquence entraînait des changements sur la totalité du domaine temporel. C'est essentiellement pour éviter ce type de désagréments que furent introduites les bases d'ondelettes $(\psi_{jk})_{j,k}$, construites par translations et dilatations dyadiques de deux fonctions ϕ et ψ, appelées respectivement fonction d'échelle et ondelette mère (voir section 2.1.1 pour plus de détails).

Outre la structure algorithmique simple, l'analyse temps-fréquence offre l'avantage de fournir des décompositions où la majorité des coefficients sont petits et où l'essentiel de l'information du signal se trouve dans les quelques grands coefficients (caractère sparse). Ainsi, il semble naturel dans le cadre de l'estimation fonctionnelle de penser privilégier les coefficients empiriques du signal suffisamment grands. C'est pourquoi se développent au milieu des années 90 les procédures dites "de seuillage", développées dans le paragraphe suivant.

1.1.3 Ondelettes et estimateurs

L'objet de ce paragraphe est de rappeler les premiers estimateurs construits par le biais des ondelettes. Pour cela, nous supposerons que le signal f admet une décomposition unique dans une base orthogonale d'ondelettes à supports compacts de $\mathbb{L}_2([0,1])$ selon l'écriture (1.1) et que nous puissions disposer d'observations $\hat{\beta}_{jk}$ des coefficients d'ondelettes β_{jk}, modélisées par des variables aléatoires indépendantes de lois Gaussiennes $\mathcal{N}(\beta_{jk}, \frac{1}{n})$ de moyenne β_{jk} et de variance $\frac{1}{n}$ $(n \in \mathbb{N}^*)$.

La famille des estimateurs linéaires est définie par :

$$\mathcal{F}_L = \left\{ \hat{f} = \sum_{j \geq -1} \sum_k \gamma_{jk} \hat{\beta}_{jk} \psi_{jk}, \quad \gamma_{jk} \in \mathbb{R} \text{ déterministe} \right\}.$$

Si f est supposé être à support compact, le risque $\mathbb{L}_2$ de l'estimateur linéaire $\hat{f}_J$ défini par $\hat{f}_J = \sum_{j=-1}^{J-1} \sum_k \hat{\beta}_{jk} \psi_{jk}$, est tel que :

$$\mathbb{E}\|\hat{f}_J - f\|_2^2 \leq \frac{2^J C}{n} + \sum_{j \geq J} \sum_k \beta_{jk}^2,$$

où C est une constante positive. Ainsi, en supposant que le signal f appartient à l'espace de Besov fort $\mathcal{B}^s_{2,\infty}$ (défini en section 2.2.1) et en choisissant la valeur optimale J^* tel que $2^{J^*} C n^{2s/(1+2s)} = n$, le risque $\mathbb{L}_2$ de l'estimateur linéaire $\hat{f}_{J^*}$ se majore par :

$$\mathbb{E}\|\hat{f}_{J^*} - f\|_2^2 \leq (1 + \|f\|_{\mathcal{B}^s_{2,\infty}}) n^{-2s/(1+2s)}, \tag{1.2}$$

avec :

$$\|f\|_{\mathcal{B}^s_{2,\infty}} := \sup_{J \geq 0} 2^{Js} \sum_{j \geq J} \sum_k \beta_{jk}^2 < \infty.$$

S'il est vrai que l'estimateur $\hat{f}_{j^*}$ est performant (voir par exemple Kerkyacharian et Picard (1992[73])), il nécessite néanmoins la connaissance explicite du paramètre s de l'espace de Besov fort supposé contenir f. En pratique, il ne semble pas très réaliste de supposer la connaissance a priori de ce paramètre de régularité. C'est pour cette raison que nous nous sommes presque exclusivement intéressés dans notre travail à des *procédures adaptatives*, c'est-à-dire des procédures dont la construction ne dépend pas de la connaissance explicite de la régularité du signal.

Par ailleurs, un grand nombre de travaux dont ceux de Nemirovski (1986[93]), Donoho, Johnstone, Kerkyacharian et Picard (1996[48]), Kerkyacharian et Picard (1993[74]), ou Rivoirard (2004[102]) ont souligné les limites des procédures linéaires. D'autres estimateurs se sont alors révélés bien plus performants, comme par exemple les estimateurs de seuillage.

Les estimateurs de seuillage furent introduits par Donoho et Johnstone (1994[42]) pour des bases arbitraires. Ils furent ensuite introduits dans les méthodes d'ondelettes au début des années 90 dans une série d'articles de Donoho et Johnstone (1994[43], 1995[44]) et de Donoho, Johnstone, Kerkyacharian et Picard (1995[47], 1996[48], 1997[49]). L'idée sous-jacente était de reconstruire le signal f uniquement à l'aide des coefficients empiriques $\hat{\beta}_{jk}$ dont la valeur absolue était supérieure à un seuil fixé λ. En particulier l'estimateur de *seuillage dur*

$$\hat{f}_h = \sum_{j \geq -1} \sum_k \hat{\beta}_{jk} \mathbb{1} \{|\hat{\beta}_{jk}| > \lambda\} \psi_{jk},$$

et l'estimateur de *seuillage doux*

$$\hat{f}_s = \sum_{j \geq -1} \sum_k \operatorname{sign}(\hat{\beta}_{jk})(|\hat{\beta}_{jk}| - \lambda)_+ \psi_{jk}$$

se sont vite montrés très performants tant au point de vue théorique que pratique.
Le choix du seuil λ est alors apparu comme un problème essentiel et fut l'objet de nombreux travaux. Citons ceux de Donoho et Johnstone (1994[43], 1995[44]), Nason (1996[92]), Abramovich et Benjamini (1995[2]), Ogden et Parzen (1996[95], 1996[96]), Jansen, Malfait et Bultheel (1997[62]).

1.2 Le point de vue minimax

1.2.1 L'approche minimax

Le but de cette section est de rappeler un point de vue théorique classique pour mesurer la performance d'une procédure d'estimation : le point de vue minimax. Notons $R_\rho^n(\hat{f}_n, f)$ le risque de tout estimateur $\hat{f}_n$, associé à une fonction de perte ρ, défini par :

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

Ainsi défini, le risque $R_\rho^n(\hat{f}_n, f)$ dépend du signal f que l'on cherche à reconstruire. En choisissant un espace fonctionnel V supposé contenir f, on peut alors définir le *risque minimax pour* V par :

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

où l'infimum est pris sur l'ensemble de tous les estimateurs possibles de f. Notons qu'il est nécessaire de choisir un espace fonctionnel V qui soit assez régulier, comme par exemple un espace de Sobolev, de Hölder ou encore de Besov, pour espérer construire par cette approche de bons estimateurs de f. En effet, sans hypothèse de régularité sur f, on ne peut pas en général obtenir des résultats de convergence pour

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

où $\mathcal{F}(\mathbb{R},\mathbb{R})$ désigne l'ensemble des applications de $\mathbb{R}$ dans $\mathbb{R}$ (Farrell (1967[52])). Si r_n est une suite qui tend vers 0 et s'il existe deux constantes positives C_1 et C_2 telles que :

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

alors r_n est appelée *vitesse minimax pour l'espace* V associée à la perte ρ. Les fonctions de perte les plus rencontrées dans la littérature statistique sont celles dérivant des normes $\mathbb{L}_p$ ou des normes associées à des espaces de Sobolev, de Hölder ou de Besov.

L'objectif principal de l'approche minimax est de fournir des estimateurs qui atteignent cette vitesse de convergence. Un estimateur $\hat{f}_n^*$ est alors dit optimal au sens minimax s'il existe une constante positive C_3 telle que :

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

1.2.2 Avantages et inconvénients de cette approche

L'approche minimax constitue donc un moyen de mesurer la performance d'une procédure statistique. Les vitesses minimax ont été calculées pour différents modèles statistiques et pour différentes classes fonctionnelles comme les espaces de Sobolev, de Hölder, de Besov forts et de Besov faibles. Citons parmi d'autres les travaux de Bretagnolle et Huber (1979[14]), Ibragimov et Khasminski (1981[59]), Stone (1982[108]), Birgé (1983[9], 1985[10]), Nemirovski (1986[93]), Kerkyacharian et Picard (1992[73]) et de Rivoirard (2002[101]). A cela, s'ajoutent les travaux de Donoho et Johnstone (1994[43], 1995[44], 1996[45] et 1998[46]), montrant l'optimalité des procédures classiques de seuillage pour estimer les fonctions appartenant aux espaces de Besov.
Mettant en valeur la décomposition biais/variance, l'approche minimax a été source de nombreuses avancées dans les vingt dernières années (méthode de pénalisations, méthode de Lepski,...) et permet de définir un critère d'optimalité théorique pour les estimateurs, relativement à un espace fonctionnel V fixé. Cependant, cette approche présente plusieurs inconvénients qu'il est intéressant de souligner. En premier lieu, l'approche minimax semble trop pessimiste pour fournir une stratégie de décision similaire à celle que l'on pourrait envisager d'un point de vue pratique, en ce sens qu'elle consiste à rechercher des estimateurs minimisant le "risque maximum". En deuxième lieu, le choix de l'espace V supposé contenir le signal f ne fait pas l'unanimité parmi la communauté statistique et reste donc très discutable. Pour finir, cette approche ne fournit pas de critère de comparaison concernant les procédures optimales au sens minimax. Ce sont, entre autres, les raisons pour lesquelles Cohen, DeVore, Kerkyacharian et Picard (2001[31]) ou Kerkyacharian et Picard (2000[75], 2002[76]) envisagèrent une alternative au point de vue minimax : le point de vue maxiset.

1.3 Le point de vue maxiset

1.3.1 L'approche maxiset

Développée au début des années 2000 et inspirée d'une approche de même type en théorie de l'approximation, l'approche maxiset permet d'envisager une nouvelle façon de mesurer la performance d'un estimateur. Ce nouveau point de vue n'a pas pour volonté de s'opposer au point de vue minimax défini juste avant, mais plutôt de fournir une approche qui viendrait compléter la première en écartant les inconvénients mentionnés plus haut. L'approche maxiset consiste à déterminer l'espace fonctionnel maximal (ou maxiset) sur lequel une procédure d'estimation atteint une vitesse de convergence donnée. Sous cette approche, une procédure statistique sera dite *plus performante au sens maxiset* qu'une autre dès lors que le maxiset de la première contiendra celui de la deuxième. Bien évidemment, l'espace maximal d'une procédure sera d'autant plus grand que la vitesse choisie sera faible et inversement. On notera $MS(\hat{f}_n, \rho, v_n)$ le maxiset de toute procédure $\hat{f}_n$ associé à la fonction de perte ρ et à la vitesse de convergence v_n, c'est-à-dire :

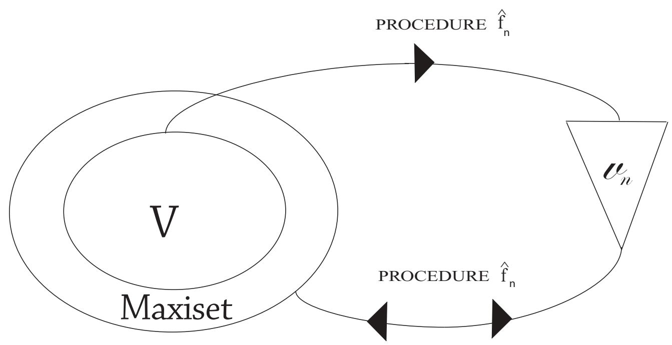
$$MS(\hat{f}_n, \rho, v_n) := \left\{ f; \quad \sup_n v_n^{-1} \mathbb{E}(\rho(\hat{f}_n, f)) < \infty \right\}.$$

En général, les vitesses choisies sont de type n^{-r} ou $\left(\frac{\log(n)}{n}\right)^r$ $(r > 0)$, bien que des vitesses plus générales peuvent aussi apparaître.

Techniques de calcul maxiset.

Bien que l'approche maxiset semble différente de l'approche minimax, les techniques de calcul relatives aux espaces maximaux d'estimateurs sont finalement assez comparables à celles que l'on est amené à faire pour prouver qu'une procédure est asymptotiquement minimax. Par exemple, face à une situation statistique particulière, la méthode standard pour prouver qu'un espace fonctionnel B est le maxiset d'une procédure $\hat{f}_n$, relativement à la fonction de perte ρ et à la vitesse de convergence v_n, s'effectue (exactement comme dans la théorique minimax) en deux étapes : en premier lieu, il s'agit de montrer que $\hat{f}_n$ atteint la vitesse v_n sur B, ce qui revient à dire que $B \subset MS(\hat{f}_n, \rho, v_n)$. Cette première étape utilise des arguments similaires à ceux utilisés pour l'obtention d'inégalités de type majoration (upper bound) dans le contexte minimax. En deuxième lieu, il s'agit de montrer l'inclusion $MS(\hat{f}_n, \rho, v_n) \subset B$. Nous verrons que cette dernière étape utilise, quant à elle, des arguments souvent plus simples que ceux employés pour l'obtention d'inégalités de type minoration (lower bound) dans le contexte minimax.

Comme le montre le schéma qui suit, l'approche maxiset est beaucoup moins pessimiste que l'approche minimax en ce sens qu'elle fournit des espaces fonctionnels directement liés à la procédure d'estimation choisie.



Ainsi, si $\hat{f}_n$ est un estimateur atteignant la vitesse minimax v_n sur un espace fonctionnel V, alors nécessairement $V \subset MS(\hat{f}_n, \rho, v_n)$.

Dans notre travail, nous utiliserons l'approche maxiset pour mesurer les performances de procédures ou de familles de procédures. Plus précisément, nous montrerons les points suivants :

a) Les estimateurs de seuillage sont robustes relativement à l'hypothèse de compacité du support de la fonction f à estimer (chapitre 3).

b) Les procédures de seuillage aléatoire, comme par exemple celle proposée par Juditsky et Lambert-Lacroix (2004[72]), peuvent se révéler plus performantes que celles de seuillage déterministe (chapitre 3).

c) L'approche maxiset nous permet de choisir les lois a priori en contexte Bayésien. Nous montrerons en particulier que si les lois à queues lourdes donnent de bonnes performances, comme l'ont montré Johnstone et Silverman (2002[68],2004[70]) ou Rivoirard (2004[103]), on peut néanmoins utiliser une loi a priori Gaussienne en compensant par l'apport d'une grande variance (chapitre 4).

d) Les procédures héréditaires, qui tiennent compte des liaisons filiales de type dyadiques et dont certaines peuvent être liées à la procédure de Lepski, fournissent des maxisets plus grands que ceux connus jusqu'alors (chapitre 5).

e) Les estimateurs de seuillage par blocs sont plus performants que les estimateurs de seuillage classique, dès lors que la longueur des blocs est assez petite (chapitre 6).

Il est intéressant de souligner que les points b) et e) confirment certaines observations faites en pratique, à savoir les meilleures performances des procédures de seuillage aléatoire par rapport aux procédures de seuillage déterministe (voir par exemple Donoho et Johnstone (1995[44])), ainsi que les meilleures performances des procédures de seuillage par blocs par rapport aux procédures de seuillage individuel (voir Hall, Penev, Kerkyacharian et Picard (1997 [56]) et Cai(1998[16], 1999[17], 2002[18])).

Avant d'exposer plus en détails les principaux résultats de cette thèse, rappelons les premiers résultats de type maxiset établis pour les estimateurs linéaires, les estimateurs de seuillage ainsi que les estimateurs Bayésiens.

1.3.2 Résultats antérieurs

L'idée de l'approche maxiset était sous-jacente dans les résultats de Kerkyacharian et Picard (1993[74]) relatifs au modèle de l'estimation d'une densité (voir section 2.3.1) . En effet, ces auteurs ont prouvé que l'espace maximal de tout estimateur linéaire, associé à la perte $\mathbb{L}_p$ ($p \geq 2$) et à la vitesse de convergence $n^{-sp/(1+2s)}$, est l'espace de Besov fort $\mathcal{B}^s_{p,\infty}$ (voir section 2.2.1).

Dans le cadre du modèle de bruit blanc Gaussien (voir section 2.3.3), Kerkyacharian et Picard (2000[75]) exhibèrent les maxisets des procédures de seuillage. Ils prouvèrent le théorème suivant :

Théorème 1.1. *Soient $1 < p < \infty$ et $0 < \alpha < 1$. On suppose donnée une fonction*

$$f = \sum_{j \geq -1} \sum_k \beta_{jk} \psi_{jk} \in \mathbb{L}_p([0,1]).$$

Sous le modèle de bruit blanc Gaussien

$$y_{jk} = \beta_{jk} + \epsilon z_{jk}, \quad z_{jk} \overset{iid}{\sim} \mathcal{N}(0,1), \quad j \geq -1,\ k \in \mathbb{Z},$$

on considère l'estimateur de seuillage dur :

$$\hat{f}^T_\epsilon = \sum_{j < j_\epsilon} \sum_k y_{jk} \mathbb{1}\ \{|y_{jk}| > \kappa\epsilon\sqrt{\log(\epsilon^{-1})}\} \psi_{jk},$$

avec $2^{-j_\epsilon} \leq \epsilon^2 \log(\epsilon^{-1}) < 2^{-j_\epsilon+1}$ *et* $\kappa > 0$.
Si κ *est une constante assez grande, alors on a l'équivalence suivante :*

$$\sup_{0<\epsilon<1} \left(\epsilon\sqrt{\log(\epsilon^{-1})}\right)^{-\alpha p} \mathbb{E}\|\hat{f}_\epsilon^T - f\|_p^p < \infty \iff f \in \mathcal{B}_{p,\infty}^{\alpha/2} \cap W((1-\alpha)p, p),$$

où

$$\mathcal{B}_{p,\infty}^s = \left\{ f = \sum_{j\geq -1}\sum_k \beta_{jk}\psi_{jk} \in \mathbb{L}_p(\mathbb{R}) : \quad \sup_{J\geq -1} 2^{Js} \sum_{j\geq J} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p < \infty \right\},$$

$$W(r,p) = \left\{ f = \sum_{j\geq -1}\sum_k \beta_{jk}\psi_{jk} : \quad \sup_{\lambda>0} \lambda^r \sum_{j=-1}^{\infty} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\left\{|\beta_{jk}| > \lambda\right\} < \infty \right\}.$$

Les espaces de Besov forts $\mathcal{B}_{p,\infty}^s$ (voir section 2.2.1) constituent une grande famille d'espaces contenant les espaces de Hölder. Les espaces $W(r,p)$, quant à eux, sont appelés espaces de Besov faibles et constituent une sous-classe des espaces de Lorentz (voir section 2.2.2) dont les normes associées permettent de mesurer la régularité et le caractère sparse d'une fonction comme on peut le voir par exemple dans le cas $r = 2$. Ce résultat, auquel on peut ajouter ceux de Cohen, DeVore, Kerkyacharian et Picard (2001[31]), montre donc que les procédures de seuillage s'avèrent performantes dès lors que le signal f est assez sparse.
La présence de tels espaces dans le cadre maxiset n'est pas étonnante puisque les travaux de Donoho (1996[40]) et de Cohen, DeVore et Hochmuth (2000[30]) avaient déjà montré le rôle important des espaces de Lorentz en codage et également en théorie de l'approximation. D'autres résultats mêlant espaces de Lorentz et théorie de l'approximation peuvent se trouver dans les travaux de DeVore (1989[50]), DeVore et Lorentz (1993[38]), Donoho (1993[39]), Johnstone (1994[63]), Donoho et Johnstone (1996[45]), DeVore, Konyagin et Temlyakov (1998[37]), Temlyakov (1999[110]) ou encore Cohen (2000[28]).

Plus récemment, Kerkyacharian et Picard (2002[76]) ont montré que les procédures consistant à sélectionner localement le pas d'un noyau (voir Lepski (1991[78])) étaient au moins aussi performantes au sens maxiset que les procédures de seuillage. De cette constatation, surgit naturellement l'idée d'exhiber des procédures adaptatives directement inspirées de celles-ci et d'en étudier les propriétés statistiques. Un tel type de travail sera effectué au cours du chapitre 5.

Les travaux de Rivoirard (2004[102], 2004[103]) reposent sur la détermination des espaces maximaux des procédures linéaires et des estimateurs bayésiens construits à partir de densités à queues lourdes. Il en résulte d'une part que les estimateurs linéaires sont sous-optimaux au sens maxiset par rapport aux procédures de seuillage et d'autre part que les espaces maximaux des procédures bayésiennes classiques sont de même type que les estimateurs de seuillage, à savoir l'intersection d'un espace de Besov fort avec un espace de Besov faible. Nous verrons au cours du chapitre 4 que les procédures de seuillage ainsi que les procédures bayésiennes relatives à la médiane et à la moyenne de la loi a posteriori sont optimales au sens maxiset parmi toute une famille de procédures : les procédures élitistes.

1.4 Principaux résultats

Après avoir défini au cours du chapitre 2 les différents outils et les diverses notions que nous serons amenés à utiliser dans notre travail nous présenterons l'intégralité des résultats obtenus ainsi que leurs preuves dans les chapitres 3 à 6. L'objet de cette section est de donner un premier aperçu de ces résultats. En particulier, la section 1.4.1 décrit les résultats du chapitre 3 relatifs aux points a) et b). La section 1.4.2 présente, quant à elle, les résultats du chapitre 4 concernant le point c). Pour finir, la section 1.4.3 aborde les résultats établis au cours des chapitres 5 et 6 relatifs aux points d) et e).

1.4.1 Seuillage déterministe contre seuillage aléatoire

Dans le chapitre 3, nous nous placerons dans le modèle de l'estimation d'une densité (voir section 2.3.1), en considérant n variables aléatoires indépendantes $X_1, \dots, X_n$ dont la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$ se décompose dans une base biorthogonale d'ondelettes (voir section 2.1.2) comme suit :

$$f = \sum_{j \geqslant -1} \sum_{k \in \mathbb{Z}} \beta_{jk} \tilde{\psi}_{jk}.$$

Les objectifs de ce chapitre seront multiples. Dans un premier temps, nous généraliserons les résultats de Cohen, DeVore, Kerkyacharian et Picard (2001[31]) relatifs aux performances maxisets des procédures de seuillage dur, en considérant le risque associé à la norme de Besov $\mathcal{B}^0_{p,p}$ $(1 \leq p < \infty)$. Nous verrons en particulier que les procédures de seuillages dur sont robustes face à l'hypothèse de compacité du support. En effet, l'espace maximal sur lequel ces procédures atteignent la vitesse $\sqrt{n^{-1}\log(n)}^{\alpha p}$ $(0 < \alpha < 1)$ est l'intersection entre un espace de Besov fort et un espace de Besov faible, non restreint aux fonctions à support compact. Par ailleurs, nous verrons que cette procédure est la plus performante parmi les procédures qui consistent à négliger les coefficients empiriques $\hat{\beta}_{jk}$, définis par

$$\hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(X_i),$$

de trop faible valeur. (Théorèmes 3.1 et 3.3).

Au vu des travaux de Donoho et Johnstone (1995[44]), Juditsky (1997[71]), Birgé et Massart (2000[12]) et de Juditsky et Lambert-Lacroix (2004[72]) sur le choix de seuils non plus déterministes mais aléatoires pour la construction de procédures optimales au sens minimax, il semblait naturel de s'interroger sur l'intérêt de tels choix. Un des objectifs du chapitre 3 sera aussi de justifier les meilleures performances (au sens maxiset) des procédures de seuillage aléatoire par rapport aux procédures de seuillage déterministe. Pour cela, nous nous intéresserons plus particulièrement au maxiset de la procédure proposée par Juditsky et Lambert-Lacroix (2004[72]), définie par

$$\bar{f}_n = \sum_{j<j_n}\sum_{k\in\mathbb{Z}} \hat{\beta}_{jk} \mathbb{1}\ \{|\hat{\beta}_{jk}| > \mu\sqrt{\frac{\log(n)}{n}}\hat{\sigma}_{jk}\}\tilde{\psi}_{jk},$$

où $2^{-j_n} \leq \frac{\log(n)}{n} < 2^{1-j_n}$, $\hat{\sigma}^2_{jk} = \frac{1}{n}\sum_{i=1}^{n}\psi^2_{jk}(X_i) - \hat{\beta}^2_{jk}$, $\mu > 0$.

Ce maxiset associé à la vitesse de convergence $v_n := \left(\frac{\log(n)}{n}\right)^{\alpha p/2}$ se révélera être plus grand que celui associé à la procédure de seuillage déterministe $\hat{f}_n$, définie par :

$$\hat{f}_n = \sum_{j<j_n}\sum_{k\in\mathbb{Z}} \hat{\beta}_{jk} \mathbb{1}\ \{|\hat{\beta}_{jk}| > \mu\sqrt{\frac{\log(n)}{n}}\}\tilde{\psi}_{jk},$$

où $2^{-j_n} \leq \frac{\log(n)}{n} < 2^{1-j_n}$ et $\mu > 0$ (assez grand). Plus précisément, nous montrerons que :

Théorème 1.2. *Soient* $0 < \alpha < 1$ *et* $1 \leqslant p < \infty$ *tels que* $\alpha p > 2$. *Pour toute valeur de* μ *assez grande, on a :*

$$\begin{aligned} MS(\hat{f}_n, \|.\|^p_{\mathcal{B}^0_{p,p}}, v_n) &= \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p), \\ MS(\bar{f}_n, \|.\|^p_{\mathcal{B}^0_{p,p}}, v_n) &= \mathcal{B}^{\alpha/2}_{p,\infty} \cap W_\sigma((1-\alpha)p, p), \end{aligned}$$

où $W(r,p)$ *caractérise l'espace de Besov faible de paramètres* r *et* p *et*

$$W_\sigma(r,p) = \left\{ f;\ \sup_{\lambda>0} \lambda^r \sum_{j\geqslant -1} 2^{j(\frac{p}{2}-1)} \sum_k \sigma^p_{jk} \mathbb{1}\ \{|\beta_{jk}| > \lambda\sigma_{jk}\} < \infty \ \textit{avec}\ \sigma^2_{jk} = \int f(t)\psi^2_{jk} dt - \beta^2_{jk} \right\}.$$

Nous prouverons aussi que les espaces maximaux de ces procédures sont emboîtés de la façon suivante :

$$\mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p) \subset \mathcal{B}^{\alpha/2}_{p,\infty} \cap W_\sigma((1-\alpha)p, p).$$

Ainsi nous pourrons conclure que la procédure de Juditsky et de Lambert-Lacroix $\bar{f}_n$ est plus performante que celle de seuillage dur classique $\hat{f}_n$.

Ce résultat permet d'apporter une justification théorique à un premier phénomène observé en pratique, à savoir que les procédures de seuillage aléatoire sont souvent de meilleure performance que celles de seuillage déterministe (Donoho, Johnstone (1995[44]).

Dans les chapitres 4 à 6, nous nous placerons dans le modèle du bruit blanc Gaussien (voir section 2.3.3) :

$$X_\epsilon(dt) = f(t)dt + \epsilon W(dt),$$

où ϵ ($\epsilon > 0$) représente le niveau de bruit. Cette fois-ci, nous supposerons que f est à support dans $[0,1]$. Nous nous intéresserons à la famille des procédures de *contraction* (shrinkage rules), définie par :

$$\mathcal{F}_{sh} = \left\{ \hat{f}_\epsilon = \sum_{j\geq -1} \sum_k \gamma_{jk} y_{jk} \psi_{jk},\ \gamma_{jk} \in [0,1],\ y_{jk} = X_\epsilon(\psi_{jk}) \right\}.$$

Pour tout $\lambda > 0$, nous noterons j_λ le plus petit entier j tel que $2^{-j} \leq \lambda^2$.

1.4.2 Maxisets et choix de la loi a priori pour les procédures Bayésiennes

Dans le chapitre 4, nous nous intéresserons aux performances maxisets associées au risque $\mathbb{L}_2$ de deux grandes familles de procédures qui reflètent des comportements standards parmi les procédures habituellement employées :

La famille des *procédures limitées* $\mathcal{L}(\lambda, a)$ regroupera les procédures attribuant de faibles poids ($\gamma_{jk} \leq a$) aux observations y_{jk} telles que $2^{-j} \leq \lambda$. Les procédures usuelles rencontrées dans la littérature statistique sont toutes limitées (procédures linéaires, de seuillage dur et de seuillage doux, procédures Bayésiennes,etc.).

La famille des *procédures élitistes* $\mathcal{E}(\lambda, a)$ regroupera les procédures attribuant de faibles poids ($\gamma_{jk} \leq a$) aux observations y_{jk} inférieures ou égales en valeur absolue au seuil λ. Les procédures de seuillage dur et de seuillage doux, par exemple, sont élitistes.

Nous verrons que limiter une procédure ou la rendre élitiste a pour conséquence de restreindre son maxiset pour certaines vitesses. Plus précisément, nous fournirons pour chacune de ces deux familles un espace fonctionnel (espace de saturation ou maxiset idéal) pour lequel l'espace maximal de toute procédure de cette famille sera contenu dans ce dernier. En particulier, nous verrons que les espaces de Besov forts (voir section 2.2.1) sont les espaces de saturation des procédures limitées (Théorème 4.1) et que les espaces de Besov faibles (voir section 2.2.2) sont les espaces de saturation des procédures élitistes (Théorème 4.2).

Nous donnerons ensuite des conditions suffisantes pour qu'une procédure d'une classe donnée soit optimale au sens maxiset (Théorèmes 4.4 et 4.5) et nous exhiberons des exemples de telles procédures. Ainsi, nous verrons que les estimateurs linéaires sont optimaux parmi les estimateurs limités. De la même façon, il sera prouvé que les estimateurs de type seuillage dur et de seuillage doux sont optimaux parmi les estimateurs élitistes.

Grâce à l'introduction de ces deux familles de procédures, nous apporterons de nouveaux résultats sur les performances des procédures bayésiennes classiques qui complète-

ront les résultats maxisets établis par Rivoirard (2004[103]). En particulier, de manière analogue à Antoniadis et al. (2002[6]) et Rivoirard (2004[103]) qui ont établi des liens entre les procédures de seuillage et certaines procédures Bayésiennes, nous montrerons que les procédures Bayésiennes classiques sont élitistes, et donc qu'elles ne peuvent pas être de meilleure performance que les procédures de seuillage habituelles.
Pour cela, comme l'ont déjà fait Abramovich, Amato et Angelini (2004[1]), Johnstone et Silverman (2002[68], 2002[69], 2004[70]) et Rivoirard (2004[103]), nous introduirons le modèle a priori suivant sur les coefficients d'ondelettes du signal :

$$\beta_{jk} \sim \pi_{j,\epsilon}\gamma_{j,\epsilon} + (1-\pi_{j,\epsilon})\delta(0), \tag{1.3}$$

où $0 \leq \pi_{j,\epsilon} \leq 1$, $\delta(0)$ représente la masse de Dirac en 0 et où les β_{jk} sont indépendants. Nous supposerons que $\gamma_{j,\epsilon}$ est la dilatation d'une densité fixée γ, continue, unimodale, symétrique et positive :

$$\gamma_{j,\epsilon}(\beta_{jk}) = \frac{1}{\tau_{j,\epsilon}}\gamma\left(\frac{\beta_{jk}}{\tau_{j,\epsilon}}\right),$$

où le paramètre de dilatation $\tau_{j,\epsilon}$ est positif. Le paramètre $\pi_{j,\epsilon}$ représente la proportion des coefficients non négligeables du signal f. Enfin, nous noterons

$$\omega_{j,\epsilon} = \frac{\pi_{j,\epsilon}}{1-\pi_{j,\epsilon}},$$

le paramètre indiquant le caractère plus ou moins sparse du signal. En effet si le signal est sparse, alors un grand nombre de coefficients $\omega_{j,\epsilon}$ sera de petite valeur.

Nous nous intéresserons à deux estimateurs bayésiens particuliers : l'estimateur de la médiane a posteriori :

$$[GaussMedian] \quad \breve{f}_\epsilon = \sum_{j<j_\epsilon}\sum_k \breve{\beta}_{jk}\psi_{jk}, \qquad \text{avec } \breve{\beta}_{jk} = \mathrm{Med}(\beta_{jk}|y_{jk}), \tag{1.4}$$

et l'estimateur de la moyenne a posteriori :

$$[GaussMean] \quad \tilde{f}_\epsilon = \sum_{j<j_\epsilon}\sum_k \tilde{\beta}_{jk}\psi_{jk}, \qquad \text{avec } \tilde{\beta}_{jk} = \mathbb{E}(\beta_{jk}|y_{jk}). \tag{1.5}$$

Dans un premier temps, nous étudierons les performances maxisets de ces deux estimateurs dans le cas où γ est la densité Gaussienne et

$$\tau_{j,\epsilon}^2 = c_1 2^{-\alpha j}, \quad \pi_{j,\epsilon} = \min(1, c_2 2^{-bj}),$$

où c_1, c_2, α et b sont des constantes positives, comme suggéré par Abramovich, Sapatinas et Silverman (1998[4]) ou encore Abramovich, Amato et Angelini (2004[1]). En particulier nous prouverons que ces estimateurs limités sont de performance médiocre lorsque $\alpha > 1 + 2s$, en ce sens que leur espace maximal ne contient aucun des espaces de Besov $\mathcal{B}^s_{p,\infty}$, $1 \leq p \leq \infty$ (Théorème 4.8).

Dans un deuxième temps, nous porterons notre étude aux cas où la fonction γ caractérise soit une densité à queue lourde soit la densité Gaussienne, en supposant cette fois-ci que les paramètres $\tau_{j,\epsilon}$ et $w_{j,\epsilon}$ ne dépendent que du niveau de bruit ϵ. Nous montrerons alors que, pour de bons choix de tels paramètres, les procédures limitées définies en (1.4) et (1.5) sont élitistes et nous mettrons en évidence leur optimalité au sens maxiset, à l'aide des Théorèmes 4.9 et 4.10 que nous pouvons résumer par le Théorème suivant :

Théorème 1.3. *Considérons le modèle (1.3), en supposant que $\tau_{j,\epsilon} = \tau(\epsilon)$ et que $\omega_{j,\epsilon} = w(\epsilon)$ sont des paramètres indépendants du niveau j et que w est une fonction continue positive. S'il existe deux constantes positives q_1 et q_2 assez grandes telles que $\epsilon^{q_1} \leq w(\epsilon) \leq \epsilon^{q_2}$, et si de plus l'une des deux hypothèses suivantes est vérifiée*

1. *il existe $M > 0$ et $M_1 > 0$ telles que $\sup_{\beta \geq M_1} \left| \frac{d}{d\beta} \log \gamma(\beta) \right| = M < \infty$ et $\tau(\epsilon) = \epsilon$*
2. *γ est la densité Gaussienne et $1 + \epsilon^{-2}\tau(\epsilon)^2 = (\epsilon\sqrt{\log(\epsilon^{-1})})^{-1}$,*

alors, on a l'équivalence suivante :

$$\sup_{0<\epsilon<1} (\epsilon\sqrt{\log(1/\epsilon)})^{-4s/(1+2s)} \mathbb{E}\|f^0_\epsilon - f\|_2^2 < \infty \iff f \in \mathcal{B}^{s/(1+2s)}_{2,\infty} \cap W(\frac{2}{1+2s}, 2),$$

avec $f^0_\epsilon \in \{\tilde{f}_\epsilon, \breve{f}_\epsilon\}$. C'est-à-dire, en reprenant la notation maxiset :

$$MS(f^0_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(\epsilon^{-1})})^{4s/(1+2s)}) = \mathcal{B}^{s/(1+2s)}_{2,\infty} \cap W(\frac{2}{1+2s}, 2),$$

Ainsi, s'il est vrai que les procédures Bayésiennes associées à des lois a priori à queues lourdes possèdent des performances équivalentes à celle des procédures de seuillage, il en est de même pour les procédures Bayésiennes où les lois a priori sont des lois Gaussiennes à grande variance. Un intérêt pratique émane de ce résultat. En effet, s'il est certaines procédures Bayésiennes qui peuvent paraître difficiles à programmer, ce n'est pas le cas pour les procédures Bayésiennes avec loi a priori Gaussienne.

Nous avons ainsi choisi de mesurer les performances d'un point de vue pratique des deux estimateurs définis en (1.4) et en (1.5) dans le cas où γ est la densité Gaussienne.

Nous avons procédé de la manière suivante. Sous le modèle de régression

$$g_i = f(\frac{i}{n}) + \sigma\epsilon_i, \quad 1 \leq i \leq n = 1024, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0,1),$$

où σ est supposé connu, nous avons appliqué la transformée en ondelettes discrète (voir section 2.3.2) pour les différents vecteurs introduits précédemment afin d'obtenir le modèle statistique suivant :

$$y_{jk} = d_{jk} + \sigma z_{jk}, \quad z_{jk} \overset{iid}{\sim} \mathcal{N}(0,1), \quad -1 \leq j \leq N-1, 0 \leq k < 2^j,$$

où $y_{jk} = (\mathcal{W}g)_{jk}$, $d_{jk} = (\mathcal{W}f^0)_{jk}$, $f^0 = (f(\frac{i}{n}), \quad 1 \leq i \leq n)^T$ et $z_{jk} = (\mathcal{W}\epsilon)_{jk}$. Le problème de l'estimation de f se substitue alors à celui de l'estimation des coefficients $(d_{jk})_{j,k}$. Nous avons ensuite munis ces coefficients d'un modèle Bayésien où la densité a priori est une densité Gaussienne de grande variance. Puis, nous avons reconstruit le signal en estimant les coefficients $(d_{jk})_{j,k}$ selon la façon voulue par le type de procédure considéré (médiane a posteriori, moyenne a posteriori) et en appliquant la transformée en ondelettes discrète inverse.

Nous avons comparé les performances des deux estimateurs pour les quatre fonctions "test" classiques de Donoho et Johnstone ("Blocks", "Bumps", "Heavisine", "Doppler") dans le cas où $\omega_\epsilon = \omega(n) = 10(\sigma n^{-1/2})^q$. Le Tableau 1.1 compare les procédures GaussMedian et GaussMean aux procédures déterministes classiques VisuShrink (Donoho et Johnstone (1994[43])) et GlobalSure (Nason (1996[92])) ainsi qu'à la procédure Bayésienne *BayesThresh* (Abramovich et al. (1998[4])) en fournissant la moyenne de l'erreur en moyenne quadratique (AMSE) calculée à partir de 100 applications de chacune de ces procédures, pour $q = 1$ et pour différents rapports signal/niveau de bruit (RSNR).

RSNR=5	Blocks	Bumps	Heavisine	Doppler
VisuShrink	2.08	2.99	0.17	0.77
GlobalSure	0.82	0.92	0.18	0.59
BayesThresh	0.67	0.74	**0.15**	0.30
GaussMedian	0.72	0.76	0.20	0.30
GaussMean	**0.62**	**0.68**	0.19	**0.29**
RSNR=7	Blocks	Bumps	Heavisine	Doppler
VisuShrink	1.29	1.77	0.12	0.47
GlobalSure	0.42	0.48	0.12	0.21
BayesThresh	0.38	0.45	**0.10**	0.16
GaussMedian	0.41	0.42	0.12	0.15
GaussMean	**0.35**	**0.38**	0.11	**0.15**
RSNR=10	Blocks	Bumps	Heavisine	Doppler
VisuShrink	0.77	1.04	0.08	0.27
GlobalSure	0.25	0.29	0.08	0.11
BayesThresh	0.22	0.25	0.06	0.09
GaussMedian	0.21	0.23	0.06	0.08
GaussMean	**0.18**	**0.20**	**0.06**	**0.07**

TAB. 1.1 – AMSEs pour VisuShrink, GlobalSure, BayesThresh, GaussMedian et GaussMean pour différentes fonctions tests et différentes valeurs de RSNR.

Les résultats obtenus dans le Tableau 1.1 indiquent que les performances des procédures GaussMedian et GaussMean sont très bonnes pour les fonctions "Blocks", "Bumps" et "Doppler" et un peu moins bonnes pour la fonction "Heavysine".
Gaussmean apparaît ici comme la procédure Bayésienne la plus performante puisque ses AMSEs sont généralement les plus faibles (10 fois sur 12). Les performances de GaussMedian, quant à elles, sont presque tout le temps meilleures que celles des procédures non Bayésiennes VisuShrink et GlobalSure et globalement meilleures que celles de BayesThresh dès lors que le rapport signal/bruit est grand (RSNR ≥ 7). Néanmoins, si les procédures GaussMedian et GaussMean s'avèrent très performantes, il faut noter l'apparition d'artefacts (voir Figure 4.1) qu'il est possible de faire disparaître en augmentant les valeurs de q (voir Figure 4.2). Toutefois, de tels choix entraînent inévitablement une aug-

mentation de l'erreur en moyenne quadratique. La valeur $q = 1$ semble alors un bon choix pour obtenir une bonne reconstruction du signal et une erreur en moyenne quadratique des meilleures.

1.4.3 Procédures héréditaires et procédures de μ-seuillage

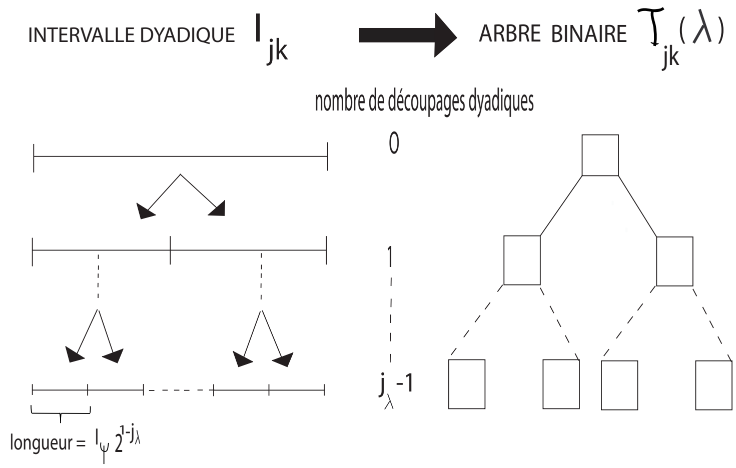
Un des principaux objectifs des chapitres 5 et 6 est de montrer l'existence de procédures adaptatives dont les performances au sens maxiset se trouvent meilleures que celles des procédures élitistes. Pour cela, on s'intéressera aux propriétés de deux nouvelles familles de procédures : les procédures héréditaires et les procédures de μ-seuillage.

Dans tout le paragraphe, on notera $t_\epsilon := \epsilon\sqrt{\log(\epsilon^{-1})}$ et on appellera intervalle dyadique tout intervalle I_{jk} $(j \geq 0, k \in \mathbb{Z})$ tel que $I_{jk} := support(\psi_{jk})$.

PROCEDURES HEREDITAIRES :

Au cours du chapitre 5, nous étudions les espaces maximaux associés à une nouvelle famille de procédures qui utilise plus profondément la structure dyadique des méthodes d'ondelettes : les *procédures héréditaires*.

Pour tout $\lambda > 0$ et pour tout intervalle dyadique I_{jk}, considérons l'ensemble des intervalles dyadiques $I_{j'k'}$ obtenus après $j_\lambda - 1$ découpages dyadiques de I_{jk}. Il est possible de construire de façon naturelle un arbre binaire $\mathcal{T}_{jk}(\lambda)$ de profondeur j_λ dont les noeuds sont justement ces intervalles (voir shéma ci-dessous).



La famille des *procédures héréditaires* $\mathcal{H}(\lambda, a)$ regroupera alors les procédures attribuant de faibles poids ($\gamma_{jk} \leq a$) aux observations y_{jk} telles que pour tout intervalle $I_{j'k'}$ de $\mathcal{T}_{jk}(\lambda)$, l'observation $y_{j'k'}$ est inférieure en valeur absolue au seuil λ.

De manière analogue au chapitre 4, nous déterminerons dans un premier temps l'espace de saturation lié aux procédures héréditaires, qui s'avérera être plus grand que celui des procédures élitistes. Dans un deuxième temps, nous exhiberons deux exemples de procédures héréditaires optimales au sens maxiset (procédures hard tree et soft tree). Ce résultat aura toute son importance, car si le plus grand maxiset rencontré dans la littérature statistique demeurait jusqu'à présent celui des procédures de seuillage classique, il n'en sera plus.

Dans la deuxième partie du chapitre 5, nous montrons que l'une des deux procédures optimales mentionnées plus haut, que l'on nommera procédure hard tree, est étroitement liée à la procédure de Lepski (1991[78]). En effet, en supposant que la base d'ondelettes choisie est celle de Haar (voir section 2.1.1), nous mettrons en évidence les similitudes entre cette procédure et celle de Lepski ainsi que les différences entre cette procédure et la procédure hybride (non héréditaire) proposée par Picard et Tribouley (2000[99]).

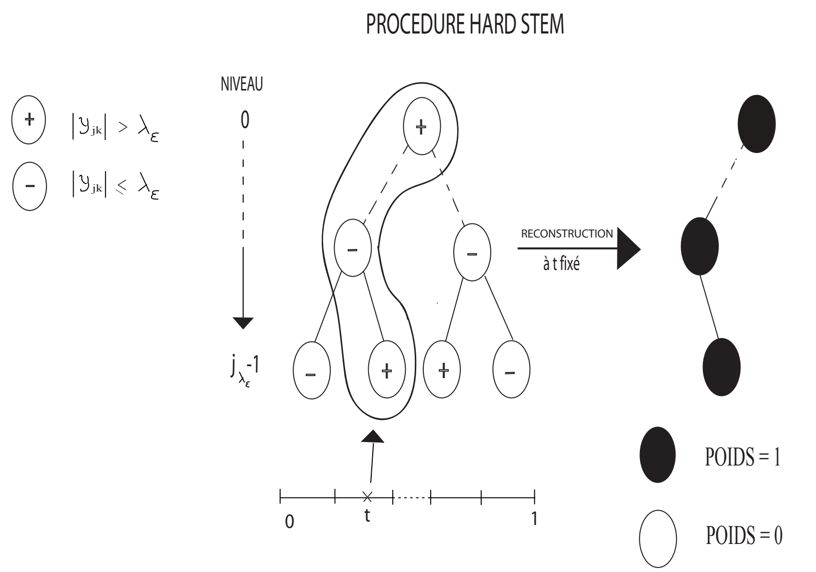
Procédure hard stem.

La procédure de Picard et Tribouley, que l'on nommera dorénavant hard stem, est basée sur une reconstruction locale du signal f. Au même titre que les procédures de seuillage dur, les observations y_{jk} dont le niveau de résolution j est trop grand ne sont pas prises en compte dans la reconstruction de f (poids 0). A t fixé, l'estimateur hard stem du signal est défini comme suit :

$$\tilde{f}_L(t) = y_{-10}\psi_{-10}(t) + \sum_{0\leq j<j_\epsilon} \sum_k \gamma_{jk}(t) y_{jk}\psi_{jk}(t) \tag{1.6}$$

où,
- $2^{-j_\epsilon} \leq (mt_\epsilon)^2 < 2^{1-j_\epsilon}, \quad m > 0$
- $\gamma_{jk}(t) = 1$ s'il existe un intervalle $I_{j'k'} = [\frac{k'}{2^{j'}}, \frac{k'+1}{2^{j'}}[$ inclus dans I_{jk} **et contenant t** tel que $2^{-j'} > (mt_\epsilon)^2$ et $|y_{j'k'}| > mt_\epsilon$, $\gamma_{jk}(t) = 0$ sinon.

Le schéma ci-dessous illustre un tel type de construction à t fixé.



Cette procédure s'est déjà avérée très efficace pour la construction d'intervalles de confiance.

Procédure hard tree.

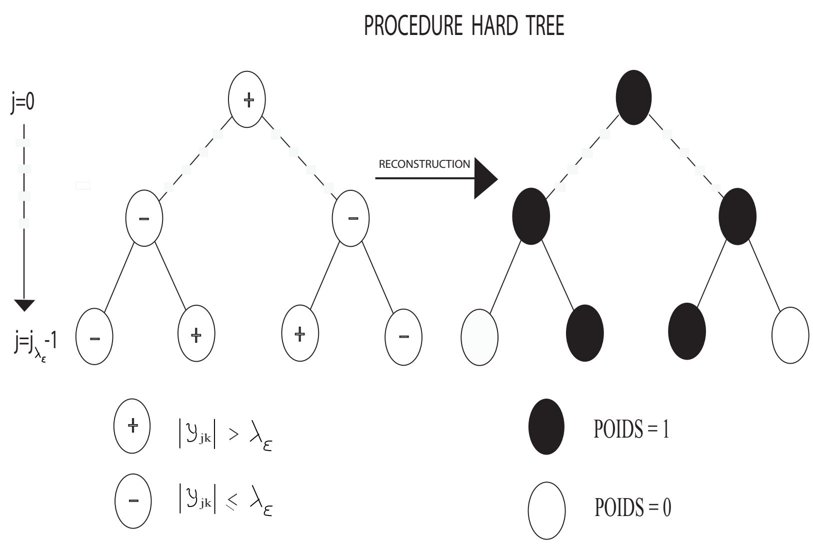
Dans le cas où la base d'ondelettes choisie est celle de Haar, la procédure héréditaire hard tree se définit comme suit :

$$\tilde{f}_T(t) = y_{-10}\psi_{-10}(t) + \sum_{0\leq j<j_\epsilon} \sum_k \gamma_{jk} y_{jk} \psi_{jk}(t) \tag{1.7}$$

où,

- $2^{-j_\epsilon} \leq (mt_\epsilon)^2 < 2^{1-j_\epsilon}, \quad m > 0$
- $\gamma_{jk} = 1$ s'il existe un intervalle $I_{j'k'} = [\frac{k'}{2^{j'}}, \frac{k'+1}{2^{j'}}[$ inclus dans I_{jk} tel que $2^{-j'} > (mt_\epsilon)^2$ et $|y_{j'k'}| > mt_\epsilon$, $\gamma_{jk} = 0$ sinon.

Nous verrons que cette procédure vérifie des contraintes d'hérédité au sens d'Engel (1994[51]) et de Donoho (1997[41]). Le schéma ci-dessous illustre un tel type de construction.



Nous comparerons alors les espaces maximaux associés aux procédures hard stem et hard tree pour le risque $\mathbb{L}_2$ et la vitesse de convergence $t_\epsilon^{4s/(1+2s)}$. Plus précisément, nous montrerons le théorème suivant (résumant les théorèmes 5.3 et 5.4) :

Théorème 1.4. *Soit $s > 0$. Pour tout $m \geq 4\sqrt{3}$:*

$$MS(\tilde{f}_{L}, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}) = \mathcal{B}_{2,\infty}^{s/(1+2s)} \cap W^{L}(\frac{2}{1+2s}, 2)$$

et

$$MS(\tilde{f}_{T}, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}) = \mathcal{B}_{2,\infty}^{s/(1+2s)} \cap W^{T}(\frac{2}{1+2s}, 2),$$

où

$$W^{L}(r,p) = \left\{ f \; ; \; \sup_{\lambda>0} \lambda^r \sum_{0 \leq j < j_\lambda} 2^{j\frac{p}{2}} \sum_k \sum_{|I|=2^{1-j_\lambda}} |\beta_{jk}|^p \mathbb{1} \; \{\forall I_{j'k'} \; / \; I \subset I_{j'k'} \subset I_{jk}, \; |\beta_{j'k'}| \leq \frac{\lambda}{2}\} < \infty \right\}$$

et

$$W^{T}(r,p) = \left\{ f \; ; \; \sup_{\lambda>0} \lambda^{r-2} \sum_{0 \leq j < j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1} \; \{\forall I_{j'k'} \subset I_{jk} \; / \; |I_{j'k'}| > \lambda^2, |\beta_{j'k'}| \leq \frac{\lambda}{2}\} \; < \infty \right\}$$

Contrairement aux espaces de Besov faibles, les espaces $W^{L}(r,p)$ et $W^{T}(r,p)$ ne sont pas invariants par permutations des coefficients de même niveau de résolution j. Nous montrerons que pour tout $0 < r < p < \infty$, $\mathbf{W(r,p) \subset W^L(r,p) \subset W^T(r,p)}$. Ainsi, il nous sera possible de comparer les performances maxisets de ces procédures. En effet, l'emboîtement de ces espaces fonctionnels prouve d'une part que les deux procédures (hard stem et hard tree) sont plus performantes au sens maxiset que les procédures classiques de seuillage, et d'autre part que la procédure hard tree est meilleure que celle proposée par Picard et Tribouley.

Ce chapitre montre donc qu'il est possible de construire des procédures héréditaires dont les performances maxisets sont meilleures que celles de toutes les procédures élitistes. Au cours du chapitre 6, nous verrons qu'une autre famille de procédures offre aussi la possibilité d'exhiber des procédures plus performantes que toutes les procédures élitistes : les procédures de μ-seuillage.

PROCEDURES DE μ-SEUILLAGE :

La famille des procédures de μ-seuillage est une généralisation des procédures usuelles de seuillage faisant intervenir une famille de fonctions décroissantes positives $(\mu_{jk})_{j,k}$ sur lesquelles reposera le choix de garder ou de rejeter les observations y_{jk} pour la reconstruction du signal f. Elle est définie de la façon suivante :

$$\mathcal{F}^{\epsilon}_{seuil} = \left\{ \hat{f}_\mu = \sum_{j<j_\epsilon} \sum_k \mathbb{1}\ \{\mu_{_{jk}}(mt_\epsilon, y_{mt_\epsilon}) > mt_\epsilon\} y_{jk}\psi_{jk},\ \forall \lambda > 0, \mu_{_{jk}}(\lambda, .) : \mathbb{R}^{\# y_\lambda} \longrightarrow \mathbb{R}^+ \right\},$$

avec $m > 0$, $2^{-j_\epsilon} \leq (mt_\epsilon)^2 < 2^{1-j_\epsilon}$ et pour tout $\lambda > 0$, $y_\lambda := (y_{jk}; j < j_\lambda, k)$.

Dans le chapitre 6, nous présenterons des résultats maxisets associés au risque $\mathcal{B}^0_{p,p}$ et à des vitesses de convergences générales. Pour être plus concis, nous nous limiterons ici à présenter ceux relatifs aux vitesses de convergence $t_\epsilon^{2sp/(1+2s)}$, $1 \leq p < \infty$.
Par ailleurs, pour tout $\lambda > 0$ et pour toute fonction f se décomposant comme (1.1), nous noterons $\beta_\lambda := (\beta_{jk}; j < j_\lambda, k)$.

Théorème 1.5. *Soit $\hat{f}_\mu$ une procédure de μ-seuillage telle que les fonctions μ_{jk} associées vérifient les conditions suivantes :*

$\forall (\lambda, t) \in \mathbb{R}^+ \times \mathbb{R}^+, |\mu_{_{jk}}(\lambda, y_\lambda) - \mu_{_{jk}}(\lambda, \beta_\lambda)| > t \Longrightarrow$ *il existe $j' < j_\lambda$ et k' tels que $|y_{j'k'} - \beta_{j'k'}| > t$.*

Si m est suffisamment grand, alors

$$MS(\hat{f}_\mu, \|.\|^p_{\mathcal{B}^0_{p,p}}, t_\epsilon^{2sp/(1+2s)}) = \mathcal{B}^{s/(1+2s)}_{p,\infty} \cap W_\mu(\frac{p}{1+2s}, p) \cap W^*_\mu(\frac{p}{1+2s}, p),$$

où

$$W_\mu(r, p) = \left\{ f; \sup_{\lambda > 0} \lambda^{r-p} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{\mu_{_{jk}}(\lambda, \beta_\lambda) \leq \frac{\lambda}{2}\} < \infty \right\}$$

et

$$W^*_\mu(r, p) = \left\{ f; \quad \sup_{\lambda > 0} \lambda^r \left(\log(\frac{1}{\lambda}) \right)^{-1} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{_{jk}}(\lambda, \beta_\lambda) > 2\lambda\} < \infty \right\}.$$

Ce théorème général permet de caractériser les espaces maximaux des procédures de μ-seuillage. Il est à noter que les espaces $W_\mu(r, p)$ (respectivement $W^*_\mu(r, p)$) sont d'autant plus grands (respectivement petits) que les fonctions μ_{jk} sont grandes. Nous établirons alors des conditions suffisantes à imposer aux fonctions μ_{jk} afin de nous assurer d'une part

que $W_\mu(r,p) \subset W^*_\mu(r,p)$ et d'autre part que la procédure de μ-seuillage associée soit plus performante que les procédures de seuillage dur classiques. En considérant alors des choix particuliers de fonctions μ_{jk}, nous prouverons la supériorité (en terme de performance maxiset) des procédures de seuillage par blocs sur celles de seuillage individuel, dès lors que la longueur des blocs n'excède pas $O(\log^{p/2}(\epsilon^{-1}))$.
Ce résultat est important dans la mesure où ces procédures ont un comportement équivalent à celui des procédures de seuillage classique sous l'approche minimax alors qu'on sait depuis longtemps que les procédures consistant à seuiller les coefficients non pas individuellement mais par blocs donnent souvent de bien meilleurs résultats en pratique, comme l'attestent les travaux de Cai (1998[16], 1999[17], 2002[18]) et Hall, Penev, Kerkyacharian et Picard (1997[56]).

1.5 Perspectives

SUR LE PLAN PRATIQUE :

A travers les différents résultats exposés, notre travail a permis d'une part de comparer les performances de diverses procédures qui étaient jusqu'alors considérées comme équivalentes au sens minimax, et d'autre part d'exhiber des procédures dont les performances au sens maxiset sont meilleures que celles des procédures classiques de seuillage. Si la comparaison en terme de performance maxiset des procédures de seuillage par blocs et des procédures héréditaires ne semble pas envisageable d'un point de vue maxiset (les maxisets associés ne s'emboîtent pas), un axe de recherche possible serait de comparer les performances numériques de ces procédures. Par ailleurs, il serait aussi intéressant de comparer les performances numériques des procédures Bayésiennes construites à partir de lois a priori Gaussiennes (GaussMedian et GaussMean) à celles des procédures Bayésiennes construites à partir de lois a priori à queues lourdes.

SUR LE PLAN THEORIQUE :

Pour chacun des modèles considérés dans notre travail, nous avons supposé que le niveau de bruit $\epsilon > 0$ était connu et que les observations des coefficients du signal étaient indépendantes. Nous pourrions par suite écarter ces hypothèses en envisageant une approche bayésienne pour estimer ϵ, comme l'ont fait Clyde, Parmigiani et Vidakovic (1998[27]), Vidakovic (1998[116]) ou Vidakovic et Ruggeri (2001[117]) d'une part, et modéliser la dépendance des coefficients comme l'ont fait Müller et Vidakovic (1995[90]), Crouse, Nowak et Baraniuk (1998[32]), Huang et Cressie (2000[58]) et Vannucci et Corradi (1999[115]) d'autre part.

Au cours des chapitres 4 et 5, nous avons étudié trois familles de procédures particulières définies en fonction de deux paramètres déterministes λ et a : les procédures limitées, élitistes et héréditaires. Un autre axe de recherche consisterait à prolonger ces travaux en supposant que ces paramètres peuvent dépendre du niveau j et être éventuellement aléatoires. Les espaces maximaux seraient sensiblement différents et pourraient dans certains cas fournir de meilleures procédures au sens maxiset que celles exposées ici.

Enfin, nous avons toujours supposé dans notre travail que le signal f se décomposait de manière unique une fois la base d'ondelettes choisie. Il serait intéressant de s'affranchir de cette hypothèse en considérant des "familles génératrices surabondantes" $(\psi_{a,b})_{a\in[1,\infty),b\in\mathbb{R}_+}$ avec

$$\psi_{a,b}(t) = a^{\frac{1}{2}}\psi(at-b),$$

pouvant offrir des décompositions plus adaptatives (voir Davis, Mallat et Zhang (1994[35]) et Chen, Donoho et Saunders (1998[23])).

L'approche maxiset nous a permis de mesurer les performances d'estimateurs adaptatifs très variés, comme les procédures de μ-seuillage, les procédures bayésiennes classiques et certaines procédures de type arbre. Cette approche semble donc très prometteuse et pourrait être envisagée pour mesurer les performances d'autres estimateurs comme par exemple les estimateurs CART ou bien les estimateurs associés à une pénalisation (voir Birgé et Massart (1997[11], 2001[13]), Loubes et van de Geer (2002[83])ou van de Geer (2000[113])). Enfin, il serait intéressant d'utiliser ce point de vue pour d'autres modèles étudiés jusqu'alors sous une approche minimax comme le modèle à données dépendantes (voir Johnstone et Silverman (1997[66]) ou Johnstone (1999[64])) ou encore le modèle de l'estimation ponctuelle (voir Picard et Tribouley (2000[99])).

Les chapitres 3, 4, 5 et 6 font l'objet d'articles soumis à des revues. Le chapitre 4 a été écrit en commun avec D. Picard et V. Rivoirard. L'étude des performances maxisets des méthodes de pénalisation est actuellement en cours, en collaboration avec J.M. Loubes et V.Rivoirard.

Chapitre 2

Préliminaires

Le but de ce chapitre est de définir les divers outils mathématiques que nous serons amenés à utiliser dans les chapitres suivants. En particulier, nous rappellerons les notions utiles liées à la théorie des ondelettes et nous définirons les différents espaces fonctionnels ainsi que les différents modèles statistiques mentionnés en introduction.

2.1 Construction de bases d'ondelettes

L'objet de cette section est de rappeler la façon de construire des bases d'ondelettes. Pour plus de détails on se référera aux ouvrages de Meyer (1992[89]), Daubechies (1992[34]) et Mallat (1998[85]).

2.1.1 Bases orthogonales d'ondelettes

La construction de bases orthogonales d'ondelettes repose sur l'analyse multirésolution.

Définition 2.1. *On appelle analyse multirésolution de* $\mathbb{L}_2(\mathbb{R})$ *toute suite croissante de sous espaces fermés de* $\mathbb{L}_2(\mathbb{R})$, $(V_j)_{j\in\mathbb{Z}}$, *vérifiant les propriétés suivantes :*

- $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$,
- $\bigcup_{j\in\mathbb{Z}} V_j$ *est dense dans* $\mathbb{L}_2(\mathbb{R})$,
- $\forall\, f \in \mathbb{L}_2(\mathbb{R}), \quad \forall\, j \in \mathbb{Z}, \quad f(x) \in V_j \iff f(2x) \in V_{j+1}$,
- $\forall\, f \in \mathbb{L}_2(\mathbb{R}), \quad \forall\, k \in \mathbb{Z}, \quad f(x) \in V_0 \iff f(x-k) \in V_0$,

– *il existe une fonction* $\phi \in V_0$*, appelée fonction d'échelle de l'analyse multirésolution, telle que* $\{\phi(x-k): \quad k \in \mathbb{Z}\}$ *soit une base orthonormée de* V_0*.*

A chaque niveau de résolution j, l'espace V_j possède une base orthonormée obtenue par translations et dilatations de la fonction d'échelle ϕ : $\{\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k),\ k \in \mathbb{Z}\}$. Ainsi, la projection de toute fonction de $\mathbb{L}_2(\mathbb{R})$ sur l'espace V_j constitue une approximation de celle-ci au niveau de résolution j. D'autre part, la projection de toute fonction f de $\mathbb{L}_2(\mathbb{R})$ sur l'espace supplémentaire orthogonal W_j de V_j correspond précisément à la différence d'approximation $P_{j+1}f - P_j f$, où P_j (respectivement P_{j+1}) représente l'opérateur de projection de $\mathbb{L}_2(\mathbb{R})$ sur l'espace V_j (respectivement V_{j+1}). Il est alors possible de construire une fonction ψ, appelée *ondelette mère*, de telle sorte que $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k) : \ k \in \mathbb{Z}\}$ soit une base orthonormée de W_j. Ainsi :

$$V_j = \overline{\text{vect}\{\phi_{jk} :\ k \in \mathbb{Z}\}} \quad \text{et} \quad W_j = \overline{\text{vect}\{\psi_{jk} :\ k \in \mathbb{Z}\}},$$

et pour tout entier naturel j_0, toute fonction f de $\mathbb{L}_2(\mathbb{R})$ peut se décomposer comme suit

$$f = \sum_{k \in \mathbb{Z}} \alpha_{j_0 k}\phi_{j_0 k} + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} \beta_{jk}\psi_{jk}, \tag{2.1}$$

où les coefficients d'ondelettes sont définis par

$$\alpha_{j_0 k} = \int f(x)\phi_{j_0 k}(x)dx \text{ et } \beta_{jk} = \int f(x)\psi_{jk}(x)dx.$$

Comme premier exemple de bases d'ondelettes, nous pouvons citer la base de Haar, construite à partir de la fonction échelle $\phi(x) = \mathbb{1}\,\{x \in [0,1]\}$ et de l'ondelette mère $\psi(x) = \mathbb{1}\,\{x \in [0,1/2]\} - \mathbb{1}\,\{x \in]1/2,1]\}$. Comme toute base d'ondelettes construite à partir de l'analyse multirésolution, les atomes de cette base sont à la fois localisés en temps et en fréquence, et construits par translations et dilatations dyadiques d'un système "fonction d'échelle/ondelette" (ϕ, ψ). Cependant, dans ce cas précis, les fonctions associées sont irrégulières et peu oscillantes. Daubechies (1988[33]) propose d'autres bases d'ondelettes à supports compacts pour lesquelles les fonction d'échelle et ondelette mère sont *r régulières*, c'est-à-dire de classe C^r. D'autres exemples de système "fonction d'échelle/ondelette" (ϕ, ψ) sont donnés dans les livres de Daubechies (1992[34]), Mallat

(1998[85]) et Härdle, Kerkyacharian, Picard et Tsybakov (1998[57]).

Dans notre travail, nous avons privilégié (sans perte de généralité) le niveau de résolution $j_0 = 0$ et avons noté $\forall\, k \in \mathbb{Z}$, $\forall\, x \in \mathbb{R}$, $\psi_{-1k}(x) = \phi_{0k}(x)$, $\beta_{-1k} = \alpha_{0k}$ par souci de simplification d'écriture. Il est aussi important de souligner que la majorité des résultats mentionnés dans cette thèse n'impose pas le choix explicite d'une base d'ondelettes.

2.1.2 Bases biorthogonales d'ondelettes

Dans cette section, nous définissons la notion de base biorthogonale d'ondelettes de $\mathbb{L}_2(\mathbb{R})$. Des exemples de telles bases sont donnés par Daubechies (1992[34]).

Définition 2.2. *Soient* (ϕ, ψ) *et* $(\tilde{\phi}, \tilde{\psi})$ *deux systèmes fonction d'échelle/ondelette. On dira que* $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ *constitue une base biorthogonale d'ondelettes de* $\mathbb{L}_2(\mathbb{R})$ *si*

$$\forall j \geq -1, \forall j' \geq -1, \quad \forall (k, k') \in \mathbb{Z}^2, \quad \int_{\mathbb{R}} \psi_{jk}\tilde{\psi}_{j'k'}(t)dt = \delta_{j-j'}\delta_{k-k'},$$

où δ *représente le symbole de Kronecker, et si toute fonction* f *de* $\mathbb{L}_2(\mathbb{R})$ *peut se décomposer de la façon suivante :*

$$f = \sum_{j \geq -1} \sum_{k \in \mathbb{Z}} \left(\int_{\mathbb{R}} f(t)\psi_{jk}(t)dt \right) \tilde{\psi}_{jk} := \sum_{j \geq -1} \sum_{k \in \mathbb{Z}} \beta_{jk}\tilde{\psi}_{jk}$$

en adoptant pour notations $\psi_{-1k} = \phi_{0k}$ *et* $\tilde{\psi}_{-1k} = \tilde{\phi}_{0k}$.

L'utilisation de ce type de base est fréquente dans le cadre du modèle de l'estimation d'une densité sur $\mathbb{R}$. Pour référence, on citera les travaux de Juditsky et Lambert-Lacroix (2004[72]) sur lesquels se sont appuyés les résultats énoncés dans le chapitre 3 mettant en évidence l'intérêt des méthodes de seuillage aléatoire par rapport aux méthodes de seuillage déterministe.

2.2 Quelques espaces fonctionnels mis en jeu

L'objet de cette section est de définir certains espaces fonctionnels souvent rencontrés sous l'approche maxiset : les espaces de Besov. Nous en rappellerons aussi certaines propriétés qui leur sont associées.

2.2.1 Les espaces de Besov forts

Dans un premier temps, nous commençons par définir les *espaces de Besov forts*. Pour plus de détails, on se référera aux travaux de Bergh et Löfström (1976[8]), Peetre (1976[98]), Meyer (1992[89]) ou DeVore et Lorentz (1993[38]).

Les espaces de Besov forts se définissent en terme de module de continuité. Notons pour tout $(x,h) \in \mathbb{R}^2$, $\Delta_h f(x) = f(x-h) - f(x)$ et $\Delta_h^2 f(x) = \Delta_h(\Delta_h f(x))$.

Pour tout $0 < s < 1$, $1 \leq p \leq \infty$, $1 \leq q < \infty$, on définit

$$\gamma_{spq}(f) = \left(\int_{\mathbb{R}} \left(\frac{\|\Delta_h f\|_p}{|h|^s} \right)^q \frac{dh}{|h|} \right)^{1/q},$$

et

$$\gamma_{sp\infty}(f) = \sup_{h \in \mathbb{R}^*} \frac{\|\Delta_h f\|_p}{|h|^s}.$$

Lorsque $s = 1$, on pose

$$\gamma_{1pq}(f) = \left(\int_{\mathbb{R}} \left(\frac{\|\Delta_h^2 f\|_p}{|h|} \right)^q \frac{dh}{|h|} \right)^{1/q},$$

et

$$\gamma_{1p\infty}(f) = \sup_{h \in \mathbb{R}^*} \frac{\|\Delta_h^2 f\|_p}{|h|}.$$

Pour tout $0 < s \leq 1$, $1 \leq p, q \leq \infty$, l' espace de Besov fort de paramètres s, p et q, noté $\mathcal{B}_{p,q}^s$ est défini par :

$$\mathcal{B}_{p,q}^s = \left\{ f \in \mathbb{L}_p(\mathbb{R}) : \quad \gamma_{spq}(f) < \infty \right\},$$

muni de la norme :

$$\|f\|_{\mathcal{J}_{p,q}^s} = \|f\|_p + \gamma_{spq}(f).$$

Dès lors que $s = [s] + \alpha$, avec $[s] \in \mathbb{N}$ et $0 < \alpha \leq 1$, on dira que $f \in \mathcal{B}_{p,q}^s$ si et seulement si $f^{(m)} \in \mathcal{B}_{p,q}^{\alpha}$, pour tout $m \leq [s]$. Cet espace est muni de la norme :

$$\|f\|_{\mathcal{J}_{p,q}^s} = \|f\|_p + \sum_{m \leq [s]} \gamma_{\alpha pq}(f^{(m)}).$$

Une caractérisation essentielle des espaces de Besov forts repose sur la notion de vitesse d'approximation. En effet, on a le résultat suivant (Donoho, Johnstone, Kerkyacharian et Picard (1996[48])) :

Théorème 2.1. *Soient* $N \in \mathbb{N}$, $0 < s < N+1$, $1 \leq p, q \leq \infty$ *et* (ϕ, ψ) *un système "fonction d'échelle/ondelette" pour lequel il existe une fonction décroissante bornée* H *telle que :*

$$\begin{aligned} &1) \quad \forall\, x, y \quad |\sum_k \phi(x-k)\phi(y-k)| \leq H(|x-y|) \\ &2) \quad \int H(u)|u|^{N+1} du < \infty \\ &3) \quad \phi^{(N+1)} \text{ existe et } \sup_{x\in\mathbb{R}} \sum_k |\phi^{(N+1)}(x-k)| < \infty. \end{aligned}$$

Notons P_j, $j \geq 0$, *les opérateurs de projection sur les espaces* V_j. *Alors* f *appartient à l'espace de Besov fort* $\mathcal{B}^s_{p,q}$ *si et seulement si* $f \in \mathbb{L}_p(\mathbb{R})$ *et s'il existe une suite de nombres positifs* $(\epsilon_j)_{j\in\mathbb{N}} \in l_q(\mathbb{N})$ *telle que :*

$$\forall\, j \in \mathbb{N}, \quad \|f - P_j f\|_p \leq 2^{-js}\epsilon_j.$$

En terme de coefficients d'ondelettes, il est alors possible de donner une nouvelle définition des espaces de Besov forts, qui présente l'avantage d'être facile d'emploi et plus adaptée à la théorie des ondelettes :

Définition 2.3. *Toute fonction* $f \in \mathbb{L}_p(\mathbb{R})$, *dont les coefficients dans une base d'ondelettes fixée sont*

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

appartient à l'espace de Besov fort $\mathcal{B}^s_{p,q}$ *si et seulement si*

$$\|f\|_{\mathcal{B}^s_{p,q}} = \|\alpha_{0.}\|_{l_p} + \left(\sum_{j\geq 0} 2^{jq(s-1/p+1/2)}\|\beta_{j.}\|^q_{l_p}\right)^{1/q} < \infty, \quad \text{si } q < \infty,$$

et

$$\|f\|_{\mathcal{B}^s_{p,q}} = \|\alpha_{0.}\|_{l_p} + \sup_{j\geq 0} 2^{j(s-1/p+1/2)}\|\beta_{j.}\|_{l_p} < \infty \quad \text{si } q = \infty.$$

Les normes $\|.\|_{\mathcal{B}_{s,p,q}}$ et $\|.\|_{\mathcal{J}_{s,p,q}}$ sont équivalentes et les inclusions suivantes sont vérifiées :

$$\mathcal{B}^s_{p,q} \subset \mathcal{B}^{s'}_{p,q'}, \text{ si } s > s' \text{ ou pour } s' = s \text{ et } q \leq q',$$

$$\mathcal{B}^s_{p,q} \subset \mathcal{B}^{s'}_{p',q}, \text{ si } p' > p \text{ et } s' - 1/p' = s - 1/p.$$

De plus, pour $s > 1/p$ et $q > 1$, $\mathcal{B}^s_{p,q}$ est inclus dans l'espace des fonctions continues et bornées.

Les espaces de Besov forts constituent une très grande famille de fonctions. En particulier, rappelons que l'espace de Sobolev H^s correspond précisément à l'espace $\mathcal{B}^s_{2,2}$ et l'espace de Hölder H^s (avec $0 < s \notin \mathbb{N}$) à l'espace $\mathcal{B}^s_{\infty,\infty}$.

Nous verrons au chapitre 4 le lien entre les espaces de Besov forts et les estimateurs limités.

Donoho et Johnstone (1996[45]), Cohen (2000[28]), Kerkyacharian et Picard (2002[76]) et Rivoirard(2004[103]) ont mis en évidence de fortes connexions entre les procédures de seuillage et une sous classe des espaces de Lorentz : les *espaces de Besov faibles*.

2.2.2 Les espaces de Besov faibles

Commençons tout d'abord par rappeler la définition des espaces de Lorentz, aussi appelés espaces $\mathbb{L}_p$ faibles, ou espaces de Marcinkiewicz (voir Lorentz (1950[81], 1966[82]), DeVore et Lorentz (1993[38])).

Définition 2.4. *Si Ω est un espace muni d'une mesure positive μ, pour tout $0 < p < \infty$, l'espace de Lorentz $\mathbb{L}_{p,\infty}(\Omega, \mu)$ est l'ensemble des fonctions $f : \Omega \longrightarrow \mathbb{R}$ μ-mesurables telles que :*

$$\sup_{\lambda > 0} \lambda^p \mu(|f| > \lambda) = \|f\|^p_{\mathbb{L}_{p,\infty}(\Omega,\mu)} < \infty.$$

Si $\Omega = \mathbb{N}^*$ et si μ est une mesure sur $\mathbb{N}^*$, on notera $wl_p(\mu) = \mathbb{L}_{p,\infty}(\mathbb{N}^*, \mu)$ et $wl_p = wl_p(\mu^*)$ si μ^* est la mesure de comptage sur $\mathbb{N}^*$.

De manière évidente,

$$wl_p = \left\{ \theta = (\theta_n;\ n \in \mathbb{N}^*; \quad \sup_{\lambda>0} \lambda^p \sum_n \mathbb{1}\ \{|\theta_n| > \lambda\} < \infty \right\}$$

peut être identifié avec l'ensemble des suites $\theta = (\theta_n;\ n \in \mathbb{N}^*)$ telles que

$$\sup_{n\in\mathbb{N}^*} n^{\frac{1}{p}} |\theta|_{(n)} < \infty, \tag{2.2}$$

où

$$|\theta|_{(1)} \geq |\theta|_{(2)} \geq \cdots \geq |\theta|_{(n)} \ldots,$$

est le réarrangement de θ dans l'ordre décroissant. Cet espace séquentiel est fortement lié à l'espace l_p et peut être vu comme une version faible de l'espace l_p. En effet :

$$l_p \subset wl_p \subset l_{p+\delta}, \quad \delta > 0.$$

La majoration (2.2) fournit un contrôle polynomial de la suite $(|\theta|_{(n)})_{n\in\mathbb{N}^*}$, et donc un contrôle de la proportion des grandes composantes de θ, relativement à p. Les espaces wl_p constituent donc une classe idéale pour mesurer le caractère sparse d'une suite. De même, en considérant les espaces $wl_p(\mu)$ avec un bon choix de μ, il sera possible de mesurer la régularité d'une suite.

Les espaces de Besov faibles de paramètres r et p sont définis par :

$$W(r,p) = \left\{ f = \sum_{j\geq -1} \sum_k \beta_{jk}\psi_{jk} : \quad \sup_{\lambda>0} \lambda^r \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\mathbb{Z}} \mathbb{1}\ \{|\beta_{jk}| > \lambda\} < \infty \right\},$$

ou la définition équivalente donnée par Cohen (2000[28])

$$W(r,p) = \left\{ f = \sum_{j\geq -1} \sum_k \beta_{jk}\psi_{jk} : \quad \sup_{\lambda>0} \lambda^{p-r} \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_{k\in\mathbb{Z}} |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq \lambda\} < \infty \right\}.$$

Ainsi définis, les espaces $W(r,p)$ constituent clairement une sous-classe des espaces de Lorentz dont la norme associée permet de mesurer la régularité (paramètre p) et le caractère sparse (paramètre r) d'une fonction. En effet, quand r diminue, le nombre de coefficients négligeables augmente mais les quelques rares coefficients non négligeables peuvent être très grands.

En utilisant la version séquentielle des espace de Besov forts, on peut remarquer que $W(r,p)$ apparaît comme une version faible de l'espace de Besov fort classique $\mathcal{B}^s_{r,r}$ avec $s = \frac{1}{2}(\frac{p}{r} - 1)$, $p > r$.

Nous verrons au chapitre 4 les liens entre ces espaces et les estimateurs élitistes. Dans ce même chapitre, nous verrons aussi que d'autres espaces, dont les définitions sont assez proches des espaces de Besov faibles, seront mis à contribution lors de l'étude des espaces maximaux associés à d'autres familles d'estimateurs.

2.3 Modèles statistiques

L'objet de cette section est de décrire les modèles statistiques sur lesquels s'appuie notre travail.

2.3.1 Le modèle de l'estimation d'une densité

Dans le chapitre 3, nous nous plaçons dans le modèle de l'estimation d'une densité. Ce modèle statistique est celui utilisé lorsqu'on désire estimer une densité f à partir d'un échantillon de variables indépendantes $X_1, \ldots, X_n$, dont la loi de probabilité admet f comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$

Soient alors (ϕ, ψ) et $(\tilde{\phi}, \tilde{\psi})$ deux systèmes fonction d'échelle/ondelette tels que, ou bien $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ constitue une base biorthogonale d'ondelettes de $\mathbb{L}_2(\mathbb{R})$, ou bien tels que $\phi = \tilde{\phi}$ et $\psi = \tilde{\psi}$. Notons

$$f = \sum_{j \geq -1} \sum_{k \in \mathbb{Z}} \beta_{jk} \tilde{\psi}_{jk} = \sum_{j \geq -1} \sum_{k \in \mathbb{Z}} \left(\int_{\mathbb{R}} f(t) \psi_{jk}(t) dt \right) \tilde{\psi}_{jk}$$

la décomposition de f associé à ce système fonction d'échelle/ondelette et, pour tout (j,k),

$$\hat{\beta}_{jk} = \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(X_i) \quad \text{et} \quad \sigma^2_{jk} = \mathbb{E}(\psi^2_{jk}(X_1)) - \beta^2_{jk}.$$

En appliquant le théorème de la limite centrale, nous avons :

$$\sqrt{n}\left(\frac{\hat{\beta}_{jk}-\beta_{jk}}{\sigma_{jk}}\right) \xrightarrow{loi} \mathcal{N}(0,1),$$

et en appliquant la loi forte des grands nombres, nous avons

$$\hat{\beta}_{jk} \xrightarrow{ps} \beta_{jk}.$$

Chaque $\hat{\beta}_{jk}$ constitue donc un estimateur naturel de β_{jk}, construit par la méthode des moments. C'est donc à partir des $\hat{\beta}_{jk}$ que seront construites les procédures étudiées dans le chapitre 3.

2.3.2 Modèle de régression et transformée en ondelettes discrète

Un des problèmes statistiques les plus classiques consiste à estimer une fonction à partir des observations bruitées des valeurs de cette fonction calculées en n points répartis de manière équidistante sur un intervalle compact. Ainsi, il est très naturel de considérer le modèle de régression non paramétrique suivant :

$$g_i = f(\frac{i}{n}) + \sigma\epsilon_i, \quad 1 \le i \le n, \tag{2.3}$$

où f est la fonction à estimer à partir des n observations $g_1, \dots, g_n$, et chaque ϵ_i suit une loi normale centrée réduite. Le niveau de bruit σ sera supposé connu et les ϵ_i indépendants.

Afin de comparer d'un point de vue pratique certaines de nos procédures, nous exploitons au chapitre 4 le modèle (2.3) en utilisant les outils de la *transformée en ondelettes discrète* : chaque vecteur de taille dyadique subit une succession de transformations linéaires orthogonales définies à partir de filtres associés à un système fonction d'échelle/ondelette (ϕ, ψ). Si $n = 2^N$, $N \in \mathbb{N}$, on construit ainsi une matrice orthogonale W qui transforme le vecteur $f^0 = (f(\frac{i}{n}), \quad 1 \le i \le n)^T$ en un vecteur de même taille noté $d = (d_{jk})_{-1\le j\le N-1, k\in\mathcal{I}_j}$, où $\mathcal{I}_j = \{k \in \mathbb{N} : \quad 0 \le k < 2^j\}$. Le vecteur f^0 est reconstruit en utilisant la formule $f^0 = W^T d$. Mallat (1989[84]) montre que l'ensemble de ces opérations pouvait être effectuées en $O(n)$ opérations. Sous certaines conditions (voir Donoho et Johnstone (1994[43])), si $W_{jk,i}$ désigne le coefficient se trouvant à l'intersection de la $([2^j]+1+k)$ème ligne et de la $i$ème colonne de W, on a l'approximation suivante :

$$n^{\frac{1}{2}}W_{jk,i} \approx 2^{\frac{j}{2}}\psi(2^j i/n - k).$$

Nous en déduisons :

$$d_{jk} \approx n^{\frac{1}{2}}\beta_{jk}, \tag{2.4}$$

où les β_{jk} désignent les coefficients d'ondelette ordinaires de la fonction f définis par :

$$\beta_{jk} = \int_0^1 f(t)\psi_{jk}(t)dt.$$

Puisque la transformation W est orthogonale, on obtient donc le modèle suivant :

$$y_{jk} = d_{jk} + \sigma z_{jk}, \quad z_{jk} \overset{iid}{\sim} \mathcal{N}(0,1), \quad -1 \le j \le N-1, k \in \mathcal{I}_j,$$

où

$$y_{jk} = (\mathcal{W}g)_{jk},$$

et

$$z_{jk} = (\mathcal{W}\epsilon)_{jk}.$$

Une présentation détaillée de l'algorithme précédent est donnée par Daubechies (1992[34]) ou par Härdle, Kerkyacharian, Picard et Tsybakov (1998[57]). Parce que cet algorithme utilise une extension périodique du vecteur f^0, il est préférable d'utiliser des fonctions de $[0,1]$ que l'on peut prolonger de manière périodique sur $\mathbb{R}$ et sans perte de régularité.
Bien que pas toujours des plus fiables, l'approximation (2.4) permet donc de relier un modèle pratique (le modèle (2.3)) et des modèles plus théoriques, comme par exemple, le modèle de bruit blanc Gaussien.

2.3.3 Le modèle du bruit blanc Gaussien

Dans les chapitres 4, 5 et 6, nous nous placerons dans le modèle du bruit blanc Gaussien. Ce modèle est construit à partir d'un processus de Wiener dimensionnel $(W_t)_t$ et s'écrit sous la forme :

$$X_\epsilon(dt) = f(t)dt + \epsilon W(dt), \quad t \in [0,1], \epsilon > 0 \tag{2.5}$$

où f représente le signal à reconstruire à l'aide des observations mises à notre dispositions, à savoir :

$$\mathcal{O} = \left\{ \int_{[0,1]^d} \phi(t) dX_t : \phi \in \mathbb{L}_2([0,1], dt) \right\}.$$

Ce modèle est important en statistique (voir Ibragimov et Khasminski (1981[59])et se trouve très présent dans la littérature. En plus de sa simplicité d'utilisation, il présente un avantage considérable. En effet, en supposant donnée une base orthonormée $\mathcal{E} = (e_k)_{k\in\mathbb{N}}$ de $\mathbb{L}_2([0,1])$, dans laquelle f se décomposerait comme suit $f(t) = \sum_{k\in\mathbb{N}} \theta_k e_k(t)$, les quantités $x_k = \int e_k(t) dX_t$, $k \in \mathbb{N}$ constitueraient alors des observations naturelles des θ_k, vérifiant :

$$x_k = \theta_k + \epsilon z_k, \quad z_k \overset{iid}{\sim} \mathcal{N}(0,1). \tag{2.6}$$

Ainsi, en prenant le cas particulier où la base orthonormée de $\mathbb{L}_2([0,1])$ est une base d'ondelettes, on peut substituer au modèle (2.5) le modèle séquentiel suivant :

$$y_{jk} = \beta_{jk} + \epsilon z_{jk}, \quad z_{jk} \overset{iid}{\sim} \mathcal{N}(0,1) \tag{2.7}$$

Dans les deux modèles, le fait de s'intéresser à la reconstruction du signal f devient analogue à celui de la reconstruction des coefficients d'ondelettes associés. Par ailleurs ce modèle peut être considéré comme une approximation au sens de la convergence des expériences de nombreux modèles classiques, comme les modèle de régression ou de densité décrits précédemment (voir Brown et Low (1996[15]) ou Nussbaum (1996[94])).

Remark 2.1. *Le modèle (2.6) est en fait un cas particulier d'un modèle séquentiel plus général souvent utilisé en statistique des problèmes inverses (voir par exemple Sudakov et Khalfin (1964[109]), Bakushinski (1969[7]), Wahba (1981[118]) et plus récemment Korostelev et Tsybakov (1993[77]) Cavalier (1998[19]), Cavalier et al. (2002[20]), Cavalier et Tsybakov (2002[22]), Johnstone (1999[64]) et Tsybakov (2000[111])).*

Chapitre 3

Maxisets for non compactly supported densities

Summary : The problem of density estimation on $\mathbb{R}$ is concerned. Adopting the maxiset point of view, we focus on adaptive procedures for which the small empirical coefficients are neglected in the reconstruction of the density goal f. Without any assumption on the compacity of the support of f, we show that hard thresholding rule is the best procedure among a large family of procedures, called elitist rules. Then, we point out the significance of data-driven thresholds in density estimation by comparing the maxiset of hard thresholding rule with the one of the procedure using proposed by Juditsky and Lambert-Lacroix.

3.1 Introduction

Dealing with the problem of estimation of compactly supported densities, Cohen, DeVore, Kerkyacharian and Picard (2001[31]) have studied the maximal space (maxiset) where hard thresholding procedure. They have shown that this maxiset is exactly the intersection of a Besov space and a weak Besov space. In this chapter we show that the hypothesis of compacity of the support of f can be kept away.

Recently, Juditsky and Lambert-Lacroix (2004[72]) have proposed a new adaptive procedure for density estimation on $\mathbb{R}$ when dealing with Hölder spaces. In their procedure, they propose to use a data-driven threshold so as to estimate the density function. A natural question arises here : with maxiset regard, is it relevant to alter the usual threshold by a data-driven one ? The main goal of this chapter is to answer this question, underlining the limits of shrinkage rules with non random thresholds in the maxiset sense. Precisely, the aim of this chapter is threefold. Calling **elitist rule** any procedure where the empirical coefficients smaller than v_n in absolute value are neglected, we prove that the maximal space where such a procedure attain the rate $v_n^{\alpha p}$ for the Besov-risk is always contained in a weak Besov space. In fact, we exhibit conditions on procedures ensuring that their maxiset is contained in the intersection of a Besov space and a weak Besov space. Secondly, without any assumption on the compactness of the density to be estimated, we prove that hard thresholding procedures are the *best procedures* among elitist ones, since their maxisets are the largest one among those of elitist rules (**ideal maxiset**). Thirdly, we point out the significance of the choice of data-driven thresholds in density estimation by proving that the maxiset of Juditsky and Lambert-Lacroix's procedure is larger than any elitist rule's one.

The chapter is organized as follows :
Section 3.2 recalls the problem of density estimation on $\mathbb{R}$ and defines the basic tools and functional spaces we shall need in the study. The aim of section 3.3 is to exhibit the ideal maxiset of elitist rules (Theorem 3.1). In section 3.4, we prove that hard thresholding rules are the best procedures (Theorems 3.2, 3.3 and 3.4) among elitist rules. Section 3.5 deals with the data-driven thresholds and section 3.6 is devoted to the proofs of technical lemmas.

3.2 Model and functional spaces

3.2.1 Density estimation model

We consider the problem of estimating an unknown density function f which is as follows. Let $X_1, \ldots, X_n$ be n independent copies of a random variable X with density f with respect to the Lebesgue measure.

To begin, let $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ be compactly supported functions of $\mathbb{L}_2(\mathbb{R})$ and denote for all $k \in \mathbb{Z}$ and $x \in \mathbb{R}$, $\psi_{-1k}(x) = \phi(x-k)$, (resp. $\tilde{\psi}_{-1k}(x) = \tilde{\phi}(x-k)$) and for all $j \in \mathbb{N}$,

$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ (resp. $\tilde{\psi}_{jk}(x) = 2^{j/2}\tilde{\psi}(2^j x - k)$).

Suppose that :

- $\{\psi_{jk};\ j \geq -1; k \in \mathbb{Z}\}$ and $\{\tilde{\psi}_{jk};\ j \geq -1; k \in \mathbb{Z}\}$ constitute a biorthogonal pair of wavelet bases of $\mathbb{L}_2(\mathbb{R})$.
- The *reconstruction wavelet* $\tilde{\psi}$ is $\mathbf{C}^{N+1}$ for some $N \in \mathbb{N}$.
- The *wavelet* ψ is orthogonal to any polynomial of degree less than N.
- $\phi(x) = \mathbb{1}\ \{-\frac{1}{2} \leq x < \frac{1}{2}\}$ and *support*$(\psi) \subset [-\frac{m}{2}, \frac{m}{2}[$ for some $m \in \mathbb{N}^*$.

The important feature of this particular basis which is intensively used throughout the chapter, is that there exists $\nu > 0$ such that $|\psi(x)| \geq \nu$ on the support of ψ. Some most popular examples of such bases are given in Daubechies (1992[34]) and Donoho and Johnstone (1994[43]).

Suppose now that f can be represented as :

$$f(t) = \sum_{j\geq -1}\sum_{k\in\mathbb{Z}} \beta_{jk}\tilde{\psi}_{jk}(t)$$

where $\forall j \geq -1, \forall k \in \mathbb{Z}$:

- $\beta_{jk} = \displaystyle\int_{I_{jk}} f(t)\psi_{jk}(t)dt.$
- $I_{jk} = \left\{x \in \mathbb{R}; -\frac{m}{2} \leq (2^j \vee 1)x - k < \frac{m}{2}\right\}.$

Remark 3.1. *As for any (j,k), the support of ψ_{jk} is contained in I_{jk}, we can easily prove that for any $j \geq -1$ and any $x \in \mathbb{R}$:*

$$\#\{I_{jk};\ x \in I_{jk}\} \leq m. \tag{3.1}$$

In the sequel, we denote :

- $p_{jk} = \displaystyle\int_{I_{jk}} f(t)dt, \qquad \forall j \geq -1 \text{ and } k \in \mathbb{Z},$
- $\sigma^2_{jk} = \displaystyle\int_{I_{jk}} f(t)\psi^2_{jk}(t)dt - \beta^2_{jk}, \qquad \forall j \geq -1 \text{ and } k \in \mathbb{Z},$
- $f_j = \displaystyle\sum_{l=-1}^{j-1}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}, \qquad \forall j > -1.$

Remark 3.2. *Since for all distinct integers* i, i', $\psi_{j,mi}$ *and* $\psi_{j,mi'}$ *have disjoint supports, one gets :*

$$\sum_k p_{jk} = \sum_{l=1}^{m} \sum_i p_{j,mi+l} \leq \sum_{l=1}^{m} \int f(x)dx = m. \tag{3.2}$$

3.2.2 Functional spaces

In this paragraph, we introduce the following sequence spaces often met when dealing with the maxiset approach (see Cohen et al. (2001[31]) and Kerkyacharian and Picard (2000[75])).

Definition 3.1. *Let* $0 < s < N + 1$ *and* $1 \leq p, q \leq \infty$*. We say that a density* f *of* $\mathbb{L}_p(\mathbb{R})$ *belongs to the Besov space* $\mathcal{B}^s_{p,q}$*, if and only if :*

$$\left(2^{j(s-\frac{1}{p}+\frac{1}{2})}\|\beta_{j.}\|_{l_p}\ ;\ j \geq -1\right) \in l_q.$$

Remark 3.3. *It is clear, using the definition above, that the following equivalence is true :*

$$f \in \mathcal{B}^s_{p,\infty} \iff \sup_{J\in\mathbb{N}} 2^{Jsp} \sum_{j\geq J} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p < \infty. \tag{3.3}$$

The Besov spaces are of statistical interest since they model important forms of spatial inhomogeneity. These spaces have been proved to play a prominent part when dealing with the maxiset approach. Indeed, Kerkyacharian and Picard (1993[74]) have proved that the maximal space where any linear procedure attains the rate of convergence $(\sqrt{n^{-1}\log(n)})^p$ for the $\mathbb{L}_p$-risk, $p \geq 2$, is contained in the Besov space $\mathcal{B}^s_{p,\infty}$. Let us recall that the scale of Besov spaces includes the Hölder spaces ($C^s = \mathcal{B}^s_{\infty,\infty}$) and the Hilbert-Sobolev spaces ($H^s_2 = \mathcal{B}^s_{2,2}$).

Definition 3.2. *Let* $0 < r < p < \infty$*. We say that a density* f *belongs to the weak Besov space* $W(r,p)$ *if and only if :*

$$\sup_{\lambda>0} \lambda^r \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{|\beta_{jk}| > \lambda\} < \infty$$

which is equivalent to (see Cohen at al.(2001 [31])) :

$$\sup_{\lambda>0} \lambda^{r-p} \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq \lambda\} < \infty.$$

These spaces naturally appeared when studying the maximal spaces of thresholding rules (see Cohen et al. (2001[31]) and Kerkyacharian and Picard (2000[75])). Weak Besov spaces constitute a large class of functions since, using Markov's inequality, it is easy to prove that for $r < p$, the Besov space $\mathcal{B}^s_{rr} \subset W(r,p)$ when $s \geq \frac{p}{2r} - \frac{1}{2}$. Under the maxiset approach, we prove in section 3.3 that weak Besov spaces are directly connected to a large family of procedures, called *elitist rules.*

Definition 3.3. *Let $0 < r < p < \infty$. We say that a density f belongs to the space $W_\sigma(r,p)$ if and only if :*

$$\sup_{\lambda>0} \lambda^r \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k \sigma^p_{jk} \mathbb{1}\ \{|\beta_{jk}| > \lambda\sigma_{jk}\} < \infty$$

which is equivalent to (see Kerkyacharian and Picard (2000[75])) :

$$\sup_{\lambda>0} \lambda^{r-p} \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq \lambda\sigma_{jk}\} < \infty.$$

$W(r,p)$ and $W_\sigma(r,p)$ are natural spaces to measure the sparsity of a sequence by controlling the proportion of non negligible β_{jk}'s. In section 3.5, we shall show the strong link between the spaces $W_\sigma(r,p)$ and procedures based on data-driven thresholds.

Definition 3.4. *Let $0 < r < p < \infty$. We say that a function f belongs to the space $\chi(r,p)$ if and only if :*

$$\sup_{\lambda>0} \lambda^{r-p} \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{p_{jk} \leq \lambda^2\} < \infty.$$

These functional spaces constitute a large family of functions. To be more precise, let us consider the following proposition, dealing with functional spaces embeddings.

Proposition 3.1. *For any $0 < \alpha < 1$ and any $1 \leq p < \infty$, we have the following inclusions spaces :*

$$\mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p,p) \subset \mathcal{B}^{\alpha/2}_{p,\infty} \cap \chi((1-\alpha)p,p) \text{ and } \mathcal{B}^{\alpha/2}_{p,\infty} \cap W_\sigma((1-\alpha)p,p) \subset \mathcal{B}^{\alpha/2}_{p,\infty} \cap \chi((1-\alpha)p,p) \quad (3.4)$$

Moreover, if $\alpha p > 2$, then :

$$\mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p,p) \subset \mathcal{B}^{\alpha/2}_{p,\infty} \cap W_\sigma((1-\alpha)p,p). \quad (3.5)$$

Proof :
Here and later, the constant C represents any constant we shall need, and can be different from one line to one other.
Denote $K_\psi = \|\psi_{-1}\|_\infty \vee \|\psi_0\|_\infty$. Let $\lambda > 0$ and u be the integer such that $2^u \leq \lambda^{-2} < 2^{1+u}$.

Clearly, if $\lambda^2 \geq \frac{\nu^2}{2K_\psi^2}$, then for any f that belonging to $\mathcal{B}^{\alpha/2}_{p,\infty}$:

$$\begin{aligned}
\sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\left\{p_{jk} \leq \lambda^2\right\} &\leq \sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\leq C\left(\frac{\nu^2}{2K_\psi^2}\right)^{\alpha p/2} \\
&\leq C\lambda^{\alpha p}.
\end{aligned}$$

Suppose now that $\lambda^2 < \frac{\nu^2}{2K_\psi^2}$. Since for any (j,k), $|\beta_{jk}| \leq K_\psi 2^{j/2} p_{jk}$, we have for any $j < u$:

$$\begin{aligned}
p_{jk} \leq \lambda^2 \Longrightarrow |\beta_{jk}| \leq K_\psi \lambda, \text{and } \sigma^2_{jk} &\geq 2^j \nu^2 p_{jk} - 2^j K_\psi^2 p_{jk}^2 \\
&= 2^j p_{jk}(\nu^2 - K_\psi^2 p_{jk}) \\
&\geq 2^{j-1}\nu^2 p_{jk}.
\end{aligned}$$

So, if f belongs to $W((1-\alpha)p, p)$ (resp. $W_\sigma((1-\alpha)p,p)$),

$$\begin{aligned}
\sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\left\{p_{jk} \leq \lambda^2\right\} &\leq \sum_{j<u} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\left\{p_{jk} \leq \lambda^2\right\} + \sum_{j\geq u} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\leq C\sum_{j=-1}^{u-1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\left\{|\beta_{jk}| \leq K_\psi\lambda\right\} + C2^{-\frac{\alpha}{2}up} \\
&\leq C\lambda^{\alpha p}.
\end{aligned}$$

(resp.

$$\begin{aligned}
\sum_{j\geq -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\left\{p_{jk} \leq \lambda^2\right\} &\leq \sum_{j<u} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\left\{p_{jk} \leq \lambda^2\right\} + \sum_{j\geq u} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\leq C\sum_{j=-1}^{u-1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\left\{|\beta_{jk}| \leq K_\psi 2^{j/2} p_{jk}\right\} + C2^{-\frac{\alpha}{2}up}
\end{aligned}$$

$$\begin{aligned}&\leq C\sum_{j=-1}^{u-1}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{1}\,\{|\beta_{jk}|\leq\frac{\sqrt{2}K_\psi}{\nu}\lambda\sigma_{jk}\}+C2^{-\frac{\alpha}{2}up}\\&\leq C\lambda^{\alpha p}.\,)\end{aligned}$$

We conclude that $f\in\chi((1-\alpha)p,p)$. So (3.4) is satisfied. Now, (3.5) is clearly satisfied since for any $1\leq p<\infty$ and any $\alpha>2/p$, $\quad f\in\mathcal{B}^{\alpha/2}_{p,\infty}\Longrightarrow\sup_{j,k}\sigma_{jk}<\infty$. □

3.3 Elitist rules

In this section, we focus on adaptive procedures (i.e which do not depend on the parameter α) concentrating on large empirical coefficients. In particular, we shall study the maxiset properties of such procedures, called *elitist rules.*

3.3.1 Definition of elitist rules

Fix $r>0$. Let $v(n)$ be a decreasing sequence of strictly positive real numbers of limit 0 when n is tending to ∞. Denote j_n the integer such that $2^{j_n}\leq v(n)^{-r}<2^{1+j_n}$ and let E_n be a sequence of statistical experiments such that for any f we can estimate β_{jk} by $\hat{\beta}_{jk}$ for all j,k.

Consider the sub-family $\mathcal{F}_K^{'}$ of Keep-Or-Kill procedures defined by :

$$\mathcal{F}_K^{'}=\left\{\hat{f}(.)=\sum_{j<j_n}\sum_k\omega_{jk}\hat{\beta}_{jk}\tilde{\psi}_{jk}(.);\quad\omega_{jk}\in\{0,1\}\;measurable\right\}.$$

Definition 3.5. *We say that $\hat{f}\in\mathcal{F}_K^{'}$ is an* **elitist rule** *if and only if for any j and any $k\in\mathbb{Z}$:*

$$|\hat{\beta}_{jk}|\leq v(n)\Longrightarrow\omega_{jk}=0$$

This definition exactly means that the "small" coefficients will be neglected.
In Chapter 4, we shall generalize the definition of elitist rules for shrinkage rules.

In the sequel, the choice for the loss function is the Besov norm. A possible alternative could be to use the $\mathbb{L}_p$ norm but this choice leads to technical difficulties avoided by choosing the Besov norm.

3.3.2 Ideal maxisets for elitist rules

The goal of this paragraph is to prove that the maximal space where any elitist rule of $\mathcal{F}_K^{'}$ attains the rate of convergence $v(n)^{\alpha p}$ is contained in the intersection of a Besov space and a weak Besov space.

We have the following theorem :

Theorem 3.1. *Let $0 < \alpha < 1$ and $\hat{f}$ be an elitist rule belonging to $\mathcal{F}_K^{'}$. Then, for any $1 \leq p < \infty$,*

$$MS(\hat{f}, \|.\|^p_{\mathcal{B}^0_{p,p}}, v(n)^{\alpha p}) \subset \mathcal{B}^{\alpha/r}_{p,\infty} \cap W((1-\alpha)p, p).$$

Hence, the intersection spaces constitutes an **ideal maxiset** for elitist rules.

Proof of Theorem 3.1 :
Fix $1 \leq p < \infty$ and let f be such that $\sup_{n>1} v(n)^{-\alpha p} \, \mathbb{E}\|\hat{f} - f\|^p_{\mathcal{B}^0_{p,p}} < \infty$.
On the one hand, for all $n > 1$, we have :

$$\begin{aligned} \sum_{j \geq j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p &\leq \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk} - \hat{\beta}_{jk} \mathbb{1} \{\omega_{jk} = 1\}|^p + \sum_{j \geq j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\ &= \mathbb{E}\|\hat{f} - f\|^p_{\mathcal{B}^0_{p,p}} \\ &\leq C \; v(n)^{\alpha p} \\ &\leq C \; 2^{-j_n \frac{\alpha p}{r}}. \end{aligned}$$

From (3.3), it comes that $f \in \mathcal{B}^{\alpha/r}_{p,\infty}$.

On the other hand, since :

$$|\beta_{jk}| \mathbb{1} \{|\beta_{jk}| \leq \frac{v(n)}{2}\} \leq |\beta_{jk} - \hat{\beta}_{jk} \mathbb{1} \{\omega_{jk} = 1\} \mathbb{1} \{|\hat{\beta}_{jk}| > v(n)\}|,$$

we have :

$$
\begin{aligned}
& \sum_{j \geqslant -1} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq \frac{v(n)}{2}\} \\
\leq\ & \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq \frac{v(n)}{2}\} + \sum_{j \geqslant j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
\leqslant\ & \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk} - \hat{\beta}_{jk} \mathbb{1}\ \{\omega_{jk} = 1\} 1\{|\hat{\beta}_{jk}| > v(n)\}|^p + \sum_{j \geqslant j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
=\ & \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk} - \hat{\beta}_{jk} \mathbb{1}\ \{\omega_{jk} = 1\}|^p + \sum_{j \geqslant j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
=\ & \mathbb{E}\|\hat{f} - f\|_p^p \\
\leqslant\ & Cv(n)^{\alpha p}.
\end{aligned}
$$

So, we have just shown that $f \in W((1-\alpha)p, p)$. □

The aim of the next section is to provide an elitist rule having an ideal maxiset, that is to say a procedure for which the maximal space where it attains the rate $v(n)^{\alpha p}$ is exactly the intersection of a Besov space and a weak Besov space as described above.

3.4 Ideal elitist rule

In this section, we decompose the study into two parts. In a first one, we recall the main result about maxisets of Cohen et al. (2001[31]) when dealing with estimation for **compactly supported densities** (see Theorem 3.2). In the second one, we generalize it for **non compactly supported densities** (see Theorem 3.4). The final outcome of this section is to prove that hard thresholding rules are optimal in the maxiset sense among elitist rules belonging to $\mathcal{F}'_K$. In the sequel, we suppose that $v(n) = \mu\sqrt{\frac{\log(n)}{n}}$, for some $m > 0$, and $r = 2$.

3.4.1 Compactly supported densities

Cohen et al. (2001[31]) have studied the maximal space of hard thresholding rules. These authors have obtained the following result :

Theorem 3.2. *[Cohen et al. (2001[31])] For any $a > 0$, let $I = [-a, a]$, and j_n be the integer such that $2^{j_n} \leq \frac{n}{\log(n)} < 2^{j_n+1}$.*
Denote $\hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(X_i)$ and let us consider the following hard thresholding estimator :

$$\hat{f}_\mu = \sum_{j<j_n}\sum_k \hat{\beta}_{jk}\mathbb{1}\ \{|\hat{\beta}_{jk}| > \mu\sqrt{\frac{\log(n)}{n}}\}\tilde{\psi}_{jk}, \tag{3.6}$$

where μ is a large enough constant. We have for any $0 < \alpha < 1$ and any $1 < p < \infty$:

$$MS(\hat{f}_\mu, \|.\|_p^p, (\frac{\log(n)}{n})^{\alpha p/2}) = \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p). \tag{3.7}$$

The proof of this theorem uses the unconditional nature of the wavelet basis $\{\tilde{\psi}_{jk}; j \geq -1; k \in \mathbb{Z}\}$. In the same way, it would be easy to prove the following similar result.

Theorem 3.3. *Let $1 \leq p < \infty$. Under the same assumptions and definitions as in Theorem 3.2, we get for any $0 < \alpha < 1$ and any $1 \leq p < \infty$:*

$$MS(\hat{f}_\mu, \|.\|^p_{\mathcal{B}^0_{p,p}}, (\frac{\log(n)}{n})^{\alpha p/2}) = \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p). \tag{3.8}$$

Thus, using Theorem 3.1, we conclude that the hard thresholding procedure is optimal in the maxiset sense between the family of elitist rules of $\mathcal{F}'_K$.
A natural question arises here : Is the hard thresholding procedure still optimal among this class of rules, when making no assumption about the compactness of the density goal f ? The answer is YES. We shall prove it in the next paragraph.

3.4.2 Non compactly supported densities

This paragraph aim at proving that hard thresholding procedures still are optimal in the maxiset sense when the density f is supposed to be non compactly supported. Let us introduce the following quantities :

- $m_n = \frac{\mu^2}{K_\psi}\left(1 \wedge \frac{\nu^2}{2K_\psi}\right)\log(n)$
- $\lambda_n = \mu\sqrt{\frac{\log(n)}{n}}$

– $n_{jk} = \sum_{i=1}^{n} \mathbb{1}\ \{X_i \in I_{jk}\}$

– $\hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(X_i)$.

The following theorem can be viewed as a generalization of Theorem 3.3, when dealing with density estimation on $\mathbb{R}$.

Theorem 3.4. *Let $0 < \alpha < 1$ and $1 \leq p < \infty$ such that $\alpha p > 2$. If μ is large enough, then :*

$$\sup_n \left(\frac{n}{\log(n)}\right)^{\alpha p/2} \mathbb{E}\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} < \infty \iff f \in \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p, p).$$

Proof of Theorem 3.4 :
$\subset$: It suffices to apply Theorem 1.
$\supset$: The Besov-risk of $\hat{f}_\mu$ can be decomposed as follows :

$$\begin{aligned} \mathbb{E}\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk} - \hat{\beta}_{jk}\mathbb{1}\ \{|\hat{\beta}_{jk}| > \lambda_n\}|^p + \|f - f_{j_n}\|^p_{\mathcal{B}^0_{p,p}} \\ &= A_0 + A_1. \end{aligned}$$

Since $f \in \mathcal{B}^{\alpha/2}_{p,\infty}$, from (3.3) :

$$A_1 = \|f - f_{j_n}\|^p_{\mathcal{B}^0_{p,p}} \leq \mathbb{E}\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} \leq C\ 2^{-j_n \alpha p/2} \leq C\ \left(\frac{\log(n)}{n}\right)^{\alpha p/2}.$$

A_0 can be decomposed into two parts :

$$\begin{aligned} A_0 &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk} - \hat{\beta}_{jk}\mathbb{1}\ \{|\hat{\beta}_{jk}| > \lambda_n\}|^p \\ &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| \leq \lambda_n\}|^p + \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk} - \hat{\beta}_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| > \lambda_n\} \\ &= A_0^{'} + A_0^{''}. \end{aligned}$$

Now :

$$\begin{aligned} A_0^{'} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| \leq \lambda_n\} \\ &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| \leq \lambda_n\} \left[\mathbb{1}\ \{|\beta_{jk}| \leq 2\lambda_n\} + \mathbb{1}\ \{|\beta_{jk}| > 2\lambda_n\}\right] \\ &= A_{01}^{'} + A_{02}^{'}. \end{aligned}$$

Using the definition of $W((1-\alpha)p, p)$:

$$\begin{aligned} A_{01}^{'} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| \leq \lambda_n\}\mathbb{1}\ \{|\beta_{jk}| \leq 2\lambda_n\} \\ &\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq 2\lambda_n\} \\ &\leq C\,(2\lambda_n)^{\alpha p} \\ &\leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}. \end{aligned}$$

Let us consider the following lemma :

Lemma 3.1. *Let $1 \leq p < \infty$. For any $\gamma > 0$, there exists $\mu(\gamma) < \infty$ and $C < \infty$ such that for any $-1 \leq j < j_n$ and any $k \in \mathbb{Z}$, $\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > \mu\sqrt{\frac{\log(n)}{n}}) \leq \frac{C}{n^\gamma}$.*

The proof is clear using the Bernstein inequality. □

Choosing $\mu(\gamma)$ such that $\gamma \geq \frac{p}{2}$, one gets :

$$\begin{aligned} A_{02}^{'} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| \leq \lambda_n\}\mathbb{1}\ \{|\beta_{jk}| > 2\lambda_n\} \\ &\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{P}_f(|\hat{\beta}_{jk} - \beta_{jk}| > \lambda_n) \\ &\leq C\,n^{-\gamma} \\ &\leq C\,\left(\frac{\log(n)}{n}\right)^{\alpha p/2}. \end{aligned}$$

Let us now consider the following lemma :

Lemma 3.2. *For any $j < j_n$ and any k, $|\hat{\beta}_{jk}| > \lambda_n \Longrightarrow n_{jk} \geq m_n$.*

The proof is given in the appendix.

So, we can decomposed $A_0^{''}$ into three parts :

$$\begin{aligned}
A_0^{''} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| > \lambda_n\} \\
&= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| > \lambda_n\}\mathbb{1}\ \{n_{jk} > m_n\} \\
&= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| > \lambda_n\}\mathbb{1}\ \{n_{jk} > m_n\}\mathbb{1}\ \{p_{jk} < \frac{m_n}{2n}\} + \\
&\quad \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}| > \lambda_n\}\mathbb{1}\ \{n_{jk} > m_n\}\mathbb{1}\ \{p_{jk} \geq \frac{m_n}{2n}\} \left[1\{|\beta_{jk}| \leq \frac{\lambda_n}{2}\} + 1\{|\beta_{jk}| > \frac{\lambda_n}{2}\}\right] \\
&= A_{01}^{''} + A_{02}^{''} + A_{03}^{''}.
\end{aligned}$$

To bound $A_{01}^{''}$, $A_{02}^{''}$ and $A_{03}^{''}$, we introduce two lemmas.

Lemma 3.3. *For any $\gamma > 0$ there exists $\mu = \mu(\gamma) < \infty$ such that for any j, k and any n large enough :*

$$\mathbb{P}_f(n_{jk} < m_n) \leq \frac{p_{jk}}{n^\gamma} \quad if \quad p_{jk} \geq \frac{2m_n}{n}$$
$$\mathbb{P}_f(n_{jk} \geq m_n) \leq \frac{p_{jk}}{n^\gamma} \quad if \quad p_{jk} < \frac{m_n}{2n}$$

where $m_n = \frac{\mu^2}{K_\psi}\log(n)$.

This lemma is a generalization of Lemma 4 of Juditsky and Lambert-Lacroix (2004[72]). Its proof is given in the appendix.

Lemma 3.4. *Let $1 \leq p < \infty$. Then :*

1. $\mathbb{E}|\hat{\beta}_{jk} - \beta_{jk}|^{2p} \leq C\left(\frac{2^j p_{jk}}{n}\right)^p$ *if* $p_{jk} \geq \frac{1}{n}$
2. $\mathbb{E}|\hat{\beta}_{jk} - \beta_{jk}|^{2p} \leq C\left(\frac{2^j}{n^2}\right)^p n p_{jk}$ *if* $p_{jk} < \frac{1}{n}$

3. $\mathbb{E}|\hat\beta_{jk}-\beta_{jk}|^{2p} \le C\left(\frac{2^j}{n}\right)^p p_{jk}$.

The proof is given in the appendix.

Using Lemma 3.3, Lemma 3.4 (*3.*) and the Cauchy-Schwartz inequality, we have :

$$\begin{aligned}
A''_{01} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat\beta_{jk}-\beta_{jk}|^p \mathbb{1}\,\{|\hat\beta_{jk}|>\lambda_n\}\mathbb{1}\,\{n_{jk}\ge m_n\}\mathbb{1}\,\{p_{jk}<\frac{m_n}{2n}\}\\
&\le \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat\beta_{jk}-\beta_{jk}|^p \mathbb{1}\,\{n_{jk}\ge m_n\}\mathbb{1}\,\{p_{jk}<\frac{m_n}{2n}\}\\
&\le \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \mathbb{E}^{1/2}|\hat\beta_{jk}-\beta_{jk}|^{2p}|\mathbb{P}_f^{1/2}(n_{jk}\ge m_n)\mathbb{1}\,\{p_{jk}<\frac{m_n}{2n}\}\\
&\le \frac{C}{n^{\gamma/2}}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k p_{jk}\left[\frac{2^j}{n}\right]^{p/2}\\
&\le C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

Last inequality is due to (3.2) and requires to choose $\mu=\mu(\gamma)$ such that $\gamma\ge 2(p-1)$.

Using the Cauchy-Schwartz inequality and Lemma 3.1 with $\gamma\ge(1+\alpha)p-1$, and Lemma 3.4 (*3.*), one gets :

$$\begin{aligned}
A''_{02} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat\beta_{jk}-\beta_{jk}|^p \mathbb{1}\,\{|\hat\beta_{jk}|>\lambda_n\}\mathbb{1}\,\{|\beta_{jk}|\le\frac{\lambda_n}{2}\}\mathbb{1}\,\{n_{jk}\ge m_n\}\mathbb{1}\,\{p_{jk}\ge\frac{m_n}{2n}\}\\
&\le \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat\beta_{jk}-\beta_{jk}|^p \mathbb{1}\,\{|\hat\beta_{jk}-\beta_{jk}|>\frac{\lambda_n}{2}\}\mathbb{1}\,\{p_{jk}\ge\frac{m_n}{2n}\}\\
&\le \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \mathbb{E}^{1/2}|\hat\beta_{jk}-\beta_{jk}|^{2p}\;\mathbb{P}_f^{1/2}(|\hat\beta_{jk}-\beta_{jk}|>\frac{\lambda_n}{2})\mathbb{1}\,\{p_{jk}\ge\frac{m_n}{2n}\}\\
&\le C\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\left(\frac{2^j}{n}\right)^{p/2}\lambda_n^{\gamma-1}\sum_k p_{jk}\\
&\le C\,\lambda_n^{\alpha p}.
\end{aligned}$$

Finally, we have :

$$\begin{aligned}
A''_{03} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p \mathbb{1}\,\{n_{jk}\geq m_n\}\mathbb{1}\,\{p_{jk}\geq \frac{m_n}{2n}\}\mathbb{1}\,\{|\hat{\beta}_{jk}|>\lambda_n\}\mathbb{1}\,\{|\beta_{jk}|>\frac{\lambda_n}{2}\}\\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \mathbb{E}|\hat{\beta}_{jk}-\beta_{jk}|^p \mathbb{1}\,\{p_{jk}\geq \frac{m_n}{2n}\}\mathbb{1}\,\{|\beta_{jk}|>\frac{\lambda_n}{2}\}\\
&\leq C\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \left(\frac{2^j p_{jk}}{n}\right)^{p/2} \mathbb{1}\,\{p_{jk}\geq \frac{m_n}{2n}\}\mathbb{1}\,\{|\beta_{jk}|>\frac{\lambda_n}{2}\}\\
&\leq C\,\frac{1}{n^{p/2}}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k \mathbb{1}\,\{|\beta_{jk}|>\frac{\lambda_n}{2}\}\\
&\leq C\,\lambda_n^{\alpha p}.
\end{aligned}$$

Last inequalities use the fact that, for any $f\in\mathcal{B}^{\alpha/2}_{p,\infty}$ (with $\alpha p>2$), $\sup_{j,k} 2^{j} p_{jk}<\infty$. □

Until now, we have focused on non random thresholds. In particular we have proved that hard thresholding estimator are the *best procedures* among elitist ones belonging to $\mathcal{F}'_K$, when dealing with the maxiset approach. It seems to be interesting to answer the following question : do there exist adaptive procedures which outperform hard thresholding rules ? Once again, the answer is YES, by considering data-driven thresholds (see Birgé and Massart (2000[12]), Donoho and Johnstone (1995[44]), Johnstone (1999[64]), Juditsky (1997[71]) and Juditsky and Lambert-Lacroix (2004[72])), as we shall prove it in the next section.

3.5 On the significance of data-driven thresholds

Adopting a maxiset point of view, the aim of this section is to prove the significance of data-driven thresholds, in the context of estimating compactly or non compactly supported densities. For this, we study the maxiset associated with the data-driven thresholding procedure described by Juditsky and Lambert Lacroix (2004[72]). Here, the decision to *keep* or to *kill* empirical coefficients $\hat{\beta}_{jk}$ is chosen by comparing them to their standard deviation. We prove that the maxiset associated with this particular data-driven thresholding procedure is larger than the ideal maxiset of elitist rules. Let us denote :

– $\hat{\gamma}_{jk} = \mu\sqrt{\frac{\log(n)}{n}}\hat{\sigma}_{jk} = \lambda_n\hat{\sigma}_{jk}$ where $\hat{\sigma}^2_{jk} = \frac{1}{n}\sum_{i=1}^{n}(\psi^2_{jk}(X_i) - \hat{\beta}^2_{jk})$,

– $\gamma_{jk} = \mu\sqrt{\frac{\log(n)}{n}}\sigma_{jk} = \lambda_n\sigma_{jk}$ where $\sigma^2_{jk} = \mathbb{E}(\psi_{jk}(X_i) - \beta_{jk})^2$.

Let us consider the **data-driven thresholding estimator** defined by Juditsky and Lambert-Lacroix (2004[72]) :

$$\bar{f}_n(t) = \sum_{j=-1}^{j_n-1}\sum_{k\in\mathbb{Z}} \hat{\beta}_{jk}\mathbb{1}\ \{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}\tilde{\psi}_{jk}(t).$$

We have the following theorem :

Theorem 3.5. *Let $0 < \alpha < 1$ and $1 \leq p < \infty$ such that $\alpha p > 2$. If μ is large enough then :*

$$MS(\bar{f}_n, \|.\|^p_{\mathcal{B}^0_{p,p}}, (\frac{\log(n)}{n})^{\alpha p/2}) = \mathcal{B}^{\alpha/2}_{p,\infty} \cap W_\sigma((1-\alpha)p, p).$$

When adding to (3.5) of Proposition 3.1, this theorem proves that the maxiset associated with the data-driven thresholding estimator $\bar{f}_n$ is larger than the maxiset of any elitist estimator $\hat{f}$ of $\mathcal{F}'_K$, building with non random threshold.

Proof of Theorem 3.5

$\subset$: Fix $1 \leq p < \infty$ and let f be such that $\sup_{n>1}\left(\frac{n}{\log(n)}\right)^{\alpha p/2}\mathbb{E}\|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} < \infty$. On one hand, with same arguments that are in the proof of Theorem 3.1, for all $n > 1$, we have :

$$\sum_{j\geq j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \leq\ \mathbb{E}\|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} \leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2} \leq C2^{-j_n\frac{\alpha p}{2}}.$$

It comes that $f \in \mathcal{B}^{\alpha/2}_{p,\infty}$.

On the other hand, for any $n > 1$ we have :

$$\begin{aligned}
&\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \le \frac{\lambda_n \sigma_{jk}}{4}\} \\
= &\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \le \frac{\gamma_{jk}}{4}\} \\
= &\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \le \frac{\gamma_{jk}}{4}\} \left[\mathbb{1}\ \{p_{jk} \le \frac{m_n}{2n}\} + \mathbb{1}\ \{\frac{m_n}{2n} \le p_{jk} \le \frac{\nu^2}{2K_\psi^2}\} + \mathbb{1}\ \{p_{jk} > \frac{\nu^2}{2K_\psi^2}\}\right] \\
= &B_0 + B_1 + B_2.
\end{aligned}$$

Let us introduce the following lemma :

Lemma 3.5. *For any $j < j_n$, any k and any n large enough, $|\hat{\beta}_{jk}| > \hat{\gamma}_{jk} \Longrightarrow n_{jk} \ge m_n$.*

The proof of this lemma is given in the appendix.

To bound B_0, we use Lemma 3.3 with $\gamma \ge \frac{p}{2}$ and Lemma 3.5 :

$$\begin{aligned}
B_0 &= \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \le \frac{\gamma_{jk}}{4}\} \mathbb{1}\ \{p_{jk} \le \frac{m_n}{2n}\} \\
&\le \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{p_{jk} \le \frac{m_n}{2n}\} \\
&= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{p_{jk} \le \frac{m_n}{2n}\} \left[\mathbb{1}\ \{n_{jk} < m_n\} + \mathbb{1}\ \{n_{jk} \ge m_n\}\right] \\
&\le \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{n_{jk} < m_n\} + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{P}(n_{jk} \ge m_n) \mathbb{1}\ \{p_{jk} \le \frac{m_n}{2n}\} \\
&\le \mathbb{E}\|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \ \frac{p_{jk}}{n^\gamma} \\
&\le \mathbb{E}\|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} + C\ n^{-\gamma} \\
&\le C\ \left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

To bound B_1, let us consider the following lemma :

Lemma 3.6. *Fix $\gamma > 0$. There exists $\mu = \mu(\gamma) < \infty$ such that :*

1. *if $p_{jk} \ge \frac{\mu^2}{2K_\psi}.\frac{\log(n)}{n}$ then : $\mathbb{P}(\hat{\gamma}_{jk} > \mu\sqrt{\frac{\log(n)}{n}}) \le \frac{p_{jk}}{n^\gamma}$.*

2. *Moreover if* $\frac{\mu^2}{2K_\psi}.\frac{\log(n)}{n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}$, *for* n *large enough, then :*

   (a) $\mathbb{P}(|\hat{\gamma}_{jk} - \gamma_{jk}| > \frac{\gamma_{jk}}{2}) \leq \frac{2p_{jk}}{n^\gamma}$

   (b) $\mathbb{P}(|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}) \leq \frac{2p_{jk}}{n^\gamma}$.

Proof : This lemma is a simple generalization of Proposition 1 in Juditsky and Lambert-Lacroix (2004[72]). The proof is omitted since it uses similar arguments to those used by there. □

Since $|\beta_{jk}|\mathbb{1}\ \{|\beta_{jk}| \leq \frac{\hat{\gamma}_{jk}}{2}\} \leq |\beta_{jk} - \hat{\beta}_{jk}\mathbb{1}\ \{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}|$, by using *2.a)* of Lemma 3.6 with $\gamma \geq \frac{p}{2}$, one gets :

$$\begin{aligned}
B_1 = & \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\}\mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
\leq & \ \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\} \left[\mathbb{1}\ \{|\beta_{jk}| \leq \frac{\hat{\gamma}_{jk}}{2}\} + \mathbb{1}\ \{|\beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\}\right] \mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
\leq & \ \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k (|\beta_{jk} - \hat{\beta}_{jk}\mathbb{1}\ \{n_{jk} \geq m_n\}\mathbb{1}\ \{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}|^p + |\beta_{jk}|^p \mathbb{1}\ \{\hat{\gamma}_{jk} < \frac{\gamma_{jk}}{2}\}\mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\}) \\
\leq & \ \mathbb{E}\|\bar{f}_n - f\|_{\mathcal{B}^0_{p,p}}^p + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{P}(|\hat{\gamma}_{jk} - \gamma_{jk}| > \frac{\gamma_{jk}}{2})\mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
\leq & \ C\left[\left(\frac{\log(n)}{n}\right)^{\alpha p/2} + \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \frac{p_{jk}}{n^\gamma}\right] \\
\leq & \ C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

Now, using the fact that $\sup_{j,k} 2^j p_{jk} < \infty$ and $\sigma_{jk}^2 \leq 2^j K_\psi^2 \, p_{jk}$:

$$\begin{aligned}
B_2 &= \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\,\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\} \mathbb{1}\,\{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq C \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\gamma_{jk}|^p \mathbb{1}\,\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\} \mathbb{1}\,\{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&= C \,\lambda_n^p \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\sigma_{jk}|^p \mathbb{1}\,\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\} \mathbb{1}\,\{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq C \,\lambda_n^p \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k (2^j \; p_{jk})^{p/2} \mathbb{1}\,\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\} \mathbb{1}\,\{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq C \,\lambda_n^p \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k (2^j \; p_{jk})^{p/2} \mathbb{1}\,\{|\beta_{jk}| \leq \frac{\gamma_{jk}}{4}\} \mathbb{1}\,\{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq C \left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

Consequently, looking at the bounds of B_i, $0 \leq i \leq 2$, we deduce that $f \in W_\sigma((1-\alpha)p, p)$.

$\supset$: Let $\mu > 0$ be such that $\gamma \geq \alpha p + \max(0, p-2, (1-\frac{\alpha}{2})p - 1)$. The Besov-risk of $\bar{f}_n$ can be decomposed as follows :

$$\begin{aligned}
\mathbb{E}\|\bar{f}_n - f\|_{\mathcal{B}^0_{p,p}}^p &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk} - \hat{\beta}_{jk} \mathbb{1}\,\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}|^p + \|f - f_{j_n}\|_{\mathcal{B}^0_{p,p}}^p \\
&= C_0 + C_1.
\end{aligned}$$

Using similar arguments as in the proof of Theorem 3.4, since $f \in \mathcal{B}^{\alpha/2}_{p,\infty}$:

$$C_1 = \|f - f_{j_n}\|_{\mathcal{B}^0_{p,p}}^p \leq C \left(\frac{\log(n)}{n}\right)^{\alpha p/2}.$$

Using Lemma 3.5, we can decompose C_0 as follows :

$$\begin{aligned}
C_0 &= \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk} - \hat{\beta}_{jk} \mathbb{1}\,\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\}|^p \\
&\leq \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\,\{n_{jk} \leq m_n\} + \mathbb{E} \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} \mathbb{1}\,\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p \mathbb{1}\,\{n_{jk} \geq m_n\} \\
&= C_0' + C_0''.
\end{aligned}$$

Since $f \in \mathcal{B}^{\alpha/2}_{p,\infty} \cap W((1-\alpha)p,p)$, $f \in \chi((1-\alpha)p,p)$. So, by using Lemma 3.3 with $\gamma \geq \frac{p}{2}$, one gets :

$$\begin{aligned}
C_0' &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{1}\,\{n_{jk}\leq m_n\}\\
&= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{1}\,\{n_{jk}\leq m_n\}[\mathbb{1}\,\{p_{jk}\leq \frac{m_n}{2n}\}+\mathbb{1}\,\{p_{jk}> \frac{m_n}{2n}\}]\\
&\leq \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{1}\,\{p_{jk}\leq \frac{m_n}{2n}\}+\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbb{P}(n_{jk}\leq m_n)\mathbb{1}\,\{p_{jk}> \frac{m_n}{2n}\}\\
&\leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}+\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\frac{p_{jk}}{n^\gamma}\\
&\leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}+Cn^{-\gamma}\\
&\leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

we have the following decomposition for C_0'' :

$$\begin{aligned}
C_0'' &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}\mathbb{1}\,\{|\hat{\beta}_{jk}|>\hat{\gamma}_{jk}\}-\beta_{jk}|^p\mathbb{1}\,\{n_{jk}\geq m_n\}\\
&= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}\mathbb{1}\,\{|\hat{\beta}_{jk}|>\hat{\gamma}_{jk}\}-\beta_{jk}|^p\mathbb{1}\,\{n_{jk}\geq m_n\}[\mathbb{1}\,\{p_{jk}< \frac{m_n}{2n}\}+\mathbb{1}\,\{p_{jk}\geq \frac{m_n}{2n}\}]\\
&= C_{01}''+C_{02}''.
\end{aligned}$$

Now, since :

$$|\hat{\beta}_{jk}\mathbb{1}\,\{|\hat{\beta}_{jk}|>\hat{\gamma}_{jk}\}-\beta_{jk}|\leq|\hat{\beta}_{jk}-\beta_{jk}|+|\beta_{jk}|,$$

C_{01}'' can be decomposed into $C_{011}''+C_{012}''$, with :

$$\begin{aligned}
C_{011}'' &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p\mathbb{1}\,\{n_{jk}\geq m_n\}\mathbb{1}\,\{p_{jk}< \frac{m_n}{2n}\}\\
C_{012}'' &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbb{1}\,\{n_{jk}\geq m_n\}\mathbb{1}\,\{p_{jk}< \frac{m_n}{2n}\}.
\end{aligned}$$

Still using Lemma 3.3 with $\gamma \geq 2(p-1)$ and *3.* of Lemma 3.4 :

$$\begin{aligned}
C''_{011} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} - \beta_{jk}|^p \mathbb{1}\ \{n_{jk} \geq m_n\} \mathbb{1}\ \{p_{jk} < \frac{m_n}{2n}\} \\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{E}^{1/2}|\hat{\beta}_{jk} - \beta_{jk}|^{2p} \mathbb{P}^{1/2}(n_{jk} \geq m_n) \mathbb{1}\ \{p_{jk} < \frac{m_n}{2n}\} \\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k \left(\frac{2^j}{n}\right)^{\frac{p}{2}} \frac{p_{jk}}{\sqrt{n^\gamma}} \\
&\leq C\ \frac{2^{j_n(\frac{p}{2}-1)}}{n^{\gamma/2}} \\
&\leq C\ \left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

$$\begin{aligned}
\text{and } C''_{012} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{n_{jk} \geq m_n\} \mathbb{1}\ \{p_{jk} < \frac{m_n}{2n}\} \\
&\leq \sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{p_{jk} \leq \frac{m_n}{2n}\} \\
&\leq C\ \left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

The last inequality uses the fact that $f \in \chi((1-\alpha)p, p)$.

We decompose C''_{02} into two parts :

$$\begin{aligned}
C''_{02} &= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} \mathbb{1}\ \{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p \mathbb{1}\ \{n_{jk} \geq m_n\} \mathbb{1}\ \{p_{jk} \geq \frac{m_n}{2n}\} \\
&= \mathbb{E}\sum_{j<j_n} 2^{j(\frac{p}{2}-1)} \sum_k |\hat{\beta}_{jk} \mathbb{1}\ \{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p \mathbb{1}\ \{n_{jk} \geq m_n\} [\mathbb{1}\ \{p_{jk} > \frac{\nu^2}{2K_\psi^2}\} + \mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2K_\psi^2}\}] \\
&= C''_{021} + C''_{022}.
\end{aligned}$$

Let us now consider this new lemma :

Lemma 3.7. *There exists a constant $C < \infty$ such that, for any $\lambda > 0$:*

$$
\begin{aligned}
|\hat{\beta}_{jk}\mathbb{1}\,\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}| &\leq C\left(|\hat{\beta}_{jk} - \beta_{jk}| + \mu\sqrt{\frac{\log(n)}{n}}\right) + |\beta_{jk}|\mathbb{1}\,\{\hat{\gamma}_{jk} > \mu\sqrt{\frac{\log(n)}{n}}\} \text{ and,}\\
|\hat{\beta}_{jk}\mathbb{1}\,\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p &\leq C\left(|\hat{\beta}_{jk} - \beta_{jk}|^p\mathbb{1}\,\{|\hat{\beta}_{jk} - \beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\} + \min(|\beta_{jk}|, \gamma_{jk})^p\right) + |\beta_{jk}|^p\mathbb{1}\,\{\hat{\gamma}_{jk} > \frac{3\gamma_{jk}}{2}\}
\end{aligned}
$$

Proof : The proof of this lemma is given in Juditsky and Lambert-Lacroix (2004[72]). □

Using Lemma 3.6 with $\gamma \geq \frac{p}{2}$ and Lemma 3.7, one gets :

$$
\begin{aligned}
C''_{021} &= \mathbb{E}\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}\mathbb{1}\,\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p\mathbb{1}\,\{n_{jk} \geq m_n\}\mathbb{1}\,\{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}\}\\
&\leq C\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k \mathbb{E}\left[|\hat{\beta}_{jk} - \beta_{jk}|^p + \mu^p\sqrt{\frac{\log^p(n)}{n^p}} + |\beta_{jk}|^p\mathbb{1}\,\{\hat{\gamma}_{jk} > \mu\sqrt{\frac{\log(n)}{n}}\}\right]\mathbb{1}\,\{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}\}\\
&\leq C\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k \left[\left(\frac{2^j p_{jk}}{n}\right)^{\frac{p}{2}} + \sqrt{\frac{\log^p(n)}{n^p}} + \frac{2^{jp/2}}{n^\gamma}\right]\mathbb{1}\,\{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}\}\\
&\leq C\left[\left(\frac{1}{n}\right)^{\frac{p}{2}} + \sqrt{\frac{\log^p(n)}{n^p}} + n^{-\gamma}\right]\\
&\leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}
$$

Still using Lemma 3.7 :

$$
\begin{aligned}
C''_{022} &= \mathbb{E}\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}\mathbb{1}\,\{|\hat{\beta}_{jk}| > \hat{\gamma}_{jk}\} - \beta_{jk}|^p\mathbb{1}\,\{n_{jk} \geq m_n\}\mathbb{1}\,\{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\}\\
&= C\,(C''_{0221} + C''_{0222} + C''_{0223}).
\end{aligned}
$$

Using the Cauchy-Schwartz inequality, (*1.*) of Lemma 3.4 and Lemma 3.6 with $\gamma \geq 2(p-1)$:

$$\begin{aligned}
C''_{0221} &= \mathbb{E}\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k |\hat{\beta}_{jk}-\beta_{jk}|^p \mathbb{1}\ \{|\hat{\beta}_{jk}-\beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2}\}\mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq K\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k \left(\frac{2^j p_{jk}}{n}\right)^{\frac{p}{2}} \mathbb{P}^{\frac{1}{2}}(|\hat{\beta}_{jk}-\beta_{jk}| > \frac{\hat{\gamma}_{jk}}{2})\mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq K\frac{2^{j_n(\frac{p}{2}-1)}}{\sqrt{n^\gamma}} \\
&\leq K\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

$$\begin{aligned}
C''_{0222} &= \mathbb{E}\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k \min(|\beta_{jk}|,\gamma_{jk})^p \mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk}\} \\
&\leq \mathbb{E}\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{1}\ \{|\beta_{jk}| \leq \gamma_{jk}\} + \mathbb{E}\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k \gamma_{jk}^p \mathbb{1}\ \{|\beta_{jk}| > \gamma_{jk}\} \\
&\leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

These inequalities are obtained using the fact that $f \in W_\sigma((1-\alpha)p, p)$.

Finally, using Lemma Lemma 3.6 :

$$\begin{aligned}
C''_{0223} &= \mathbb{E}\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{1}\ \{\hat{\gamma}_{jk} > \frac{3\gamma_{jk}}{2}\}\mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq \sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{P}(|\hat{\gamma}_{jk}-\gamma_{jk}| > \frac{\gamma_{jk}}{2})\mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq C\sum_{j=-1}^{j_n-1} 2^{j(\frac{p}{2}-1)}\sum_k \left(2^{\frac{j}{2}} p_{jk}\right)^p \frac{p_{jk}}{n^\gamma}\mathbb{1}\ \{\frac{m_n}{2n} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}\} \\
&\leq C\frac{2^{j_n(p-1)}}{n^\gamma} \\
&\leq C\left(\frac{\log(n)}{n}\right)^{\alpha p/2}.
\end{aligned}$$

Consequently, looking at the bounds of C_0 and C_1, we deduce that :

$$\sup_{n>1}\left(\frac{n}{\log(n)}\right)^{\alpha p/2}\mathbb{E}\|\bar{f}_n - f\|^p_{\mathcal{B}^0_{p,p}} < \infty.$$

□

3.6 Appendix

Proof of Lemma 3.2 :

$$\begin{aligned}
\mu\sqrt{\frac{\log(n)}{n}} < |\hat{\beta}_{jk}| &= \frac{1}{n}|\sum_{i=1}^{n}\psi_{jk}(X_i)| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}2^{j/2}K_\psi \mathbb{1}\,\{X_i\in I_{jk}\} \\
&\leq \frac{1}{\mu.n}\sum_{i=1}^{n}\sqrt{\frac{n}{\log(n)}}K_\psi \mathbb{1}\,\{X_i\in I_{jk}\} \\
&\leq \frac{1}{\mu.n}\sqrt{\frac{n}{\log(n)}}K_\psi n_{jk}.
\end{aligned}$$

Finally, one gets :

$$|\hat{\beta}_{jk}| > \mu\sqrt{\frac{\log(n)}{n}} \Longrightarrow n_{jk} > \frac{\mu^2}{K_\psi}\log(n).\Box\Box$$

Proof of Lemma 3.3 :

Step 1 : suppose that $np_{jk}\geq 2\rho\log(n)$.
Since $\tau^2_{jk} = \mathbb{V}ar_f(\mathbb{1}\,\{X_1\in I_{jk}\}) = np_{jk}(1-p_{jk})$ then $2\tau^2_{jk}\leq\frac{n^2p^2_{jk}}{\rho\log(n)}$. Using the Bernstein inequality, we have :

$$\begin{aligned}
\mathbb{P}_f(n_{jk}<\rho\log(n)) &= \mathbb{P}_f(np_{jk}-n_{jk} > np_{jk}-\rho\log(n)) \\
&\leq \mathbb{P}_f(np_{jk}-n_{jk} > \frac{n}{2}p_{jk}) \\
&\leq \exp\left(-\frac{n^2p^2_{jk}}{8(\tau^2_{jk}+\frac{np^2_{jk}}{6})}\right)
\end{aligned}$$

$$\begin{aligned}
&\leq \exp\left(-\frac{n^2 p_{jk}^2}{8n^2 p_{jk}^2(\frac{1}{2\rho\log(n)} + \frac{1}{6n})}\right) \\
&\leq \exp\left(-K\rho\log(n)\right) \\
&= n^{-K\rho} \\
&\leq \frac{p_{jk}}{n^{\gamma}}.
\end{aligned}$$

The last inequality is obtained by taking ρ such that $K\rho \geq 1+\gamma$.

Step 2 : suppose now that $\frac{1}{n^{\gamma+1}} \leq np_{jk} \leq 2\rho\log(n)$. Using the Bernstein inequality, one gets :

$$\begin{aligned}
\mathbb{P}_f(n_{jk} \geq \rho\log(n)) &= \mathbb{P}_f(n_{jk} - np_{jk} \geq \rho\log(n)) - np_{jk}) \\
&\leq \mathbb{P}_f(n_{jk} - np_{jk} \geq \frac{\rho\log(n)}{2}) \\
&\leq \exp\left(-\frac{\rho^2\log(n)^2}{8(\tau_{jk}^2 + \frac{\rho\log(n)}{6})}\right) \\
&\leq \exp\left(-\frac{\rho^2\log(n)^2}{8(np_{jk} + \frac{\rho\log(n)}{6})}\right) \\
&\leq \exp\left(-K\rho\log(n)\right) \\
&= n^{-K\rho} \\
&\leq \frac{p_{jk}}{n^{\gamma}}.
\end{aligned}$$

The last inequality requires that ρ satisfies $K\rho \geq 2(1+\gamma)$.

Step 3 : consider that $np_{jk} \leq \frac{1}{n^{\gamma+1}}$. Using simple bounds on the tails of the binomial distribution (see inequality 1 page 482 in Shorack & Wellner (1986 [105])) :

$$\begin{aligned}
\mathbb{P}_f(n_{jk} \geq \rho \log(n)) &\leq \frac{(1-p_{jk})}{1-\frac{(n+1)p_{jk}}{2}}\, \mathcal{C}_n^2\, p_{jk}^2(1-p_{jk})^{n-2} \\
&\leq \frac{n^2 p_{jk}^2}{2(1-\frac{(n+1)p_{jk}}{2})} \\
&\leq \frac{n^2 p_{jk}}{n^{\gamma+2}} \\
&= \frac{p_{jk}}{n^{\gamma}}.
\end{aligned}$$

□

Proof of Lemma 3.4 :

1. and *2.*. By the Rosenthal inequality, for any j,k :

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_{jk}-\beta_{jk})^{2p} &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\psi_{jk}(X_i)-\beta_{jk}\right)^{2p} \\
&\leq \frac{C}{n^{2p}}\left[\sum_{i=1}^{n}\mathbb{E}(\psi_{jk}(X_i)-\beta_{jk})^{2p}+\left(\sum_{i=1}^{n}\mathbb{E}(\psi_{jk}(X_i)-\beta_{jk})^2\right)^p\right] \\
&\leq \frac{C}{n^{2p}}(D_0+D_1)
\end{aligned}$$

where :

$$\begin{aligned}
D_0=\sum_{i=1}^{n}\mathbb{E}\left(\psi_{jk}(X_i)-\beta_{jk}\right)^{2p} &\leq C\ n\ \left(\mathbb{E}(\psi_{jk}^{2p}(X_1))+(\beta_{jk})^{2p}\right) \\
&\leq C\ n\left(2^{jp}p_{jk}+(2^{j/2}p_{jk})^{2p}\right) \\
&\leq C\ 2^{jp}np_{jk}
\end{aligned}$$

$$\begin{aligned}
D_1=\left(\sum_{i=1}^{n}\mathbb{E}\left(\psi_{jk}(X_i)-\beta_{jk}\right)^2\right)^p &= \left(\sum_{i=1}^{n}\mathbb{V}ar(\psi_{jk}(X_i))\right)^p \\
&\leq \left(\sum_{i=1}^{n}\mathbb{E}(\psi_{jk}^2(X_i))\right)^p \\
&\leq C\ n^p\left(2^j p_{jk}\right)^p \\
&\leq C\ 2^{jp}(np_{jk})^p.
\end{aligned}$$

Now, if $np_{jk} \geq 1$ then $np_{jk} \leq (np_{jk})^p$. So :

$$\mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^{2p} \leq C \left(\frac{2^j p_{jk}}{n}\right)^p.$$

If $np_{jk} < 1$ then $np_{jk} > (np_{jk})^p$. So

$$\mathbb{E}(\hat{\beta}_{jk} - \beta_{jk})^{2p} \leq C \; np_{jk} \left(\frac{2^j}{n^2}\right)^p.$$

Finally, *3.* is just a consequence of *1.* and *2.*. □

Proof of Lemma 3.5 :

Suppose that $\hat{\gamma}_{jk} < |\hat{\beta}_{jk}|$. Then :

$$\begin{aligned}
\mu^2 \frac{\log(n)}{n} . \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(X_i)^2 &< \mu^2 \frac{\log(n)}{n} \hat{\beta}_{jk}^2 + \hat{\beta}_{jk}^2 \\
&= (\mu^2 \frac{\log(n)}{n} + 1) \hat{\beta}_{jk}^2.
\end{aligned}$$

By using bounds on the left and the right parts, one gets for n large enough :

$$\mu^2 \frac{\log(n)}{n^2} 2^j \nu^2 n_{jk} < 2\hat{\beta}_{jk}^2.$$

And since $n|\hat{\beta}_{jk}| \leq 2^{j/2} K_\psi n_{jk}$,

$$\mu^2 \nu^2 \log(n) < 2K_\psi^2 n_{jk}.$$

Finally, one gets :

$$|\hat{\beta}_{jk}| > \hat{\gamma}_{jk} \Longrightarrow n_{jk} > \frac{\mu^2 \nu^2}{2K_\psi^2} \log(n).$$

□

Chapitre 4

Maxisets and choice of priors for Bayesian rules

Summary : In this chapter our aim is twofold. First, we provide tools for easily calculating the maxisets of several procedures. Then, we apply these results to perform a comparison between several Bayesian estimators in a non parametric setting. We obtain that many Bayesian rules can be described through a general behavior such as being shrinkage rules, limited and/or elitist rules. This has consequences on their maxisets which happen to be automatically included in some Besov or weak Besov spaces, whereas other properties such as cautiousness imply that their maxiset conversely contains some of the spaces quoted above.
Secondly, we compare Bayesian rules taking into account the sparsity of the signal with priors which are combination of a Dirac with a standard distribution. We consider the case of Gaussian and heavy tail priors and we prove that the heavy tail assumption is not necessary to attain maxisets equivalent to the thresholding methods. Finally, simulated examples of Bayesian rules are used and comparisons are made with other thresholding methods.

4.1 Introduction and model

In the first part of the chapter (sections 4.3 and 4.4), we provide tools for easily calculating the maxisets of several procedures. To be more precise, we provide conditions

ensuring that the maxiset of a procedure is necessarily larger than some fixed space, and conversely prove that other conditions restrict a procedure to have its maxiset smaller than a fixed space. This study is performed on the class of shrinkage procedures in a white noise model. Among these procedures we investigate the consequences for a procedure to be limited, elitist and/or cautious (see the definitions in paragraph 4.2.2).
It is important to notice that this study can obviously be generalized to different models (since the conditions on the model are in fact not very restrictive), and one can easily imagine conditions on kernel methods (for instance) translating the notions of shrinkage, limited, elitist or cautious although it is certainly less natural.

The second part of the chapter (section 4.5) uses the results of the first one to perform a comparison among Bayesian estimates. We choose to focus on Bayes rules precisely because Bayesian techniques have now become very popular to estimate signals decomposed on wavelet bases. From the practical point of view, many authors have built Bayes estimates that outperform classical procedures and in particular thresholding procedures. See for instance, Chipman et al.(1997[24]), Abramovich et al (1998[4]), Clyde et al. (1998[27]), Johnstone and Silverman (1998[67]), Vidakovic (1998[116])or Clyde and George (1998[25], 1998[26]) who discussed the choice of the Bayes model to capture the sparsity of the signal to be estimated and the choice of the Bayes rule (and among others, posterior mean or median). We also refer the reader to the very complete review paper of Antoniadis et al. (2001[5]) who provide descriptions and comparisons of various Bayesian wavelet shrinkage and wavelet thresholding estimators.
From the minimax point of view, recent works have proved that Bayes rules can achieve optimal rates of convergence. Abramovich et al. (2004[1]) investigated theoretical performance of the procedures introduced by Abramovich et al. (1998[4]). More precisely, they considered a prior model based on a combination of a point mass at zero and a normal density. For the mean squared error, they proved that the non adaptive posterior mean and posterior median achieve optimal rates up to a logarithmic factor on the Besov space $\mathcal{B}^s_{p,q}$ when $p \geq 2$. When $p < 2$, these estimators can achieve only the best possible rates for linear estimates. As Abramovich et al. (2004[1]), Johnstone et Silverman (2002[68],2004[70]) investigated minimax properties of Bayes rules, but the prior is based on heavy-tailed distributions and they consider an empirical Bayes setting. In this case, the posterior mean and median are optimal. Other more sophisticated results concerning

minimax properties of Bayes rules have been established by Zhang (2002[120]).
The goal of section 4.5 is to study some Bayesian procedures from the maxiset point of view in the light of the results of sections 4.3 and 4.4. To capture the sparsity of the signal, we introduce the following prior model on the wavelet coefficients :

$$\beta_{jk} \sim \pi_{j,\epsilon}\gamma_{j,\epsilon} + (1-\pi_{j,\epsilon})\delta(0), \tag{4.1}$$

where $0 \leq \pi_{j,\epsilon} \leq 1$, $\delta(0)$ is a point mass at zero and the β_{jk}'s are independent. The nonzero part of the prior $\gamma_{j,\epsilon}$ is assumed to be the dilation of a fixed symmetric, positive, unimodal and continuous density γ :

$$\gamma_{j,\epsilon}(\beta_{jk}) = \frac{1}{\tau_{j,\epsilon}}\gamma\left(\frac{\beta_{jk}}{\tau_{j,\epsilon}}\right),$$

where the dilation parameter $\tau_{j,\epsilon}$ is positive. The parameter $\pi_{j,\epsilon}$ can be interpreted as the proportion of non negligible coefficients. We also introduce the parameter

$$w_{j,\epsilon} = \frac{\pi_{j,\epsilon}}{1-\pi_{j,\epsilon}}.$$

When the signal is sparse, most of the $w_{j,\epsilon}$ are small. These priors or very close forms have extensively been used by the authors cited above and especially Abramovith et al. (2004[1]), Johnstone and Silverman (2002[68],2002[70]). To complete the definition of the prior model, we have to fix the hyperparameters $\tau_{j,\epsilon}$ and $w_{j,\epsilon}$ and the density γ. The most popular choice for γ is the normal density. However priors with heavy tails have proved also to work extremely well. One of our results is to show that if some Bayesian procedures using Gaussian priors behave quite unwell (in terms of maxisets) compared to those with heavy tails, it is nevertheless possible to attain a maxiset as good as thresholding estimates, among procedures based on Gaussian priors, under the condition that the hyperparameter $\tau_{j,\epsilon}$ is "large". Under this assumption, the density $\gamma_{j,\epsilon}$ is then more spread around 0, which enables us to avoid considering heavy-tailed densities.

Finally, in section 4.6, we give simulations of Bayesian rules with Gaussian priors and we show that such estimators have excellent numerical performances relative to more traditional wavelet estimators when using the mean-squarred error.

4.2 Model and shrinkage rules.

4.2.1 Model

We consider a white noise setting : $X_\epsilon(.)$ is a random measure satisfying on $[0,1]$ the following equation :

$$X_\epsilon(dt) = f(t)dt + \epsilon W(dt)$$

where $0 < \epsilon < 1/e$ is the noise level and f is a function defined on $[0,1]$, $W(.)$ is a Brownian motion on $[0,1]$. As usual, to connect with the standard framework of sequences of experiments we put $\epsilon = n^{-1/2}$.
Let $\{\psi_{jk}(\cdot),\ j \geq -1, k \in \mathbb{Z}\}$ be a compactly supported wavelet basis of $\mathbb{L}_2([0,1])$, such that any $f \in \mathbb{L}_2([0,1])$ can be represented as :

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

where $\beta_{jk} = (f, \psi_{jk})_{\mathbb{L}_2}$. (As usual, ψ_{-1k} denotes the translations of the scaling function.) The model is reduced to a sequence space model if we put : $y_{jk} = X_\epsilon(\psi_{jk}) = \int f\psi_{jk} + \epsilon Z_{jk}$ where Z_{jk} are i.i.d $\mathcal{N}(0,1)$. Let us note that at each level $j \geq 0$, the number of non-zero wavelet coefficients is smaller than or equal to $2^j + l_\psi - 1$, where l_ψ is the maximal size of the supports of the scaling function and the wavelet. So, there exists a constant S_ψ such that at each level $j \geq -1$, there are less than or equal to $S_\psi \times 2^j$ coefficients to be estimated. In the sequel, we shall not distinguish between f and $\beta = (\beta_{jk})_{jk}$ its sequence of wavelet coefficients.

4.2.2 Classes of Estimators

Let us first consider the following very general class of shrinkage estimators :

$$\mathcal{F}_\epsilon = \left\{ \hat{f}_\epsilon(.) = \sum_{j \geq -1} \sum_k \gamma_{jk} y_{jk} \psi_{jk}(.); \quad \gamma_{jk}(\varepsilon) \in [0,1],\ \text{measurable} \right\}.$$

Let us observe here that the γ_{jk} may be constant (linear estimators) or data dependent. Among this class, we'll particularly focus on the following classes of estimators :

Definition 4.1. *We say that* $\hat{f}_\epsilon \in \mathcal{F}_\epsilon$ *is a* **limited rule** *if there exist a determinist function of* ϵ, λ_ϵ, *and a constant* $a \in [0,1[$ *such that, for any* j,k,

$$\gamma_{jk} > a \Longrightarrow 2^{-j} > \lambda_\epsilon.$$

We note $\hat{f}_\epsilon \in \mathcal{L}(\lambda_\epsilon, a)$.

The simplest example to illustrate limited rules is provided by the projection estimator :

$$\gamma_{jk}(\epsilon) = \gamma_j^{(1)}(\lambda_\epsilon) = \mathbb{1}\ \{2^{-j} > \lambda_\epsilon\},$$

which obviously belongs to $\mathcal{L}(\lambda_\epsilon, 0)$. But, more generally, the class of linear shrinkage estimates provides natural limited procedures. For instance, linear estimates associated with Tikhonov-Phillips weights :

$$\gamma_{jk}(\epsilon) = \gamma_j^{(2)}(\lambda_\epsilon) = \frac{1}{1 + (2^j \lambda_\epsilon)^\alpha}, \quad \alpha > 0,$$

or with Pinsker weights :

$$\gamma_{jk}(\epsilon) = \gamma_j^{(3)}(\lambda_\epsilon) = (1 - (2^j \lambda_\epsilon)^\alpha)_+, \quad \alpha > 0,$$

are limited rules respectively belonging to $\mathcal{L}(\lambda_\epsilon, 1/2)$ and $\mathcal{L}(\lambda_\epsilon, 0)$.
To detail other examples, let us introduce

$$\begin{aligned} t_\epsilon &= \epsilon\sqrt{\log(\epsilon^{-1})} \\ j_\epsilon \in \mathbb{N},\ 2^{-j_\epsilon} &\leq t_\epsilon^2 < 2^{1-j_\epsilon}. \end{aligned}$$

This will be denoted in the sequel by $2^{j_\epsilon} \sim t_\epsilon^{-2}$. We recall the hard thresholding $\hat{f}^T$ and the soft thresholding $\hat{f}^S$ rules respectively defined by

$$\hat{f}^T = \sum_{-1 \leq j < j_\epsilon} \sum_k y_{jk} \mathbb{1}\ \{|y_{jk}| > m t_\epsilon\} \psi_{jk}, \tag{4.2}$$

$$\hat{f}^S = \sum_{-1 \leq j < j_\epsilon} \sum_k \left(1 - \frac{m t_\epsilon}{|y_{jk}|}\right) \mathbb{1}\ \{|y_{jk}| > m t_\epsilon\} y_{jk} \psi_{jk}, \tag{4.3}$$

where m is a positive constant. It is obvious that these procedures belong to $\mathcal{L}(t_\epsilon^2, 0)$. In sections 4.5, we shall provide many more examples of limited rules.

Definition 4.2. *We say that $\hat{f}_\epsilon \in \mathcal{F}_\epsilon$ is an* **elitist rule** *if there exist a determinist function of ϵ, λ_ϵ, and a constant $a \in [0,1[$ such that, for any j,k*

$$\gamma_{jk} > a \Longrightarrow |y_{jk}| > \lambda_\epsilon.$$

In the sequel, we note $\hat{f}_\epsilon \in \mathcal{E}(\lambda_\epsilon, a)$.

Remark 4.1. *This definition generalizes the notion of elitist rules introduced in Chapter 3 for the model of density estimation.*

To give some examples of elitist rules, consider $\hat{f}^T$ and $\hat{f}^S$ defined in (4.2) and (4.3) that belong to $\mathcal{E}(mt_\epsilon, 0)$. Other examples of elitist rules will be given in section 4.5 by considering Bayesian procedures.

Definition 4.3. *We say that $\hat{f}_\epsilon \in \mathcal{F}_\epsilon$ is a* **cautious rule** *if there exist a determinist function of ϵ, λ_ϵ and a constant $a \in]0,1]$ such that, for any $j < j_\epsilon$ and any k*

$$\gamma_{jk} \le a \Longrightarrow |y_{jk}| \le \lambda_\epsilon,$$

where $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$. In the sequel, we note $\hat{f}_\epsilon \in \mathcal{C}(\lambda_\epsilon, a)$.

Remark 4.2. *For instance, $\hat{f}^T$ and $\hat{f}^S$ defined in (4.2) and (4.3) belong respectively to $\mathcal{C}(mt_\epsilon, \frac{1}{2})$ and $\mathcal{C}(2mt_\epsilon, \frac{1}{2})$.*

Remark 4.3. *The limited rules as well as the elitist rules are forming a non decreasing class with respect to a. The cautious rules are forming a non increasing class with respect to a. We also have that any of the classes introduced above are convex. So they are obviously stable if we consider aggregation of procedures or as in learning algorithms, if we build a procedure averaging the opinions of different experts all belonging to one of the previous class.*

4.3 Ideal maxisets for particular classes of estimators.

Proving lower bound inequalities in minimax theory consists in showing that if we consider the class of all estimators on a functional spaces, there exists a best achievable rate α_n. In this section our tactic will be of the same spirit, but somewhat different since we will fix the rate α_n, consider classes of procedures and prove that they have a best achievable maxiset. More precisely, we will prove that when a procedure belongs to one of the classes considered above, its maxiset is necessarily smaller than a simple functional class. Here, for simplicity, we shall restrict to the case where ρ is the square of the $\mathbb{L}_2$ norm, even though if a large majority of the following results can be extended to more general norms.

4.3.1 Functional spaces

We recall the definitions of the following functional spaces. They will play an important role in the sequel. Note that, here, they appear with definitions depending on the wavelet basis. However, as has been remarked in Meyer(1990[89]) and Cohen et al. (2001[31]), most of them also have different definitions proving that this dependence in the basis is not crucial at all. Here and later we set for all $\lambda > 0$, $2^{j_\lambda} \sim \lambda^{-2}$.

Definition 4.4. *Let $s > 0$. We say that a function $f \in \mathbb{L}_2([0,1])$ belongs to the Besov space $\mathcal{B}^s_{2,\infty}$, if and only if :*

$$\sup_{J \geq -1} 2^{2Js} \sum_{j \geq J} \sum_k \beta_{jk}^2 \ < \infty.$$

We denote by $\mathcal{B}^s_{2,\infty}(R)$ the ball of radius R in this space.

Definition 4.5. *Let $0 < r < 2$. We say that a function f belongs to the weak Besov space $W(r,2)$ if and only if :*

$$\|f\|_{W_r} := [\sup_{\lambda > 0} \lambda^{r-2} \sum_{j \geq -1} \sum_k \beta_{jk}^2 \mathbb{1}\ \{|\beta_{jk}| \leq \lambda\}]^{1/2} < \infty.$$

We denote by $W(r,2)(R)$, the ball of radius R in this space.

Definition 4.6. *Let $0 < r < 2$. We say that a function f belongs to the space $W^*(r,2)$ if and only if :*

$$\|f\|_{W^*_r} := [\sup_{0<\lambda<1} \lambda^r [\log(\frac{1}{\lambda})]^{-1} \sum_{-1 \leq j < j_\lambda} \sum_k \mathbb{1}\ \{|\beta_{jk}| > \lambda\}]^{1/2} < \infty.$$

Remark 4.4. *If $\subsetneq$ denotes the strict inclusion between two functional spaces, for all $0 < r < 2$, it is easy to see using Markov inequality that $\mathcal{B}^s_{2,\infty} \subsetneq W(r,2)$ as soon as $s \geq \frac{1}{r} - \frac{1}{2}$ and $W(r,2) \subsetneq W^*(r,2)$.*

For sake of simplicity, the result presented in the following section emphasizes the cases where the rate of convergence is linked in a direct way to either the limitation or to the threshold bound for elitist or cautious rules. This constraint can be relaxed. For instance, there are many cases where either the threshold bound or the rate contain logarithmic factors. In these cases the link is not so direct. Results can also be obtained in these cases, which may be less aesthetic, but still useful. These results are given in the appendix.

Notation: For $\mathcal{A}$, a given normed space, the following notations :

$$\begin{aligned} MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s}) &\subset \mathcal{A} \\ (resp.) \quad \mathcal{A} &\subset MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s}) \end{aligned}$$

will mean in the sequel

$$\begin{aligned} \forall\, M \;\exists\, M',\; MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s})(M) &\subset \mathcal{A}(M') \\ (resp.) \quad \forall\, M' \;\exists\, M,\; \mathcal{A}(M') &\subset MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s})(M), \end{aligned}$$

where M and M' respectively denote the radii of balls of $MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s})$ and $\mathcal{A}$. $\diamond$

4.3.2 Ideal maxisets for limited rules

In this section, we study the ideal maxisets for limited procedures. For this purpose, let us give a sequence $(\lambda_\epsilon)_\epsilon$ going to 0 as ϵ tending to 0.

Theorem 4.1 (Ideal maxiset for limited rules)**.** *Let $\sigma > 0$ and $\hat{f}_\epsilon$ be a limited rule in $\mathcal{L}(\lambda_\epsilon, a)$, with $a \in [0,1[$. Then, if λ_ε is a non decreasing, continuous function such that $\lambda_0 = 0$,*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2\sigma}) \subset \mathcal{B}^\sigma_{2,\infty}$$

(with $M' = \frac{\sqrt{2M}}{(1-a)}$.)

Proof of Theorem 4.1 : Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2\sigma})(M)$. If we observe that if $2^{-j} \leq \lambda_\epsilon$ then $\gamma_{jk} \leq a$, we have :

$$\begin{aligned}
& (1-a)^2 \sum_{j,k} \beta_{jk}^2 \mathbb{1}\, \{2^{-j} \leq \lambda_\epsilon\} \\
= \;& 2(1-a)^2 \sum_{j,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0) \mathbb{1}\, \{\beta_{jk} \geq 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0) \mathbb{1}\, \{\beta_{jk} < 0\}\right] \mathbb{1}\, \{2^{-j} \leq \lambda_\epsilon\} \\
\leq \;& 2\mathbb{E} \sum_{j,k} \left[(\gamma_{jk} y_{jk} - \beta_{jk})^2 \mathbb{1}\, \{\beta_{jk} \geq 0\} + (\gamma_{jk} y_{jk} - \beta_{jk})^2 \mathbb{1}\, \{\beta_{jk} < 0\}\right] \mathbb{1}\, \{2^{-j} \leq \lambda_\epsilon\} \\
\leq \;& 2\mathbb{E} \sum_{j,k} (\gamma_{jk} y_{jk} - \beta_{jk})^2 \\
\leq \;& 2M\, \lambda_\epsilon^{2\sigma}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce

$$\sup_{J \geq -1} 2^{2J\sigma} \sum_{j \geq J} \sum_k \beta_{jk}^2 \;\leq \frac{2M}{(1-a)^2},$$

and f belongs to $\mathcal{B}^\sigma_{2,\infty}$. □

We have proved here that $\mathcal{B}^\sigma_{2,\infty}$ is a good candidate for an ideal maxiset among limited rules. We will prove in section 4.4 that it is reached by standard and well known limited procedures. So, as a consequence, $\mathcal{B}^\sigma_{2,\infty}$ is the ideal maxiset among limited rules with the relation between the limiting parameter and the rate of convergence above prescribed.
In the next subsection, we focus on elitist procedures.

4.3.3 Ideal maxisets for elitist rules

Theorem 4.2 (Ideal maxiset for elitist rules)**.** *Let $\hat{f}_\epsilon$ be an elitist rule in $\mathcal{E}(\lambda_\epsilon, a)$ with $a \in [0,1[$. Then, if λ_ε is a non decreasing, continuous function such that $\lambda_0 = 0$, and $0 < r < 2$ is a real number,*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r}) \subset W(r,2)$$

(with $M' = \frac{\sqrt{2M}}{(1-a)}$.)

Remark 4.5. *It is important to notice that this inclusion will be mostly used for $\lambda_\epsilon = t_\epsilon$, $r = \frac{2}{1+2s}$, $2 - r = \frac{4s}{1+2s}$, where we find back the usual rates of convergence.*

Proof of Theorem 4.2 : Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r})(M)$. If we observe that if $|y_{jk}| \leq \lambda_\epsilon$ then $\gamma_{jk} \leq a$, we have :

$$
\begin{aligned}
& (1-a)^2 \sum_{j,k} \beta_{jk}^2 \mathbb{1}\ \{|\beta_{jk}| \leq \lambda_\epsilon\} \\
= \ & 2(1-a)^2 \sum_{j,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0) \mathbb{1}\ \{\beta_{jk} \geq 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0) \mathbb{1}\ \{\beta_{jk} < 0\}\right] \mathbb{1}\ \{|\beta_{jk}| \leq \lambda_\epsilon\} \\
\leq \ & 2\mathbb{E} \sum_{j,k} \left[(\beta_{jk} - \gamma_{jk} y_{jk})^2 \mathbb{1}\ \{\beta_{jk} \geq 0\} + (\beta_{jk} - \gamma_{jk} y_{jk})^2 \mathbb{1}\ \{\beta_{jk} < 0\}\right] \mathbb{1}\ \{|\beta_{jk}| \leq \lambda_\epsilon\} \\
\leq \ & 2\mathbb{E} \sum_{j,k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 \\
\leq \ & 2M\ \lambda_\epsilon^{2-r}.
\end{aligned}
$$

So, using the continuity of λ_ϵ in 0, we deduce that

$$
\sup_{\lambda > 0} \lambda^{r-2} \sum_{j \geq -1} \sum_k \beta_{jk}^2 \mathbb{1}\ \{|\beta_{jk}| \leq \lambda\} \ \leq \frac{2M}{(1-a)^2},
$$

and f belongs to $W(r, 2)$. □

In the next subsection, we focus on cautious procedures.

4.3.4 Ideal maxisets for cautious rules

Theorem 4.3 (Ideal maxiset for cautious rules)**.** *Let $\hat{f}_\epsilon$ be a cautious rule in $\mathcal{C}(\lambda_\epsilon, a)$ with $a \in]0, 1]$. Let us suppose that $0 < r < 2$ is a real number and λ_ε is a non decreasing, continuous function such that $\lambda_0 = 0$. Suppose that*

$$
\exists\, c > 0, \quad \forall\, \epsilon > 0,\ \frac{\lambda_\epsilon}{\sqrt{\log(\frac{1}{\lambda_\epsilon})}} \leq c\epsilon. \tag{4.4}
$$

Then

$$
MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r}) \subset W^*(r, 2)
$$

(with $M' = \frac{2c\sqrt{2M}}{a}$.)

Remark 4.6. *Note that the case $\lambda_\epsilon = t_\epsilon$ (resp. $\lambda_\epsilon = \epsilon$) satisfies (4.4) with $c = \sqrt{2}$ (resp. $c = 1$)*

Proof of Theorem 4.3 : It is a consequence of the following lemma :

Lemma 4.1. *Let $\epsilon > 0$ and suppose that $|\beta_{jk}| > \lambda_\epsilon$ and $sign(\beta_{jk})y_{jk} < |\beta_{jk}|$. Then,*

$$a|\beta_{jk} - y_{jk}| \leq 2|\beta_{jk} - \gamma_{jk}y_{jk}|.$$

Proof : We only prove the case $\beta_{jk} > \lambda_\epsilon$ and $y_{jk} < \beta_{jk}$ since the case $\beta_{jk} < -\lambda_\epsilon$ and $y_{jk} > \beta_{jk}$ can be proved with the same arguments.
It is clear that,

a) if $y_{jk} \geq 0$, then, $a(\beta_{jk} - y_{jk}) \leq a(\beta_{jk} - \gamma_{jk}y_{jk})$

b) if $y_{jk} < -\lambda_\epsilon$, then, because the rule is cautious, $\gamma_{jk} > a$ and $a(\beta_{jk} - y_{jk}) \leq \gamma_{jk}(\beta_{jk} - y_{jk}) \leq (\beta_{jk} - \gamma_{jk}y_{jk})$

c) if $-\lambda_\epsilon \leq y_{jk} < 0$, then $a(\beta_{jk} - y_{jk}) \leq 2a\beta_{jk} \leq 2a(\beta_{jk} - \gamma_{jk}y_{jk})$.

Since $0 < a < 1$ we deduce from a) b) and c) that $a(\beta_{jk} - y_{jk}) \leq 2(\beta_{jk} - \gamma_{jk}y_{jk})$.

□

Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r})(M)$. Using (4.4),

$$a^2\lambda_\epsilon^2 \left[\log(\frac{1}{\lambda_\epsilon})\right]^{-1} \sum_{j<j_\epsilon,k} \mathbb{1}\ \{|\beta_{jk}| > \lambda_\epsilon\} \leq a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} \mathbb{1}\ \{|\beta_{jk}| > \lambda_\epsilon\}.$$

Now, let us recall that if X is a zero-mean Gaussian variable with variance ϵ^2, then

$$\mathbb{E}(X^2 I_{\{X<0\}}) = \mathbb{E}(X^2 I_{\{X>0\}}) = \frac{\epsilon^2}{2}.$$

So, from Lemma 4.1

$$\begin{aligned}
& a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} \mathbb{1}\ \{|\beta_{jk}| > \lambda_\epsilon\} \\
= \ & a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} [\mathbb{1}\ \{\beta_{jk} > \lambda_\epsilon\} + \mathbb{1}\ \{\beta_{jk} < -\lambda_\epsilon\}] \\
= \ & 2a^2c^2\ \mathbb{E} \sum_{j<j_\epsilon,k} (\beta_{jk} - y_{jk})^2\, [\mathbb{1}\ \{y_{jk} - \beta_{jk} < 0\}\mathbb{1}\ \{\beta_{jk} > \lambda_\epsilon\} + \mathbb{1}\ \{y_{jk} - \beta_{jk} > 0\}\mathbb{1}\ \{\beta_{jk} < -\lambda_\epsilon\}] \\
\leq \ & 8c^2\ \mathbb{E} \sum_{j<j_\epsilon,k} (\beta_{jk} - \gamma_{jk}y_{jk})^2 \\
\leq \ & 8c^2M\ \lambda_\epsilon^{2-r}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that

$$\sup_{\lambda>0} \lambda^r \left[\log(\frac{1}{\lambda})\right]^{-1} \sum_{j<j_\lambda,k} \mathbb{1}\left\{|\beta_{jk}| > \lambda\right\} \leq \frac{8c^2 M}{a^2}$$

and f belongs to $W^*(r,2)$. □

4.4 Rules ensuring that their maxiset contains a prescribed subset

In this section we prove two types of conditions ensuring that the maxiset of a given shrinkage rule contains either a Besov space or a weak Besov space. This part is obviously strongly linked with upper bounds inequalities in minimax theory. Indeed, our technique of proof here will be to show that some classes of estimators satisfy an upper bound inequality associated with the considered subset.

4.4.1 When does the maxiset contain a Besov space ?

We have the following result, which is a converse result to Theorem 4.1 with respect to the ideal maxiset result for limited rules :

Theorem 4.4. *Let $s > 0$ and $(\gamma_j(\epsilon))_{jk}$ a non increasing sequence of weights lying in $[0,1]$ such that $\hat{\beta}^L_\epsilon = (\gamma_j(\epsilon) y_{jk})_{jk}$ belongs to $\mathcal{L}(\lambda_\epsilon, a)$, with $a \in [0,1[$, λ_ε is continuous and $\lambda_0 = 0$. If there exist C_1 and C_2 in $\mathbb{R}$ such that, with $\gamma_{-2} = 1$, $\forall\, \epsilon > 0$,*

$$\sum_{j\geq -1} (\gamma_{j-1} - \gamma_j)(1-\gamma_j) 2^{-2js} \mathbb{1}\left\{2^j < \lambda_\epsilon^{-1}\right\} \leq C_1 \lambda_\epsilon^{2s}$$

$$\sum_{j\geq -1} 2^j \gamma_j(\epsilon)^2 \leq C_2\, \epsilon^{-2}\, \lambda_\epsilon^{2s}$$

then,

$$\mathcal{B}^s_{2,\infty} \subset MS(\hat{\beta}^L_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s}).$$

Proof of the Theorem 4.4 : This result is a simple consequence of Theorem 2 of Rivoirard (2004[102]). A more general result is established in the appendix. □

Combining Theorems 4.1 and 4.4, by straightforward computations, we obtain :

Corollary 4.1. *If we consider linear estimates associated with the weights $\gamma_j^{(1)}(\lambda_\epsilon)$, $\gamma_j^{(2)}(\lambda_\epsilon)$ with $\alpha > (s \vee 1/2)$ or $\gamma_j^{(3)}(\lambda_\epsilon)$ with $\alpha > s$ (see section 4.2.2), then for $i \in \{1,2,3\}$*

$$MS((\gamma_j^{(i)}(\lambda_\epsilon) y_{jk})_{jk}, \|.\|_2^2, \lambda_\epsilon^{2s}) = \mathcal{B}^s_{2,\infty},$$

as soon as $(\epsilon^2 \lambda_\epsilon^{-(1+2s)})_\epsilon$ is bounded. In particular, for the polynomial rate $\epsilon^{4s/(1+2s)}$, corresponding to $\lambda_\epsilon = \epsilon^{2/(1+2s)}$, $\mathcal{B}^s_{2,\infty}$ is exactly the maxiset of these estimates.

Remark 4.7. *Rivoirard (2004[103]) extended these results for a more general statistical model : the heteroscedastic white noise model that naturally appears in the literature of inverse problems. This last result illustrates the strong link between linear procedures (and more generally limited procedures) and Besov spaces. This has already been pointed out by Kerkyacharian and Picard (1993[74]) who studied maxisets for linear procedures for the model of density estimation.*

4.4.2 When does the maxiset contain a weak Besov space ?

We have the following result, which is a converse result to Theorems 4.1 and 4.2 with respect to the ideal maxiset results for limited and elitist rules :

Theorem 4.5. *Let $s > 0$ and $\gamma_{jk}(\epsilon)$ a sequence of random weights lying in $[0,1]$. We assume that there exist positive constants c, m and $K(\gamma)$ such that for any $\epsilon > 0$*

$$\hat{\beta}(\epsilon) = (\gamma_{jk}(\epsilon) y_{jk})_{jk} \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(m t_\epsilon, c t_\epsilon), \tag{4.5}$$

$$(1 - \gamma_{jk}(\epsilon)) \leq K(\gamma) \left(\frac{t_\varepsilon}{|y_{jk}|} + t_\epsilon \right), \quad a.e. \quad \forall\, j < j_\epsilon, \ \forall\, k. \tag{4.6}$$

Then, as soon as $m \geq 8$,

$$\mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap W(\frac{2}{1+2s}, 2) \subset MS(\hat{f}_\epsilon, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}).$$

Remark 4.8. *It is worthwhile to note that (4.6) is a condition implying that the procedure belongs to $\mathcal{C}(t_\epsilon, Dt_\epsilon)$, and can be considered as a refinement of the cautiousness condition. It is enough to verify condition (4.6) for ϵ small enough without modifying the conclusion of the theorem. This remark will be useful in sections 4.5.2 and 4.5.3, where we apply Theorem 4.5 to Bayesian procedures.*

This theorem, is an obvious consequence of the following two propositions concerning functional spaces inclusions and general upper bound results for shrinkage procedures.

Proposition 4.1. *Let $0 < r < 2$, $C > 0$ and $f \in W(r,2)$. Then,*

$$\sup_{\lambda>0} \lambda^r \sum_{j,k} \mathbb{1}\ \{|\beta_{jk}| > \lambda\} \leq \frac{2^{2-r}\|f\|^2_{W_r}}{1-2^{-r}}.$$

The proof of this proposition is standard, see for instance in Kerkyacharian and Picard (2000[75]), where it is proved that the condition above is in fact equivalent to the fact that $f \in W(r,2)$.

Proposition 4.2. *Under the conditions of Theorem 4.5, we have the following inequality :*

$$\begin{aligned}\mathbb{E}\|\hat{f}_\epsilon - f\|_2^2 \ \leq\ & \Big[4c^2 S_\psi + 4(1+K(\gamma)^2)\|f\|_2^2 + 4\sqrt{3}S_\psi + 2(2^{\frac{4s}{1+2s}} + 2^{\frac{-4s}{1+2s}})m^{\frac{4s}{1+2s}}\|f\|^2_{W_{\frac{2}{1+2s}}} + \\ & + \tfrac{8m^{-2/(1+2s)}}{(1-2^{-2/(1+2s)})}(1+8K(\gamma)^2)\|f\|^2_{W_{\frac{2}{1+2s}}} + \|f\|^2_{B^{\frac{s}{1+2s}}_{2,\infty}}\Big]\, t_\epsilon^{\frac{4s}{1+2s}}.\end{aligned}$$

Proof : Let $f \in \mathcal{B}^{\frac{s}{1+2s}}_{2,\infty} \cap W(\frac{2}{1+2s}, 2)$. Obviously, using the limitation assumption, we have for j_ϵ such that $2^{j_\epsilon} \sim t_\epsilon^{-2}$

$$\mathbb{E}\|\hat{f}_\epsilon - f\|_2^2 = \mathbb{E}\|\sum_{j<j_\epsilon,k}(\gamma_{jk}(\epsilon)y_{jk} - \beta_{jk})\psi_{j,k}\|_2^2 + \sum_{j\geq j_\epsilon,k}\beta_{jk}^2.$$

The second term is a bias term bounded by $t_\epsilon^{\frac{4s}{1+2s}}\|f\|^2_{B^{\frac{s}{1+2s}}_{2,\infty}}$, by definition of the Besov norm.
We split $\mathbb{E}\sum_{j<j_\epsilon,k}(\gamma_{jk}(\epsilon)y_{jk} - \beta_{jk})^2$ into $2(A+B)$ with

$$\begin{aligned} A &= \mathbb{E}\sum_{j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ \mathbb{1}\ \{|y_{jk}| \leq mt_\epsilon\}, \\ B &= \mathbb{E}\sum_{j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ \mathbb{1}\ \{|y_{jk}| > mt_\epsilon\}.\end{aligned}$$

Again, we split A into $A_1 + A_2$, and using $\hat{\beta}(\epsilon) \in \mathcal{E}(mt_\epsilon, ct_\epsilon)$, we have on $\{|y_{jk}| \leq mt_\epsilon\}$, $\gamma_{jk} \leq ct_\epsilon$. So,

$$\begin{aligned}
A_1 &= \mathbb{E} \sum_{j<j_\epsilon,k} \gamma_{jk}(\epsilon)^2 (y_{jk} - \beta_{jk})^2 \ \mathbb{1}\ \{|y_{jk}| \leq mt_\epsilon\} \\
&\leq c^2 S_\psi 2^{j_\epsilon} t_\epsilon^2 \epsilon^2 \\
&\leq 2c^2 S_\psi t_\epsilon^2.
\end{aligned}$$

$$\begin{aligned}
A_2 &= \mathbb{E} \sum_{j<j_\epsilon,k} (1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2 \ \mathbb{1}\ \{|y_{jk}| \leq mt_\epsilon\} \\
&\leq \mathbb{E} \sum_{j<j_\epsilon,k} \beta_{jk}^2 \ \mathbb{1}\ \{|y_{jk}| \leq mt_\epsilon\}[\mathbb{1}\ \{|\beta_{jk}| \leq 2mt_\epsilon\} + \mathbb{1}\ \{|\beta_{jk}| > 2mt_\epsilon\}] \\
&\leq (2mt_\epsilon)^{4s/(1+2s)} \|f\|^2_{W_{\frac{2}{1+2s}}} + \sum_{j<j_\epsilon,k} \beta_{jk}^2 \mathbb{P}(|\beta_{jk} - y_{jk}| \geq mt_\epsilon) \\
&\leq (2mt_\epsilon)^{4s/(1+2s)} \|f\|^2_{W_{\frac{2}{1+2s}}} + \|f\|_2^2 \epsilon^{m^2/2} \\
&\leq (2mt_\epsilon)^{4s/(1+2s)} \|f\|^2_{W_{\frac{2}{1+2s}}} + \|f\|_2^2 t_\epsilon^2.
\end{aligned}$$

We have used here the concentration property of the Gaussian distribution and the fact that $m^2 \geq 4$.

$$\begin{aligned}
B &:= B_1 + B_2 \\
&= \mathbb{E} \sum_{j<j_\epsilon,k} [\gamma_{jk}(\epsilon)^2 (y_{jk} - \beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2] \ \mathbb{1}\ \{|y_{jk}| > mt_\epsilon\}[\mathbb{1}\ \{|\beta_{jk}| \leq mt_\epsilon/2\} \\
&\quad + \mathbb{1}\ \{|\beta_{jk}| > mt_\epsilon/2\}].
\end{aligned}$$

For B_1 we use the Schwartz inequality :

$$\mathbb{E}(y_{jk} - \beta_{jk})^2 \mathbb{1}\ \{|y_{jk} - \beta_{jk}| > mt_\epsilon/2\} \leq (\mathbb{P}(|y_{jk} - \beta_{jk}| > mt_\epsilon/2))^{1/2} (\mathbb{E}(y_{jk} - \beta_{jk})^4)^{1/2}.$$

Now, observing that $\mathbb{E}(y_{jk} - \beta_{jk})^4 = 3\epsilon^4$ and that $\mathbb{P}(|y_{jk} - \beta_{jk}| > mt_\epsilon/2) \leq \epsilon^{\frac{m^2}{8}}$, we have for $m^2 \geq 32$:

$$\begin{aligned}
B_1 &\leq \sqrt{3} \sum_{j<j_\epsilon,k} \epsilon^2 \mathbb{1}\ \{|\beta_{jk}| \leq mt_\epsilon/2\} \epsilon^{\frac{m^2}{16}} + \sum_{j<j_\epsilon,k} \beta_{jk}^2 \mathbb{1}\ \{|\beta_{jk}| \leq mt_\epsilon/2\} \\
&\leq 2\sqrt{3} S_\psi t_\epsilon^2 + \left(\frac{m}{2} t_\epsilon\right)^{4s/(1+2s)} \|f\|^2_{W_{\frac{s}{1+2s}}}.
\end{aligned}$$

For B_2, we use Proposition 4.1,

$$\begin{aligned} B_2 &= \mathbb{E}\sum_{j<j_\epsilon,k} [\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2]\, \mathbb{1}\,\{|y_{jk}| > mt_\epsilon\}\mathbb{1}\,\{|\beta_{jk}| > mt_\epsilon/2\} \\ &\leq \sum_{j<j_\epsilon,k} [\epsilon^2 \mathbb{1}\,\{|\beta_{jk}| > mt_\epsilon/2\} + B_3 \\ &\leq \frac{4m^{-2/(1+2s)}}{(1-2^{-2/(1+2s)})}\|f\|^2_{W_{\frac{2}{1+2s}}} t_\epsilon^{4s/(1+2s)} + B_3. \end{aligned}$$

$$\begin{aligned} B_3 &:= \sum_{j<j_\epsilon,k} \mathbb{E}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\, \mathbb{1}\,\{|y_{jk}| > mt_\epsilon\}\mathbb{1}\,\{|\beta_{jk}| > mt_\epsilon/2\}[\mathbb{1}\,\{|y_{jk}| \geq |\beta_{jk}|/2\} + \mathbb{1}\,\{|y_{jk}| < |\beta_{jk}|/2\} \\ &:= B_3' + B"_3. \\ B"_3 &\leq \sum_{j<j_\epsilon,k} \beta_{jk}^2\mathbb{P}(|y_{jk}-\beta_{jk}| \geq mt_\varepsilon/4) \\ &\leq \|f\|_2^2 t_\epsilon^2. \end{aligned}$$

since $m^2 \geq 64$. We have used in the line above the concentration property of the Gaussian distribution. Now using (4.6) and Proposition 4.1, we get,

$$\begin{aligned} B_3' &\leq \sum_{j<j_\epsilon,k} \mathbb{E}\beta_{jk}^2(1-\gamma_{jk}(\epsilon))^2\mathbb{1}\,\{|y_{jk}| \geq |\beta_{jk}|/2\}\mathbb{1}\,\{|\beta_{jk}| > mt_\epsilon/2\}\mathbb{1}\,\{|y_{jk}| \geq mt_\epsilon\}] \\ &\leq \sum_{j<j_\epsilon,k} \mathbb{E}\beta_{jk}^2 K(\gamma)^2\left(\frac{t_\varepsilon}{|y_{jk}|} + t_\epsilon\right)^2 \mathbb{1}\,\{|y_{jk}| \geq |\beta_{jk}|/2\}I\{|\beta_{jk}| > mt_\epsilon/2\}) \\ &\leq K(\gamma)^2\frac{32m^{-2/(1+2s)}}{1-2^{-2/(1+2s)}}\,\|f\|^2_{W_{\frac{2}{1+2s}}}\, t_\epsilon^{4s/(1+2s)} + 2K(\gamma)^2\|f\|_2^2 t_\epsilon^2. \end{aligned}$$

□

We deduce as a corollary the following results.

Corollary 4.2. *The hard thresholding $\hat{f}_T$ and the soft thresholding $\hat{f}_S$ rules as defined in (4.2) and (4.3) with $m \geq 8$ are satisfying :*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}) = \mathcal{B}^{s/(1+2s)}_{2,\infty} \cap W(\frac{2}{1+2s}, 2).$$

The proof of this corollary is an elementary consequence of Theorems 4.1, 4.2 and 4.5. It proves that these procedures are optimal in the maxiset sense among elitist rules which are limited.

4.5 Maxisets for Bayesian procedures

In this section, we focus on the study of Bayes rules. We recall that we consider the prior model defined in Introduction.

4.5.1 Gaussian priors : a first approach

Let us consider the Bayes model (4.1) where γ is the Gaussian density, which is the most classical choice. In this case, we easily derive the Bayes rules of β_{jk} associated with the l^1-loss and the l^2-loss :

$$\breve{\beta}_{jk} = \mathrm{Med}(\beta_{jk}|y_{jk}) = \mathrm{sign}(y_{jk}) \max(0, \xi_{jk}),$$

$$\tilde{\beta}_{jk} = \mathbb{E}(\beta_{jk}|y_{jk}) = \frac{b_j}{1+\eta_{jk}} y_{jk},$$

where

$$\xi_{jk} = b_j|y_{jk}| - \epsilon\sqrt{b_j}\Phi^{-1}\left(\frac{1+\min(\eta_{jk},1)}{2}\right),$$

$$b_j = \frac{\tau_{j,\epsilon}^2}{\epsilon^2+\tau_{j,\epsilon}^2},$$

$$\eta_{jk} = \frac{1}{w_{j,\epsilon}}\frac{\sqrt{\epsilon^2+\tau_{j,\epsilon}^2}}{\epsilon}\exp\left(-\frac{\tau_{j,\epsilon}^2 y_{jk}^2}{2\epsilon^2(\epsilon^2+\tau_{j,\epsilon}^2)}\right),$$

and Φ is the normal cumulative distributive function. Both rules are then shrinkage rules. We also note that $\breve{\beta}_{jk}$ is zero whenever y_{jk} falls in an implicitly defined interval $[-\lambda_{j,\epsilon}, \lambda_{j,\epsilon}]$. So it is a thresholding rule. In the following, we study the maxisets of the previous estimates associated with the following very classical form for the hyperparameters :

$$\tau_{j,\epsilon}^2 = c_1 2^{-\alpha j}, \quad \pi_{j,\epsilon} = \min(1, c_2 2^{-bj}),$$

where c_1, c_2, α and b are positive constants. This particular form for the hyperparameters was suggested by Abramovich et al. (1998[4]) and then used by Abramovich et al. (2004[1]). A nice interpretation was provided by these authors who explained how α, b, c_1 and c_2 can be derived for applications.

Remark 4.9. *An alternative for eliciting these hyperparameters consists in using empirical Bayes methods and EM algorithm (see Clyde and George (1998[25],2000[26]) or Johnstone et Silverman (1998[67])).*

In a minimax setting, Abramovich et al. (2004[1]) obtained the following result :

Theorem 4.6. *Let β^0 be $\breve{\beta}$ or $\tilde{\beta}$. With $\alpha = 2s+1$ and any $0 \leq b < 1$, there exist two positive constants C_1 and C_2 such that $\forall\ \epsilon > 0$,*

$$C_1(\epsilon\sqrt{\log(1/\epsilon)})^{4s/(2s+1)} \leq \sup_{\beta \in \mathcal{B}^s_{2,\infty}(M)} \mathbb{E}\|\beta^0 - \beta\|_2^2 \leq C_2 \log(1/\epsilon)\epsilon^{4s/(2s+1)}.$$

Now, let us consider the maxiset setting. Both previous Bayesian procedures are limited. Indeed, as soon as $\tau^2_{j,\epsilon} \leq \epsilon^2$ we have $b_j \leq 1/2$. So, each of these procedures belongs to $\mathcal{L}((c_1^{-1}\epsilon^2)^{1/\alpha}, 1/2)$. So, if $\alpha > 1$, by using Theorem 4.1, for $\beta^0 \in \{\breve{\beta}, \tilde{\beta}\}$,

$$MS(\beta^0, \|.\|_2^2, \epsilon^{2(\alpha-1)/\alpha}) \subset \mathcal{B}^{(\alpha-1)/2}_{2,\infty}.$$

With $s > 0$ and $\alpha = 1 + 2s$,

$$MS(\beta^0, \|.\|_2^2, \epsilon^{4s/(1+2s)}) \subset \mathcal{B}^s_{2,\infty}. \tag{4.7}$$

Actually, we have the following theorem :

Theorem 4.7. *For $s > 0$, $\alpha = 2s+1$, any $0 \leq b < 1$, and if β^0 is $\breve{\beta}$ or $\tilde{\beta}$,*

1. *for the rate $\epsilon^{4s/(1+2s)}$,*
$$MS(\beta^0, \|.\|_2^2, \epsilon^{4s/(1+2s)}) \subsetneq \mathcal{B}^s_{2,\infty},$$
2. *for the rate $(\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}$,*
$$MS(\beta^0, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) \subset \mathcal{B}^{*s}_{2,\infty},$$
3. *for the rate $\epsilon^{4s/(1+2s)}\log(1/\epsilon)$,*
$$\mathcal{B}^s_{2,\infty} \subset MS(\beta^0, \|.\|_2^2, \epsilon^{4s/(1+2s)}\log(1/\epsilon)).$$

with

$$\mathcal{B}^{*s}_{2,\infty} = \left\{ f \in \mathbb{L}^2 : \quad \sup_{J>0} 2^{2Js} J^{-2s/(1+2s)} \sum_{j \geq J} \sum_k \beta^2_{jk} < \infty \right\}.$$

Proof : The first point is a simple consequence of equation (4.7) and Theorem 4.6. The second one is easily obtained by using similar arguments as for the proof of Theorem 4.1. Finally, the proof of the last one is provided by Theorem 4.6. □

If we consider limited procedures, this theorem shows that the maxiset of these Bayesian procedures is not the ideal one. The first point of Theorem 4.7 and Corollary 4.1 show that they are also outperformed by linear estimates for polynomial rates of convergence. Furthermore, these procedures do not achieve the same performance as classical non linear procedures, since, obviously, $\mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W(\frac{2}{2s+1}, 2)$ is not included in $\mathcal{B}_{2,\infty}^{*s}$. The following theorem even reinforces this bad sentence by proving that these procedures are highly non robust with respect to the choice of α, which is a serious drawback in practise since s is generally unknown.

Theorem 4.8. *With the previous choice for the hyperparameters, for $s > 0$ and $\beta^0 \in \{\breve{\beta}, \tilde{\beta}\}$,*

- *$\alpha > 2s+1$ implies $\mathcal{B}_{p,\infty}^s$ is not included in $MS(\beta^0, \|.\|_2^2, t_\epsilon^{4s/(1+2s)})$ for any $1 \leq p \leq \infty$.*
- *$\alpha = 2s+1$ implies $\mathcal{B}_{p,\infty}^s$ is not included in $MS(\beta^0, \|.\|_2^2, t_\epsilon^{4s/(1+2s)})$ if $p < 2$,*

where

$$\mathcal{B}_{p,\infty}^s = \left\{ f : \quad \sup_{j \geq -1} 2^{jp(s+\frac{1}{2}-\frac{1}{p})} \sum_k |\beta_{jk}|^p < \infty \right\}.$$

Remark 4.10. *Theorem 4.8 is established for the rate $t_\epsilon^{4s/(1+2s)}$ but it can be generalized for any rate of convergence of the form $\epsilon^{4s/(1+2s)}(\log(1/\epsilon))^m$, with $m \geq 0$.*

The proof of Theorem 4.8 is based on the following result :

Proposition 4.3. *If $\beta \in MS(\beta^0, \|.\|_2^2, t_\epsilon^{4s/(1+2s)})$ then there exists a constant C such that, for ϵ small enough :*

$$\sum_{j,k} \beta_{jk}^2 \mathbb{1}\, \{\tau_{j,\epsilon}^2 \leq \epsilon^2\} \mathbb{1}\, \{|\beta_{jk}| > t_\epsilon\} \leq C t_\epsilon^{\frac{4s}{1+2s}} \tag{4.8}$$

Proof :

Here we shall distinguish the cases of the posterior mean and median.

The posterior median can be written as follows :

$$\breve{\beta}_{jk} = \text{sign}(y_{jk})(b_j|y_{jk}| - g(\epsilon, \tau_{j,\epsilon}, y_{jk})),$$

with $0 \le g(\epsilon, \tau_{j,\epsilon}, y_{jk}) \le b_j |y_{jk}|$.
Let us assume that $b_j |y_{jk} - \beta_{jk}| \le (1-b_j)|\beta_{jk}|/2$ and $\tau_{j,\epsilon}^2 \le \epsilon^2$, so $b_j \le 1/2$.
First, let us suppose that $y_{jk} \ge 0$ so $\breve{\beta}_{jk} \ge 0$. If $\beta_{jk} \ge 0$, then

$$\begin{aligned} |\breve{\beta}_{jk} - \beta_{jk}| &= |b_j(y_{jk} - \beta_{jk}) - (1-b_j)\beta_{jk} - g(\epsilon, \tau_{j,\epsilon}, y_{jk})| \\ &= (1-b_j)\beta_{jk} - b_j(y_{jk} - \beta_{jk}) + g(\epsilon, \tau_{j,\epsilon}, y_{jk}) \\ &\ge \frac{1}{2}(1-b_j)\beta_{jk} \\ &\ge \frac{1}{4}\beta_{jk}. \end{aligned}$$

If $\beta_{jk} \le 0$, then

$$|\breve{\beta}_{jk} - \beta_{jk}| \ge \frac{1}{4}|\beta_{jk}|.$$

The case $y_{jk} \le 0$ is handled by using similar arguments and the particular form of the posterior median. So, we obtain :

$$\begin{aligned} \mathbb{E}(\breve{\beta}_{jk} - \beta_{jk})^2 \mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\} &\ge \frac{1}{16}\beta_{jk}^2\, \mathbb{P}(b_j|y_{jk} - \beta_{jk}| \le (1-b_j)|\beta_{jk}|/2)\mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\} \\ &\ge \frac{1}{16}\beta_{jk}^2\, \mathbb{P}(|y_{jk} - \beta_{jk}| \le |\beta_{jk}|/2)\mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\}. \end{aligned}$$

So, we obtain :

$$\begin{aligned} \mathbb{E}(\breve{\beta}_{jk} - \beta_{jk})^2 \mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\} &\ge \frac{1}{16}\beta_{jk}^2 \mathbb{P}(|y_{jk} - \beta_{jk}| \le |\beta_{jk}|/2)\mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\} \\ &\ge \frac{1}{16}\beta_{jk}^2 (1 - \mathbb{P}(|y_{jk} - \beta_{jk}| > |\beta_{jk}|/2))\mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\} \end{aligned}$$

Using the large deviations inequalities for the Gaussian variables, we obtain for ϵ small enough :

$$\begin{aligned} \mathbb{E}(\breve{\beta}_{jk} - \beta_{jk})^2 \mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\}\mathbb{1}\,\{|\beta_{jk}| > t_\epsilon\} &\ge \frac{1}{16}\beta_{jk}^2 (1 - \mathbb{P}(|y_{jk} - \beta_{jk}| > t_\epsilon/2))\mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\}\mathbb{1}\,\{|\beta_{jk}| > t_\epsilon\} \\ &\ge \frac{1}{32}\beta_{jk}^2 \mathbb{1}\,\{\tau_{j,\epsilon}^2 \le \epsilon^2\}\mathbb{1}\,\{|\beta_{jk}| > t_\epsilon\} \end{aligned}$$

This implies (4.8).

For the posterior mean, we have :

$$\begin{aligned}
\mathbb{E}(\tilde{\beta_{jk}} - \beta_{jk})^2 &= \mathbb{E}\left(\frac{b_j}{1+\eta_{jk}}(y_{jk}-\beta_{jk}) - (1-\frac{b_j}{1+\eta_{jk}})\beta_{jk}\right)^2 \\
&\geq \frac{1}{4}\mathbb{E}\left((1-\frac{b_j}{1+\eta_{jk}})\beta_{jk}\right)^2 \mathbb{1}\left\{\frac{b_j}{1+\eta_{jk}}|y_{jk}-\beta_{jk}| \leq (1-\frac{b_j}{1+\eta_{jk}})|\beta_{jk}|/2\right\}
\end{aligned}$$

So, we obtain :

$$\begin{aligned}
\mathbb{E}(\tilde{\beta_{jk}} - \beta_{jk})^2 \mathbb{1}\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} &\geq \frac{1}{16}\beta_{jk}^2 \mathbb{P}(|y_{jk}-\beta_{jk}| \leq |\beta_{jk}|/2)\mathbb{1}\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} \\
&\geq \frac{1}{16}\beta_{jk}^2 (1-\mathbb{P}(|y_{jk}-\beta_{jk}| > |\beta_{jk}|/2))\mathbb{1}\{\tau_{j,\epsilon}^2 \leq \epsilon^2\}
\end{aligned}$$

Finally, using similar arguments as those used for the posterior median, we obtain (4.8). Proposition 4.3 is proved. □

Now, let us prove Theorem 4.8. Let us first investigate the case $\alpha > 2s+1$.
Let us take β such that all the β_{jk}'s are zero, except 2^j coefficients at each level j that are equal to $2^{-j(s+\frac{1}{2})}$. Then, $\beta \in \mathcal{B}^s_{p,\infty}$. Since $\tau_{j,\epsilon}^2 = c_1 2^{-j\alpha}$, if we put $2^{J_\alpha} \sim c_1^{\frac{1}{\alpha}} \epsilon^{-\frac{2}{\alpha}}$ and $2^{J_s} \sim t_\epsilon^{\frac{-2}{2s+1}}$, we observe that asymptotically $J_\alpha < J_s$. So, for ϵ small enough :

$$\begin{aligned}
\sum_{j,k} \beta_{jk}^2 \mathbb{1}\{\tau_{j,\epsilon}^2 \leq \epsilon^2\}\mathbb{1}\{|\beta_{jk}| > t_\epsilon\} &= \sum_{J_\alpha \leq j < J_s} 2^{-2js} \\
&\geq c\epsilon^{\frac{4s}{\alpha}},
\end{aligned}$$

with c a positive constant. Using Proposition 4.3, β does not belong to $MS(\beta^0, \|.\|_2^2, t_\epsilon^{4s/(1+2s)})$.
Let us then investigate the case $\alpha = 2s+1$.
Let us take β such that all the β_{jk}'s are zero, except 1 coefficient at each level j that is equal to $2^{-j(s+\frac{1}{2}-\frac{1}{p})}$. Then, $\beta \in \mathcal{B}^s_{p,\infty}$. Similarly, we put $2^{J_\alpha} \sim c_1^{\frac{1}{\alpha}} \epsilon^{-\frac{2}{\alpha}}$ and $2^{\tilde{J}_s} \sim t_\epsilon^{-1/(s+\frac{1}{2}-\frac{1}{p})}$, we observe that asymptotically $J_\alpha < \tilde{J}_s$. So, for ϵ small enough :

$$\begin{aligned}
\sum_{j,k} \beta_{jk}^2 \mathbb{1}\{\tau_{j,\epsilon}^2 \leq \epsilon^2\}\mathbb{1}\{|\beta_{jk}| > t_\epsilon\} &= \sum_{J_\alpha \leq j < \tilde{J}_s} 2^{-2j(s+\frac{1}{2}-\frac{1}{p})} \\
&\geq \tilde{c}\epsilon^{4(s+\frac{1}{2}-\frac{1}{p})/\alpha},
\end{aligned}$$

with $\tilde{c}$ a positive constant. Using Proposition 4.3, β does not belong to $MS(\beta^0, \|.\|_2^2, t_\epsilon^{4s/(1+2s)})$, since $p < 2$. □

The goal of the following subsections is to investigate a different choice for the hyperparameters $\tau_{j,\epsilon}$ and $w_{j,\epsilon}$ and for the density γ. Indeed, as in Johnstone et Silverman (2002[68],2004[70]) in the minimax setting, we would like to point out posterior Bayes estimates stemmed from the prior model (4.1) that achieve the same performance as non linear ones in the maxiset approach. It is all the more natural since Bayesian procedures can achieve better performances than classical non linear ones from a practical point of view. More precisely, we investigate a choice for the hyperparameters and for the density γ that enables us to obtain maxisets at least as large as $\mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W(\frac{2}{2s+1}, 2)$. Two different ways will be investigated. In section 4.5.2, we give up Gaussian densities and we consider heavy-tailed densities γ, as in Johnstone et Silverman (2002[68],2004[70]). Not surprisingly, the modified Bayesian procedures achieve very good performances. We show this result by proving that the Bayesian procedures are both limited and elitist. Then, in section 4.5.3, we wonder whether heavy-tailed priors are unavoidable and we consider, once more, Gaussian priors but with a different choice for the hyperparameters.

4.5.2 Heavy-tailed priors

In this section, we still consider the prior model (4.1), but the density γ is no longer Gaussian. We assume that there exist two positive constants M and M_1 such that

$$\sup_{\beta \geq M_1} \left| \frac{d}{d\beta} \log \gamma(\beta) \right| = M < \infty. \tag{4.9}$$

The hypothesis (4.9) means that the tails of γ have to be exponential or heavier. Indeed, under (4.9), we have :

$$\forall\, u \geq M_1, \quad \gamma(u) \geq \gamma(M_1) \exp(-M(u - M_1)).$$

In the minimax approach of Johnstone et Silverman (2002[68],2004[70]), the priors also verified (4.9). To complete the prior model, we assume that $\tau_{j,\epsilon} = \epsilon$ and $w_{j,\epsilon}$ depends only on ϵ with

$$w_{j,\epsilon} = w(\epsilon) \to 0, \text{ as } \epsilon \to 0$$

and w a positive continuous function. Using these assumptions, the following proposition describes the properties of the posterior median and mean :

Proposition 4.4. *We have :*

1. *The estimates* $\breve{\beta}_{jk} = Med(\beta_{jk}|y_{jk})$ *and* $\tilde{\beta}_{jk} = \mathbb{E}(\beta_{jk}|y_{jk})$ *are shrinkage rules : for* $\beta^0_{jk} \in \{\breve{\beta}_{jk}, \tilde{\beta}_{jk}\}$, $y_{jk} \longrightarrow \beta^0_{jk}$ *is antisymmetric, increasing on* $(-\infty, +\infty)$ *and*
$$0 \leq \beta^0_{jk} \leq y_{jk}, \quad \forall\, y_{jk} \geq 0.$$
2. $\breve{\beta}_{jk}$ *is a thresholding rule : there exists* $\breve{t}_\epsilon$ *such that*
$$\breve{\beta}_{jk} = 0 \iff |y_{jk}| \leq \breve{t}_\epsilon,$$
where the threshold $\breve{t}_\epsilon$ *verifies for* ϵ *small enough,* $\breve{t}_\epsilon \geq \epsilon\sqrt{2\log(1/w(\epsilon))}$ *and*
$$\lim_{\epsilon\to 0} \frac{\breve{t}_\epsilon}{\epsilon\sqrt{2\log(1/w(\epsilon))}} = 1.$$
3. *There exists a positive constant* C *such that*
$$\tilde{\beta}_{jk} = \tilde{\gamma}_{jk} y_{jk},$$
with
$$0 \leq \tilde{\gamma}_{jk} \leq Cw(\epsilon)\exp(\frac{y^2_{jk}}{2\epsilon^2}).$$
4. *Let us consider the threshold* $\breve{t}_\epsilon$ *introduced previously. There exists a positive constant* K *such that for* $\beta^0_{jk} \in \{\breve{\beta}_{jk}, \tilde{\beta}_{jk}\}$
$$\limsup_{\epsilon\to 0} |\epsilon^{-1} y_{jk} - \epsilon^{-1}\beta^0_{jk}| \mathbb{1}_{\,|y_{jk}|>2\breve{t}_\epsilon} \leq K. \quad a.s.$$

Proof : The first point has been established by Johnstone et Silverman (2002[68],2004[70]). The second point is an immediate consequence of Proposition 3 of Rivoirard (2004[103]). To prove the third point, we use Proposition 4 and Remark 1 of Rivoirard (2004[102]) yielding that there exist two positive constants C_1 and C_2 and two positive functions $\tilde{e}_1$ and $\tilde{e}_2$ such that

$$\tilde{\beta}_{jk} \;=\; y_{jk} \times \frac{\tilde{e}_1(\epsilon^{-1} y_{jk})}{1 + w(\epsilon)^{-1}\exp(-\frac{y^2_{jk}}{2\epsilon^2})\gamma(\epsilon^{-1}y_{jk})^{-1}\tilde{e}_2(\epsilon^{-1}y_{jk})},$$

where

$$\forall\, x \geq 0, \quad C_1 \leq \tilde{e}_1(x), \tilde{e}_2(x) \leq C_2$$

So,

$$\tilde{\gamma}_{jk} \leq \frac{C_2 \Gamma}{C_1} w(\epsilon) \exp(\frac{y_{jk}^2}{2\epsilon^2}),$$

where Γ is an upper bound for γ. The fourth point is easily derived by using Propositions 3 and 4 of Rivoirard (2004[102]). □

Now, let us introduce the following procedures. Given the previous prior model, we set

$$\breve{f}_\epsilon = \sum_{j<j_\epsilon} \sum_k \breve{\beta}_{jk} \psi_{jk}, \quad \breve{\beta}_{jk} = \mathrm{Med}(\beta_{jk} | y_{jk}), \tag{4.10}$$

and

$$\tilde{f}_\epsilon = \sum_{j<j_\epsilon} \sum_k \tilde{\beta}_{jk} \psi_{jk}, \quad \tilde{\beta}_{jk} = \mathbb{E}(\beta_{jk} | y_{jk}), \tag{4.11}$$

where j_ϵ is such that $2^{j_\epsilon} \sim t_\epsilon^{-2}$. Using the first two points of Proposition 4.4, we immediately obtain :

Corollary 4.3. *With C and $\breve{t}_\epsilon$ that have been introduced in Proposition 4.4, and $a \in]0,1[$, we have :*

$$\breve{f}_\epsilon \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(\breve{t}_\epsilon, 0),$$

$$\tilde{f}_\epsilon \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(\tilde{t}_\epsilon, a),$$

as soon as $\tilde{t}_\epsilon \leq \epsilon \sqrt{2 \log(\frac{a}{C w(\epsilon)})}$.

Remark 4.11. *Proposition 4.4 also shows that the posterior median is a cautious procedure. Using a proper choice of the hyperparameters, we can easily prove that the procedure associated with the posterior mean is also cautious.*

We have the following consequences on the maxisets of the procedures :

Theorem 4.9. *Let $s > 0$. We suppose that there exist two positive constants ρ_1 and ρ_2 such that for $\epsilon > 0$ small enough,*

$$\epsilon^{\rho_1} \leq w(\epsilon) \leq \epsilon^{\rho_2}.$$

Then, we have :

$$MS(f^0_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) = \mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W(\frac{2}{2s+1}, 2),$$

where $f^0_\epsilon \in \{\tilde{f}_\epsilon, \breve{f}_\epsilon\}$*, as soon as* $\rho_2 \geq 32$ *for the posterior median and* $\rho_2 \geq 33$ *for the posterior mean.*

Proof of Theorem 4.9 : The inclusions

$$MS(\breve{f}_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) \subset \mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W(\frac{2}{2s+1}, 2)$$

and

$$MS(\tilde{f}_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) \subset \mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W(\frac{2}{2s+1}, 2)$$

are provided by Theorems 4.1 and 4.2 and Corollary 4.3.
The inclusions

$$\mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W(\frac{2}{2s+1}, 2) \subset MS(\breve{f}_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)})$$

and

$$\mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W(\frac{2}{2s+1}, 2) \subset MS(\tilde{f}_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)})$$

are provided by the fourth point of Proposition 4.4, Corollary 4.3 and Theorem 4.5. □
So, the adaptive Bayesian procedures based on heavy-tailed prior densities are optimal among the class of limited and elitist procedures. We can also note that they outperform the Bayesian procedures of section 4.5.1 from the maxiset point of view.

4.5.3 Gaussian priors with large variance

The previous subsection has shown the power of the Bayes procedures built from heavy-tailed prior models in the maxiset setting. The goal of this section is then to answer the following questions. Are heavy-tailed priors unavoidable ? Can we simultaneously consider Gaussian densities and ignore the empirical Bayes setting to build optimal Bayesian procedures ? In other words, if γ is the Gaussian density, does there exist a fixed and adaptive choice of the hyperparameters $\pi_{j,\epsilon}$ and $w_{j,\epsilon}$ such that

$$MS(f^0_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) = \mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W(\frac{2}{2s+1}, 2),$$

where $f_\epsilon^0 \in \{\breve{f}_\epsilon, \tilde{f}_\epsilon\}$ (see (4.10) and (4.11)) ?

This is a very important issue since calculation using Gaussian priors are mostly direct and obviously much easier than heavy tails priors.
The answers are provided by the following theorem :

Theorem 4.10. *We consider the prior model (4.1), where γ is the Gaussian density. We assume that $\tau_{j,\epsilon} = \tau(\epsilon)$ and $w_{j,\epsilon} = w(\epsilon)$ are independent of j with w a continuous positive function. We consider $\breve{f}_\epsilon$ and $\tilde{f}_\epsilon$ introduced in (4.10) and (4.11). If*

$$1 + \epsilon^{-2}\tau(\epsilon)^2 = t_\epsilon^{-1} \text{ with } c_2 > 0$$

and there exist $q_1 > 0$ and $q_2 > 0$ such that for ϵ small enough

$$\epsilon^{q_1} \leq w(\epsilon) \leq \epsilon^{q_2}, \tag{4.12}$$

we have :

$$MS(f_\epsilon^0, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) = \mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W(\frac{2}{2s+1}, 2),$$

where $f_\epsilon^0 \in \{\tilde{f}_\epsilon, \breve{f}_\epsilon\}$ as soon as $q_2 > 63/2$ for the posterior median and $q_2 \geq 65/2$ for the posterior mean.

Whereas we usually consider $\tau_{j,\epsilon}^2 = \epsilon^2$ or $\tau_{j,\epsilon}^2 = 2^{-j\alpha}$, here we impose a "larger" variance. It is the key point of the proof of Theorem 4.10. In a sense, we re-create the heavy tails by increasing the variance.

Before giving it, let us prove that both Bayesian procedures belong to the class of limited and elitist procedures :

Proposition 4.5. *Under the assumptions of Theorem 4.10, we have for any $m > 0$ and for ϵ small enough,*

- *if $q_2 > \frac{m^2-1}{2}$, $\breve{f}_\epsilon \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(mt_\epsilon, 0)$,*
- *if $q_2 \geq \frac{m^2+1}{2}$, $\tilde{f}_\epsilon \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(mt_\epsilon, t_\epsilon)$.*

Proof : Using the definition of j_ϵ, each Bayesian procedure belongs to $\mathcal{L}(t_\epsilon^2, 0)$. Now, let us assume that $|y_{jk}| \leq mt_\epsilon$. Then,

$$\begin{aligned}\eta_{jk} &= \frac{1}{w(\epsilon)} \frac{\sqrt{\epsilon^2+\tau(\epsilon)^2}}{\epsilon} \exp\left(-\frac{\tau(\epsilon)^2 y_{jk}^2}{2\epsilon^2(\epsilon^2+\tau(\epsilon)^2)}\right) \\ &\geq \frac{1}{w(\epsilon)} t_\epsilon^{-1/2} \exp\left(-\frac{m^2 t_\epsilon^2}{2\epsilon^2}\right) \\ &\geq \epsilon^{\frac{m^2}{2}-\frac{1}{2}} \frac{1}{w(\epsilon)} (\log(1/\epsilon))^{-1/4}.\end{aligned}$$

If $q_2 > \frac{m^2-1}{2}$, for ϵ small enough, $\eta_{jk} \geq 1$ and $\breve{\beta}_{jk} = 0$. So, $\breve{f}_\epsilon \in \mathcal{E}(mt_\epsilon, 0)$.
If $q_2 \geq \frac{m^2+1}{2}$, for ϵ small enough, $\eta_{jk} \geq t_\epsilon$ and $\frac{b_j}{1+\eta_{jk}} \leq t_\epsilon$. So, $\tilde{f}_\epsilon \in \mathcal{E}(mt_\epsilon, 1/2)$ for $\epsilon < 1$. □

Now let us prove the theorem :
Proof of Theorem 4.10 : The inclusion

$$MS(f_\epsilon^0, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) \subset \mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W(\frac{2}{2s+1}, 2)$$

is a direct consequence of Proposition 4.5 and Theorems 4.1 and 4.2.
Now, let us prove that

$$\mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W(\frac{2}{2s+1}, 2) \subset MS(f_\epsilon^0, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}).$$

For this purpose, let us prove (4.6). Let us fix a constant $M \geq \sqrt{6+4q_1}$. We assume $|y_{jk}| > Mt_\epsilon$. Then, for ϵ small enough,

$$\begin{aligned}\eta_{jk} &= \frac{1}{w(\epsilon)} \frac{\sqrt{\epsilon^2+\tau(\epsilon)^2}}{\epsilon} \exp\left(-\frac{\tau(\epsilon)^2 y_{jk}^2}{2\epsilon^2(\epsilon^2+\tau(\epsilon)^2)}\right) \\ &\leq \frac{1}{w(\epsilon)} \frac{\sqrt{\epsilon^2+\tau(\epsilon)^2}}{\epsilon} \epsilon^{\frac{M^2}{4}} \\ &\leq \frac{1}{w(\epsilon)} t_\epsilon^{-1/2} \epsilon^{\frac{M^2}{4}} \\ &\leq t_\epsilon.\end{aligned}$$

Let us prove (4.6) for $\breve{\beta}_{jk}$. Using the previous inequality, we have for ϵ small enough, and for any $j < j_\epsilon$ and any k,

$$\epsilon\sqrt{b_j}\Phi^{-1}\left(\frac{1+\min(\eta_{jk},1)}{2}\right) \leq t_\epsilon.$$

So,

$$\begin{aligned} |y_{jk} - \breve{\beta_{jk}}| &= |y_{jk} - \breve{\beta_{jk}}| \mathbb{1}\{|y_{jk}| > Mt_\epsilon\} + |y_{jk} - \breve{\beta_{jk}}| \mathbb{1}\{|y_{jk}| \leq Mt_\epsilon\} \\ &\leq ((1-b_j)|y_{jk}| + t_\epsilon) \mathbb{1}\{|y_{jk}| > Mt_\epsilon\} + 2|y_{jk}| \mathbb{1}\{|y_{jk}| \leq Mt_\epsilon\} \\ &\leq t_\epsilon |y_{jk}| + (1+2M)t_\epsilon, \end{aligned}$$

which implies the required inequality. Now, let us deal with the posterior mean. For ϵ small enough, and for any $j < j_\epsilon$ and any k,

$$\begin{aligned} |y_{jk} - \tilde{\beta_{jk}}| &= |y_{jk} - \tilde{\beta_{jk}}| \mathbb{1}\{|y_{jk}| > Mt_\epsilon\} + |y_{jk} - \tilde{\beta_{jk}}| \mathbb{1}\{|y_{jk}| \leq Mt_\epsilon\} \\ &\leq \left(1 - \frac{b_j}{1+\eta_{jk}}\right) |y_{jk}| \mathbb{1}\{|y_{jk}| > Mt_\epsilon\} + 2|y_{jk}| \mathbb{1}\{|y_{jk}| \leq Mt_\epsilon\} \\ &\leq (1 - b_j + \eta_{jk})|y_{jk}| \mathbb{1}\{|y_{jk}| > Mt_\epsilon\} + 2|y_{jk}| \mathbb{1}\{|y_{jk}| \leq Mt_\epsilon\} \\ &\leq 2t_\epsilon |y_{jk}| + 2Mt_\epsilon, \end{aligned}$$

which implies (4.6) for the posterior mean.
Now, using Proposition 4.5 and Theorem 4.5, we obtain the required inclusion. □
So, Theorem 4.10 provides optimal Bayesian procedures among limited and elitist procedures, based on Gaussian priors, under the condition that the hyperparameter $\tau_{j,\epsilon}$ is "large". Under this assumption, the density $\gamma_{j,\epsilon}$ is then more spread around 0, which enables us to avoid considering heavy-tailed densities. Since the maxiset of these estimates is the intersection of the Besov space $\mathcal{B}^{s/(2s+1)}_{2,\infty}$ and the Lorentz space $W(\frac{2}{2s+1}, 2)$, they achieve the same performance as thresholding ones.

4.6 Simulations

Dealing with the prior model (4.1), we compare in this section the performances of the two Bayesian rules described in (4.10) and (4.11), where the prior is a Gaussian density with a large variance (see Theorem 4.10) with thresholding rules of Donoho and Johnstone called VisuShrink, GlobalSure (Nason (1996[92])) as well as the Bayesian thresholding procedures of Abramovich et al. (1998[4]) denoted as BayesThresh. For this purpose, we use the mean-squared error. But before this, let us precise our statistical model.

4.6.1 Model and discrete wavelet transform

Let us consider the standard regression problem :

$$g_i = f(\frac{i}{n}) + \sigma\epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0,1), \quad 1 \leq i \leq n, \tag{4.13}$$

where $n = 1024$. We introduce the discrete wavelet transform (denoted DWT) of the vector $f^0 = (f(\frac{i}{n}), \quad 1 \leq i \leq n)^T$:

$$d := \mathcal{W} f^0.$$

The DWT matrix $\mathcal{W}$ is orthogonal. Therefore, we can reconstruct f^0 by the relation

$$f^0 = \mathcal{W}^T d.$$

These transformations performed by Mallat's fast algorithm require only $O(n)$ operations (see Mallat (1998[85])). The DWT provides n discrete wavelet coefficients d_{jk}, $-1 \leq j \leq N-1, k \in \mathcal{I}_j$. They are related to the wavelet coefficients β_{jk} of f by the simple relation

$$d_{jk} \approx \beta_{jk} \times \sqrt{n}.$$

Using the DWT, the regression model (4.13) is reduced to the following one :

$$y_{jk} = d_{jk} + \sigma z_{jk}, \quad -1 \leq j \leq N-1, \quad k \in \mathcal{I}_j,$$

where

$$y := (y_{jk})_{j,k} = \mathcal{W} g$$

and

$$z := (z_{jk})_{j,k} = \mathcal{W}\epsilon.$$

Since $\mathcal{W}$ is orthogonal, z is a vector of independent $\mathcal{N}(0,1)$ variables. Now, instead of estimating f, we estimate the d_{jk}'s.

We suppose in the following that σ is known. Nevertheless, it could robustly be estimated by the median absolute deviation of the $(d_{N-1,k})_{k \in \mathcal{I}_{N-1}}$ divided by 0.6745 (see Donoho and Johnstone (1994[43])).

For the reconstruction of d_{jk}'s, we used the posterior median and the posterior mean of a prior having the following form :

$$d_{jk} \sim \frac{\omega_n}{1+\omega_n}\,\gamma_{j,n} + \frac{1}{1+\omega_n}\,\delta(0),$$

where $\omega_n = \omega^* = 10(\frac{\sigma}{\sqrt{n}})^q$ $(q > 0)$, $\delta(0)$ is a point mass at zero and γ is assumed to be the Gaussian density and

$$\gamma_{j,n}(d_{jk}) = \frac{1}{\tau_n}\,\gamma(\frac{d_{jk}}{\tau_n}),$$

with τ_n is such that $\frac{n\tau_n^2}{\sigma^2+n\tau_n^2} = 0,999$.

Dealing with this prior model, we respectively denote *GaussMedian* and *GaussMean*, the two Bayesian rules described in (4.10) and (4.11).

The Symmlet 8 wavelet basis (as described on page 198 of Daubechies (1992[34]) is used for all the methods of reconstruction. In Table 4.1 we measure the performances of the two estimators by using the four test functions : "Blocks", "Bumps", "Heavisine" and "Doppler" thanks to the mean-squared error defined by :

$$\mathrm{MSE}(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}(\frac{i}{n}) - f(\frac{i}{n})\right)^2.$$

Remark 4.12. *Recall that the test functions functions have been chosen by Donoho and Johnstone (1994[43]) to represent a large variety of inhomogeneous signals.*

4.6.2 Simulations and discussion

Table 4.1 shows the average mean-squared error (denoted AMSE) using 100 replications for VisuShrink, GlobalSure, BayesThresh, GaussMedian and GaussMean (for $q = 1$) with different values for the root signal to noise ration (RSNR).

RSNR=5	Blocks	Bumps	Heavisine	Doppler
VisuShrink	2.08	2.99	0.17	0.77
GlobalSure	0.82	0.92	0.18	0.59
BayesThresh	0.67	0.74	**0.15**	0.30
GaussMedian	0.72	0.76	0.20	0.30
GaussMean	**0.62**	**0.68**	0.19	**0.29**
RSNR=7	Blocks	Bumps	Heavisine	Doppler
VisuShrink	1.29	1.77	0.12	0.47
GlobalSure	0.42	0.48	0.12	0.21
BayesThresh	0.38	0.45	**0.10**	0.16
GaussMedian	0.41	0.42	0.12	0.15
GaussMean	**0.35**	**0.38**	0.11	**0.15**
RSNR=10	Blocks	Bumps	Heavisine	Doppler
VisuShrink	0.77	1.04	0.08	0.27
GlobalSure	0.25	0.29	0.08	0.11
BayesThresh	0.22	0.25	0.06	0.09
GaussMedian	0.21	0.23	0.06	0.08
GaussMean	**0.18**	**0.20**	**0.06**	**0.07**

TAB. 4.1 – AMSEs pour VisuShrink, GlobalSure, BayesThresh, GaussMedian and GaussMean with various test functions and various values of the RSNR.

According to Table 4.1, we remark that "purely Bayesian" procedures (BayesThresh, GaussMedian and GaussMean) are preferable to "purely deterministic" ones (VisuShrink and GlobalSure) under the AMSE approach for inhomogeneous signals. Looking at this Table, we note that GaussMedian and GaussMean often outperform the others procedures. In particular GaussMean constitutes the best procedures considered here since its AMSEs are globally the smallest (10 times on 12). Although the performances of GaussMedian are worse than BayesThresh for large σ (RNSR ≤ 5) they are better when σ is small (RNSR ≥ 7).

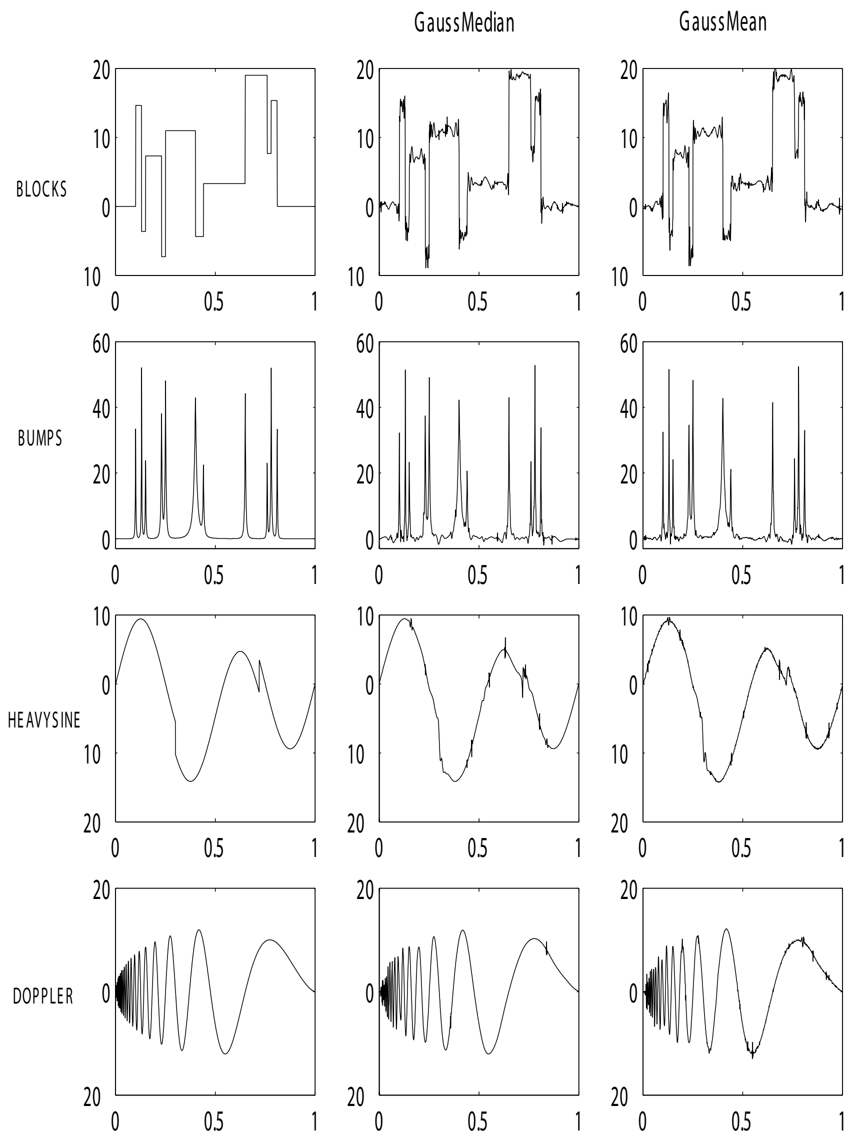


FIG. 4.1 – Original test functions and reconstructions using GaussMedian and GaussMean with $q = 1$ (RSNR=5).

In Figure 4.1, we note that in our two Bayesian procedures high-frequency some artefacts appear. However, these artefacts disappear if we take large values of q. Figure 4.2 show an example of reconstructions using GaussMedian and GaussMean when the RSNR is equal to 5 ($\sigma = 7/5$) for different values of q.

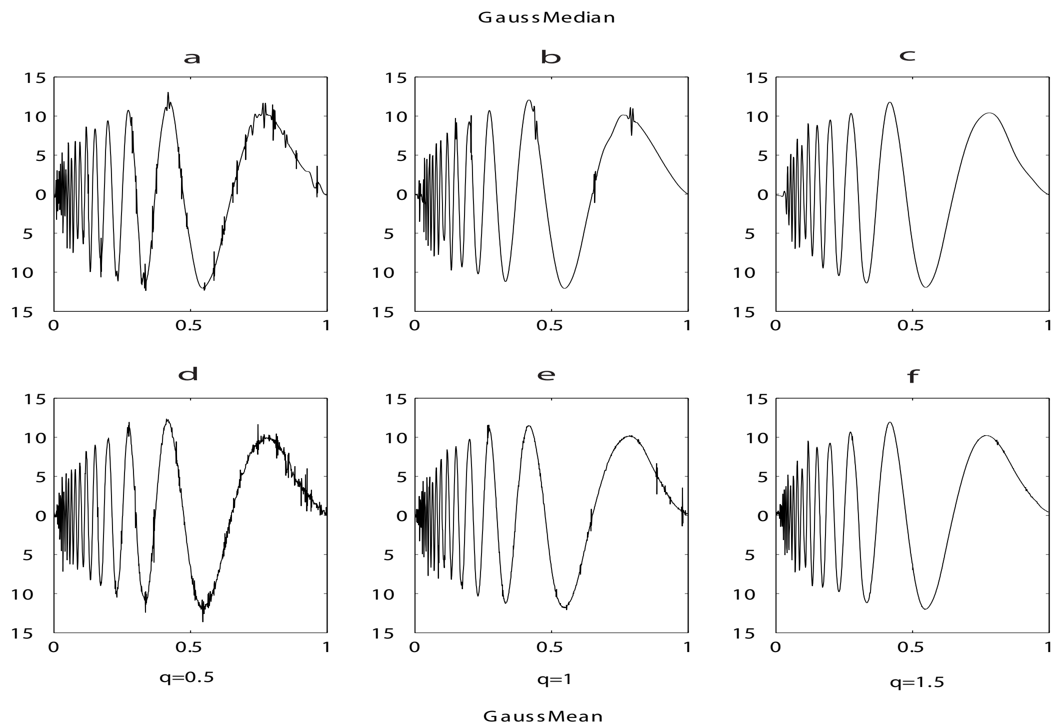


FIG. 4.2 – Reconstructions with GaussMedian (schémas a,b et c) and GaussMean (schémas d,e et f) for various values of q when RSNR=5 ; a : AMSE=0.37. b : AMSE=0.30. c : AMSE=0.33. d : AMSE=0.39. e : AMSE=0.29. f : AMSE=0.30.

As we can see in Figure 4.2, the artefacts are less numerous when q increases . But this improvement has a cost : in general the AMSE increases when q is around 0 or strictly greater than 1. Consequently, the value $q = 1$ appears as a good compromise to obtain good reconstructions and good AMSE with the GaussMedian and GaussMean procedures.

4.7 Appendix

In the previous sections, for sake of simplicity, the choice of the rates of convergence was often restricted. Indeed, the rate was linked in a direct way to either the limitation or to the threshold bound for elitist or cautious rules. But generally, it is not necessary and we show in this section how this constraint can be relaxed.

Maxisets for limited rules.

Definition 4.7. *Let $s > 0$ and u be an increasing continuous map of $\mathbb{R}_+$ such that $u(0) = 0$. We shall say that a function $f \in \mathbb{L}_2([0,1])$ belongs to the space $\mathcal{B}^s_{2,\infty}(u)$, if and only if :*

$$\sup_{\lambda>0}(u(\lambda))^{-2s}\sum_j\sum_k \beta_{jk}^2 \mathbb{1}\ \{2^{-j} \le \lambda\}\ < \infty.$$

Of course, when $u(x) = x$, $\mathcal{B}^s_{2,\infty}(u)$ is the classical Besov space $\mathcal{B}^s_{2,\infty}$. In this section, we study the ideal maxisets for limited procedures. We also provide estimates that are optimal among the class of limited ones. For this purpose, let λ_ε be a increasing continuous function with $\lambda_0 = 0$,

Théorème 4.1 (Ideal maxiset for limited rules)**.** *Let $s > 0$ and $\hat{f}_\epsilon$ be a limited rule belonging to $\mathcal{L}(\lambda_\epsilon, a)$, with $a \in [0, 1[$. Then*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{2s}) \subset \mathcal{B}^s_{2,\infty}(u).$$

Proof of Theorem 4.1 : Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{2s})$. We have :

$$\begin{aligned}
& (1-a)^2 \sum_{j,k} \beta_{jk}^2 \mathbb{1}\ \{2^{-j} \le \lambda_\epsilon\} \\
= \ & 2(1-a)^2 \sum_{j,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0)\mathbb{1}\ \{\beta_{jk} \ge 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0)\mathbb{1}\ \{\beta_{jk} < 0\}\right] \mathbb{1}\ \{2^{-j} \le \lambda_\epsilon\} \\
\le \ & 2\mathbb{E} \sum_{j,k} \left[(\gamma_{jk} y_{jk} - \beta_{jk})^2 \mathbb{1}\ \{\beta_{jk} \ge 0\} + (\gamma_{jk} y_{jk} - \beta_{jk})^2 \mathbb{1}\ \{\beta_{jk} < 0\}\right] \mathbb{1}\ \{2^{-j} \le \lambda_\epsilon\} \\
\le \ & 2\mathbb{E} \sum_{j,k} (\gamma_{jk} y_{jk} - \beta_{jk})^2 \\
\le \ & C\ (u(\lambda_\epsilon))^{2s},
\end{aligned}$$

where C is a positive constant. So, f belongs to $\mathcal{B}^s_{2,\infty}(u)$. □

Conversely, we have the following result :

Théorème 4.2. *Let $s > 0$ and $(\gamma_j(\epsilon))_{jk}$ be a non increasing sequence of weights lying in $[0,1]$ such that $\hat{\beta}^L_\epsilon = (\gamma_j(\epsilon)y_{jk})_{jk}$ belongs to $\mathcal{L}(\lambda_\epsilon, a)$, with $a \in [0,1[$. If there exist C_1 and C_2 in $\mathbb{R}$ such that, with $\gamma_{-2} = 1$, $\forall\ \epsilon > 0$,*

$$\sum_{j\geq -1}(\gamma_{j-1}-\gamma_j)(1-\gamma_j)(u(2^{-j}))^{2s}\mathbb{1}\ \{2^j < \lambda_\epsilon^{-1}\} \leq C_1(u(\lambda_\epsilon))^{2s} \tag{4.14}$$

$$\sum_{j\geq -1} 2^j\gamma_j(\epsilon)^2 \leq C_2\ \epsilon^{-2}\ (u(\lambda_\epsilon))^{2s} \tag{4.15}$$

then,

$$\mathcal{B}^s_{2,\infty}(u) \subset MS(\hat{\beta}^L_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{2s}).$$

Proof of Theorem 4.2 : With $s_l = \sum_j\sum_k \beta^2_{jk}\mathbb{1}\ \{2^{-j} \leq 2^{-l}\}$, we have, using (4.14) and (4.15) :

$$\begin{aligned}
\sum_{j,k}\mathbb{E}(\gamma_j y_{jk}-\beta_{jk})^2 &= \sum_{j,k}\mathbb{E}(\gamma_j(y_{jk}-\beta_{jk})-(1-\gamma_j)\beta_{jk})^2\\
&= \sum_{j,k}\gamma_j^2\epsilon^2 + \sum_{j,k}(1-\gamma_j)^2\beta^2_{jk}\\
&\leq S_\psi\epsilon^2\sum_j 2^j\gamma_j^2 + \sum_{j,k}\beta^2_{jk}\mathbb{1}\ \{2^{-j}\leq\lambda_\epsilon\} + \sum_{j,k}(1-\gamma_j)^2\beta^2_{jk}\mathbb{1}\ \{2^{-j}>\lambda_\epsilon\}\\
&\leq (S_\psi C_2 + M'^2)\ (u(\lambda_\epsilon))^{2s} + \sum_{j\geq -1}(1-\gamma_j)^2(s_j - s_{j+1})\mathbb{1}\ \{2^{-j}>\lambda_\epsilon\}\\
&\leq (S_\psi C_2 + M'^2)\ (u(\lambda_\epsilon))^{2s} + 2\sum_{j\geq -1}(\gamma_{j-1}-\gamma_j)(1-\gamma_j)s_j\mathbb{1}\ \{2^{-j}>\lambda_\epsilon\}\\
&\leq (S_\psi C_2 + M'^2)\ (u(\lambda_\epsilon))^{2s} + 2M'^2\sum_{j\geq -1}(\gamma_{j-1}-\gamma_j)(1-\gamma_j)(u(2^{-j}))^{2s}\mathbb{1}\ \{2^{-j}>\lambda_\epsilon\}\\
&\leq (S_\psi C_2 + M'^2 + 2M'^2C_1)(u(\lambda_\epsilon))^{2s}.
\end{aligned}$$

□

Combining Theorems 4.1 and 4.2, by straightforward computations, we obtain :

Corollary 4.4. *If we assume that* $u(x) = x\tilde{u}(x)$ *where*

$$\tilde{u}(x)^{-1} = O(1) \text{ as } x \text{ goes to } 0$$

and if we consider linear estimates associated with the weights $\gamma_j^{(1)}(\lambda_\epsilon)$, $\gamma_j^{(2)}(\lambda_\epsilon)$ *with* $\alpha > (s \vee 1/2)$ *or* $\gamma_j^{(3)}(\lambda_\epsilon)$ *with* $\alpha > s$ *(see section 4.2.2), then for* $i \in \{1,2,3\}$

$$MS((\gamma_j^{(i)}(\lambda_\epsilon)y_{jk})_{jk}, \|.\|_2^2, (u(\lambda_\epsilon))^{2s}) = \mathcal{B}^s_{2,\infty}(u),$$

as soon as $(\epsilon^2\lambda_\epsilon^{-1}u(\lambda_\epsilon)^{-2s})_\epsilon$ *is bounded.*

To shed light on this result, let us take $\lambda_\epsilon = \epsilon^{2/(1+2s)}$. So, $(\epsilon^2\lambda_\epsilon^{-1}u(\lambda_\epsilon)^{-2s})_\epsilon$ is bounded as soon as $(\epsilon^{4s/(1+2s)}\, u(\lambda_\epsilon)^{-2s})_\epsilon$ is bounded. So, for the rate $\epsilon^{4s/(1+2s)}(\log(1/\epsilon))^{2sm}$, $m \geq 0$, the maxisets of the linear estimates mentioned in Corollary 4.4 are the spaces $\mathcal{B}^s_{2,\infty}(u)$, where $u(x) = x(\log(1/x))^m$.

Maxisets for elitist rules.

Definition 4.8. *Let* $0 < r < 2$ *and* u *be an increasing continuous map of* $\mathbb{R}_+$ *such that* $u(0) = 0$. *We shall say that a function* $f \in \mathbb{L}_2([0,1])$ *belongs to the space* $W_u(r,2)$ *if and only if :*

$$\sup_{\lambda>0}(u(\lambda))^{r-2}\sum_j\sum_k |\beta_{jk}|^2 I_{\{|\beta_{jk}|\leqslant\lambda\}} < \infty.$$

Théorème 4.3 (Ideal maxiset for elitist rules)**.** *Let* $s > 0$ *and* $\hat{f}_\epsilon$ *be an elitist rule that belongs to* $\mathcal{E}(\lambda_\epsilon, a)$ *with* $a \in [0,1[$, *where* λ_ϵ *is an increasing continuous function of* ϵ, *such that* $\lambda_0 = 0$. *Then*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{4s/(1+2s)}) \subset W_u(\frac{2}{1+2s}, 2).$$

Proof of Theorem 4.3 : Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{4s/(1+2s)})(M)$. We have :

$$
\begin{aligned}
& (1-a)^2 \sum_{j,k} \beta_{jk}^2 \mathbb{1}\ \{|\beta_{jk}| \le \lambda_\epsilon\} \\
= \ & 2(1-a)^2 \sum_{j,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0)\mathbb{1}\ \{\beta_{jk} \ge 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0)\mathbb{1}\ \{\beta_{jk} < 0\}\right] \mathbb{1}\ \{|\beta_{jk}| \le \lambda_\epsilon\} \\
\le \ & 2\mathbb{E} \sum_{j,k} \left[(\beta_{jk} - \gamma_{jk} y_{jk})^2 \mathbb{1}\ \{\beta_{jk} \ge 0\} + (\beta_{jk} - \gamma_{jk} y_{jk})^2 \mathbb{1}\ \{\beta_{jk} < 0\}\right] \mathbb{1}\ \{|\beta_{jk}| \le \lambda_\epsilon\} \\
\le \ & 2\mathbb{E} \sum_{j,k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 \\
\le \ & 2M\ (u(\lambda_\epsilon))^{4s/(1+2s)}.
\end{aligned}
$$

So, using the continuity of λ_ϵ in 0, we deduce that $f \in W_u(\frac{2}{1+2s}, 2)$.

Maxisets for cautious rules.

Definition 4.9. *Let $0 < r < 2$ and u be a increasing continuous map of $\mathbb{R}_+$ such that $u(0) = 0$. We shall say that a function $f \in \mathbb{L}_2([0,1])$ belongs to the space $W_u^*(r, 2)$ if and only if :*

$$\sup_{\lambda > 0} (u(\lambda))^{r-2} \lambda^2 \left[\log(\frac{1}{\lambda})\right]^{-1} \sum_{j < j_\lambda, k} I_{\{|\beta_{jk}| > \lambda\}} < \infty.$$

Théorème 4.4 (Ideal maxiset for cautious rules)**.** *Let $s > 0$ and $\hat{f}_\epsilon$ be a cautious rule that belongs to $\mathcal{C}(\lambda_\epsilon, a)$ with $a \in]0,1]$. Let λ_ε be an increasing continuous function with $\lambda_0 = 0$ such that :*

$$\exists\ c > 0, \quad \forall\ \epsilon > 0,\ \frac{\lambda_\epsilon}{\sqrt{\log(\frac{1}{\lambda_\epsilon})}} \le c\epsilon. \tag{4.16}$$

Then

$$MS(\hat{f}_\epsilon, \|.\|_2^2, u(\lambda_\epsilon))^{4s/(1+2s)} \subset W_u^*(\frac{2}{1+2s}, 2).$$

Remark 4.13. *Note that the case $\lambda_\epsilon = t_\epsilon$ (resp. $\lambda_\epsilon = \epsilon$) satisfies (4.16) with $c = \sqrt{2}$ (resp. $c = 1$)*

Proof of Theorem 4.4 : Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, u(\lambda_\epsilon))^{4s/(1+2s)})(M)$. Using (4.16),

$$a^2\lambda_\epsilon^2 \left[\log(\frac{1}{\lambda_\epsilon})\right]^{-1} \sum_{j<j_\epsilon,k} \mathbb{1}\ \{|\beta_{jk}| > \lambda_\epsilon\} \leq a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} \mathbb{1}\ \{|\beta_{jk}| > \lambda_\epsilon\}.$$

Now, let us recall that if X is a zero-mean Gaussian variable with variance ϵ^2, then

$$\mathbb{E}(X^2 I_{\{X<0\}}) = \mathbb{E}(X^2 I_{\{X>0\}}) = \frac{\epsilon^2}{2}.$$

From Lemma 4.1,

$$\begin{aligned}
& a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} \mathbb{1}\ \{|\beta_{jk}| > \lambda_\epsilon\} \\
= \ & a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} [\mathbb{1}\ \{\beta_{jk} > \lambda_\epsilon\} + \mathbb{1}\ \{\beta_{jk} < -\lambda_\epsilon\}] \\
= \ & 2a^2c^2\ \mathbb{E} \sum_{j<j_\epsilon,k} (\beta_{jk} - y_{jk})^2 [\mathbb{1}\ \{y_{jk} - \beta_{jk} < 0\}\mathbb{1}\ \{\beta_{jk} > \lambda_\epsilon\} + \mathbb{1}\ \{y_{jk} - \beta_{jk} > 0\}\mathbb{1}\ \{\beta_{jk} < -\lambda_\epsilon\}] \\
\leq \ & 8c^2\ \mathbb{E} \sum_{j<j_\epsilon,k} (\beta_{jk} - \gamma_{jk}y_{jk})^2 \\
\leq \ & 8c^2M\ (u(\lambda_\epsilon))^{4s/(1+2s)}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that f belongs to $W_u^*(\frac{2}{1+2s}, 2)$. □

Up to now the largest maxiset that we encountered is of the form $\mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W(\frac{2}{2s+1}, 2)$, when dealing with the rate $t_\epsilon^{4s/(1+2s)}$. A natural question arises here. Does there exist a non linear procedure that outperforms the thresholding procedures in terms of maxiset comparisons ? The purpose of the following chapter is to prove that the answer to this question is YES and provide examples of procedures yielding larger maxisets. By making use of the dyadic structure of the wavelet bases (which has not been used before in fact) and performing algorithm with tree properties, we can prove that this provides a first way of enlarging the maxisets.

Chapitre 5

Hereditary rules and Lepski's procedure

Summary : In this chapter we focus on a new large class of procedures, called *hereditary rules*. Based on tree structure, these procedures are proved to outperform elitist rules in the maxiset sense. In particular, we exhibit an optimal hereditary estimator (hard tree rule) having some connections with the procedure of Lepski (1991[78]). Then, we compare it to the hybrid version of Lepski's procedure proposed by Picard and Tribouley (2000[99]), assuming that the wavelet basis is the Haar one.

5.1 Introduction and model

In the previous chapter, we have shown that thresholding rules and many Bayesian procedures achieve the same performance under the maxiset approach. Precisely, the maximal space where these procedures attain the rate $(\epsilon\sqrt{\log(\epsilon^{-1})})^{4s/(1+2s)}$ was proved to be the intersection of the Besov space $\mathcal{B}_{2,\infty}^{s/(1+2s)}$ and the Lorentz space $W(\frac{2}{1+2s}, 2)$. Up to now, this maxiset constitutes the largest maxiset we encountered dealing with non random thresholds. The aim of this chapter is to prove the existence of adaptive rules for which the maxiset is larger than this intersection of Besov spaces.

The first part of the paper (sections 5.2 and 5.3) deals with a sub-class of cautious rules : the hereditary rules. Analogously to the previous chapter, we provide a functional space which contains all the maximal spaces associated with such rules. Then, we exhibit two examples of hereditary rules which are optimal in the maxiset sense. These shrinkage procedures, called respectively the hard tree rule and the soft tree rule, are based on thresholding properties combined with heredity constraints (in the sense of Engel (1994[51])).

In the second part of the paper (section 5.4), we show that the hard tree rule is connected to the local bandwidth selection's procedure of Lepski (1991[78]) when the wavelet basis considered for the reconstruction is the Haar one. Then, we compare this procedure with the hybrid version of the Lepski's procedure which has been proposed by Picard and Tribouley (2000[99]) for the construction of adaptive confidence intervals. We prove that the maximal space where these two procedures attain the rate $(\epsilon\sqrt{\log(\epsilon^{-1})})^{4s/(1+2s)}$ for the $\mathbb{L}_2$-risk, is larger than the one of any elitist estimator (including hard and soft thresholding rules). This result is closely akin to the one of Kerkyacharian and Picard (2002[76]), who prove by the way of oracle inequalities that maxisets of local bandwidth selection procedures are larger than thresholding procedures.

Let us notice that although the results presented here emphasize in a direct way to the threshold bound for hereditary rules, there is no doubt that similar results could be easily obtained when relaxing this constraint.

The model is the following : we will consider a white noise setting : $X_\epsilon(.)$ is a random measure satisfying on $[0,1[$ the following equation :

$$X_\epsilon(dt) = f(t)dt + \epsilon W(dt)$$

where

- $0 < \epsilon < 1/e$ is the noise level,
- f is a function defined on $[0,1]$,
- $W(.)$ is a Brownian motion on $[0,1]$.

Let $\{\psi_{jk}(\cdot),\ j \geq -1, k \in \mathbb{N}\}$ be a compactly supported wavelet basis of $\mathbb{L}_2([0,1])$. $f \in \mathbb{L}_2([0,1])$ can be represented as :

$$f = \sum_{j\geq -1}\sum_k \beta_{jk}\psi_{jk} = \sum_{j\geq -1}\sum_k (f,\psi_{jk})_{\mathbb{L}_2}\psi_{jk}. \tag{5.1}$$

At each level $j \geq 0$, the number of non-zero wavelet coefficients is smaller than or equal to $2^j + l_\psi - 1$, where l_ψ is the maximal size of the supports of the scaling function and the wavelet. So, there exists a constant S_ψ such that at each level $j \geq -1$, there are less than or equal to $S_\psi \times 2^j$.

Let us suppose that we dispose of observations : $y_{jk} = X_\epsilon(\psi_{jk}) = \beta_{jk} + \epsilon Z_{jk}$ where Z_{jk} are independent Gaussian variables $\mathcal{N}(0,1)$.

In the sequel we shall say that I is a dyadic interval if and only if $I = I_{jk} = Support(\psi_{jk})$, for some j and some k. In this case, we shall note y_I (resp. β_I) instead of y_{jk} (resp. β_{jk}) and we shall set $|I| = l_\psi 2^{-j}$, its length.

Along the chapter, we set $2^{j_\lambda} \sim \lambda^{-2}$ to design the integer j_λ such that $2^{-j_\lambda} \leq \lambda^2 < 2^{1-j_\lambda}$ and we denote for any ϵ, $t_\epsilon := \epsilon\sqrt{\log(\epsilon^{-1})}$.

5.2 Hereditary rules

This section aim at studying the maxiset a new class of procedures : the hereditary rules. As in the previous chapter, we firstly point out the ideal maxiset of this class for the rate $(\epsilon\sqrt{\log(\epsilon^{-1})})^{4s/(1+2s)}$ (Theorem 5.1). Then we give sufficient conditions over hereditary procedures to ensure that their maxiset is the ideal one and we propose two examples of such rules (Theorem 5.2).

5.2.1 Definitions

Definition 5.1. *Let $\lambda > 0$ and I_{jk} be a dyadic interval such that $0 \leq j < j_\lambda$. We denote $\mathcal{T}_{jk}(\lambda)$ the binary tree containing the set of dyadic intervals such that the following properties are satisfied :*

- $I_{jk} \in \mathcal{T}_{jk}(\lambda)$.
- $I \in \mathcal{T}_{jk}(\lambda) \Longrightarrow I \subset I_{jk}$ *and* $|I| > l_\psi \lambda^2$.
- *Two distinct dyadic intervals of $\mathcal{T}_{jk}(\lambda)$ with same length have their interiors disjointed.*
- *The numbers of dyadic intervals of $\mathcal{T}_{jk}(\lambda)$ of length $l_\psi 2^{-j'}$ ($j \leq j' < j_\lambda$) is equal to* $2^{j'-j}$

– *Any set of all dyadic intervals of $\mathcal{T}_{jk}(\lambda)$ with same length is forming a partition of I_{jk}.*

Let us now introduce the following class of procedures :

Definition 5.2. *Let $\hat{f}_\epsilon \in \mathcal{F}_\epsilon$ (see paragraphe 4.2.2). We say that $\hat{f}_\epsilon$ is a* **hereditary rule** *if there exist a determinist function of ϵ, λ_ϵ, and a constant $a \in [0,1[$ such that for any $0 \leq j < j_\epsilon$ and any k*

$$\gamma_{jk} > a \Longrightarrow \exists I \in \mathcal{T}_{jk}(\lambda_\epsilon) \text{ such that } |y_I| > \lambda_\epsilon, \tag{5.2}$$

where $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$. In the sequel, we note $\hat{f}_\epsilon \in \mathcal{H}(\lambda_\epsilon, a)$.

Some examples are given in the paragraph 5.3.2.

Remark 5.1. *As the limited rules and the elitist rules, the hereditary rules are forming a non decreasing class with respect to a. Undoubtedly any hereditary rule $\hat{f}_\epsilon$ belonging to $\mathcal{H}(\lambda_\epsilon, a)$ is a cautious rule belonging to $\mathcal{C}(\lambda_\epsilon, a)$.*

5.2.2 Functional spaces

In this paragraph, we will prove that the maximal space of any hereditary rule is necessarily smaller than a simple functional class. For sake of simplicity, we shall restrict to the case where ρ is the square of the $\mathbb{L}_2$ norm, even though if a large majority of the following results can be extended to more general norms.

Let us define the functional spaces which shall play an important role in the sequel.

Definition 5.3. *Let $s > 0$. We say that a function $f \in \mathbb{L}_2([0,1])$ belongs to the Besov space $\mathcal{B}^s_{2,\infty}$, if and only if :*

$$\sup_{J \geq -1} 2^{2Js} \sum_{j \geq J} \sum_k \beta_{jk}^2 < \infty.$$

We denote by $\mathcal{B}^s_{2,\infty}(R)$ the ball of radius R in this space.

In chapter 4, we have shown that Besov spaces naturally appear when dealing with the maxisets of limited rules.

Definition 5.4. *Let $0 < r < 2$. We say that a function f belongs to the weak Besov space $W(r,2)$ if and only if :*

$$\sup_{\lambda>0} \lambda^{r-2} \sum_{\leq j\geq 0} \sum_{k} \beta_{jk}^2 \mathbf{1}\{|\beta_{jk}| \leq \lambda\} \ < \infty.$$

We denote by $W(r,2)(R)$ the ball of radius R in this space.

Weak Besov spaces compose a sub-family of Lorentz spaces (see Lorentz (1950[81], 1966[82]) or DeVore and Lorentz (1993[38])). Many results in approximation theory deals with weak Besov spaces (see DeVore (1989[50]), DeVore et Lorentz (1993[38]), DeVore, Konyagin and Temlyakov (1998[37])). In chapter 4, we have shown that weak Besov spaces naturally appear when dealing with the maxisets of elitist rules.

As far as the hereditary rules are concerned, we shall see in the next paragraph that the maxisets of such procedures are always contained in large functional spaces : the tree-Besov spaces.

Definition 5.5. *Let $0 < r < 2$. We say that a function f belongs to the tree-Besov space $W^{T}(r,2)$ if and only if :*

$$\|f\|_{W_r^T} := [\sup_{\lambda>0} \lambda^{r-2} \sum_{0\leq j<j_\lambda} \sum_{k} \beta_{jk}^2 \mathbb{1}\ \{\forall I \in \mathcal{T}_{jk}(\lambda),\ |\beta_I| \leq \frac{\lambda}{2}\}]^{1/2} \ < \infty.$$

We denote by $W^{T}(r,2)(R)$ the ball of radius R in this space.

Remark 5.2. *Obviously, $W(r,2) \subset W^{T}(r,2)$. These spaces taking account of the dyadic structure of the wavelet bases are very close to the oscillation spaces introduced by Jaffard (1998[60], 2004[61]).*

5.2.3 Ideal maxisets for hereditary rules

The result presented here emphasizes the cases where the rate of convergence is linked in a direct way to the threshold bound for hereditary rules. But there are many cases where either the threshold bound or the rate contain logarithmic factors. Analogously to Chapter 4 , we could easily obtain similar result when relaxing this constraint.

Théorème 5.1. *Let $\hat{f}_\epsilon$ be a hereditary rule that belongs to $\mathcal{H}(\lambda_\epsilon, a)$ with $a \in [0,1[$. Let $0 < r < 2$ be a real number and λ_ε be a non decreasing, continuous function with $\lambda_0 = 0$ such that there exists a constant $C > 0$ which satisfies for any $\epsilon > 0$,*

$$\mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}) \leq C\lambda_\epsilon^4 \tag{5.3}$$

with $Z \sim \mathcal{N}(0,1)$. Then :

$$MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r}) \subset W^T(r,2).$$

Remark 5.3. *For instance, for $\lambda_\epsilon = mt_\epsilon$, condition (5.3) is satisfied for any $m \geq 4\sqrt{2}$.*

Proof of Theorem 5.1 : Let $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$ and $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r})(M)$. Denote :

- $|\bar{y}_{jk}(\lambda_\epsilon)| := \max\{|y_I|;\ I \in \mathcal{T}_{jk}(\lambda_\epsilon)\}$,
- $|\bar{\beta}_{jk}(\lambda_\epsilon)| := \max\{|\beta_I|;\ I \in \mathcal{T}_{jk}(\lambda_\epsilon)\}$,
- $|\bar{\delta}_{jk}(\lambda_\epsilon)| := \max\{|y_I - \beta_I|;\ I \in \mathcal{T}_{jk}(\lambda_\epsilon)\}$.

We have the two following lemmas :

Lemma 5.1. *Let $\lambda > 0$ and I_{jk} be a dyadic interval such that $0 \leq j < j_\lambda$. The numbers of elements of the binary tree $\mathcal{T}_{jk}(\lambda)$ is exactly*

$$\#\mathcal{T}_{jk}(\lambda) = 2^{j_\lambda - j} - 1.$$

This lemma is easy to prove that's why we omit the proof.

Lemma 5.2. *If λ_ϵ satisfies (5.3) then, for any $0 \leq j < j_\epsilon$ and any k :*

$$\mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon)\mathbb{1}\ \{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} \leq 2C\ \lambda_\epsilon^2.$$

Proof : Let $Z \sim \mathcal{N}(0,1)$. Using Lemma 5.1, we have for any $0 \le j < j_\epsilon$ and any k :

$$\begin{aligned}
\mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon)\mathbb{1}\,\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2}\} &\le \mathbb{P}(|\bar{\delta}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon/2) \\
&\le \sum_{I \in \mathcal{T}_{jk}(\lambda_\epsilon)} \mathbb{P}(|y_I - \beta_I| > \frac{\lambda_\epsilon}{2}) \\
&\le \#\mathcal{T}_{jk}(\lambda_\epsilon)\mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}) \\
&\le 2^{j_\epsilon}\mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}) \\
&\le 2\lambda_\epsilon^{-2}\,\mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}) \\
&\le 2C\,\lambda_\epsilon^2
\end{aligned}$$

□

Now, using the fact that the rule is hereditary and Lemma 5.2 :

$$\begin{aligned}
&(1-a)^2 \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 \mathbb{1}\,\{\forall I \in \mathcal{T}_{jk}(\lambda_\epsilon),\ |\beta_I| \le \frac{\lambda_\epsilon}{2}\} \\
= \ &(1-a)^2 \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 \mathbb{1}\,\{\bar{\beta}_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2}\} \\
= \ &2(1-a)^2 \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 \,[\mathbb{P}(y_{jk}-\beta_{jk}<0)\mathbb{1}\,\{\beta_{jk}>0\} + \mathbb{P}(y_{jk}-\beta_{jk}>0)\mathbb{1}\,\{\beta_{jk}<0\}]\,\mathbb{1}\,\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2}\} \\
\le \ &2(1-a)^2\mathbb{E} \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 \,[\mathbb{1}\,\{y_{jk}-\beta_{jk}<0\}\mathbb{1}\,\{\beta_{jk}>0\} + \mathbb{1}\,\{y_{jk}-\beta_{jk}>0\}\mathbb{1}\,\{\beta_{jk}<0\}]\,\mathbb{1}\,\{|\bar{y}_{jk}(\lambda_\epsilon)| \le \lambda_\epsilon\} \\
&+2\mathbb{E} \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 \mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon)\mathbb{1}\,\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2}\} \\
\le \ &2\,\mathbb{E} \sum_{0\le j<j_\epsilon,k} (\beta_{jk}-\gamma_{jk}y_{jk})^2 \mathbb{1}\,\{|\bar{y}_{jk}(\lambda_\epsilon)| \le \lambda_\epsilon\} + \frac{\lambda_\epsilon^2}{2} \sum_{0\le j<j_\epsilon,k} \mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon\}\mathbb{1}\,\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2}\} \\
\le \ &2\,\mathbb{E} \sum_{j,k} (\beta_{jk}-\gamma_{jk}y_{jk})^2 + 2S_\psi C\lambda_\epsilon^2 \\
\le \ &2(M+S_\psi C)\,\lambda_\epsilon^{2-r}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that

$$\sup_{\lambda>0} \lambda^{r-2} \sum_{0\le j<j_\lambda,k} \beta_{jk}^2 \mathbb{1}\,\{\forall I \in \mathcal{T}_{jk}(\lambda), |\beta_I| \le \frac{\lambda}{2}\} \ \le \frac{2(M+S_\psi C)}{(1-a)^2}.$$

It comes that $f \in W^T(r,2)$. □

5.3 Optimal hereditary rules

In this section we prove conditions ensuring that the maxiset of a given shrinkage rule contains a tree-Besov space. This part is strongly linked with upper bounds inequalities in minimax theory and our technique of proof is the same as the one in paragraph 4.4.2.

5.3.1 When does the maxiset contains a tree-Besov space ?

In this paragraph, we give a converse result to Theorem 4.1 and Theorem 5.1 with respect to the ideal maxiset results for limited and hereditary rules.

Théorème 5.2. *Let $s > 0$, $m > 0$, $c > 0$ and $\gamma_{jk}(\epsilon)$ a sequence of weights lying in $[0,1]$ such that $\hat{\beta}(\epsilon) = (\gamma_{jk}(\epsilon)y_{jk})_{jk}$ belongs to $\mathcal{L}((mt_\epsilon)^2, 0) \cap \mathcal{H}(mt_\epsilon, ct_\epsilon)$. Suppose in addition that for any k, $\gamma_{-1k} = 1$ and that there exists a constant $K(\gamma)$ such that for any $\epsilon > 0$, any $0 \le j < j_\epsilon$ and any k :*

$$\max\{|y_I|;\ I \in \mathcal{T}_{jk}(mt_\epsilon)\} > mt_\epsilon \Longrightarrow (1 - \gamma_{jk}(\epsilon)) \le K(\gamma)[t_\epsilon + \frac{\varepsilon}{|y_{jk}| \vee mt_\epsilon}], \quad a.e. \tag{5.4}$$

where $2^{j_\epsilon} \sim (mt_\epsilon)^{-2}$. Then, as soon as $m \ge 4\sqrt{3}$:

$$MS(\hat{f}_\epsilon, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}) \supset \mathcal{B}_{2,\infty}^{s/(1+2s)} \cap W^T(\frac{2}{1+2s}, 2).$$

To prove this result, let us introduce the two following propositions.

Proposition 5.1. *For any $0 < r < 2$ and any $f \in \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W_r^T$, then*

$$\sup_{0<\lambda<1/e} \lambda^r \left[\log(\frac{1}{\lambda})\right]^{-1} \sum_{0 \le j < j_\lambda, k} \mathbb{1}\{\exists I \in \mathcal{T}_{jk}(\lambda),\ |\beta_I| > \frac{\lambda}{2}\} \le \frac{2^{6-r}\left(\|f\|^2_{W_r^T} + \|f\|^2_{\mathcal{B}_{2,\infty}^{(2-r)/4}}\right)}{(1-2^{-r})\log(2)} \tag{5.5}$$

Moreover, we have the following inclusion spaces :

$$W(r,2) \subset W^T(r,2) \quad and \quad \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W^T(r,2) \subset \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W^*(r,2). \tag{5.6}$$

Proof : The inclusion $W(r,2)\subset W^{T}(r,2)$ is easy to prove using the definitions of $W(r,2)$ and $W^{T}(r,2)$. The second inclusion $\mathcal{B}_{2,\infty}^{(2-r)/4}\cap W^{T}(r,2)\subset\mathcal{B}_{2,\infty}^{(2-r)/4}\cap W^{*}(r,2)$ is just a consequence of (5.5). To prove (5.5), let us introduce the following definition :

Definition 5.6. *Let $\lambda>0$ and I_{jk} be a dyadic interval such that $0\le j<j_\lambda$. We say that a dyadic interval $I_{j'k'}$ is a λ-ancestor of I_{jk} if and only if $I_{jk}\in\mathcal{T}_{j'k'}(\lambda)$.*

Let $f\in\mathcal{B}_{2,\infty}^{(2-r)/4}\cap W^{T}(r,2)$ and $0<\lambda<1/e$. We recall that $2^{j_\lambda}\sim\lambda^{-2}$ and we set for any $u\in\mathbb{N}$, $2^{j_{\lambda,u}}\sim(2^{1+u}\lambda)^{-2}$. Since for any $\lambda>0$ and any I_{jk}, $\mathcal{T}_{jk}(\lambda)$ is a binary tree, there exist at most $j+1$ λ-ancestors of I_{jk}. So,

$$
\begin{aligned}
&\sum_{0\le j<j_\lambda,k}\mathbb{1}\left\{\exists I\in\mathcal{T}_{jk}(\lambda),\ |\beta_I|>\frac{\lambda}{2}\right\}\\
\le\ &\sum_{0\le j<j_\lambda,k}(j+1)\mathbb{1}\left\{|\beta_{jk}|>\frac{\lambda}{2},\quad\forall I\in\mathcal{T}_{jk}(\lambda),I\neq I_{jk},\ |\beta_I|\le\frac{\lambda}{2}\right\}\\
\le\ &\sum_{0\le j<j_\lambda,k}(j+1)\mathbb{1}\left\{|\beta_{jk}|>\frac{\lambda}{2},\quad\forall I\in\mathcal{T}_{jk}(\lambda),\ |\beta_I|\le|\beta_{jk}|\right\}\\
\le\ &\sum_{u\ge0}\sum_{0\le j<j_\lambda,k}(j+1)\mathbb{1}\left\{|\beta_{jk}|>2^{u-1}\lambda,\quad\forall I\in\mathcal{T}_{jk}(\lambda),\ |\beta_I|\le2^u\lambda\right\}\\
\le\ &\frac{2^4}{\log(2)}\log(\frac{1}{\lambda})\sum_{u\ge0}(2^u\lambda)^{-2}\sum_{0\le j<j_\lambda,k}\beta_{jk}^2\mathbb{1}\left\{\forall I\in\mathcal{T}_{jk}(2^{1+u}\lambda),\ |\beta_I|\le2^u\lambda\right\}\\
\le\ &\frac{2^4}{\log(2)}\log(\frac{1}{\lambda})\sum_{u\ge0}(\lambda2^u)^{-2}\sum_{0\le j<j_{\lambda,u},k}\beta_{jk}^2\mathbb{1}\left\{\forall I\in\mathcal{T}_{jk}(2^{1+u}\lambda),\ |\beta_I|\le2^u\lambda\right\}\\
&+\frac{2^4}{\log(2)}\log(\frac{1}{\lambda})\sum_{u\ge0}(\lambda2^u)^{-2}\sum_{j\ge j_{\lambda,u},k}\beta_{jk}^2\\
\le\ &\frac{2^{6-r}}{(1-2^{-r})\log(2)}\left(\|f\|_{W_r^T}^2+\|f\|_{\mathcal{B}_{2,\infty}^{(2-r)/4}}^2\right)\log(\frac{1}{\lambda})\lambda^{-r}.
\end{aligned}
$$

The last inequalities use the fact that $f\in\mathcal{B}_{2,\infty}^{(2-r)/4}\cap W^{T}(r,2)$. This ends the proof of the proposition.

□

Proposition 5.2. *Under the conditions of Theorem 5.2, we have the following inequality :*

$$\mathbb{E}\|\hat{f}_\epsilon - f\|_2^2 \leq t_\epsilon^{\frac{4s}{1+2s}} \Bigg[\Big(\frac{4c^2 S_\psi}{m^2} + S_\psi\Big) + 2\Big(\frac{2}{m^2} + 1 + 2K(\gamma)^2\Big)\|f\|_2^2 + \frac{4\sqrt{6}S_\psi}{m^3} + 2(4^{\frac{4s}{1+2s}} + 1)m^{\frac{4s}{1+2s}}\|f\|^2_{W_{\frac{2}{1+2s}}} + \frac{2^{7-2/(1+2s)}m^{-2/(1+2s)}}{(1-2^{-2/(1+2s)})\log(2)}(1+8K(\gamma)^2)\Big(\|f\|^2_{W_{\frac{2}{1+2s}}} + \|f\|^2_{B^{\frac{s}{1+2s}}_{2,\infty}}\Big) + m^{\frac{4s}{1+2s}}(1 + 2\times 4^{\frac{4s}{1+2s}})\|f\|^2_{B^{\frac{s}{1+2s}}_{2,\infty}} \Bigg].$$

Proof : Obviously because of the limitation assumption, we have for $2^{j_\epsilon} \sim (mt_\epsilon)^{-2}$,

$$\mathbb{E}\|\hat{f}_\epsilon - f\|_2^2 \leq S_\psi \epsilon^2 + \mathbb{E}\|\sum_{0\leq j<j_\epsilon,k} (\gamma_{jk}(\epsilon)y_{jk} - \beta_{jk})\psi_{j,k}\|_2^2 + \sum_{j\geq j_\epsilon,k} \beta_{jk}^2.$$

The third term can be bounded by $(mt_\epsilon)^{4s/(1+2s)}\|f\|^2_{B^{\frac{s}{1+2s}}_{2,\infty}}$, by using the definition of the Besov norm.

Let us recall, for any $\lambda > 0$, the following notations :

- $|\bar{y}_{jk}(\lambda)| := \max\{|y_I|;\ I \in \mathcal{T}_{jk}(\lambda)\}$,
- $|\bar{\beta}_{jk}(\lambda)| := \max\{|\beta_I|;\ I \in \mathcal{T}_{jk}(\lambda)\}$,
- $|\bar{\delta}_{jk}(\lambda)| := \max\{|y_I - \beta_I|;\ I \in \mathcal{T}_{jk}(\lambda)\}$.

The term $\mathbb{E}\sum_{0\leq j<j_\epsilon,k}(\gamma_{jk}(\epsilon)y_{jk} - \beta_{jk})^2$ can be bounded by $2(A+B)$, where

$$\begin{aligned} A + B &= \mathbb{E}\sum_{0\leq j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)| \leq mt_\epsilon\} \\ &+ \mathbb{E}\sum_{0\leq j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} \end{aligned}$$

Again we split A into $A_1 + A_2$, and because of the condition $\mathcal{H}(mt_\epsilon, ct_\epsilon)$, we have that, on $\{|\bar{y}_{jk}(mt_\epsilon)| \leq mt_\epsilon\}$, $\gamma_{jk} \leq ct_\epsilon$. So,

$$\begin{aligned} A_1 &= \mathbb{E}\sum_{0\leq j<j_\epsilon,k}\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)| \leq mt_\epsilon\} \\ &\leq c^2 2^{j_\epsilon} S_\psi t_\epsilon^2 \epsilon^2 \\ &\leq \frac{2c^2 S_\psi}{m^2} t_\epsilon^2. \end{aligned}$$

As for the proof of Proposition 5.1, and by using lemma 5.2, we obtain

$$\begin{aligned}
A_2 &\leq \mathbb{E} \sum_{0\leq j<j_\epsilon,k} \beta_{jk}^2 \, \mathbb{1}\, \{|\bar{y}_{jk}(mt_\epsilon)| \leq mt_\epsilon\}[\mathbb{1}\, \{|\bar{\beta}_{jk}(mt_\epsilon)| \leq 2mt_\epsilon\} + \mathbb{1}\, \{|\bar{\beta}_{jk}(mt_\epsilon)| > 2mt_\epsilon\}] \\
&\leq (4mt_\epsilon)^{4s/(1+2s)}(\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{B^{s/(1+2s)}_{2,\infty}}) + \sum_{0\leq j<j_\epsilon,k} \beta_{jk}^2 \mathbb{P}(|\bar{\delta}_{jk}(mt_\epsilon)| > mt_\epsilon) \\
&\leq (4mt_\epsilon)^{4s/(1+2s)}(\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{B^{s/(1+2s)}_{2,\infty}}) + 2^{j_\epsilon}\|f\|_2^2 \epsilon^{m^2/2} \\
&\leq (4mt_\epsilon)^{4s/(1+2s)}(\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{B^{s/(1+2s)}_{2,\infty}}) + \frac{2\|f\|_2^2}{m^2} t_\epsilon^2
\end{aligned}$$

We have used the fact that $m^2 \geq 8$.

$$\begin{aligned}
B &= \mathbb{E} \sum_{0\leq j<j_\epsilon,k} [\gamma_{jk}(\epsilon)^2 (y_{jk} - \beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2] \, \mathbb{1}\, \{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\}[\mathbb{1}\, \{|\bar{\beta}_{jk}(mt_\epsilon)| \leq mt_\epsilon/2\} \\
&\quad + \mathbb{1}\, \{|\bar{\beta}_{jk}(mt_\epsilon) > mt_\epsilon/2\}] \\
&:= B_1 + B_2
\end{aligned}$$

For B_1 we use the Schwartz inequality :

$$\mathbb{E}(y_{jk} - \beta_{jk})^2 \mathbb{1}\, \{|\bar{\delta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} \leq (\mathbb{P}(|\bar{\delta}_{jk}(mt_\epsilon)| > mt_\epsilon/2)^{1/2} (\mathbb{E}(y_{jk} - \beta_{jk})^4)^{1/2}$$

where $\mathbb{E}(y_{jk} - \beta_{jk})^4 = 3\epsilon^4$ and $\mathbb{P}(|\bar{\delta}_{jk}(mt_\epsilon)| > mt_\epsilon/2) \leq \epsilon^{m^2/8}$ (using lemma 5.2). So, choosing m such that $m^2 \geq 48$,

$$\begin{aligned}
B_1 &\leq \sqrt{3}\; 2^{\frac{j_\epsilon}{2}} \textstyle\sum_{0\leq j<j_\epsilon,k} \epsilon^2 \mathbb{1}\, \{|\bar{\beta}_{jk}(mt_\epsilon)| \leq mt_\epsilon/2\} \epsilon^{m^2/16} + \sum_{0\leq j<j_\epsilon,k} \beta_{jk}^2 \mathbb{1}\, \{|\bar{\beta}_{jk}(mt_\epsilon)| \leq mt_\epsilon/2\} \\
&\leq \sqrt{3} S_\psi \; 2^{\frac{3j_\epsilon}{2}} t_\epsilon^{2+m^2/16} + \|f\|^2_{W^T_{\frac{2}{1+2s}}} (mt_\epsilon)^{4s/(1+2s)} \\
&\leq \tfrac{2\sqrt{6}S_\psi}{m^3} t_\epsilon^2 + \|f\|^2_{W^T_{\frac{2}{1+2s}}} (mt_\epsilon)^{4s/(1+2s)}
\end{aligned}$$

For B_2, we use, Proposition 5.1 :

$$\begin{aligned}
B_2 &= \mathbb{E}\textstyle\sum_{0\leq j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2+(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)|>mt_\epsilon\}\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}\\
&\leq \textstyle\sum_{0\leq j<j_\epsilon,k}[\epsilon^2\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}+B_3\\
&\leq \tfrac{2^{6-2/(1+2s)}}{(1-2^{-2/(1+2s)})\log(2)}\left(\|f\|^2_{W^T_{\frac{2}{1+2s}}}+\|f\|^2_{\mathcal{B}^{s/(1+2s)}_{2,\infty}}\right)\epsilon^2\log(\tfrac{1}{mt_\epsilon})(mt_\epsilon)^{-\frac{2}{1+2s}}+B_3\\
&\leq \tfrac{2^{6-2/(1+2s)}m^{-2/(1+2s)}}{(1-2^{-2/(1+2s)})\log(2)}\left(\|f\|^2_{W^T_{\frac{2}{1+2s}}}+\|f\|^2_{\mathcal{B}^{s/(1+2s)}_{2,\infty}}\right)t_\epsilon^{4s/(1+2s)}+B_3
\end{aligned}$$

$$\begin{aligned}
B_3 &:= \sum_{0\leq j<j_\epsilon,k}\mathbb{E}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)|>mt_\epsilon\}\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}\\
&\leq \sum_{0\leq j<j_\epsilon,k}\mathbb{E}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)|>mt_\epsilon\}\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}\mathbb{1}\ \{|\beta_{jk}|<|y_{jk}|+mt_\epsilon\}\\
&\quad+\sum_{0\leq j<j_\epsilon,k}\mathbb{E}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)|>mt_\epsilon\}\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}\mathbb{1}\ \{|y_{jk}-\beta_{jk}|\geq mt_\epsilon\}\\
&\leq \sum_{0\leq j<j_\epsilon,k}\mathbb{E}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)|>mt_\epsilon\}\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}\mathbb{1}\ \{|\beta_{jk}|<2(|y_{jk}|\vee mt_\epsilon)\}\\
&\quad+\sum_{0\leq j<j_\epsilon,k}\mathbb{E}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\ \mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)|>mt_\epsilon\}\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}\mathbb{1}\ \{|y_{jk}-\beta_{jk}|\geq mt_\epsilon\}\\
&:= B'_3+B"_3\\
B"_3 &\leq \sum_{0\leq j<j_\epsilon,k}\beta_{jk}^2\mathbb{P}(|y_{jk}-\beta_{jk}|\geq mt_\epsilon)\leq\|f\|_2^2\epsilon^{\frac{m^2}{2}}\leq\|f\|_2^2\,{t_\epsilon}^2
\end{aligned}$$

since $m^2\geq 4$. Now, using (5.4) and Proposition 5.1 we get,

$$\begin{aligned}
B'_3 &\leq \sum_{0\leq j<j_\epsilon,k}\mathbb{E}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\mathbb{1}\ \{|\bar{y}_{jk}(mt_\epsilon)|>mt_\epsilon\}\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}\mathbb{1}\ \{|\beta_{jk}|<2(|y_{jk}|\vee mt_\epsilon)\}\\
&\leq K(\gamma)^2\sum_{0\leq j<j_\epsilon,k}\mathbb{E}[t_\epsilon+\frac{\varepsilon}{|y_{jk}|\vee mt_\epsilon}]^2\beta_{jk}^2\mathbb{1}\ \{|\bar{\beta}_{jk}(mt_\epsilon)|>mt_\epsilon/2\}\mathbb{1}\ \{|\beta_{jk}|<2(|y_{jk}|\vee mt_\epsilon)\}\\
&\leq 2K(\gamma)^2\left[t_\epsilon^2\|f\|_2^2+\frac{2^{8-2/(1+2s)}m^{-2/(1+2s)}}{(1-2^{-2/(1+2s)})\log(2)}(\|f\|^2_{W^T_{\frac{2}{1+2s}}}+\|f\|^2_{\mathcal{B}^{s/(1+2s)}_{2,\infty}})t_\epsilon^{4s/(1+2s)}\right].
\end{aligned}$$

□

The next paragraph aim at giving two examples of hereditary rules which satisfy condition (5.4) of Theorem 5.2.

5.3.2 Two examples of optimal hereditary rules

A first example of optimal hereditary rule is given by the following procedure (**hard tree rule**) :

$$\tilde{f}_{T}(t) = \sum_k y_{-1k}\psi_{-1k}(t) + \sum_{0\leq j<j_\epsilon} \sum_k \gamma^{H}_{jk} y_{jk}\psi_{jk}(t) \tag{5.7}$$

where $2^{j_\epsilon} \sim (mt_\epsilon)^{-2}$, $\gamma^{H}_{jk} = 1$ if $|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon$ and $\gamma^{H}_{jk} = 0$ otherwise.
It is obvious that

$$\tilde{f}_{T} \in \mathcal{L}((mt_\epsilon)^2, 0) \cap \mathcal{H}(mt_\epsilon, t_\epsilon).$$

Remark 5.4. *In the paragraph 5.4.2 we show that this procedure can be viewed as an hybrid version of Lepski's procedure in the particular case where the wavelet basis is the Haar one. In Chapter 6, we shall see that the hard tree estimator belongs to a large class of estimates : the μ-thresholding estimators with, for any $\epsilon > 0$, any $0 < j < j_\epsilon$ and any k :*

$$\mu_{jk}(mt_\epsilon, y_{mt_\epsilon}) = \max\{|y_{I}|,\ I \in \mathcal{T}_{jk}(mt_\epsilon)\}.$$

To point out a second example of hereditary rule, let us consider the following procedure (**soft tree rule**) defined by :

$$\tilde{f}_{ST}(t) = \sum_k y_{-1k}\psi_{-1k}(t) + \sum_{0\leq j<j_\epsilon} \sum_k \gamma^{S}_{jk} y_{jk}\psi_{jk}(t) \tag{5.8}$$

where $2^{j_\epsilon} \sim (mt_\epsilon)^{-2}$, $\gamma^{S}_{jk} = 1 - \frac{\epsilon}{|\bar{y}_{jk}(mt_\epsilon)|}$ if $|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon$ and $\gamma^{S}_{jk} = 0$ otherwise.
It is obvious that

$$\tilde{f}_{ST} \in \mathcal{L}((mt_\epsilon)^2, 0) \cap \mathcal{H}(mt_\epsilon, t_\epsilon).$$

Hard tree rule and soft tree rule are optimal in the maxiset sense since the following theorem holds :

Théorème 5.3. *If m is large enough, then*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}) = \mathcal{B}^{s/(1+2s)}_{2,\infty} \cap W^{T}(\frac{2}{1+2s}, 2),$$

with $\hat{f}_\epsilon \in \{\tilde{f}_{T}, \tilde{f}_{ST}\}$.

The proof is an elementary consequence of Theorems 4.1, 5.1 and 5.2. It proves that these two procedures are optimal in the maxiset sense among limited and hereditary rules. Consequently there exist hereditary rules which outperform elitist rules.

In the following section, we focus on the case where the compactly supported wavelet is the Haar one (see the definition in section 2.1.1). We show that, in this particular case, the hard tree rule can be viewed as an hybrid version of Lepski(1991[78])'s rule, somewhat different to the one proposed by Picard and Tribouley (2000[99]).

5.4 Lepski's procedure adapted to wavelet methods

In this section we suppose that the wavelet basis in which the unknown signal f is decomposed is the **Haar wavelet basis** ($S_\psi = l_\psi = 1$). According to this choice of wavelet basis, any dyadic interval I is of the form $I = I_{jk} = [\frac{k}{2^j}, \frac{k+1}{2^j}]$.
The aim of this section is twofold. Firstly we prove that hard tree rule is connected to Lepski's procedure and we show the difference between this adaptive procedure and the hybrid version of Lepski's procedure proposed by Picard and Tribouley (2000[99]), denoted from now on as hard stem rule. Secondly, we prove that the maximal space of the hard tree rule is larger than the one of hard stem rule when dealing with the rate $t_\epsilon^{4s/(1+2s)}$.

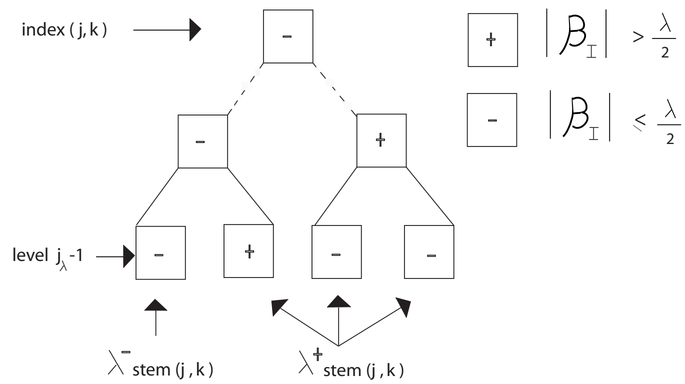
5.4.1 Hard stem rule and hard tree rule

Before recalling the definition of the hard stem rule (Kerkyacharian and Picard (2002[76]), let us introduce the following definitions :

Definition 5.7. *for any $j \in \mathbb{N}$, any $k \in \{0, \dots, 2^j - 1\}$ and any $\lambda > 0$, we say that a dyadic interval I of size 2^{1-j_λ} is*

- *a λ^-stem(j,k) if $I \subset I_{jk}$ and for any $I \subset I' \subset I_{jk}$, $|\beta_{I'}| \leq \frac{\lambda}{2}$,*
- *a λ^+stem(j,k) if $I \subset I_{jk}$ and there exists $I \subset I' \subset I_{jk}$, $|\beta_{I'}| > \frac{\lambda}{2}$.*

Let us give the following scheme to illustrate this new definition :



In the sequel of the chapter, we shall set $\lambda_\epsilon = m\epsilon\sqrt{\log(\epsilon^{-1})}$ with m is an absolute constant that will be chosen later.

A) Hard stem rule

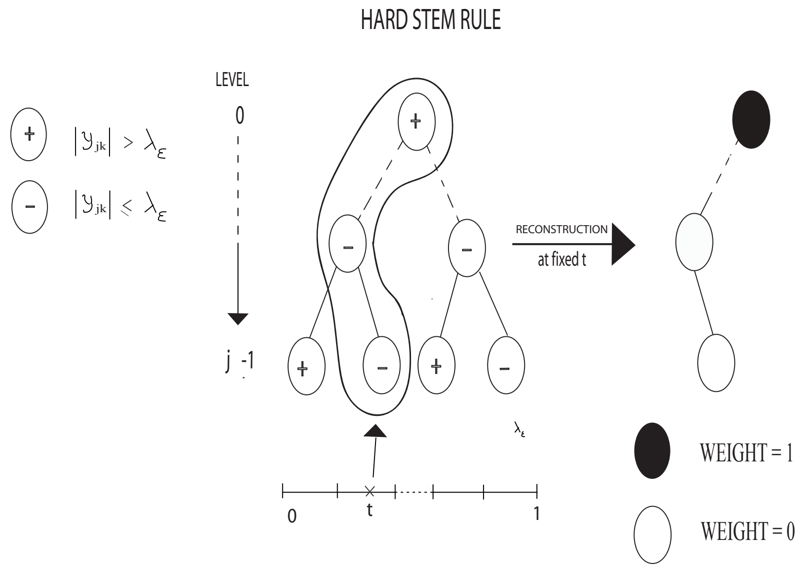
Let us consider the following procedure defined by :

$$\tilde{f}_L(t) = y_{-10}\psi_{-10}(t) + \sum_{0\leq j<j_\epsilon} \sum_k \gamma_{jk}(t) y_{jk}\psi_{jk}(t) \tag{5.9}$$

where $2^{j_\epsilon} \sim (\lambda_\epsilon)^{-2}$ and

- $\gamma_{jk}(t) = 1$ if there exists $I \subset I_{jk}$ containing t such that $|I| > \lambda_\epsilon^2$ and $|y_I| > \lambda_\epsilon$,
- $\gamma_{jk}(t) = 0$ otherwise.

This construction has also been suggested by Picard and Tribouley (2000[99]) so as to construct confidence intervals in the model of density estimation. At fixed t, this estimator is not very different from the hard thresholding one. It consists in keeping the empirical coefficients larger than λ_ϵ and somehow "in filling the holes", as we can see in the scheme below.



In the model of density estimation, Kerkyacharian and Picard (2002[76]) have shown that this rule satisfies $\mathbb{L}_2$-oracle inequalities which prove that its maxiset for the rate $(\frac{\log(n)}{n})^{2s/(1+2s)}$ is at least as good as the hard thresholding's one, but they don't characterize it. In paragraph 5.4.3, we give a precise characterization of their maxiset in the white noise model.

B) Hard tree rule

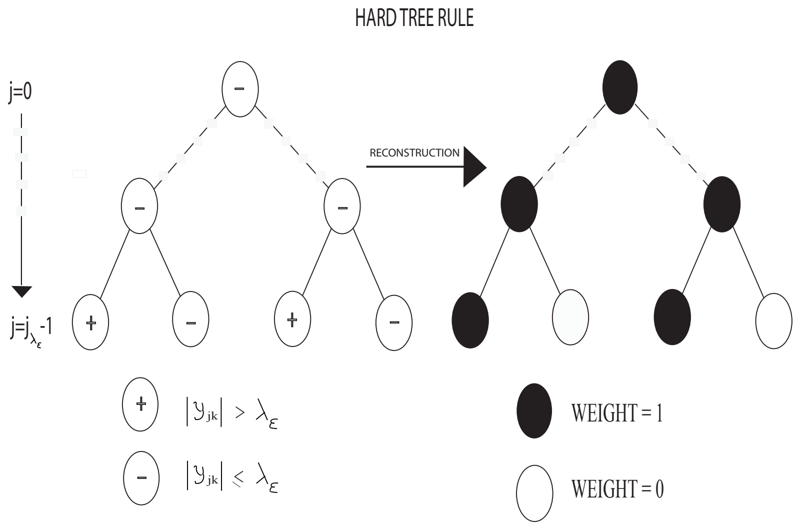
In this paragraph, we adapt the definition of hard tree rule when the wavelet basis is the Haar one. According to the paragraph 5.3.2 it is clear that the definition of the hard tree rule in this case is given by :

$$\tilde{f}_T(t) = y_{-10}\psi_{-10}(t) + \sum_{0\leq j<j_\epsilon} \sum_k \gamma_{jk} y_{jk} \psi_{jk}(t) \tag{5.10}$$

where $2^{j_\epsilon} \sim (\lambda_\epsilon)^{-2}$ and

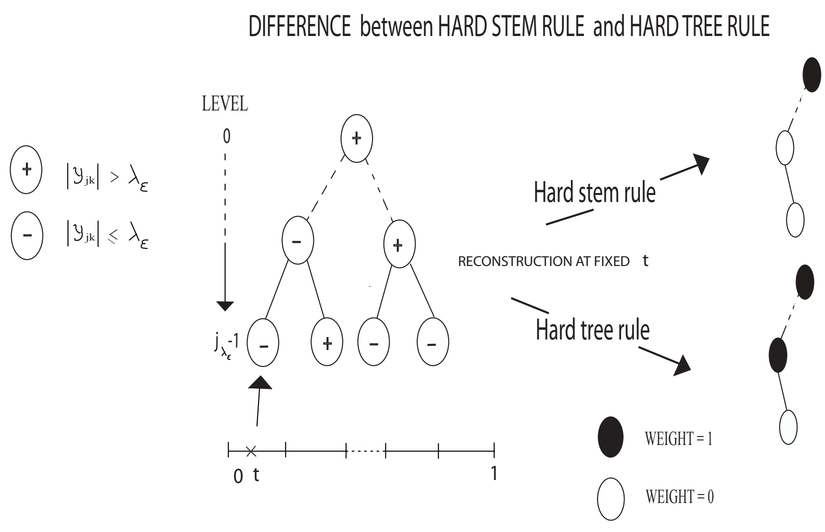
- $\gamma_{jk} = 1$ if there exists $I \subset I_{jk}$ such that $|I| > \lambda_\epsilon^2$ and $|y_I| > \lambda_\epsilon$,
- $\gamma_{jk} = 0$ otherwise.

The following scheme give an example of reconstruction using the hard tree rule :



Remark 5.5. *This estimator has a tree structure (see Engel (1994[51])) since it satisfies the following hereditary constraints :*

- $\gamma_{jk} = 1 \Longrightarrow \forall I \supset I_{jk}, \quad \gamma_{_I} = 1,$
- $\gamma_{jk} = 0 \Longrightarrow \forall I \subset I_{jk}, \quad \gamma_{_I} = 0.$



As we can see in the scheme above, this procedure is different from the first one. The difference is about the empirical coefficients between the levels 0 and $j_\epsilon - 1$ which are below the threshold λ_ϵ. In particular, in the hard tree rule, the weights γ_{jk} do not depend on t, contrary to the hard stem rule ones.

5.4.2 Connection with Lepski's procedure

In this paragraph, we show that the hard stem rule and the hard tree rule can be viewed as *wavelet-versions* of the bandwidth selection procedure of Lepski (1991[78]).
First of all, let us briefly recall the definition of the local bandwidth selection (see Lepski (1991[78]) or Lepski, Mammen and Spokoiny (1997[79]) for more details).

Local bandwidth selection

Let K be a compactly supported, bounded kernel such that $\|K\|_{\mathbb{L}_2} = 1$. For any $j \in \mathbb{N}$ and any $(t, u) \in [0, 1[^2$, let us denote

$$K_j(t, u) = 2^j K(2^j t, 2^j u) \text{ and } \hat{K}_j(t) = \int_0^1 K_j(t, u) dX_\epsilon(u).$$

Let us define the index $\hat{j}(t)$ as the minimum of *admissible* j's at the point t, where $j < j_\epsilon$ is admissible at the point t if

$$|\hat{K}_{j'+1}(t) - \hat{K}_{j'}(t)| \leq 2^{j'/2}\lambda_\epsilon \qquad \forall j \leq j' < j_\epsilon. \tag{5.11}$$

The local bandwidth selection estimator $\hat{f}_L$ is defined by :

$$\hat{f}_L(t) = \hat{K}_{\hat{j}(t)}(t).$$

The definitions of the hard stem rule and the hard tree rule are close to the definition of the local bandwidth selection procedure. Indeed, let us adapt the notion of *admissibility* from kernel estimates to wavelet estimates by considering the family of estimates $(\hat{f}_j)_{j\in\mathbb{N}}$ defined as follows :

– $\hat{f}_0(t) = y_{-10}\psi_{-10}(t)$

– $\hat{f}_{j+1}(t) = \hat{f}_j(t) + \sum_k y_{jk}\psi_{jk}(t)$.

If for any $t \in [0,1[$ we denote $\mathbf{I_j^t}$ the dyadic interval containing t such that $|I_j^t| = 2^{-j}$, then

$$|\hat{f}_{j+1}(t) - \hat{f}_j(t)| = |\sum_k y_{jk}\psi_{jk}(t)| := 2^{j/2}|y_{I_j^t}|. \tag{5.12}$$

Definition 5.8. *Say that an integer j is (t,L)-admissible if :*
either $j = j_\epsilon$ or, for all $j \le j' < j_\epsilon$, for all $\mathbf{t'} \in \mathbf{I_{j'}^t}$: $|\hat{f}_{j'+1}(t') - \hat{f}_{j'}(t')| \le 2^{j'/2}\lambda_\epsilon$.

Denote $\hat{j}_L(t) = \inf\{j;\ j$ is (t,L)-admissible$\}$. Using (5.12) we can observe that :

$$\hat{f}_{\hat{j}_L(t)}(t) = \tilde{f}_L(t). \tag{5.13}$$

Thus, this estimator can be viewed as an hybrid version of the local bandwidth selection by using the particular choice of K :

$$K_j(x,y) = \sum_k \phi_{jk}(x)\phi_{jk}(y).$$

In the same way, by introducing some modifications on the admissibility's definition, the hard tree rule can be associated with the local bandwidth selection procedure too.

Definition 5.9. *Say that an integer j is (t,T)-admissible if :*
either $j = j_\epsilon$ or, for all $j \le j' < j_\epsilon$, for all $\mathbf{t'} \in \mathbf{I_j^t}$: $|\hat{f}_{j'+1}(t') - \hat{f}_{j'}(t')| \le 2^{j'/2}\lambda_\epsilon$.

Denote $\hat{j}_T(t) = \inf\{j;\ j$ is (t,T)-admissible$\}$. Still using (5.12) we can observe that :

$$\hat{f}_{\hat{j}_T(t)}(t) = \tilde{f}_T(t). \tag{5.14}$$

So, by adapting in many ways the notion of *admissibility* from kernel estimates to wavelet estimates, we have shown that the two adaptive procedures (hard stem and hard tree rules) and the Lepski one have similitude. In the sequel of the chapter, we adopt a maxiset approach so as to compare the performances of these two rules.

5.4.3 Comparison of procedures with maxiset point of view

In this paragraph, we compare the performances of hard stem and hard tree rules. The maximal space of the hard tree rule has been established in paragraph 5.3.2. We give a new definition of the space $W^{T}(r,2)$ adapted to the case where the wavelet basis is the Haar one. Then we exhibit the maximal space of the hard stem rule.

Let us introduce the functional spaces that will be useful in the characterization of the maximal spaces associated with hard stem rule and hard tree rule.

Definition 5.10. *Let $0 < r < 2$. We shall say that a function f belongs to the space $W^{L}(r,2)$ if and only if :*

$$\sup_{\lambda>0} \lambda^r \sum_{0\leq j<j_\lambda} 2^j \sum_k \beta_{jk}^2 \#\{I \ / I \ is \ a \ \lambda^{-} stem(j,k)\} < \infty.$$

Definition 5.11. *Let $0 < r < 2$. We say that a function f belongs to the space $W^{T}(r,2)$ if and only if :*

$$\sup_{\lambda>0} \lambda^{r-2} \sum_{0\leq j<j_\lambda} \sum_k \beta_{jk}^2 \mathbf{1}\{\forall I' \subset I_{jk} \ / \ |I'| > \lambda^2, |\beta_{I'}| \leq \frac{\lambda}{2}\} \ < \infty.$$

The following proposition shows that these functional spaces associated with the same parameter r $(0 < r < 2)$ are embedded. Thanks to this result, the comparison between the maximal sets of such rules is possible, as we shall see in the end of the chapter.

Proposition 5.3. *For any $0 < r < 2$, we have the following inclusion spaces : $W(r,2) \subset W^{L}(r,2) \subset W^{T}(r,2)$.*

Proof : For any $\lambda > 0$, $0 \leq j < j_\lambda$ and any k, we have :

- $0 \leq \#\{I \ /\text{I is a } \lambda^{-} stem(j,k)\} \leq \lambda^{-2}2^{-j}$,
- $|\beta_{jk}| > \frac{\lambda}{2} \Longrightarrow \#\{I \ / \text{ I is a } \lambda^{-} stem(j,k)\} = 0$,
- $\forall I' \subset I_{jk} \ / \ |I'| > \lambda^2, |\beta_{I'}| \leq \frac{\lambda}{2} \Longrightarrow \#\{I \ / \text{ I is a } \lambda^{-} stem(j,k)\} = 2^{j_\lambda - 1 - j} \geq \lambda^{-2}2^{-(j+1)}$.

So $\frac{1}{2}\,\mathbf{1}\{\forall I' \subset I_{jk}/|I'| > \lambda^2, |\beta_{I'}| \leq \frac{\lambda}{2}\} \leq \lambda^2 2^j \#\{I$ / I is a $\lambda^- stem(j,k)\} \leq \frac{\lambda}{2}\} \leq \mathbf{1}\{|\beta_{jk}| \leq \frac{\lambda}{2}\}$, and $W(r,2) \subset W^L(r,2) \subset W^T(r,2)$. □

To point out the maxiset of the hard stem rule, let us introduce the following proposition :

Proposition 5.4. *For any $0 < r < 2$ and any $f \in \mathcal{B}^{(2-r)/4}_{2,\infty} \cap W^L(r,2)$, then :*

$$\sup_{0<\lambda<1} \lambda^{2+r}\left[\log(\frac{1}{\lambda})\right]^{-1} \sum_{0\leq j<j_\lambda} 2^j \sum_k \#\{I\ /I\ is\ a\ \lambda^+ stem(j,k)\} < \infty. \tag{5.15}$$

Remark 5.6. *We shall denote C to design absolute constants which can be different from one line to one other.*

Proof : Let $f \in \mathcal{B}^{(2-r)/4}_{2,\infty} \cap W^L(r,2)$ and $0 < \lambda < 1$. We set for any $u \in \mathbb{N}$, $2^{j_{\lambda,u}} \sim (2^{1+u}\lambda)^{-2}$. Observing that for any $j \geq 0$, any k there exist exactly $j+1$ dyadic intervals I containing I_{jk}, we have

$$\begin{aligned}
&\lambda^2 \sum_{0\leq j<j_\lambda} 2^j \sum_k \#\{I\ /\text{I is a } \lambda^+ stem(j,k)\}\\
\leq\ & \sum_{|I|=2^{1-j_\lambda}} \sum_{0\leq j<j_\lambda} 2^j \sum_k \int_I \mathbf{1}\{\text{I is a } \lambda^+ stem(j,k)\}dt\\
\leq\ & C \sum_{|I|=2^{1-j_\lambda}} \sum_{0\leq j<j_\lambda} (j+1) \sum_k \int_I \mathbf{1}\{|\beta_{jk}| > \frac{\lambda}{2},\ \ \forall I \subset I' \subsetneq I_{jk},\ |\beta_{I'}| \leq \frac{\lambda}{2}\}\psi^2_{jk}(t)dt
\end{aligned}$$

$$
\begin{aligned}
&\leq C j_\lambda \sum_{u\geq 0} \sum_{|I|=2^{1-j_\lambda}} \sum_{0\leq j<j_\lambda} \sum_k \int_I \mathbf{1}\{2^{u-1}\lambda < |\beta_{jk}| \leq 2^u\lambda, \quad \forall I \subset I' \subsetneq I_{jk},\ |\beta_{I'}| \leq \frac{\lambda}{2}\}\psi_{jk}^2(t)dt \\
&\leq C j_\lambda \sum_{u\geq 0} (2^{u-1}\lambda)^{-2} \sum_{|I|=2^{1-j_\lambda}} \sum_{0\leq j<j_\lambda} \sum_k \int_I \beta_{jk}^2 \mathbf{1}\{\forall I \subset I' \subset I_{jk},\ /\ |I'| > 4^{1+u}\lambda^2, |\beta_{I'}| \leq 2^u\lambda\}\psi_{jk}^2(t)dt \\
&\leq C\log(\frac{1}{\lambda}) \sum_{u\geq 0} (2^{u-1}\lambda)^{-2} \sum_{|I|=2^{1-j_\lambda}} \sum_{0\leq j<j_{\lambda,u}} \sum_k \int_I \beta_{jk}^2 \mathbf{1}\{\forall I \subset I' \subset I_{jk},\ /\ |I'| > 4^{1+u}\lambda^2, |\beta_{I'}| \leq 2^u\lambda\}\psi_{jk}^2(t)dt \\
&\quad + C\log(\frac{1}{\lambda}) \sum_{u\geq 0} (2^{u-1}\lambda)^{-2} \sum_{j\geq j_{\lambda,u}} \sum_k \beta_{jk}^2 \\
&\leq C\log(\frac{1}{\lambda}) \sum_{u\geq 0} \sum_{0\leq j<j_{\lambda,u}} 2^j \sum_k \beta_{jk}^2 \#\{I\ /I \text{ is a } (2^{1+u}\lambda)^{-} stem(j,k)\} \\
&\quad + C\log(\frac{1}{\lambda}) \sum_{u\geq 0} (2^{u-1}\lambda)^{-2} \sum_{j\geq j_{\lambda,u}} \sum_k \beta_{jk}^2 \\
&\leq C\ \log(\frac{1}{\lambda})\lambda^{-r}.
\end{aligned}
$$

The last inequalities use the fact that $f \in \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W^L(r,2)$. It ends the proof. □

The previous proposition will be used in the proof of the following theorem, dealing with the maxiset of the hard stem rule.

Théorème 5.4. *Let $s>0$. For any $m \geq 4\sqrt{2}$, we have the following equivalence :*

$$\sup_{0<\epsilon<1} \left(\epsilon\sqrt{\log(\epsilon^{-1})}\right)^{-4s/(1+2s)} \mathbb{E}\|\tilde{f}_L - f\|_2^2 < \infty \iff f \in \mathcal{B}_{2,\infty}^{s/(1+2s)} \cap W^L_{\frac{2}{1+2s}},$$

that is to say

$$MS(\tilde{f}_L, \|.\|_2^2, \lambda_\epsilon^{4s/(1+2s)}) = \mathcal{B}_{2,\infty}^{s/(1+2s)} \cap W^L(\frac{2}{1+2s}, 2).$$

Proof of Theorem 5.4 :
$\Longrightarrow$ Let $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$ and $f \in MS(\tilde{f}_L, \|.\|_2^2, \lambda_\epsilon^{4s/(1+2s)})$. We have,

$$\sum_{j\geq j_\epsilon,k} \beta_{jk}^2 \leq \mathbb{E}\sum_j \sum_k \|\tilde{f}_L - f\|_2^2 \leq C\lambda_\epsilon^{4s/(1+2s)} \leq C2^{-\frac{2j_\epsilon s}{1+4s}}.$$

So, using the continuity of λ_ϵ in 0, we deduce that

$$\sup_{J\geq -1} 2^{\frac{2Js}{1+2s}} \sum_{j\geq J}\sum_k \beta_{jk}^2 < \infty.$$

It comes that $f \in \mathcal{B}_{2,\infty}^{s/(1+2s)}$.

Let us denote for any $\lambda > 0$ and any I such that $|I| = 2^{1-j_\lambda}$

- $|\bar{y}_{jk}^I(\lambda)| := \max\{|y_{I'}|; I \subset I' \subset I_{jk} \text{ and } |I'| > \lambda^2\}$,
- $|\bar{\beta}_{jk}^I(\lambda)| := \max\{|\beta_{I'}|;\ I \subset I' \subset I_{jk} \text{ and } |I'| > \lambda^2\}$,
- $|\bar{\delta}_{jk}^I(\lambda)| := \max\{|y_{I'} - \beta_{I'}|; I \subset I' \subset I_{jk} \text{ and } |I'| > \lambda^2\}$.

Remark 5.7. *For any $\lambda > 0$ and any dyadic interval I,*

$$|\bar{\beta}_{jk}^I(\lambda)| \leq \frac{\lambda}{2} \iff I \textit{ is a } \lambda^- \textit{stem}(j,k),$$

$$|\bar{\beta}_{jk}^I(\lambda)| > \frac{\lambda}{2} \iff I \textit{ is a } \lambda^+ \textit{stem}(j,k).$$

Note that $|\bar{y}_{jk}^I(.)|$, $|\bar{\beta}_{jk}^I(.)|$ and $|\bar{\delta}_{jk}^I(.)|$ are decreasing functions with respect to λ and to the size of the support of I.

So, choosing $m^2 \geq 32$, we have

$$\begin{aligned}
& \lambda_\epsilon^2 \sum_{0\leq j<j_\epsilon} 2^j \sum_k \beta_{jk}^2 \#\{I \ /\ \text{I is a } \lambda_\epsilon^- stem(j,k)\} \\
\leq\ & \mathbb{E} \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\leq j<j_\epsilon} \sum_k \int_I \beta_{jk}^2 \mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} \left[\mathbf{1}\{|\bar{y}_{jk}^I(\lambda_\epsilon)| \leq \lambda_\epsilon\} + \mathbf{1}\{|\bar{y}_{jk}^I(\lambda_\epsilon)| > \lambda_\epsilon\}\right] \psi_{jk}^2(t)dt \\
\leq\ & \mathbb{E}\sum_j\sum_k \|\tilde{f}_L - f\|_2^2 + \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\leq j<j_\epsilon} \sum_k 2^{j-j_\epsilon} \beta_{jk}^2 \mathbb{P}(|\bar{y}_{jk}^I(\lambda_\epsilon)| > \lambda_\epsilon) \mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} \\
\leq\ & \mathbb{E}\sum_j\sum_k \|\tilde{f}_L - f\|_2^2 + C\lambda_\epsilon^2 \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\leq j<j_\epsilon} \sum_k 2^{j-j_\epsilon} \mathbb{P}(|\bar{\delta}_{jk}^I(\lambda_\epsilon)| > \frac{\lambda_\epsilon}{2}\} \\
\leq\ & \mathbb{E}\sum_j\sum_k \|\tilde{f}_L - f\|_2^2 + C\epsilon^{\frac{m^2}{8}-2} \\
\leq\ & C\ \lambda_\epsilon^{4s/(1+2s)}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that

$$\sup_{\lambda>0} \lambda_\epsilon^{2/(1+2s)} \sum_{0\leq j<j_\epsilon} 2^j \sum_k \beta_{jk}^2 \#\{I \ / \ \text{I is a } \lambda_\epsilon^- \, stem(j,k)\} \ < \infty.$$

It comes that $f \in W^L(\frac{2}{1+2s}, 2)$.

$\Longleftarrow$ For any $\epsilon > 0$, we have

$$\mathbb{E}\|\tilde{f}_L - f\|_2^2 = \epsilon^2 + \mathbb{E}\| \sum_{0\leq j<j_\epsilon} \sum_k (\gamma_{jk} y_{jk} - \beta_{jk})\psi_{j,k}\|_2^2 + \sum_{j\geq j_\epsilon} \sum_k \beta_{jk}^2.$$

The third term can be bounded by $C\lambda_\epsilon^{4s/(1+2s)}$, by using the definition of the Besov space $\mathcal{B}_{2,\infty}^{s/(1+2s)}$.

The term $\mathbb{E}\| \sum_{0\leq j<j_\epsilon} \sum_k (\gamma_{jk} y_{jk} - \beta_{jk})\psi_{j,k}\|_2^2$ can be bounded by $A+B$.

$$\begin{aligned} A+B &= \mathbb{E} \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\leq j<j_\epsilon} \sum_k \int_I \beta_{jk}^2 \psi_{jk}^2(t) \mathbf{1}\{|\bar{y}_{jk}^I(\lambda_\epsilon)| \leq \lambda_\epsilon\} dt \\ &+ \mathbb{E} \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\leq j<j_\epsilon} \sum_k \int_I (y_{jk} - \beta_{jk})^2 \psi_{jk}^2(t) \mathbf{1}\{|\bar{y}_{jk}^I(\lambda_\epsilon)| > \lambda_\epsilon\} dt. \end{aligned}$$

We split A into $A_1 + A_2$.

$$\begin{aligned} A &= \mathbb{E} \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\leq j<j_\epsilon} \sum_k \int_I \beta_{jk}^2 \psi_{jk}^2(t) \mathbf{1}\{|\bar{y}_{jk}^I(\lambda_\epsilon)| \leq \lambda_\epsilon\} \left[\mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| \leq 2\lambda_\epsilon\} + \mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| > 2\lambda_\epsilon\}\right] dt \\ &= A_1 + A_2. \end{aligned}$$

Since $f \in W^L(\frac{2}{1+2s}, 2)$ and $f \in \mathcal{B}^{s/(1+2s)}_{2,\infty}$,

$$\begin{aligned} A_1 &= \mathbb{E} \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\le j<j_\epsilon} \sum_k \int_I \beta_{jk}^2 \psi_{jk}^2(t) \mathbf{1}\{|\bar{y}_{jk}^I(\lambda_\epsilon)| \le \lambda_\epsilon\} \mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| \le 2\lambda_\epsilon\} dt \\ &\le \mathbb{E} \sum_{|I|=2^{5-j_\epsilon}} \sum_{0\le j<j_\epsilon-4} \sum_k \int_I \beta_{jk}^2 \psi_{jk}^2(t) \mathbf{1}\{|\bar{\beta}_{jk}^I(4\lambda_\epsilon)| \le 2\lambda_\epsilon\} dt + \sum_{j\ge j_\epsilon-4} \sum_k \beta_{jk}^2 \\ &\le C\ \lambda_\epsilon^2 \sum_{0\le j<j_\epsilon-4} 2^j \sum_k \beta_{jk}^2 \#\{I\ /\ \text{I is a } (4\lambda_\epsilon)^- stem(j,k)\} + \sum_{j\ge j_\epsilon-4} \sum_k \beta_{jk}^2 \\ &\le C\ \lambda_\epsilon^{4s/(1+2s)} \end{aligned}$$

and

$$\begin{aligned} A_2 &= \mathbb{E} \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\le j<j_\epsilon} \sum_k \int_I \beta_{jk}^2 \psi_{jk}^2(t) \mathbf{1}\{|\bar{y}_{jk}^I(\lambda_\epsilon)| \le \lambda_\epsilon\} \mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| > 2\lambda_\epsilon\} dt \\ &\le \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\le j<j_\epsilon} \sum_k \int_I \beta_{jk}^2 \psi_{jk}^2(t)\ \mathbb{P}(|\bar{\delta}_{jk}^I(\lambda_\epsilon)| > \lambda_\epsilon\} dt \\ &\le C\ j_\epsilon\ \epsilon^{m^2/2} \\ &\le C\ \lambda_\epsilon^2 \epsilon^{m^2/2-2} \\ &\le C\ \lambda_\epsilon^{4s/(1+2s)}. \end{aligned}$$

We have used here the concentration property of the Gaussian distribution and the fact that $m^2 \ge 4$.

We split B into $B_1 + B_2$ as follows.

$$\begin{aligned} B &= \mathbb{E} \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\le j<j_\epsilon} \sum_k \int_I (y_{jk} - \beta_{jk})^2 \psi_{jk}^2(t) \mathbf{1}\{|\bar{y}_{jk}^I(\lambda_\epsilon)| > \lambda_\epsilon\} [\mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| \le \lambda_\epsilon/2\} + \mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| > \lambda_\epsilon/2\}] dt \\ &= B_1 + B_2. \end{aligned}$$

For B_1 we use the Schwartz inequality :

$$B_1 \le \mathbb{E}(y_{jk} - \beta_{jk})^2 \mathbf{1}\{|\bar{\delta}_{jk}^I(\lambda_\epsilon)| > \lambda_\epsilon/2\} \le \sqrt{j_\epsilon} (\mathbb{P}(|y_{jk} - \beta_{jk}| > \lambda_\epsilon/2)^{1/2} (\mathbb{E}(y_{jk} - \beta_{jk})^4)^{1/2}$$

where $\mathbb{E}(y_{jk}-\beta_{jk})^4 = 3\epsilon^4$ and that $\mathbb{P}(|y_{jk}-\beta_{jk}| > \lambda_\epsilon/2) \leq \epsilon^{m^2/8}$ (using the concentration properties of the Gaussian distribution). So, choosing m such that $m^2 \geq 32$:

$$\begin{aligned} B_1 &\leq C\ \sqrt{j_\epsilon} \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\leq j<j_\epsilon} \sum_k \int_I \epsilon^2 \psi_{jk}^2(t)\mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| \leq \lambda_\epsilon/2\}\epsilon^{m^2/16} \\ &\leq C\ \sqrt{j_\epsilon}\ 2^{j_\epsilon}\epsilon^{2+m^2/16} \\ &\leq C\ \lambda_\epsilon^{-1}\ \epsilon^{1+m^2/16} \\ &\leq C\lambda_\epsilon^{4s/(1+2s)}. \end{aligned}$$

For B_2, we use Proposition 5.4 with $r = \frac{2}{1+2s}$:

$$\begin{aligned} B_2 &\leq \sum_{|I|=2^{1-j_\epsilon}} \sum_{0\leq j<j_\epsilon} \sum_k \int_I \epsilon^2 \psi_{jk}^2(t)\mathbf{1}\{|\bar{\beta}_{jk}^I(\lambda_\epsilon)| > \lambda_\epsilon/2\} \\ &\leq C\ \epsilon^2\lambda_\epsilon^2 \sum_{0\leq j<j_\epsilon} 2^j \sum_k \#\{I\ /\ \text{I is a } \lambda_\epsilon^+ stem(j,k)\} \\ &\leq C\lambda_\epsilon^{4s/(1+2s)}. \end{aligned}$$

□

Since the maximal spaces of the hard tree and the hard stem rules have been established, we can compare it :

Théorème 5.5. *In the maxiset sense, the hard tree rule and the hard stem rule have better performances than the hard thresholding rule. Moreover, the hard tree rule is the best procedure which has been considered here since its maxiset is larger than the hard stem rule one.*

Proof of Theorem 5.5 : This theorem is just a consequence of Theorem 5.3, Theorem 5.4 and Proposition 5.3. □

The key point of this chapter was to prove that the maxisets of elitist rules - as thresholding rules or classical Bayesian rules (see chapter 4) - don't provide the "maxi" maxisets. Indeed, according to the hard tree rule, a way of enlarging the maxisets consists in using for the reconstruction of the signal, not only the empirical coefficients y_{jk} larger than the threshold λ_ϵ in absolute value, but also their λ_ϵ-ancestors, that is to say the empirical coefficients $y_{j'k'}$ such that $I_{jk} \in \mathcal{T}_{j'k'}(\lambda_\epsilon)$.

In the next chapter, we shall see that there exist other rules, not necessary hereditary rules (for instance block thresholding rules), which provide larger maxisets than those of elitist rules.

Chapitre 6

Maxisets for μ-thresholding rules

Summary : By introducing a new large class of procedures, called μ-thresholding rules, we prove that procedures consisting in keeping or killing all the coefficients within a group provide better maxisets than those associated with elitist rules. In particular, this chapter bring a theorical explication on some phenomena appearing in the practical framework, as for instance the good performances of block thresholding rules for which the length of the blocks are not too large.

6.1 Introduction and model

Thanks to the maxiset point of view, we have successfully proved in the previous chapter that hereditary rules can outperform hard and soft thresholding rules and more than this, any elitist rule. The present chapter aims at providing other examples of adaptive procedures which outperform the elitist ones in the maxiset sense. To reach this goal, we extend the notion of thresholding rules to the notion of μ-thresholding rules which contains all the procedures $\hat{f}_\mu$ which consist in thresholding empirical coefficients individually or by groups. The class of μ-thresholding rules also contains the well-known thresholding procedures as the hard thresholding, the global thresholding and the block thresholding rules.

First, we exhibit the maximal space where these procedures attain a given rate of convergence for the Besov-risk $\mathcal{B}^0_{p,p}$ (Theorem 6.1). Then, we prove that block thresholding rules can outperform hard thresholding rules in the maxiset sense on condition that the length of their blocks are small enough (Proposition 6.3). Therefore, this result is important since it allows to give a theorical explication about the good performances of these estimates often observed in the practical setting (see Hall, Penev, Kerkyacharian and Picard (1997[56]) and Cai (1998[16], 1999[17], 2002[18])).

The chapter is organized as follows. Section 6.2 is devoted to the model and to the definition of μ-thresholding rules, illustrated by some examples. In section 6.3 we exhibit the maximal space associated with such procedures and discuss around. In section 6.4 we compare the performances of some particular μ-thresholding rules and point out the good performances of some block-thresholding rules.

We will consider a white noise setting : $X_\epsilon(.)$ is a random measure satisfying on $[0,1]$ the following equation :

$$X_\epsilon(dt) = f(t)dt + \epsilon W(dt),$$

where

- $0 < \epsilon < 1/2$ is the noise level,
- f is a function defined on $[0,1]$,
- $W(\cdot)$ is a Brownian motion on $[0,1]$.

Let $\{\phi_{0k}(\cdot),\ \psi_{jk}(\cdot),\ j \geq 0, k \in \mathbb{Z}\}$ be a compactly supported wavelet basis of $\mathbb{L}_2([0,1])$. For sake of simplicity, we shall suppose that for some $a \in \mathbb{N}^*$, the supports of ϕ and ψ are included in $[0,a]$, and we shall denote ψ_{-1k} to design ϕ_{0k}.

Any $f \in \mathbb{L}_2([0,1])$ can be represented as :

$$f = \sum_{j\geq -1} \sum_{k=1-a}^{2^j-1} \beta_{jk}\psi_{jk} = \sum_{j\geq -1} \sum_{k=1-a}^{2^j-1} (f,\psi_{jk})_{\mathbb{L}_2}\psi_{jk}. \tag{6.1}$$

Let us suppose that we dispose of observations : $y_{jk} = X_\epsilon(\psi_{jk}) = \beta_{jk} + \epsilon\xi_{jk}$ where ξ_{jk} are independent Gaussian variables $\mathcal{N}(0,1)$.

Recall that we set $2^{j_\lambda} \sim \lambda^{-2}$ to denote the integer j_λ such that $2^{-j_\lambda} \leq \lambda^2 < 2^{1-j_\lambda}$, and

$$t_\epsilon = \epsilon\sqrt{\log(\epsilon^{-1})}.$$

In following section, we define the class of procedures we shall study along the chapter : the μ-thresholding rules.

6.2 Definition of μ-thresholding rules and examples

For any $\lambda > 0$, let us denote for any sequence $(y_{jk})_{j,k}$ and any sequence $(\beta_{jk})_{j,k}$:

$$y_\lambda = (y_{jk};\ (j,k) \in I_\lambda)\,,$$

$$\beta_\lambda = (\beta_{jk};\ (j,k) \in I_\lambda),$$

where $I_\lambda = ((j,k);\ -1 \leq j < j_\lambda, -a < k < 2^j)$ and $2^{j_\lambda} \sim \lambda^{-2}$.

Remark 6.1. *For any $0 < \lambda < \sqrt{2}$, the number $\# I_\lambda$ of elements belonging to I_λ satisfies :*

$$\# I_\lambda = (a-1)(1+j_\lambda) + 2^{j_\lambda} \leq a 2^{j_\lambda}.$$

Let us consider the following class of *Keep-Or-Kill estimators* :

$$\mathcal{F}_K = \left\{ \hat{f} = \sum_j \sum_k \gamma_{jk} y_{jk} \psi_{jk}\ ;\quad \gamma_{jk}(\varepsilon) \in \{0,1\}\ measurable \right\}.$$

Definition 6.1. *We say that $\hat{f}_\mu \in \mathcal{F}_K$ is a μ-thresholding rule if :*

$$\hat{f}_\mu = \sum_{j=-1}^{j_\epsilon - 1} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\} y_{jk} \psi_{jk}, \tag{6.2}$$

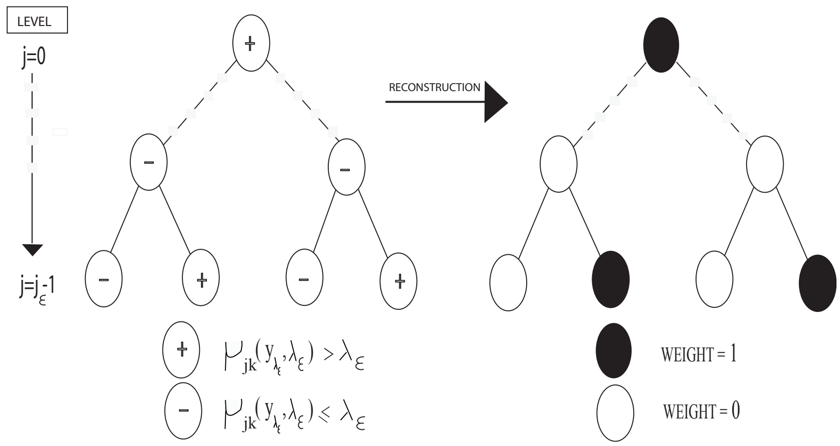
where $\lambda_\epsilon = m t_\epsilon$, $m > 0$, $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$ and for any $\lambda > 0$, $\left(\mu_{jk}(\lambda, \cdot) : \mathbb{R}^{\# I_\lambda} \longrightarrow \mathbb{R}^+\right)_{j,k}$ is a sequence of positive functions such that for any $t \in \mathbb{R}$ and any $(y_\lambda, \beta_\lambda) \in \mathbb{R}^{\# I_\lambda} \times \mathbb{R}^{\# I_\lambda}$:

$$|\mu_{jk}(\lambda, y_\lambda) - \mu_{jk}(\lambda, \beta_\lambda)| > t \Longrightarrow \exists (j_o, k_o) \in I_\lambda \ such\ that\ |y_{j_o k_o} - \beta_{j_o k_o}| > t \tag{6.3}$$

Let us notice that any μ-thresholding rule is a *limited procedure* (see Chapter 4), in the sense that the reconstruction of f by such a procedure does not use the empirical coefficients y_{jk} for which $j \geq j_\epsilon$. Moreover, any $\hat{f}_\mu$ minimizes a penalized criterion depending on the sequence of functions $(\mu_{jk})_{j,k}$. Indeed,

$$\hat{f}_\mu = \sum_{j=-1}^{j_\epsilon - 1} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\} y_{jk} \psi_{jk} \Longrightarrow \hat{f}_\mu = Arg \min_{\hat{f} \in \mathcal{F}_K} \sum_{j=-1}^{j_\epsilon - 1} \sum_k (\gamma_{jk} - 1)^2 \mu_{jk}^2(\lambda_\epsilon, y_{\lambda_\epsilon}) + \lambda_\epsilon^2 \gamma_{jk}^2.$$

The reconstruction of the signal f by a μ-thresholding rule consists in keeping the empirical coefficient y_{jk} at level strictly less than j_ϵ for which $\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon})$ is strictly larger than the threshold λ_ϵ, as we can see in the following scheme :



There is no doubt that μ-thresholding estimates constitute a large sub-family of Keep-or-Kill estimates. Let us give some examples of such procedures, by choosing different choices of functions μ_{jk} :

1) The hard thresholding procedure belongs to the family of μ-thresholding rules. It corresponds to the choice :

$$\mu_{jk}^{(1)}(\lambda_\epsilon, y_{\lambda_\epsilon}) = |y_{jk}|.$$

This procedure has been proved to have good performances in the minimax point of view (see Donoho, Johnstone, Kerkyacharian and Picard (1995[47],1996[48],1997[49])) and in the maxiset point of view (see Cohen, De Vore, Kerkyacharian and Picard (2001[31]) and Kerkyacharian and Picard (2000[75])).

2) Block thresholding procedures belong to the family of μ-thresholding rules. They correspond to the choices :

$$\mu_{jk}^{(2)}(\lambda_\epsilon, y_{\lambda_\epsilon}) = \left(\frac{1}{l_j} \sum_{k' \in \mathcal{P}_j(k)} |y_{jk'}|^p\right)^{1/p}, \quad \text{[Mean-block(p) thresholding]}$$

$$\mu_{jk}^{(3)}(\lambda_\epsilon, y_{\lambda_\epsilon}) = \max_{k' \in \mathcal{P}_j(k)} |y_{jk'}| \quad \text{[Maximum-block thresholding]}$$

and :

$$\mu_{jk}^{(4)}(\lambda_\epsilon, y_{\lambda_\epsilon}) = \max(|y_{jk}|, \mu_{jk}^{(2)}(\lambda_\epsilon, y_{\lambda_\epsilon})), \quad \text{[Maximean-block(p) thresholding]}$$

where for any (j,k) and any $0 < \epsilon < \frac{1}{2}$, $\quad k \in \mathcal{P}_j(k), \quad \mathcal{P}_j(k) \subset \{1-a, \dots, 2^j - 1\}$, $\#\mathcal{P}_j(k) = l_j$ and

$$k \in \mathcal{P}_j(k) \cap \mathcal{P}_j(k') \Longrightarrow \mathcal{P}_j(k) = \mathcal{P}_j(k').$$

Block thresholding estimators are known to have good performances in the practical setting. For example Hall, Penev, Kerkyacharian and Picard (1997[56]) considered mean-block thresholding. The goal was to increase estimation precision by utilizing information about neighboring wavelet coefficients. The method they proposed was to first obtain a near unbiased estimate of the sum of squares of the true coefficients within a block and then to keep or kill all the coefficient within the block based on the magnitude of the estimate. As well as the family blockwise James-Stein estimators (see Cai (1998[16], 1999[17], 2002[18])), on condition that the length of blocks is not exceeding $C\log(n)$ $(C > 0)$ this estimator was shown to have good performances in the practical setting (see Hall, Penev, Kerkyacharian and Picard (1997[56])) and was proved to attain the exact minimax rate of convergence for the $\mathbb{L}_2$-risk without the logarithmic penalty over a range of perturbed Hölder classes (Hall, Kerkyacharian and Picard (1999[55])).

3) The hard tree procedure belongs to the family of μ-thresholding rules, with the choice :

$$\mu_{jk}^{(5)}(\lambda_\epsilon, y_{\lambda_\epsilon}) = \max\{|y_{j'k'}|; I_{j'k'} \in \mathcal{T}_{jk}(\lambda_\epsilon)\}.$$

This procedure, which has been studied in the previous chapter when dealing with hereditary rules, is directly inspired from tree methods in approximation theory (Cohen, Dahmen, Daubechies and DeVore (2001[29])).

6.3 Maxisets associated with μ-thresholding rules

In this section, we aim at exhibiting the maximal spaces where the μ-thresholding rules attain the rate of convergence $(u(\lambda_\epsilon))^{2sp/(1+2s)}$ $(1 \leq p < \infty)$, where u is an increasing transformation map of $\mathbb{R}^+$ in $\mathbb{R}^+$ that is continuous and satisfies :

$$\forall 0 < \epsilon < 1/2, \quad \epsilon \leq u(\lambda_\epsilon). \tag{6.4}$$

Remark 6.2. *Even if the choice $u(\lambda) = \lambda$ is often used, we choose here more general rates of convergence so as to integrate, for example, logarithmic terms.*

6.3.1 Functional spaces

To begin, we introduce the functional spaces that will be useful throughout the paper when studying the maximal spaces of μ-thresholding rules.

Definition 6.2. *Let $s > 0$ and $1 \leq p < \infty$. We shall say that a function $f \in \mathbb{L}_p([0,1])$ belongs to the Besov space $\mathcal{B}^s_{p,\infty}(u)$, if and only if :*

$$\sup_{\lambda>0} (u(\lambda))^{-2sp} \sum_{j \geq j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p < \infty.$$

Notice that $\mathcal{B}^s_{p,\infty}(Id_{\mathbb{R}^+})$ is the classical Besov space, which has been proved to contain the maximal space of any *limited rule* for the rate λ_ϵ^{2sp} (see Chapter 4).

Definition 6.3. *Let $0 < r < p < \infty$. We shall say that a function f belongs to the space $W_{\mu,u}(r,p)$ if and only if :*

$$\sup_{\lambda>0} (u(\lambda))^{r-p} \sum_{j < j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1} \, \{\mu_{jk}(\lambda, \beta_\lambda) \leq \frac{\lambda}{2}\} < \infty.$$

The definitions of such spaces in the case $u = Id_{\mathbb{R}^+}$ are close to the ones of weak Besov spaces. Weak Besov spaces have been proved to be directly connected with hard and soft thresholding rules (see Cohen, De Vore, Kerkyacharian and Picard (2001[31]) and Kerkyacharian and Picard (2002[76])). In this paper, we shall see the strong relation between $W_{\mu,u}(r,p)$ and μ-thresholding rules.

Definition 6.4. *Let $0 < r < p < \infty$. We shall say that a function f belongs to the space $W^*_{\mu,u}(r,p)$ if and only if :*

$$\sup_{\lambda>0} \lambda^p (u(\lambda))^{r-p} \left(\log(\lambda^{-1})\right)^{-\frac{p}{2}} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\, \{\mu_{jk}(\lambda,\beta_\lambda) > 2\lambda\} < \infty.$$

The aim of the following paragraph is to exhibit the maxisets associated with the μ-thresholding rules. Undoubtedly, these maximal spaces depend on the choice of the transformation map u.

6.3.2 Main result

Théorème 6.1. *Let $1 \leq p < \infty$ and $m \geq 4\sqrt{p+1}$. Denote $\lambda_\epsilon = mt_\epsilon$ and suppose that $\hat{f}_\mu$ is a μ-thresholding rule such that $(\mu_{jk})_{jk}$ are decreasing functions with respect to λ. If there exist $K_m > 0$ and $\lambda_{seuil} > 0$ such that :*

$$\forall\ 0 < \lambda < \lambda_{seuil}, \quad u(4m\lambda) \leq K_m u(\lambda), \tag{6.5}$$

then :

$$\sup_{0<\epsilon<1/2} (u(\lambda_\epsilon))^{-2sp/(1+2s)} \mathbb{E}\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} < \infty \iff f \in \mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu,u}(\frac{p}{1+2s},p) \cap W^*_{\mu,u}(\frac{p}{1+2s},p).$$

Remark 6.3. *When $u(t_\epsilon) = t_\epsilon$ (resp. $u(t_\epsilon) = \epsilon$), notice that (6.5) is satisfied by taking $K_m = 4m$ (resp. $K_m = 4\sqrt{2}m$) and $\epsilon_{seuil} = \frac{1}{2}$ (resp. $\epsilon_{seuil} = \frac{1}{32m^2}$).*

Proof of Theorem 6.1 :
Here and later, we shall note C to design a constant which may be different from one line to the other.

$\Longrightarrow$ Notice that it suffices to prove the result for $0 < \epsilon < \epsilon_{seuil}$ where ϵ_{seuil} is such that $t_{\epsilon_{seuil}} = \lambda_{seuil}$. For any $0 < \epsilon < \epsilon_{seuil}$, we have,

$$\sum_{j\geq j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \leq \mathbb{E}\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} \leq C(u(\lambda_\epsilon))^{2sp/(1+2s)}.$$

So, using the continuity of t_ϵ in 0, we deduce that

$$\sup_{\lambda>0}(u(\lambda))^{-2sp/(1+2s)}\sum_{j\geq j_\lambda}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p<\infty.$$

It comes that $f\in\mathcal{B}^{s/(1+2s)}_{p,\infty}(u)$.

Moreover,

$$\begin{aligned}&\sum_{j<j_\epsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon})\leq\frac{\lambda_\epsilon}{2}\}\\ =\ &\mathbb{E}\sum_{j<j_\epsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon})\leq\frac{\lambda_\epsilon}{2}\}[\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon})\leq\lambda_\epsilon\}+\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon})>\lambda_\epsilon\}]\\ =\ &A_1+A_2.\end{aligned}$$

We have

$$\begin{aligned}A_1\ &=\ \mathbb{E}\sum_{j<j_\epsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon})\leq\frac{\lambda_\epsilon}{2}\}\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon})\leq\lambda_\epsilon\}\\ &\leq\ \mathbb{E}\sum_{j<j_\epsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon})\leq\lambda_\epsilon\}\\ &\leq\ \mathbb{E}\|\hat{f}_\mu-f\|^p_{\mathcal{B}^0_{p,p}}\\ &\leq\ C\ (u(\lambda_\epsilon))^{2sp/(1+2s)}.\end{aligned}$$

Using (6.3) and the concentration properties of the Gaussian distribution, one gets :

$$\begin{aligned}A_2\ &=\ \mathbb{E}\sum_{j<j_\epsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon})\leq\frac{\lambda_\epsilon}{2}\}\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon})>\lambda_\epsilon\}\\ &=\ \sum_{j<j_\epsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{P}(\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon})>\lambda_\epsilon)\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon})\leq\frac{\lambda_\epsilon}{2}\}\\ &\leq\ \sum_{j<j_\epsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{P}(|\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon})-\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon})|>\frac{\lambda_\epsilon}{2})\\ &=\ \sum_{j<j_\epsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{P}(\exists(j_o,k_o)\in I_{\lambda_\epsilon}\mid|y_{j_ok_o}-\beta_{j_ok_o}|>\frac{\lambda_\epsilon}{2})\\ &\leq\ C\ 2^{j_\epsilon}\epsilon^{\frac{m^2}{8}}\\ &\leq\ C\ (u(\lambda_\epsilon))^{2sp/(1+2s)}.\end{aligned}$$

Last inequality is due to the fact that $m^2 \geq 8(p+2)$.

Using the continuity of t_ϵ in 0, we deduce that

$$\sup_{\lambda>0}(u(\lambda))^{-2sp/(1+2s)} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) \leq \frac{\lambda}{2}\} < \infty.$$

It comes that $f \in W_{\mu,u}(\frac{p}{1+2s}, p)$.

Finally we have,

$$\begin{aligned} & \epsilon^p \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\} \\ = \ & C\ \mathbb{E} \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\}[\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon}) > \lambda_\epsilon\} + \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon}) \leq \lambda_\epsilon\}] \\ = \ & C(A_3 + A_4). \end{aligned}$$

$$\begin{aligned} A_3 &= \mathbb{E} \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\} \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon}) > \lambda_\epsilon\} \\ &\leq \mathbb{E} \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon}) > \lambda_\epsilon\} \\ &\leq \mathbb{E}\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} \\ &\leq C\ (u(\lambda_\epsilon))^{2sp/(1+2s)}. \end{aligned}$$

Using the Cauchy-Schwartz inequality and (6.3),

$$\begin{aligned} & (\mathbb{E}|y_{jk}-\beta_{jk}|^p)^2 \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon}) \leq \lambda_\epsilon\} \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\} \\ \leq \ & \mathbb{E}|y_{jk}-\beta_{jk}|^{2p} \mathbb{P}(|\mu_{jk}(\lambda_\epsilon,y_{\lambda_\epsilon}) - \mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon})| > \lambda_\epsilon) \\ \leq \ & \mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\ \mathbb{P}(\exists (j_o,k_o) \in I_{\lambda_\epsilon}\ |\ |y_{j_ok_o} - \beta_{j_ok_o}| > \lambda_\epsilon) \\ \leq \ & a2^{j_\epsilon} \mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\ \mathbb{P}(|y_{jk}-\beta_{jk}| > \lambda_\epsilon), \end{aligned}$$

where $\mathbb{E}|y_{jk}-\beta_{jk}|^{2p} = C\epsilon^{2p}$ and that $\mathbb{P}(|y_{jk}-\beta_{jk}| > \lambda_\epsilon) \leq \epsilon^{m^2/2}$
So, since $m^2 \geq 4(1+2p)$, from the concentration properties of the Gaussian distribution one gets

$$\begin{aligned}
A_4 &= \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\}\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) \leq \lambda_\epsilon\} \\
&\leq \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{E}^{1/2}|y_{jk}-\beta_{jk}|^{2p}\, \mathbb{P}^{1/2}(\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) \leq \lambda_\epsilon)\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\} \\
&\leq \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{E}^{1/2}|y_{jk}-\beta_{jk}|^{2p}\, \mathbb{P}^{1/2}(\exists (j_o,k_o)\in I_{\lambda_\epsilon} \mid |y_{j_ok_o}-\beta_{j_ok_o}| > \lambda_\epsilon) \\
&\leq C\ 2^{j_\epsilon/2}\epsilon^{m^2/4-p} \\
&\leq C\ (u(\lambda_\epsilon))^{2sp/(1+2s)}.
\end{aligned}$$

Using the continuity of t_ϵ in 0, we deduce that

$$\sup_{\lambda>0} \lambda^p (u(\lambda))^{r-p}\left(\log(\lambda^{-1})\right)^{-p/2} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) > 2\lambda\} < \infty.$$

It comes that $f \in W^*_{\mu,u}(\frac{p}{1+2s}, p)$.

$\Longleftarrow$ For any $0<\epsilon<\epsilon_{seuil}$, we have

$$\mathbb{E}\|\bar{f}_\epsilon - f\|^p_{\mathcal{B}^0_{p,p}} = \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\} - \beta_{jk}|^p + \sum_{j\geq j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p.$$

Since $f \in \mathcal{B}^{s/(1+2s)}_{p,\infty}(u)$, the second term can be bounded by $C\ (u(\lambda_\epsilon))^{2sp/(1+2s)}$.

The first term $\mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\} - \beta_{jk}|^p$ can be bounded by $C(B_1 + B_2)$, where

$$B_1 + B_2 = \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p\{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) \leq \lambda_\epsilon\} + \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\}.$$

We split B_1 into $B_1' + B_1''$.

$$\begin{aligned}
B_1 &= \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p\ \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) \leq \lambda_\epsilon\}[\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon}) \leq 2\lambda_\epsilon\} + \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon,\beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\}] \\
&= B_1' + B_1''.
\end{aligned}$$

Since $f \in \mathcal{B}^{s/(1+2s)}_{2,\infty}(u) \cap W_{\mu,u}(\frac{p}{1+2s}, p)$ and $(\mu_{jk})_{j,k}$ are decreasing functions with respect to λ, using (6.5) one gets :

$$\begin{aligned}
B_1' &= \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \ \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) \leq \lambda_\epsilon\}\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) \leq 2\lambda_\epsilon\} \\
&\leq \sum_{j<j_\epsilon-4} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \ \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) \leq 2\lambda_\epsilon\} + \sum_{j\geq j_\epsilon-4} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \\
&\leq \sum_{j<j_\epsilon-4} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \ \mathbb{1}\ \{\mu_{jk}(4\lambda_\epsilon, \beta_{4\lambda_\epsilon}) \leq 2\lambda_\epsilon\} + \sum_{j\geq j_\epsilon-4} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \\
&\leq C(u(4\lambda_\epsilon))^{2sp/(1+2s)} \\
&\leq C(u(\lambda_\epsilon))^{2sp/(1+2s)}.
\end{aligned}$$

Using (6.3)

$$\begin{aligned}
B_1'' &= \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \ \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) \leq \lambda_\epsilon\}\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\} \\
&= \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \ \mathbb{P}(\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) \leq \lambda_\epsilon)\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) > 2\lambda_\epsilon\} \\
&= \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \ \mathbb{P}(\exists (j_o,k_o) \in I_{\lambda_\epsilon} \mid |y_{j_ok_o} - \beta_{j_ok_o}| > \lambda_\epsilon) \\
&\leq C\ 2^{j_\epsilon}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \ \mathbb{P}(|y_{jk} - \beta_{jk}| > \lambda_\epsilon\} \\
&\leq C\ 2^{j_\epsilon}\epsilon^{m^2/2} \\
&\leq C\ \epsilon^{m^2/2-2} \\
&\leq C\ (u(\lambda_\epsilon))^{2sp/(1+2s)}.
\end{aligned}$$

We have used here the concentration property of the Gaussian distribution and the fact that $m^2 \geq 2(p+2)$.

We split B_2 into $B_2' + B_2''$ as follows.

$$\begin{aligned} B_2 &= \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\}[\mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) \leq \frac{\lambda_\epsilon}{2}\} + \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) > \frac{\lambda_\epsilon}{2}\}] \\ &= B_2' + B_2''. \end{aligned}$$

For B_2' we use the Cauchy-Schwartz inequality :

$$\begin{aligned} & (\mathbb{E}|y_{jk}-\beta_{jk}|^p)^2 \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\} \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) \leq \frac{\lambda_\epsilon}{2}\} \\ \leq\ & \mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\, \mathbb{P}(|\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) - \mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon})| > \frac{\lambda_\epsilon}{2}) \\ \leq\ & \mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\, \mathbb{P}(\exists (j_o, k_o) \in I_{\lambda_\epsilon} \mid |y_{j_o k_o} - \beta_{j_o k_o}| > \frac{\lambda_\epsilon}{2}) \\ \leq\ & a 2^{j_\epsilon} \mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\, \mathbb{P}(|y_{jk}-\beta_{jk}| > \frac{\lambda_\epsilon}{2}), \end{aligned}$$

where $\mathbb{E}|y_{jk}-\beta_{jk}|^{2p} = C\epsilon^{2p}$ and that $\mathbb{P}(|y_{jk}-\beta_{jk}| > \frac{\lambda_\epsilon}{2}) \leq \epsilon^{m^2/8}$ (using the concentration properties of the Gaussian distribution). So, choosing m such that $m^2 \geq 16(p+1)$,

$$\begin{aligned} B_2' &= \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\} \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon,}) \leq \frac{\lambda_\epsilon}{2}\} \\ &\leq C\ 2^{j_\epsilon/2} \epsilon^p \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon,}) \leq \frac{\lambda_\epsilon}{2}\} \epsilon^{m^2/16} \\ &\leq C\ 2^{j_\epsilon(p+1)/2} \epsilon^{m^2/16+p} \\ &\leq C(u(\lambda_\epsilon))^{2sp/(1+2s)}. \end{aligned}$$

Since $f \in W^*_{\mu,u}(\frac{p}{1+2s}, p)$, we can bounded B_2'' as follows.

$$\begin{aligned} B_2'' &= \mathbb{E}\sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, y_{\lambda_\epsilon}) > \lambda_\epsilon\} \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) > \frac{\lambda_\epsilon}{2}\} \\ &\leq C\ \epsilon^p \sum_{j<j_\epsilon} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) > \frac{\lambda_\epsilon}{2}\} \\ &\leq C\ (\frac{\lambda_\epsilon}{4})^p \left(\log(\frac{4}{\lambda_\epsilon})\right)^{-p/2} \sum_{j<j_\epsilon+4} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda_\epsilon, \beta_{\lambda_\epsilon}) > \frac{\lambda_\epsilon}{2}\} \\ &\leq C(u(\frac{\lambda_\epsilon}{4}))^{2sp/(1+2s)} \\ &\leq C(u(\lambda_\epsilon))^{2sp/(1+2s)}. \end{aligned}$$

□

The previous theorem point out the maximal spaces where μ-thresholding rules attain the rate of convergence $(u(\lambda_\epsilon))^{2sp/(1+2s)}$. Notice that so bigger are the functions μ_{jk}, so larger are the spaces $W_{\mu,u}(\frac{p}{1+2s}, p)$ and so thinner are the spaces $W^*_{\mu,u}(\frac{p}{1+2s}, p)$.
In the next section, we give assumptions on the choices of u and μ_{jk} to be sure that we have the following embedding :

$$\mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu,u}(\frac{p}{1+2s}, p) \subset \mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W^*_{\mu,u}(\frac{p}{1+2s}, p).$$

6.3.3 Conditions for embedding inside maximal spaces

Théorème 6.2. *Let $0 < r < p < \infty$ and $(\mu_{jk})_{jk}$ be a sequence of decreasing functions with respect to λ. Assume that there exist $C_{seuil} > 0$ and $\lambda_{seuil} > 0$ such that, for any $0 < \lambda < \lambda_{seuil}$, the following conditions are satisfied :*

$$\sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\left\{\mu_{jk}(\lambda, \beta_\lambda) > \lambda\right\} \leq C_{seuil}\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{j=-1}^{j_\lambda - 1} 2^{j(\frac{p}{2}-1)} \sum_k \sum_{n\in\mathbb{N}} \mathbb{1}\left\{|\beta_{jk}| > \lambda 2^n\right\} \mathbb{1}\left\{\mu_{jk}(\lambda, \beta_\lambda) \leq 2^{1+n}\lambda\right\} \quad (6.6)$$

$$\forall n \in \mathbb{N},\ \exists C_n > 0 \text{ (not depending on } \lambda); \quad u(2^{2+n}\lambda) \leq C_n u(\lambda), \text{ and } \sum_{n\in\mathbb{N}} C_n^{p-r} 2^{-np} < \infty \quad (6.7)$$

Then,

$$\mathcal{B}^{(p-r)/2p}_{p,\infty}(u) \cap W_{\mu,u}(r, p) \subset \mathcal{B}^{(p-r)/2p}_{p,\infty}(u) \cap W^*_{\mu,u}(r, p).$$

Remark 6.4. *It is easy to see that condition (6.7) implies condition (6.5). Once again, condition (6.7) is clearly satisfied when $u(t_\epsilon) = t_\epsilon$ or $u(t_\epsilon) = \epsilon$.*

Proof of Theorem 6.2 :
For any (j, k), let μ_{jk} and u satisfy respectively the conditions (6.6) and (6.7).
Fix $0 < \lambda < \lambda_{seuil}$ and set for any $n \in \mathbb{N}$, $2^{j_{\lambda,n}} \sim (2^{2+n}\lambda)^{-2}$. Using (6.6),

$$
\begin{aligned}
& \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) > 2\lambda\} \\
\leq\ & \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) > \lambda\} \\
\leq\ & C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \sum_{n\in\mathbb{N}} \mathbb{1}\ \{|\beta_{jk}| > 2^n\lambda\}\mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) \leq 2^{1+n}\lambda\} \\
\leq\ & C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) \leq 2^{1+n}\lambda\} \\
\leq\ & C_1 + C_2,
\end{aligned}
$$

where

$$C_1 = C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j<j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) \leq 2^{2+n}\frac{\lambda}{2}\}$$

and

$$C_2 = C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j\geq j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p.$$

Since $f \in W_{\mu,u}(r,p)$,

$$
\begin{aligned}
C_1 &= C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j<j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) \leq 2^{2+n}\frac{\lambda}{2}\} \\
&\leq C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j<j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\ \{\mu_{jk}(2^{2+n}\lambda,\beta_{2^{2+n}\lambda}) \leq 2^{2+n}\frac{\lambda}{2}\} \\
&\leq C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} (u(2^{2+n}\lambda))^{p-r} \\
&\leq C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \lambda^{-p}(u(\lambda))^{p-r} \sum_{n\in\mathbb{N}} C_n^{p-r} 2^{-np} \\
&\leq C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \lambda^{-p}(u(\lambda))^{p-r}.
\end{aligned}
$$

Last inequalities use condition (6.7).

Now, since $f \in \mathcal{B}^{(p-r)/2p}_{p,\infty}(u)$,

$$\begin{aligned} C_2 &= C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j\geq j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\ &\leq C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} (u(2^{2+n}\lambda))^{p-r} \\ &\leq C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \lambda^{-p}(u(\lambda))^{p-r} \sum_{n\in\mathbb{N}} C_n^{p-r} 2^{-np} \\ &\leq C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \lambda^{-p}(u(\lambda))^{p-r}. \end{aligned}$$

Last inequalities use condition (6.7).

By adding up C_1 and C_2, we have

$$\sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbb{1}\ \{\mu_{jk}(\lambda,\beta_\lambda) > 2\lambda\} \leq C\left(\log(\lambda^{-1})\right)^{\frac{p}{2}} \lambda^{-p}(u(\lambda))^{p-r},$$

which proves that $f \in W^*_{\mu,u}(r,p)$ and ends the proof. □

Corollary 6.1. *Let $s > 0$, $1 \leq p < \infty$ and $m \geq 4\sqrt{p+1}$. Let $MS(\hat{f}_\mu, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)})$ be the maximal set of any μ-thresholding rule $\hat{f}_\mu$ for the rate of convergence $(u(\lambda_\epsilon))^{2sp/(1+2s)}$. Under conditions of Theorem 6.2, we have :*

$$MS(\hat{f}_\mu, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}) = \mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu,u}(\frac{p}{1+2s}, p).$$

To prove it, it suffices to apply Theorem 6.1 and Theorem 6.2 (with $r = \frac{p}{1+2s}$).

Let us give two examples of such embeddings. It is clear that $\mu^{(1)}_{jk}$ and $\mu^{(5)}_{jk}$ satisfy condition (6.6) of Theorem 6.2. Consequently, the maximal space where the procedures $\hat{f}_{\mu^{(i)}}$, $i \in \{1,5\}$, attain the rate of convergence $(u(\lambda_\epsilon))^{2sp/(1+2s)}$ is

$$\mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu^{(i)},u}(\frac{p}{1+2s}, p).$$

Notice that for $u = Id_{\mathbb{R}^+}$, we identify $\mathcal{B}^{s/(1+2s)}_{p,\infty}(u)$ with the usual Besov space

$$\mathcal{B}^{s/(1+2s)}_{p,\infty} = \left\{ f; \quad \sup_{J\geq -1} 2^{J(sp+\frac{p}{2}-1)} \sum_{j\geq J}\sum_k |\beta_{jk}|^p < \infty \right\}.$$

For the same choice of u, $W_{\mu^{(1)},u}(\frac{p}{1+2s},p)$ represents the weak Besov space

$$W(\frac{p}{1+2s},p)=\left\{f;\quad \sup_{\lambda>0}\lambda^{r-p}\sum_{j<j_\lambda}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{1}\left\{|\beta_{jk}|\leq\lambda\right\}<\infty\right\},$$

and the space $W_{\mu^{(5)},u}(\frac{p}{1+2s},p)$ represents the space

$$W^T(\frac{p}{1+2s},p)=\left\{f;\quad \sup_{\lambda>0}\lambda^{r-p}\sum_{0\leq j<j_\lambda}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbb{1}\left\{\forall I_{j'k'}\in\mathcal{T}_{jk}(\lambda_\epsilon),|\beta_{j'k'}|\leq\frac{\lambda}{2}\right\}<\infty\right\}.$$

Let us recall that the maxiset of the hard thresholding rule $\hat{f}_{\mu^{(1)}}$ for the rate $\lambda_\epsilon^{2sp/(1+2s)}$ has been studied by Cohen, De Vore, Kerkyacharian and Picard (2001[31]) and Kerkyacharian and Picard (2000[75]). In the previous chapter, we have studied the maxiset of the hard tree rule $\hat{f}_{\mu^{(5)}}$ for $p=2$ and the same rate of convergence. In particular, we have proved that the maxiset performance of this rule is better than the hard thresholding one, in the sense that

$$\mathcal{B}^{s/(1+2s)}_{2,\infty}\cap W(\frac{2}{1+2s},2)\subset\mathcal{B}^{s/(1+2s)}_{2,\infty}\cap W^T(\frac{2}{1+2s},2).$$

Theorems 6.1 and 6.2 allow to exhibit the maximal space of any μ-thresholding rule, dealing with the rate of convergence $(u(\lambda_\epsilon))^{2sp/(1+2s)}$. Let us notice that the comparison of two such procedures is not always possible, since it could be plausible that their maxisets are not embedded.

6.4 On block thresholding and hard tree rules

The aim of this section is twofold. First of all, we give a way to construct μ-thresholding rules with better performances (in the maxiset sense) than the hard thresholding one $\hat{f}_{\mu^{(1)}}$. Thanks to this, we prove that block thresholding rules and the hard tree rule can outperform hard thresholding rules.
Let us state the following proposition :

Proposition 6.1. *Let $1\leq p<\infty$. Under conditions of Theorem 6.2, the maximal space for the rate $(u(\lambda_\epsilon))^{2sp/(1+2s)}$ of any μ-thresholding rule satisfying for any $\lambda>0$, any*

$\beta_\lambda \in \mathbb{R}^{\#I_\lambda}$,

$$\mu_{jk}(\lambda, \beta_\lambda) \leq \frac{\lambda}{2} \Longrightarrow |\beta_{jk}| \leq \frac{\lambda}{2}, \tag{6.8}$$

is larger than the hard thresholding one. Moreover, if for all $n \in \mathbb{N}$, $C_n \leq O(2^n)$, *then the maximal space contains the Besov space* $\mathcal{B}^s_{p,\infty}(u)$.

Remark 6.5. *For* $u(t_\epsilon) = t_\epsilon$ *(resp.* $u(t_\epsilon) = \epsilon$*), notice that, using Remark 6.3, the last condition on* C_n *is satisfied by taking* $C_n = 2^{2+n}$ *(resp.* $C_n = 2^{\frac{5}{2}+n}$*).*

Proof :
If $\hat{f}_\mu$ is a μ-thresholding rule satisfying (6.8), then we have for any $0 < r < p : W_{\mu,u}(r,p) \supset W_{\mu^{(1)},u}(r,p)$. So, using Corollary 6.1 to characterize the maxisets for the rate $(u(\lambda_\epsilon))^{2sp/(1+2s)}$ associated with $\hat{f}_\mu$ and $\hat{f}_{\mu^{(1)}}$, one gets that the maximal space for the rate $(u(\lambda_\epsilon))^{2sp/(1+2s)}$ of $\hat{f}_\mu$ is larger than the hard thresholding one.

To prove now that the Besov space $\mathcal{B}^s_{p,\infty}(u)$ is contained in the maxiset of $\hat{f}_\mu$, it suffices to prove that :

$$\mathcal{B}^s_{p,\infty}(u) \subset W_{\mu^{(1)},u}(\frac{p}{1+2s}, p).$$

Fix $0 < \lambda < \lambda_{seuil}$ and set $2^{j_{\lambda,u}} \sim \lambda_u^{-2} := (u(\lambda))^{4s/(1+2s)} \lambda^{-2}$ (resp. $2^{j_\lambda} \sim \lambda^{-2}$). Since $f \in \mathcal{B}^s_{p,\infty}(u)$ we have,

$$\begin{aligned}
\sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbb{1}\,\{|\beta_{jk}| \leq \frac{\lambda}{2}\} &\leq C 2^{j_{\lambda,u}p/2}\lambda^p + \sum_{j \geq j_{\lambda,u}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\leq C\left((u(\lambda))^{2sp/(1+2s)} + (u(\lambda_u))^{2sp}\right) \\
&= C\left((u(\lambda))^{2sp/(1+2s)} + D_1\right).
\end{aligned}$$

Since $n \in \mathbb{N}$, $C_n \leq O(2^n)$, one gets $u(\lambda_u) = u(u(\lambda)^{-2s/(1+2s)}\lambda) \leq Cu(\lambda)^{1/(1+2s)}$. So :

$$\begin{aligned}
D_1 &= (u(\lambda_u))^{2sp} \\
&\leq C(u(\lambda))^{2sp/(1+2s)}.
\end{aligned}$$

So $f \in W_{\mu,u}(r,p)$. □

In the sequel, we prove that under conditions of Theorem 6.2, the μ-thresholding rules $\hat{f}_{\mu^{(i)}}$ $(1 \le i \le 5)$ can be discriminated in the maxiset sense.

In the following proposition, we compare the maximal spaces associated with the five examples of μ-thresholding rules defined in paragraph 6.2. In particular we prove that hard thresholding rules are outperformed by hard tree rules and by block thresholding rules when the length of the blocks are correctly chosen. Indeed :

Proposition 6.2. *For any $1 \le p < \infty$ and any $m \ge 4\sqrt{p+1}$, let*

$$MS(\hat{f}_{\mu^{(i)}}, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}), \quad 1 \le i \le 5,$$

be respectively the maximal sets of procedures $\hat{f}_{\mu^{(i)}}$ for the rate of convergence $(u(\lambda_\epsilon))^{2sp/(1+2s)}$. Under conditions of Theorem 6.2, we have the following inclusions spaces :

$$MS(\hat{f}_{\mu^{(1)}}, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}) \subset MS(\hat{f}_{\mu^{(i)}}, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}), \text{ for } i \in \{3,4,5\} \quad (6.9)$$

$$MS(\hat{f}_{\mu^{(2)}}, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}) \subset MS(\hat{f}_{\mu^{(4)}}, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}), \quad (6.10)$$

and :

$$MS(\hat{f}_{\mu^{(4)}}, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}) \subset MS(\hat{f}_{\mu^{(3)}}, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}). \quad (6.11)$$

Proof :
Using Corollary 6.1, we have for any $1 \le i \le 5$:

$$MS(\hat{f}_{\mu^{(i)}}, \|.\|^p_{\mathcal{B}^0_{p,p}}, (u(\lambda_\epsilon))^{2sp/(1+2s)}) = \mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu^{(i)},u}(\frac{p}{1+2s}, p).$$

Now, for any f as (6.1) we have :

$$\max_{k' \in \mathcal{P}_j(k)} |\beta_{jk'}| \le \lambda \Longrightarrow |\beta_{jk}| \le \lambda,$$

$$\max(|\beta_{jk}|^p, \frac{1}{l_j} \sum_{k' \in \mathcal{P}_{jk}} |\beta_{jk'}|^p) \le \lambda^p \Longrightarrow |\beta_{jk}| \le \lambda,$$

and

$$\forall I_{j'k'} \in \mathcal{T}_{jk}(\lambda_\epsilon), |\beta_{j'k'}| \le \frac{\lambda}{2} \Longrightarrow |\beta_{jk}| \le \lambda.$$

So, using Proposition 6.1 the inclusion spaces (6.9) holds. In the same way, since :

$$\max(|\beta_{jk}|^p, \frac{1}{l_j}\sum_{k'\in\mathcal{P}_{jk}}|\beta_{jk'}|^p) \leq \lambda^p \Longrightarrow \frac{1}{l_j}\sum_{k'\in\mathcal{P}_{jk}}|\beta_{jk'}|^p \leq \lambda^p$$

and :

$$\max_{k'\in\mathcal{P}_j(k)}|\beta_{jk'}| \leq \lambda \Longrightarrow \max(|\beta_{jk}|^p, \frac{1}{l_j}\sum_{k'\in\mathcal{P}_{jk}}|\beta_{jk'}|^p) \leq \lambda^p,$$

the inclusions spaces (6.10) and (6.11) hold too. □

The previous proposition is important. Indeed, we see that hard tree rules and block thresholding rules with length of blocks small enough can outperform hard thresholding ones. More precisely,

Proposition 6.3. *Under the maxiset approach associated with the rate* $(u(\lambda_\epsilon))^{2sp/(1+2s)}$, *we have the following results :*

[Hard tree rules] *For any* $p \geq 2$, *the hard tree rule* $\hat{f}_{\mu^{(5)}}$ *outperform the hard thresholding rule in the maxiset sense.*

[Block thresholding rules] *For any* $1 \leq p < \infty$, *maximean- and maximum-block(p) thresholding rules such that the lengths* l_j *of the blocks* $\mathcal{P}_{jk}$ *does not exceed* $C\left(\log(\epsilon^{-1})\right)^{\frac{p}{2}}$, *for some* $C > 0$, *outperform hard thresholding rules in the maxiset sense.*

Proof :
It is just a consequence of the previous proposition. The condition $p \geq 2$ (resp. $l_j \leq C\left(\log(\epsilon^{-1})\right)^{\frac{p}{2}}$) ensures that condition (6.6) of Theorem 6.2 is satisfied when dealing with hard tree rules (resp. block thresholding rules). □

The first part of Proposition 6.3 generalizes the maxiset result of chapter 5 for the hard tree rule. The second part of Proposition 6.3 allows to give a theorical explication about the good performances of block thresholding rules which have been observed in the practical setting (see Hall, Penev, Kerkyacharian and Picard (1997[56]) and Cai (1998[16], 1999[17], 2002[18])).

Bibliographie

[1] Abramovich, F., Amato, U. and Angelini, C. (2004). On optimality of Bayesian wavelet estimators. *Scand. J. Statist.*, **31**(2), 217-234.

[2] Abramovich, F. and Benjamini, Y. (1995). Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In *Wavelets and Statistics,* pages 5-14, Springer, New York.

[3] Abramovich, F., Benjamini, Y., Donoho, D. L. and Johnstone, I. M. (2000). Adapting to unknown sparsity by controlling the false discovery rate. Technical report.

[4] Abramovich, F., Sapatinas, T. and Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **60**(4), 725-749.

[5] Antoniadis, A., Bigot, J. and Sapatinas (2001). Wavelet estimators in nonparametric regression : a comparative simulation study. *Journal of Statistical Software*, **6**(6), 1-83.

[6] Antoniadis, A., Leporini, D. and Pesquet, J.-C (2002). Wavelet thresholding for some classes of non-Gaussian noise. *Statist. Neerlandica*, **56**(4), 434-453.

[7] Bakushinski, A.B. ((1969). On the construction of regularizing algorithms under random noise. *Soviet Math. Doklady*, 189 : 231-233.

[8] Bergh, J. and Löfström, J. (1976). *Interpolation spaces. An introduction.* Springer-Verlag, Berlin. Grundlehren der Mathematischen Wissenschaften, No. 223.

[9] Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete*, **65**(2), 181-237.

[10] Birgé, L. (1985). Nonasymptotic minimax risk for Hellinger balls. *Probab. Math. Statist.*, **5**(1), 21-29.

[11] Birgé, L. and Massart, P. (1997) From model selection to adaptive estimation. *Festschrift for Lucien Le Cam*, pages 55-87. Springer, New York (1997).

[12] Birgé, L. and Massart, P. (2000) An adaptive compression algorithm in Besov spaces. *Constr. Approx.*, **16**(1), 1-36.

[13] Birgé, L. and Massart, P. (2001) Gaussian model selection. *J. Eur. Math. Soc. (JEMS)*, **3**(3), 203-268.

[14] Bretagnolle, J. and Huber, C. (1979). Estimation des densités : risque minimax. *Z. Wahrsch. Verw. Gebiete*, **47**(2), 119-137.

[15] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, **24**(6), 2384-2398.

[16] Cai, T. (1998). Numerical comparisons of BlockJS estimator with conventional wavelet methods. Unpublished manuscript.

[17] Cai, T. (1999). Adaptive wavelet estimation : a block thresholding and oracle inequality approach. *Ann. Statist.*, **27**(3), 898-924.

[18] Cai, T. (2002). On block thresholding in wavelet regression : adaptivity, block size and threshold level. *Statist. Sinica*, **12**(4), 1241-1273.

[19] Cavalier, L. (1998). Asymptotically efficient estimation in a problem related to tomography. *Math. Methods Statist.*, **7**(4), 445-456.

[20] Cavalier, L., Golubev, G.K., Picard, D. and Tsybakov A.B. (2002). Oracle inequalities for inverse problems. *Ann. Statist.*, **30**(3), 843-874.

[21] Cavalier, L. and Tsybakov A.B. (2001). Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation. *Math. Methods Statist.*, **10**(3), 247-282.

[22] Cavalier, L. and Tsybakov A.B. (2002). Sharp adaptation for inverse problems. *Probab. Theory Rel. Fields*, **123**(3), 323-354.

[23] Chen, S. S., Donoho, D. L. and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput*, **20**(1), 33-61.

[24] Chipman, H. A., Kolaczyk, E. D. and McCulloch, R. E. (1997). Adaptive bayesian wavelet shrinkage. *Journal of the American Statistical Association*, **92**, 1413-1421.

[25] Clyde, M. and George, E.I (1998). Robust empirical Bayes estimation in wavelets. Restricted nonlinear approximation. *Technical Report.*

[26] Clyde, M. and George, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **62**(4), 681-698.

[27] Clyde, M., Parmigiani, G. and Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika*, **85**(2), 391-401.

[28] Cohen, A. (2000). Wavelet methods in numerical analysis. In *Handbook of numerical analysis, Vol. VII*, pages 417-711. North-Holland, Amsterdam.

[29] Cohen, A., Dahmen W., Daubechies I., and DeVore, R. (2001). Tree Approximation and Optimal Encoding. *Appl. Comput. Harmon. Anal.*, **11**(2), 192-226.

[30] Cohen, A., DeVore, R. A. and Hochmuth, R. (2000). Restricted nonlinear approximation. *Constr. Appro.*, **16**(1), 85-113.

[31] Cohen, A., DeVore, R., Kerkyacharian, G., and Picard, D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, **11**, 167-191.

[32] Crouse, M.S., Nowak, R.D. and Baraniuk, R.G. (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Process.*, **46**(4), 886-902.

[33] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, **41**(7), 909-996.

[34] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

[35] Davis, G., Mallat, S. and Zhang, Z. (1994). Adaptive time frequency approximations with matching pursuits. In *Wavelets : theory, algorithms, and applications (Taormina, 1993)*, pages 271-293. Academic Press, San Diego, CA.

[36] DeVore, R.A. (1989). Degree of nonlinear approximation. In *Approximation theory VI,* **1**(College Station, TX, 1989) 175-201. Academic Press, Boston, MA.

[37] DeVore, R.A., Konyagin, S.V., and Temlyakov, V.N. (1998). Hyperbolic wavelet approximation. *Constr. Approx.*, **14**(1), 1-26.

[38] DeVore, R.A., and Lorentz, G.G. (1993). Constructive approximation. *Springer-Verlag*, Berlin.

[39] Donoho, D.L. (1993). Unconditional bases are optimal bases for data compression and for statistical estimation. *Appl. Comput. Harmon. Anal.*, **1**(1), 100-115.

[40] Donoho, D.L. (1996). Unconditional bases and bit level compression. *Appl. Comput. Harmon. Anal.*, **3**(4), 388-392.

[41] Donoho, D.L. (1997). CART and Best-ortho-basis. *Ann. Statist.*, **25**(5), 1870-1911.

[42] Donoho, D.L., and Johnstone, I.M. (1994). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields*, **99**(2), 277-303.

[43] Donoho, D.L., and Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**(3), 425-455.

[44] Donoho, D.L., and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, **90**(432), 1200-1224.

[45] Donoho, D.L., and Johnstone, I.M. (1996). Neo-classical minimax problems, thresholding and adaptive function estimation. *Bernoulli*, **2**(1), 39-62.

[46] Donoho, D.L., and Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**(3), 879-921.

[47] Donoho, D.L., Johnstone, I.M, Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage : asymptotia ? *J. Roy. Statist. Soc. Ser. B.*, **57**(2), 301-369. With discussion and a reply by the authors.

[48] Donoho, D.L., Johnstone, I.M, Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.*, **24**(2), 508-539.

[49] Donoho, D.L., Johnstone, I.M, Kerkyacharian, G., and Picard, D. (1997). Universal near minimaximity of wavelet shrinkage. In *Festschrift for Lucien Le Cam*, 183-218. Springer, New-York.

[50] DeVore, R.A. (1989). Degree of nonlinear approximation. In *Approximation theory VI,* **1**(College Station, TX, 1989) 175-201. Academic Press, Boston, MA.

[51] Engel, J. (1994). A simple wavelet approach to nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.*, **49**(2), 242-254.

[52] Farrell, R. H. (1967). On the lack of a uniformly consistent sequence of estimators of a density function in certain cases. *Ann. Math. Statist.*, **38**, 471-474.

[53] Golubev, G.K. (1987). Adaptive asymptotically minimax estimates of smooth signals. *Problems of Information Trans.*, **23**, 57-67.

[54] Golubev, G.K., Levit B.Y. (1996). Distribution function estimation : adapting smoothing. *Math. Methods Statist.*, **5**(4), 383-403.

[55] Hall, P., Kerkyacharian, G., and Picard, D. (1999). On the minimax optimality of Block thresholded wavelet estimators. *Statist. Sinica*, **9**(1), 33-49.

[56] Hall, P., Penev, S., Kerkyacharian, G., and Picard, D. (1997). Numerical performance of block thresholded wavelet estimators. *Statist. Comput.*, **7**, 115-124.

[57] Härdle W., Kerkyacharian, G., Picard, D., and Tsybakov, A.B. (1998). Wavelets, approximation, and statistical applications. *Springer Verlag*, New-York.

[58] Huang,H.-C. and Cressie, N. (2000). Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data. *Technometrics*, **42**(3), 262-276.

[59] Ibragimov, I. A. and Khasminski, R. Z. (1981). *Statistical estimation.* Springer-Verlag, New-York. Asymptotic theory, translated from the Russian by Samuel Kotz.

[60] Jaffard, S. (1998). Oscillation spaces : properties and applications to fractal and multifractral functions. *J. Math. Phys.*, **39**(8), 4129-4141.

[61] Jaffard, S. (2004). Beyond Besov spaces : Oscilllation spaces. *Technical Report. To appear in Constructive Approximation.*

[62] Jansen, M., Malfait, M. and Bultheel, A. (1997). Generalized cross validation for wavelet thresholding. *Signal Processing*, **56**, 33-44.

[63] Johnstone, I. (1994). Minimax Bayes, asymptotic minimax and sparse wavelet priors. In *Statistical decision theory and related topics, V (West Lafayette, IN, 1992)*, pages 303-326. Springer, New-York.

[64] Johnstone, I. (1999). Wavelet shrinkage for correlated data and inverse problems : adaptivity results. *Statistica Sinica*, **9**, 51-83.

[65] Johnstone, I. M. and Silverman, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.*, **18**(1), 251-280.

[66] Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B*, **59**(2), 319-351.

[67] Johnstone, I. M. and Silverman, B. W. (1998). Empirical Bayes approaches to mixture problems and wavelet regression. *Technical Report.*

[68] Johnstone, I. M. and Silverman, B. W. (2002). Empirical Bayes selection of wavelet thresholds. *Technical Report.*

[69] Johnstone, I. M. and Silverman, B. W. (2002). Risk bounds for empirical bayes estimates of sparse sequences. *Technical Report.*

[70] Johnstone, I. M. and Silverman, B. W. (2004). Needles and hay in haystacks : Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.*, **32**, 1594-1649.

[71] Juditsky, A. (1997) Wavelet estimators : Adapting to unknown smoothness.*Math. Methods Stat.*, **6**(1), 1-25.

[72] Juditsky, A., Lambert-Lacroix, S. (2004) On minimax density estimation on $\mathbb{R}$. *Bernoulli*, **10**(2), 187-220.

[73] Kerkyacharian, G., and Picard, D. (1992). Density estimation in Besov space. *Statist. Probab. Lett.*, **13**(1), 15-24.

[74] Kerkyacharian, G., and Picard, D. (1993). Density estimation by kernel and wavelets methods : optimality of Besov spaces. *Statist. Probab. Lett.*, **18**(4), 327-336.

[75] Kerkyacharian, G., and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases. *Test*, **9**(2), 283-344. With comments, and a rejoinder by the authors.

[76] Kerkyacharian, G., and Picard, D. (2002). Minimax or maxisets ? *Bernoulli*, **8**(2), 219-253.

[77] Korostelev, A.P. and Tsybakov, A.B. (1993). *Minimax theory of image reconstruction.* Springer-Verlag, New York.

[78] Lepski, O.V. (1991). Asymptotically minimax adaptive estimation I : Upperbounds. Optimally adaptive estimates. *Theory Probab. Appl.*, **36**, 682-697.

[79] Lepski, O.V., Mammen, E. and Spokoiny, V.G. (1997). Optmal spatial adaptation to inhomogeneous smoothness : an approach based on kernel estimates with variable bandwidth selection. *Ann. Statist.*, **25**(3), 929-947.

[80] Lepski, O.V. and Spokoiny, V.G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, **25**(6), 2512-2546.

[81] Lorentz, G.G. (1950). Some new functional spaces. *Ann. of Math.*, **51**(2), 37-55.

[82] Lorentz, G.G. (1966). Metric entropy and approximation. *Bull. Amer. Math. Soc.*, **72**, 903-937.

[83] Loubes, J.M. and van de Geer, S. (2002). Adaptive estimation with soft thresholding penalties. *Statist. Neerlandica*, **56**(4), 454-479.

[84] Mallat, S. (1989). Multiresolution approximations and wavelet orthonormal bases of $\mathbb{L}^2(r)$. *Trans. Amer. Math. Soc.*, **315**(1), 69-87.

[85] Mallat, S. (1998). *A wavelet tour of signal processing.* Academic Press Inc., San Diego, CA.

[86] Mammen, E. (1990). A short note on optimal bandwidth selection for kernel estimators. *Statist. Probab. Lett.*, **9**(1), 23-25.

[87] Mammen, E. (1995). On qualitative smoothness of kernel density estimates. *Statistics*, **26**(3), 253-267.

[88] Mammen, E. (1998). Local adaptivity of kernel estimates with plug-in local bandwidth selectors. *Scand. J. Statist.*, **25**(3), 503-520.

[89] Meyer, Y. (1992). *Wavelets and operators.* Cambridge University Press, Cambridge. Translated from the 199 French original by D. H. Salinger.

[90] Müller, P. and Vidakovic, B. (1995). Wavelet shrinkage with affine bayes rules with applications. *Technical report.*

[91] Nadaraya, E.A. (1992). Limit distribution of a square deviation of a generalized kernel estimator for the density. *Theory Probab. Appl.*, **37**(2), 383-392.

[92] Nason, G. P. (1996). Wavelet shrinkage using cross-validation. *J. Roy. Statist. Syst. Sci. B*, **23**(6), 1-11.

[93] Nemirovski, A. S. (1986). Nonparametric estimation of smooth regression functions. *J. Comput. Statist. Soc. Ser. B*, **58**(2), 463-479.

[94] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, **24**(6), 2399-2430.

[95] Ogden, T. and Parzen, E. (1996). Change-point approach to data analytic wavelet thresholding. *Statistics and Computing*, **6**, 93-99.

[96] Ogden, T. and Parzen, E. (1996). Data dependent wavelet thresholding in nonparametric regression with change-point application. *Computational Statistics and Data Analysis*, **22**, 53-70.

[97] Parzen, E. (1962). On the estimation of a probability density function and mode. *Annals of Math. Statist.*, **33**, 1065-1076.

[98] Peetre, J. (1976). *New thoughts on Besov spaces.* Mathematics Department, Duke University, Durham, N.C. Duke University Mathematics Series, No. 1.

[99] Picard, D., and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.*, **28**(1), 298-335.

[100] Rejtö, L. and Rèvész, P. (1973). Density estimation and pattern classification. *Prob. of control and Information Theory*, **2**(1), 67-80.

[101] Rivoirard, V. (2002). Non linear estimation over weak Besov spaces. *Technical report.*

[102] Rivoirard, V. (2004). Maxisets for linear procedures. *Statist. Probab. Lett.*, **67**, 267-275.

[103] Rivoirard, V. (2004). Bayesian modelization of sparse sequences and maxisets for Bayes rules *Technical Report. Submitted to Math. Methods Statist.*

[104] Rivoirard, V. (2004). Thresholding Procedure with Priors Based on Pareto Distributions. *Test*, **13**(1), 213-246.

[105] Shorack, G. and Wellner, J. (1986) Empirical processes with Applications to Statistics.

[106] Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Roy. Statist. Soc. Ser. B*, **47**(1), 1-52. With discussion.

[107] Steinberg, D. M. (1990). A Bayesian approach to flexible modeling of multivariable response functions. *J. Multivariate Anal.*, **34**(2), 157-172.

[108] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**(4), 1040-1053.

[109] Sudakov, V.N. and Khalfin, L.A (1964). Statistical approach to ill-posed problems in mathematical physics. *Soviet Math. Doklady*, 157 : 1094-1096.

[110] Temlyakov, V. N. (1999). Greedy algorithms and m-term approximation with regard to redundant dictionaries. *J. Approx. Theory*, **98**(1), 117-145.

[111] Tsybakov, A. B. (2000). On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci. Paris Sér. I Math.*, **330**(9), 835-840.

[112] Tsybakov, A. B. (2004). Introduction à l'estimation non-paramétrique. *Mathématiques & Applications (Berlin),41.* Springer-Verlag, Berlin, 2004.

[113] van de Geer, S. (2000). Least squares estimation with complexity penalties. *Math. Methods Statist.* , **10**(3), 355-374.

[114] van de Geer, S. (2003). Asymptotic theory for maximum likelihood in nonparametric mixture models. *Comput. Statist. Data Anal.* , **41**(3-4), 453-464.

[115] Vannucci, M. and Corradi, F. (1999). Modeling dependence in the wavelet domai. In *Bayesian inference in wavelet-based models*, pages 173-186, Springer, New York.

[116] Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *J. Amer. Statist. Assoc.*, **93**(441), 173-179.

[117] Vidakovic, B. and Ruggeri, F. (2001). BAMS method : theory and simulations. *Sankhya Ser. B,* **63**(2), 234-249. Special issue on wavelets.

[118] Wahba, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *Ann. Statist.*, **9**(1), 146-156.

[119] Young, A. S. (1977). A Bayesian approach to prediction using polynomials. *Biometrika*, **64**(2), 309-317.

[120] Zhang, C.-H. (2002). General empirical bayes wavelet methods. *Technical report.*