Test (2008) 17: 332–349
DOI 10.1007/s11749-006-0035-5

ORIGINAL PAPER

Maxisets for μ-thresholding rules

Florent Autin

Received: 3 January 2005 / Accepted: 18 June 2006 /
Published online: 1 March 2007

Abstract In this paper, we study the performances of a large class of procedures, called μ-thresholding rules. At first, we exhibit the maximal spaces (or maxisets) where these rules attain given rates of convergence when considering the Besov-risk. Then, we point out a way to construct μ-thresholding rules for which the maxiset contains the hard thresholding rule's one. In particular, we prove that procedures which consist in thresholding coefficients by groups, as block thresholding rules or thresholding rules with tree structure, outperform in the maxiset sense procedures which consist in thresholding coefficients individually.

Keywords Adaptive procedures · Besov spaces and weak Besov spaces · Maximal space · Minimax risk · Thresholding rules

Mathematics Subject Classification (2000) Primary 62G05 · Secondary 62G20 · 65T60

1 Introduction

Wavelet methods in nonparametric estimation of functions are renown for their adaptivity across a wide range of function classes. Frequently, the performance of a wavelet estimator is measured thanks to the minimax approach. The final outcome of this approach is to exhibit estimates which attain the minimax rate of convergence of a functional space V containing f. However, this point of view seems to be pessimistic, since it requires the knowledge of V.

Recently, Cohen et al. (2001) have suggested an alternative approach to measure the performance of an estimation procedure which consists in exhibiting the functional space (maxiset) over which an estimator attains a given rate of convergence.

F. Autin (✉)
Centre de Mathématiques et Informatique, Université d'Aix-Marseille 1, 39 rue F. Joliot Curie,
13453 Marseille Cedex 13, France
e-mail: autin@cmi.univ-mrs.fr

This approach is more optimistic than the minimax one, since it draws the strong connection between a given procedure and a functional set for which the procedure is well adapted. In this way, Cohen et al. (2001) and Kerkyacharian and Picard (2000) have shown that Besov spaces and weak Besov spaces are directly connected with the hard thresholding rules out. More than this, they have proved that thresholding rules outperform linear procedures, dealing with the density estimation model.

In this paper, we investigate the performances of a large class of procedures: the μ-thresholding rules which can be viewed as a generalization of usual (hard, global, and block) thresholding rules. Each μ-thresholding rule is associated with a sequence of positive functions $(\mu_{jk})_{j,k}$ and consists in only keeping the empirical wavelet coefficients y_{jk} for which $\mu_{jk}(\lambda, y_\lambda)$—where y_λ represents a particular set of empirical coefficients depending on λ—are strictly larger than a threshold λ. Under the maxiset approach and choosing the Besov risk, we exhibit the maximal spaces where these procedures attain given rates of convergence (Theorem 3.1). Then, we give a way to construct μ-thresholding rules at least as good as the hard thresholding rule in the maxiset sense (Proposition 4.1) and give many examples of such rules. Particularly, we prove that block thresholding rules outperform in the maxiset sense the hard thresholding rule on condition that the length of their blocks are small enough (Proposition 4.3). This result is important, since it allows us to give a theoretical explanation about the good performances of these estimators often observed in the practical setting (see Hall et al. 1997, 1999 and Cai 1998, 1999, 2002). In the same way, we show that the thresholding rule using tree structure proposed by Autin (2004) is another example of a μ-thresholding rule which outperforms hard thresholding rule. In other words, thanks to the maxiset approach, we prove in this paper that rules constructed with thresholding methods applied to *groups of empirical coefficients* are preferable to rules based on thresholding methods applied to *individual empirical coefficients*.

The paper is organized as follows. Section 2 is devoted to the model and to the definition of μ-thresholding rules illustrated by some examples. In Sect. 3, we exhibit the maximal spaces associated with such procedures and discuss around. Section 4 aims at comparing the performances of some particular μ-thresholding rules.

2 Model and classes of estimators

2.1 Model

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

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

where

- $0 < \varepsilon < 1/2$ is the noise level
- f is a function defined on [0, 1]
- $W(\cdot)$ is a Brownian motion

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 the sake of simplicity, we suppose that, for some $a \in \mathbb{N}^*$, the supports of ϕ and ψ are included in $[0, a]$ and we denote ψ_{-1k} to design ϕ_{0k}.

Any $f \in \mathbb{L}_2([0,1])$ can be represented as

$$f = \sum_{j \geq -1} \sum_{-a<k<2^j} \beta_{jk} \psi_{jk} = \sum_{j \geq -1} \sum_{-a<k<2^j} (f, \psi_{jk})_{L_2} \psi_{jk}. \tag{2.1}$$

Let us suppose that we dispose of observations $y_{jk} = X_\varepsilon(\psi_{jk}) = \beta_{jk} + \varepsilon \xi_{jk}$, where ξ_{jk} are independent Gaussian variables $\mathcal{N}(0,1)$.

In the sequel, we set $2^{j_\lambda} \sim \lambda^{-2}$ to design the smallest integer j_λ such that $2^{-j_\lambda} \leq \lambda^2$ and we denote

$$t_\varepsilon = \varepsilon \sqrt{\log\left(\varepsilon^{-1}\right)}.$$

In the following paragraph, we define the class of procedures we shall study along the paper, the μ-thresholding rules.

2.2 Definition of μ-thresholding rules and examples

For any $\lambda > 0$, let us denote for any sequence $(y_{jk})_{j,k}$ and any sequence $(\beta_{jk})_{j,k}$:

$$y_\lambda = \left(y_{jk};\ (j,k) \in I_\lambda\right),$$
$$\beta_\lambda = \left(\beta_{jk};\ (j,k) \in I_\lambda\right),$$

where $I_\lambda = ((j,k);\ -1 \leq j < j_\lambda, -a < k < 2^j)$ and $2^{j_\lambda} \sim \lambda^{-2}$.

Remark 2.1 For any $\lambda > 0$, the number $\# I_\lambda$ of elements belonging to I_λ is less than or equal to $a 2^{j_\lambda}$.

This remark is important since it shall be often used in the proofs of results presented in the paper.

Let us consider the following class of *Keep-or-Kill estimators*:

$$\mathcal{F}_K = \left\{ \hat{f} = \sum_j \sum_k \gamma_{jk} y_{jk} \psi_{jk};\ \gamma_{jk}(\varepsilon) \in \{0,1\}\ \textit{measurable} \right\}.$$

Definition 2.1 We say that $\hat{f}_\mu \in \mathcal{F}_K$ is a μ-thresholding rule if

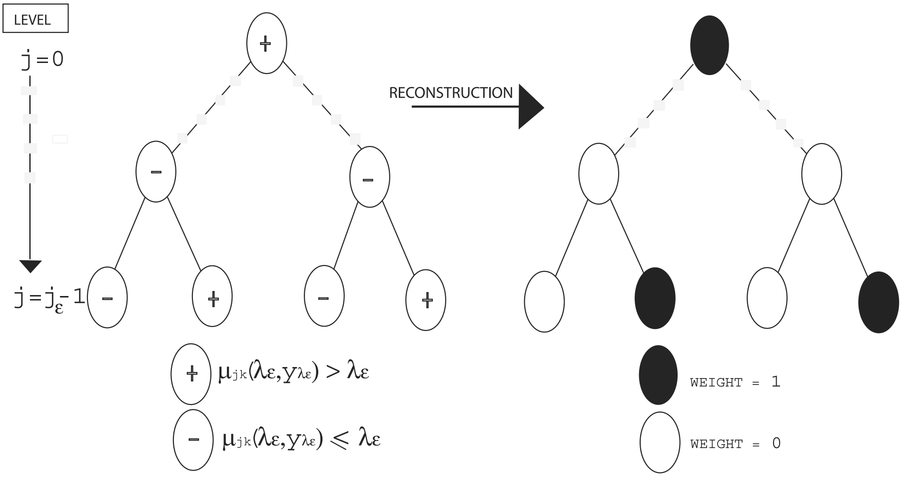
$$\hat{f}_\mu = \sum_{j=-1}^{j_\varepsilon - 1} \sum_k \mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) > \lambda_\varepsilon\right\} y_{jk} \psi_{jk}, \tag{2.2}$$

where $\lambda_\varepsilon = m t_\varepsilon$, $m > 0$, $2^{j_\varepsilon} \sim \lambda_\varepsilon^{-2}$, and $(\mu_{jk}(\lambda, \cdot) : \mathbb{R}^{\# I_\lambda} \longrightarrow \mathbb{R}^+)_{j,k}$ is, for any $\lambda > 0$, a sequence of positive functions such that, for any $t \in \mathbb{R}$ and any $(y_\lambda, \beta_\lambda) \in \mathbb{R}^{\# I_\lambda} \times \mathbb{R}^{\# I_\lambda}$:

$$\left|\mu_{jk}(\lambda, y_\lambda) - \mu_{jk}(\lambda, \beta_\lambda)\right| > t \quad \Longrightarrow \quad \exists (j_o, k_o) \in I_\lambda / |y_{j_o k_o} - \beta_{j_o k_o}| > t. \tag{2.3}$$

Let us notice that any μ-thresholding rule is a *limited procedure* (see Autin et al. 2006) in the sense that the reconstruction of f by such a procedure does not use the empirical coefficients y_{jk} for which $j \geq j_\varepsilon$.

The reconstruction of the signal f by a μ-thresholding rule consists in keeping the empirical coefficients y_{jk} at level strictly less than j_ε for which $\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon})$ are 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. **Hard thresholding procedure** belongs to the family of μ-thresholding rules. It corresponds to the choice

$$\mu^{(1)}_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) = |y_{jk}|.$$

   This procedure has been proved to have good performances in the minimax point of view (see Donoho et al. 1995, 1996, 1997) and in the maxiset point of view (see Cohen et al. 2001 and Kerkyacharian and Picard 2000).
2. **Block thresholding procedures** belong to the family of μ-thresholding rules. They correspond to the following choices:

   Mean-block(p) thresholding

$$\mu^{(2)}_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) = \left(\frac{1}{l_j}\sum_{k' \in \mathcal{P}_j(k)} |y_{jk'}|^p\right)^{\frac{1}{p}},$$

   Maximum-block thresholding

$$\mu^{(3)}_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) = \max_{k' \in \mathcal{P}_j(k)} |y_{jk'}|,$$

Maximean-block(p) thresholding

$$\mu_{jk}^{(4)}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) = \max\big(|y_{jk}|, \mu_{jk}^{(2)}(\lambda_\varepsilon, y_{\lambda_\varepsilon})\big),$$

where, for any (j,k) and any $0<\varepsilon<\frac{1}{2}$,

- $k \in \mathcal{P}_j(k) \subset \{1-a, \ldots, 2^j - 1\}$
- $\#\mathcal{P}_j(k) = l_j$ 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 et al. (1997) 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 coefficients within the block based on the magnitude of the estimate. As well as the family blockwise James–Stein estimators (see Cai 1998, 1999, 2002), on condition that the length of blocks is not exceeding $C\log(\varepsilon^{-1})$ $(C>0)$, this estimator was shown to have good performances in the practical setting (see Hall et al. 1997) and was proved to attain exactly the minimax rate of convergence for the $\mathbb{L}_2$-risk without the logarithmic penalty over a range of perturbed Hölder classes (Hall et al. 1999).

3 Maxisets associated with μ-thresholding rules

Dealing with the Besov risk $\mathcal{B}^0_{p,p}$, we aim at exhibiting in this section the maximal spaces where the μ-thresholding rules attain the rate of convergence $(u(\lambda_\varepsilon))^{2sp/(1+2s)}$ $(s>0$ and $1\le p<\infty)$, where u is an increasing transformation map of $\mathbb{R}^+$ into $\mathbb{R}^+$ that is continuous and satisfies:

$$\forall 0<\varepsilon<1/2, \quad \varepsilon \le u(\lambda_\varepsilon). \tag{3.1}$$

Remark 3.1 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.

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 3.1 Let $s>0$ and $1\le p<\infty$. We 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

$$\sup_{\lambda>0}\big(u(\lambda)\big)^{-2sp}\sum_{j\ge j_\lambda} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p < \infty.$$

Note that $\mathcal{B}^s_{p,\infty}(Id_{\mathbb{R}^+})$ is the classical Besov space, which has been proved to contain the maximal space of arbitrary *limited rule* for the rate $\lambda_\varepsilon^{2sp}$ (see Autin et al. 2006).

Definition 3.2 Let $0 < r < p < \infty$. We say that a function f belongs to the space $W_{\mu,u}(r, p)$ if and only if

$$\sup_{\lambda>0}\big(u(\lambda)\big)^{r-p}\sum_{j<j_\lambda}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbf{1}\Big\{\mu_{jk}(\lambda,\beta_\lambda)\le\frac{\lambda}{2}\Big\}<\infty.$$

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 et al. 2001 and Kerkyacharian and Picard 2000).

Definition 3.3 Let $0 < r < p < \infty$. We say that a function f belongs to the space $W^*_{\mu,u}(r, p)$ if and only if:

$$\sup_{\lambda>0}\lambda^p\big(u(\lambda)\big)^{r-p}\big(\log\big(\lambda^{-1}\big)\big)^{-\frac{p}{2}}\sum_{j<j_\lambda}2^{j(\frac{p}{2}-1)}\sum_k\mathbf{1}\big\{\mu_{jk}(\lambda,\beta_\lambda)>2\lambda\big\}<\infty.$$

In this paper, we shall see the strong relation between $W_{\mu,u}(r, p)$, $W^*_{\mu,u}(r, p)$ and μ-thresholding rules.

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.

3.2 Main result

The following theorem deals with the maximal spaces associated to μ-thresholding rules.

Theorem 3.1 *Let $1 \le p < \infty$ and $m \ge 4\sqrt{p+1}$. Denote $\lambda_\varepsilon = mt_\varepsilon$ 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 $\lambda_{\mathrm{seuil}} > 0$ such that*

$$\forall 0<\lambda<\lambda_{\mathrm{seuil}},\qquad u(4m\lambda)\le K_m u(\lambda), \tag{3.2}$$

then we have equivalence between:

(i) $\sup_{0<\varepsilon<1/2}(u(\lambda_\varepsilon))^{-2sp/(1+2s)}\mathbb{E}\|\hat{f}_\mu-f\|^p_{\mathcal{B}^0_{p,p}}<\infty,$

(ii) $f\in\mathcal{B}^{s/(1+2s)}_{p,\infty}(u)\cap W_{\mu,u}\big(\frac{p}{1+2s},p\big)\cap W^*_{\mu,u}\big(\frac{p}{1+2s},p\big).$

With the usual maxiset notation, the previous equivalence can be formulated as follows:

$$\mathrm{MS}\big(\hat{f}_\mu,\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big)=\mathcal{B}^{s/(1+2s)}_{p,\infty}(u)\cap W_{\mu,u}\bigg(\frac{p}{1+2s},p\bigg)\cap W^*_{\mu,u}\bigg(\frac{p}{1+2s},p\bigg)$$

Remark 3.2 When $u(t_\varepsilon) = t_\varepsilon$ (resp. $u(t_\varepsilon) = \varepsilon$), notice that (3.2) is satisfied by taking $K_m = 4m$ (resp. $K_m = 4\sqrt{2}m$) and $\lambda_{\mathrm{seuil}} = \frac{\sqrt{\log(2)}}{2}$ $\big($resp. $\lambda_{\mathrm{seuil}} = \frac{\sqrt{\log(32m^2)}}{32m^2}\big)$.

Proof of Theorem 3.1 Here and later, we shall denote by C a constant which may be different from line to line.

(i) $\Rightarrow$ (ii) It suffices to prove the result for $0<\varepsilon<(\varepsilon_{\text{seuil}}\vee m^{-4})$, where $\varepsilon_{\text{seuil}}$ is such that $t_{\varepsilon_{\text{seuil}}}=\lambda_{\text{seuil}}$. For any $0<\varepsilon<(\varepsilon_{\text{seuil}}\vee m^{-4})$, we have

$$\sum_{j\geq j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \leq \mathbb{E}\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}} \leq C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}.$$

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

$$\sup_{\lambda>0}\big(u(\lambda)\big)^{-2sp/(1+2s)}\sum_{j\geq j_\lambda} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p<\infty.$$

It comes that $f\in\mathcal{B}^{s/(1+2s)}_{p,\infty}(u)$.

Moreover,

$$\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})\leq\frac{\lambda_\varepsilon}{2}\right\}=A_1+A_2,$$

where

$$\begin{aligned}
A_1&=\mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})\leq\frac{\lambda_\varepsilon}{2}\right\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})\leq\lambda_\varepsilon\big\}\\
&\leq\mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})\leq\lambda_\varepsilon\big\}\\
&\leq\mathbb{E}\|\hat{f}_\mu - f\|^p_{\mathcal{B}^0_{p,p}}\leq C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)},
\end{aligned}$$

and, using (2.3), the concentration properties of the Gaussian distribution and Remark 2.1,

$$\begin{aligned}
A_2&=\mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})\leq\frac{\lambda_\varepsilon}{2}\right\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})>\lambda_\varepsilon\big\}\\
&=\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbb{P}\big(\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})>\lambda_\varepsilon\big)\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})\leq\frac{\lambda_\varepsilon}{2}\right\}\\
&\leq\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbb{P}\left(\big|\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})-\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})\big|>\frac{\lambda_\varepsilon}{2}\right)\\
&=\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p\mathbb{P}\left(\exists(j_o,k_o)\in I_{\lambda_\varepsilon}\mid |y_{j_ok_o}-\beta_{j_ok_o}|>\frac{\lambda_\varepsilon}{2}\right)\\
&\leq C2^{j_\varepsilon}\varepsilon^{\frac{m^2}{8}}\leq C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}.
\end{aligned}$$

The last inequality is due to the fact that $m^2\geq 8(p+2)$.

Using the continuity of λ_ε in 0, we deduce that

$$\sup_{\lambda>0}\big(u(\lambda)\big)^{-2sp/(1+2s)}\sum_{j<j_\lambda}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p\mathbf{1}\Big\{\mu_{jk}(\lambda,\beta_\lambda)\le\frac{\lambda}{2}\Big\}<\infty.$$

It comes that $f\in W_{\mu,u}\big(\frac{p}{1+2s},p\big)$.

Let us now consider the following lemma:

Lemma 3.1 *For all* $0<\varepsilon\le m^{-4}$, $\lambda_\varepsilon^p(\log(\lambda_\varepsilon^{-1}))^{-p/2}\le(2m)^p\varepsilon^p$.

Proof Let $0<\varepsilon\le m^{-4}$. Since $\log(\varepsilon^{-1})\le\varepsilon^{-1}$, we have $m\varepsilon\sqrt{\log(\varepsilon^{-1})}\le m\sqrt{\varepsilon}$. Hence,

$$\begin{aligned}\log\big(\lambda_\varepsilon^{-1}\big)&=\log\bigg(\frac{1}{m\varepsilon\sqrt{\log(\varepsilon^{-1})}}\bigg)\ge\log\bigg(\frac{1}{m\sqrt{\varepsilon}}\bigg)\\&\ge\frac{1}{2}\log\bigg(\frac{1}{m^2\varepsilon}\bigg)\ge\frac{1}{4}\log\big(\varepsilon^{-1}\big).\end{aligned}$$

So,

$$\lambda_\varepsilon^p\big(\log\big(\lambda_\varepsilon^{-1}\big)\big)^{-p/2}\le2^p\lambda_\varepsilon^p\big(\log\big(\varepsilon^{-1}\big)\big)^{-p/2}\le(2m)^p\varepsilon^p.$$ □

Using Lemma 3.1, we have

$$\begin{aligned}&\lambda_\varepsilon^p\big(\log\big(\lambda_\varepsilon^{-1}\big)\big)^{-p/2}\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\\&\le C\varepsilon^p\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\\&=C\mathbb{E}\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k|y_{jk}-\beta_{jk}|^p\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\\&=C(A_3+A_4)\end{aligned}$$

with

$$A_3=\mathbb{E}\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k|y_{jk}-\beta_{jk}|^p\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})>\lambda_\varepsilon\big\},$$

$$A_4=\mathbb{E}\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k|y_{jk}-\beta_{jk}|^p\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})\le\lambda_\varepsilon\big\},$$

$$\begin{aligned}A_3&=\mathbb{E}\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k|y_{jk}-\beta_{jk}|^p\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})>\lambda_\varepsilon\big\}\\&\le\mathbb{E}\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k|y_{jk}-\beta_{jk}|^p\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})>\lambda_\varepsilon\big\}\\&\le\mathbb{E}\|\hat{f}_\mu-f\|_{\mathcal{B}^0_{p,p}}^p\le C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}.\end{aligned}$$

Using the Cauchy–Schwartz inequality, (2.3), and Remark 2.1, we get

$$\begin{aligned}
&\big(\mathbb{E}|y_{jk}-\beta_{jk}|^p\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})\le\lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\big)^2\\
&\quad\le\mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\mathbb{P}\big(\big|\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})-\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})\big|>\lambda_\varepsilon\big)\\
&\quad\le\mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\mathbb{P}\big(\exists(j_o,k_o)\in I_{\lambda_\varepsilon}\mid|y_{j_ok_o}-\beta_{j_ok_o}|>\lambda_\varepsilon\big)\\
&\quad\le a2^{j_\varepsilon}\mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\mathbb{P}\big(|y_{jk}-\beta_{jk}|>\lambda_\varepsilon\big),
\end{aligned}$$

where $\mathbb{E}|y_{jk}-\beta_{jk}|^{2p}=C\varepsilon^{2p}$ and $\mathbb{P}(|y_{jk}-\beta_{jk}|>\lambda_\varepsilon)\le\varepsilon^{m^2/2}$.

So, since $m^2\ge 4(1+p)$, from the concentration properties of the Gaussian distribution and Remark 2.1 one gets

$$\begin{aligned}
A_4&=\mathbb{E}\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k|y_{jk}-\beta_{jk}|^p\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})\le\lambda_\varepsilon\big\}\\
&\le\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k\mathbb{E}^{1/2}|y_{jk}-\beta_{jk}|^{2p}\mathbb{P}^{1/2}\big(\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})\le\lambda_\varepsilon\big)\\
&\quad\times\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,\beta_{\lambda_\varepsilon})>2\lambda_\varepsilon\big\}\\
&\le\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k\mathbb{E}^{1/2}|y_{jk}-\beta_{jk}|^{2p}\cdot\mathbb{P}^{1/2}\big(\exists(j_o,k_o)\in I_{\lambda_\varepsilon}\mid|y_{j_ok_o}-\beta_{j_ok_o}|>\lambda_\varepsilon\big)\\
&\le C2^{j_\varepsilon/2}\varepsilon^{m^2/4}\le C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}.
\end{aligned}$$

Using the continuity of λ_ε in 0, we deduce that

$$\begin{aligned}
&\sup_{\lambda>0}\lambda^p\big(u(\lambda)\big)^{-2sp/(1+2s)}\big(\log\big(\lambda^{-1}\big)\big)^{-p/2}\\
&\qquad\times\sum_{j<j_\lambda}2^{j(\frac{p}{2}-1)}\sum_k\mathbf{1}\big\{\mu_{jk}(\lambda,\beta_\lambda)>2\lambda\big\}<\infty.
\end{aligned}$$

It comes that $f\in W^*_{\mu,u}\big(\frac{p}{1+2s},p\big)$.

(ii) $\Rightarrow$ (i) For any $0<\varepsilon<(\varepsilon_{\text{seuil}}\vee m^{-4})$, we have

$$\begin{aligned}
\mathbb{E}\|\hat f_\mu-f\|^p_{\mathcal{B}^0_{p,p}}&=\mathbb{E}\sum_{j<j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k\big|y_{jk}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon,y_{\lambda_\varepsilon})>\lambda_\varepsilon\big\}-\beta_{jk}\big|^p\\
&\quad+\sum_{j\ge j_\varepsilon}2^{j(\frac{p}{2}-1)}\sum_k|\beta_{jk}|^p.
\end{aligned}$$

Since $f\in\mathcal{B}^{s/(1+2s)}_{p,\infty}(u)$, the second term can be bounded by $C(u(\lambda_\varepsilon))^{2sp/(1+2s)}$.

The first term $\mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |y_{jk}\mathbf{1}\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) > \lambda_\varepsilon\} - \beta_{jk}|^p$ can be decomposed in $B_1 + B_2$, where

$$B_1 = \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) \le \lambda_\varepsilon\big\},$$

$$B_2 = \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |y_{jk} - \beta_{jk}|^p \mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) > \lambda_\varepsilon\big\}.$$

We split B_1 into $B_1' + B_1''$ as follows:

$$B_1' = \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) \le \lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) \le 2\lambda_\varepsilon\big\},$$

$$B_1'' = \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) \le \lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) > 2\lambda_\varepsilon\big\}.$$

Since $f \in \mathcal{B}^{s/(1+2s)}_{2,\infty}(u) \cap W_{\mu,u}\big(\frac{p}{1+2s}, p\big)$ and $(\mu_{jk})_{j,k}$ are decreasing functions with respect to λ, using (3.2), one gets:

$$\begin{aligned}
B_1' &= \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) \le \lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) \le 2\lambda_\varepsilon\big\} \\
&\le \sum_{j<j_\varepsilon-4} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) \le 2\lambda_\varepsilon\big\} + \sum_{j\ge j_\varepsilon-4} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \\
&\le \sum_{j<j_\varepsilon-4} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\big\{\mu_{jk}(4\lambda_\varepsilon, \beta_{4\lambda_\varepsilon}) \le 2\lambda_\varepsilon\big\} + \sum_{j\ge j_\varepsilon-4} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \\
&\le C\big(u(4\lambda_\varepsilon)\big)^{2sp/(1+2s)} \le C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}.
\end{aligned}$$

Using (2.3) and Remark 2.1, we have

$$\begin{aligned}
B_1'' &= \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) \le \lambda_\varepsilon\big\}\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) > 2\lambda_\varepsilon\big\} \\
&= \sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{P}\big(\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) \le \lambda_\varepsilon\big)\mathbf{1}\big\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) > 2\lambda_\varepsilon\big\} \\
&= \sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{P}\big(\exists (j_o, k_o) \in I_{\lambda_\varepsilon} \mid |y_{j_o k_o} - \beta_{j_o k_o}| > \lambda_\varepsilon\big) \\
&\le C2^{j_\varepsilon}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)}\sum_k |\beta_{jk}|^p \mathbb{P}\big(|y_{jk} - \beta_{jk}| > \lambda_\varepsilon\big) \\
&\le C2^{j_\varepsilon}\varepsilon^{m^2/2} \le C\varepsilon^{m^2/2-2} \le C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}.
\end{aligned}$$

We have used here the concentration property of the Gaussian distribution and the fact that $m^2 \ge 2(p+2)$.

We split B_2 into $B_2' + B_2''$ as follows:

$$B_2' = \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbf{1}\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) > \lambda_\varepsilon\}\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) \le \frac{\lambda_\varepsilon}{2}\right\},$$

$$B_2'' = \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbf{1}\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) > \lambda_\varepsilon\}\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) > \frac{\lambda_\varepsilon}{2}\right\}.$$

For B_2' we use the Cauchy–Schwartz inequality and Remark 2.1:

$$\begin{aligned}
&\left(\mathbb{E}|y_{jk}-\beta_{jk}|^p \mathbf{1}\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) > \lambda_\varepsilon\}\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) \le \frac{\lambda_\varepsilon}{2}\right\}\right)^2 \\
&\quad \le \mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\mathbb{P}\left(\left|\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) - \mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon})\right| > \frac{\lambda_\varepsilon}{2}\right) \\
&\quad \le \mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\mathbb{P}\left(\exists (j_o, k_o) \in I_{\lambda_\varepsilon} \mid |y_{j_o k_o} - \beta_{j_o k_o}| > \frac{\lambda_\varepsilon}{2}\right) \\
&\quad \le a2^{j_\varepsilon}\mathbb{E}|y_{jk}-\beta_{jk}|^{2p}\mathbb{P}\left(|y_{jk}-\beta_{jk}| > \frac{\lambda_\varepsilon}{2}\right),
\end{aligned}$$

where $\mathbb{E}|y_{jk}-\beta_{jk}|^{2p} = C\varepsilon^{2p}$ and $\mathbb{P}\big(|y_{jk}-\beta_{jk}| > \frac{\lambda_\varepsilon}{2}\big) \le \varepsilon^{m^2/8}$ (using the concentration properties of the Gaussian distribution). So, choosing m such that $m^2 \ge 16(p+1)$,

$$\begin{aligned}
B_2' &= \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbf{1}\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) > \lambda_\varepsilon\}\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) \le \frac{\lambda_\varepsilon}{2}\right\} \\
&\le C2^{j_\varepsilon/2}\varepsilon^p \sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)} \sum_k \mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) \le \frac{\lambda_\varepsilon}{2}\right\}\varepsilon^{m^2/16} \\
&\le C2^{j_\varepsilon(p+1)/2}\varepsilon^{m^2/16+p} \le C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}.
\end{aligned}$$

Since $f \in W^*_{\mu,u}\big(\frac{p}{1+2s}, p\big)$, we can bound B_2'' as follows:

$$\begin{aligned}
B_2'' &= \mathbb{E}\sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)} \sum_k |y_{jk}-\beta_{jk}|^p \mathbf{1}\{\mu_{jk}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) > \lambda_\varepsilon\}\cdot\mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) > \frac{\lambda_\varepsilon}{2}\right\} \\
&\le C\varepsilon^p \sum_{j<j_\varepsilon} 2^{j(\frac{p}{2}-1)} \sum_k \mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) > \frac{\lambda_\varepsilon}{2}\right\} \\
&\le C\left(\frac{\lambda_\varepsilon}{4}\right)^p\left(\log\left(\frac{4}{\lambda_\varepsilon}\right)\right)^{-p/2} \sum_{j<j_\varepsilon+4} 2^{j(\frac{p}{2}-1)} \sum_k \mathbf{1}\left\{\mu_{jk}(\lambda_\varepsilon, \beta_{\lambda_\varepsilon}) > \frac{\lambda_\varepsilon}{2}\right\} \\
&\le C\left(u\left(\frac{\lambda_\varepsilon}{4}\right)\right)^{2sp/(1+2s)} \le C\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}.
\end{aligned}$$

□

The previous theorem point out the maximal spaces where μ-thresholding rules attain the rate of convergence $(u(\lambda_\varepsilon))^{2sp/(1+2s)}$. We notice that the bigger are the functions μ_{jk}, the larger are the spaces $W_{\mu,u}(\frac{p}{1+2s}, p)$ and so the 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 $(\mu_{jk})_{j,k}$ to be sure that we have the following embedding:

$$\mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu,u}\left(\frac{p}{1+2s}, p\right) \subset \mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W^*_{\mu,u}\left(\frac{p}{1+2s}, p\right).$$

3.3 Conditions for embedding inside maximal spaces

Theorem 3.2 *Let* $0 < r < p < \infty$, *and let* $(\mu_{jk})_{jk}$ *be a sequence of decreasing functions with respect to* λ. *Assume that there exist* $C_{\text{seuil}} > 0$ *and* $\lambda_{\text{seuil}} > 0$ *such that, for any* $0 < \lambda < \lambda_{\text{seuil}}$, *the following conditions are satisfied*:

$$\sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbf{1}\{\mu_{jk}(\lambda, \beta_\lambda) > \lambda\} \le C_{\text{seuil}}\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \times \cdots$$
$$\times \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \sum_{n\in\mathbb{N}} \mathbf{1}\{|\beta_{jk}| > 2^n\lambda\}\mathbf{1}\{\mu_{jk}(\lambda, \beta_\lambda) \le 2^{1+n}\lambda\}. \tag{3.3}$$

$\forall n \in \mathbb{N}$, $\exists C_n > 0$ (*not depending on* λ),

$$u(2^{2+n}\lambda) \le C_n u(\lambda), \quad \textit{and} \quad \sum_{n\in\mathbb{N}} C_n^{p-r} 2^{-np} < \infty. \tag{3.4}$$

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 3.3 It is easy to see that condition (3.4) implies condition (3.2). Once again, condition (3.4) is clearly satisfied when $u(t_\varepsilon) = t_\varepsilon$ or $u(t_\varepsilon) = \varepsilon$.

Proof of Theorem 3.2 Let u satisfy condition (3.4), and let $(\mu_{jk})_{j,k}$ satisfy condition (3.3). Fix $0 < \lambda < \lambda_{\text{seuil}}$ and set, for any $n \in \mathbb{N}$, $2^{j_{\lambda,n}} \sim (2^{2+n}\lambda)^{-2}$.

Let f in (2.1) be such that $f \in \mathcal{B}^{(p-r)/2p}_{p,\infty}(u) \cap W_{\mu,u}(r, p)$. Using (3.3), we have

$$\begin{aligned}
&\sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbf{1}\{\mu_{jk}(\lambda, \beta_\lambda) > 2\lambda\} \\
&\quad\le \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbf{1}\{\mu_{jk}(\lambda, \beta_\lambda) > \lambda\} \\
&\quad\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \sum_{n\in\mathbb{N}} \mathbf{1}\{|\beta_{jk}| > 2^n\lambda\}\mathbf{1}\{\mu_{jk}(\lambda, \beta_\lambda) \le 2^{1+n}\lambda\} \\
&\quad\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\{\mu_{jk}(\lambda, \beta_\lambda) \le 2^{1+n}\lambda\} \\
&\quad\le C_1 + C_2,
\end{aligned}$$

where

$$C_1 = C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j<j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{\mu_{jk}(\lambda,\beta_\lambda) \le 2^{2+n}\frac{\lambda}{2}\Big\}$$

and

$$C_2 = C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j\ge j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p.$$

Since $f \in W_{\mu,u}(r,p)$,

$$\begin{aligned}
C_1 &= C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j<j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{\mu_{jk}(\lambda,\beta_\lambda) \le 2^{2+n}\frac{\lambda}{2}\Big\}\\
&\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j<j_{\lambda,n}} 2^{j(\frac{p}{2}-1)}\\
&\quad\times \sum_k |\beta_{jk}|^p \mathbf{1}\Big\{\mu_{jk}\big(2^{2+n}\lambda,\beta_{2^{2+n}\lambda}\big) \le 2^{2+n}\frac{\lambda}{2}\Big\}\\
&\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \big(u(2^{2+n}\lambda)\big)^{p-r}\\
&\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \lambda^{-p}\big(u(\lambda)\big)^{p-r} \sum_{n\in\mathbb{N}} C_n^{p-r} 2^{-np}\\
&\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \lambda^{-p}\big(u(\lambda)\big)^{p-r}.
\end{aligned}$$

The two last inequalities use condition (3.4).

Now, since $f \in \mathcal{B}^{(p-r)/2p}_{p,\infty}(u)$,

$$\begin{aligned}
C_2 &= C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \sum_{j\ge j_{\lambda,n}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p\\
&\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \sum_{n\in\mathbb{N}} (2^n\lambda)^{-p} \big(u(2^{2+n}\lambda)\big)^{p-r}\\
&\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \lambda^{-p}\big(u(\lambda)\big)^{p-r} \sum_{n\in\mathbb{N}} C_n^{p-r} 2^{-np}\\
&\le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \lambda^{-p}\big(u(\lambda)\big)^{p-r}.
\end{aligned}$$

Last inequalities use condition (3.4).

By adding up C_1 and C_2, we have

$$\sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k \mathbf{1}\big\{\mu_{jk}(\lambda,\beta_\lambda) > 2\lambda\big\} \le C\big(\log(\lambda^{-1})\big)^{\frac{p}{2}} \lambda^{-p}\big(u(\lambda)\big)^{p-r},$$

which proves that $f \in W^*_{\mu,u}(r,p)$ and ends the proof. □

Corollary 3.1 *Let $s > 0$, $1 \le p < \infty$, and $m \ge 4\sqrt{p+1}$.*

Let $\mathrm{MS}(\hat{f}_\mu, (u(\lambda_\varepsilon))^{2sp/(1+2s)})$ be the maximal set of any μ-thresholding rule $\hat{f}_\mu$ for the rate of convergence $(u(\lambda_\varepsilon))^{2sp/(1+2s)}$. Under the conditions of Theorem 3.2, we have

$$\mathrm{MS}\big(\hat{f}_\mu, \big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big) = \mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu,u}\left(\frac{p}{1+2s}, p\right).$$

To prove this, it suffices to apply Theorems 3.1 and 3.2 $\big($with $r = \frac{p}{1+2s}\big)$.

It is clear that $\mu^{(1)}_{jk}$ satisfies condition (3.3) of Theorem 3.2. Consequently, if u satisfies (3.4), the maximal space where the procedure $\hat{f}_{\mu^{(1)}}$ attains the rate of convergence $(u(\lambda_\varepsilon))^{2sp/(1+2s)}$ is

$$\mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu^{(1)},u}\left(\frac{p}{1+2s}, p\right).$$

Notice that, for $u = Id_{\mathbb{R}^+}$, we identify $\mathcal{B}^{s/(1+2s)}_{p,\infty}(u)$ with the usual Besov space

$$\mathcal{B}^{s/(1+2s)}_{p,\infty} = \left\{ f;\ \sup_{J \ge -1} 2^{J(sp+\frac{p}{2}-1)} \sum_{j \ge J} \sum_k |\beta_{jk}|^p < \infty \right\}.$$

For the same choice of u, $W_{\mu^{(1)},u}\big(\frac{p}{1+2s}, p\big)$ represents the weak Besov space

$$W\left(\frac{p}{1+2s}, p\right) = \left\{ f; \sup_{\lambda > 0} \lambda^{r-p} \sum_{j < j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\big\{|\beta_{jk}| \le \lambda\big\} < \infty \right\}.$$

Let us recall that the maxiset of the hard thresholding rule $\hat{f}_{\mu^{(1)}}$ for the rate $\lambda_\varepsilon^{2sp/(1+2s)}$ has already been studied by Cohen et al. (2001) and Kerkyacharian and Picard (2000).

4 Applications of results for particular μ-thresholding 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 $\hat{f}_{\mu^{(1)}}$. Then, we give some examples of such rules.

Let us state the following proposition:

Proposition 4.1 *Let $1 \le p < \infty$. Under conditions of Theorem 3.2, the maximal space for the rate $(u(\lambda_\varepsilon))^{2sp/(1+2s)}$ of any μ-thresholding rule satisfying, for any $\lambda > 0$ and any $\beta_\lambda \in \mathbb{R}^{\#I_\lambda}$,*

$$\mu_{jk}(\lambda, \beta_\lambda) \le \frac{\lambda}{2} \quad \Longrightarrow \quad |\beta_{jk}| \le \frac{\lambda}{2}, \tag{4.1}$$

is larger than the hard thresholding one. Moreover, if $C_n \le O(2^n)$ for all $n \in \mathbb{N}$, then the maximal space contains the Besov space $\mathcal{B}^s_{p,\infty}(u)$.

Remark 4.1 For $u(t_\varepsilon) = t_\varepsilon$ (resp. $u(t_\varepsilon) = \varepsilon$), notice that, using Remark 3.2, 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 (4.1), then we have for any $0 < r < p : W_{\mu,u}(r, p) \supset W_{\mu^{(1)},u}(r, p)$. So, using Corollary 3.1 to characterize the maxisets for the rate $(u(\lambda_\varepsilon))^{2sp/(1+2s)}$ associated with $\hat{f}_\mu$ and $\hat{f}_{\mu^{(1)}}$, one gets that the maximal space for the rate $(u(\lambda_\varepsilon))^{2sp/(1+2s)}$ of $\hat{f}_\mu$ is larger than the hard thresholding one.

To prove now that if, for any $n \in \mathbb{N}$, $C_n \leq O(2^n)$, the Besov space $\mathcal{B}^s_{p,\infty}(u)$ is contained in the maxiset of $\hat{f}_\mu$, it suffices to prove that

$$\mathcal{B}^s_{p,\infty}(u) \subset W_{\mu^{(1)},u}\left(\frac{p}{1+2s}, p\right).$$

Fix $0 < \lambda < \lambda_{\text{seuil}}$ and set $2^{j_{\lambda,u}} \sim \lambda_u^{-2} := (u(\lambda))^{4s/(1+2s)}\lambda^{-2}$ (resp. $2^{j_\lambda} \sim \lambda^{-2}$). For any $f \in \mathcal{B}^s_{p,\infty}(u)$, we have

$$\begin{aligned}
\sum_{j<j_\lambda} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \mathbf{1}\left\{|\beta_{jk}| \leq \frac{\lambda}{2}\right\} &\leq C 2^{j_{\lambda,u} p/2}\lambda^p + \sum_{j \geq j_{\lambda,u}} 2^{j(\frac{p}{2}-1)} \sum_k |\beta_{jk}|^p \\
&\leq C\Big(\big(u(\lambda)\big)^{2sp/(1+2s)} + \big(u(\lambda_u)\big)^{2sp}\Big) \\
&= C\Big(\big(u(\lambda)\big)^{2sp/(1+2s)} + D_1\Big).
\end{aligned}$$

Now, there exists $n_\lambda \in \mathbb{N}$ such that $2^{1+n_\lambda} \leq u(\lambda)^{-2s/(1+2s)} \leq 2^{2+n_\lambda}$.

Therefore, since $C_n \leq O(2^n)$ for all $n \in \mathbb{N}$, one gets:

$$u(\lambda_u) = u\big(u(\lambda)^{-2s/(1+2s)}\lambda\big) \leq u\big(2^{2+n_\lambda}\lambda\big) \leq C 2^{n_\lambda} u(\lambda) \leq C u(\lambda)^{1/(1+2s)}.$$

So

$$D_1 = \big(u(\lambda_u)\big)^{2sp} \leq C\big(u(\lambda)\big)^{2sp/(1+2s)}.$$

Finally, we deduce that $f \in W_{\mu,u}(r, p)$. □

4.1 On block thresholding rules

In the following proposition, we compare the maximal spaces associated with the four examples of μ-thresholding rules defined in Sect. 2.2. We prove that hard thresholding rules are outperformed by block thresholding rules when the length of the blocks are correctly chosen. Indeed:

Proposition 4.2 *For any* $1 \leq p < \infty$ *and any* $m \geq 4\sqrt{p+1}$, *let*

$$\text{MS}\big(\hat{f}_{\mu^{(i)}}, \big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big), \quad 1 \leq i \leq 4,$$

be respectively the maximal sets of procedures $\hat{f}_{\mu^{(i)}}$ *for the rate of convergence* $(u(\lambda_\varepsilon))^{2sp/(1+2s)}$. *Under the conditions of Theorem* 3.2, *we have the following inclusions of spaces*:

$$\mathrm{MS}\big(\hat{f}_{\mu^{(1)}}, \big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big) \subset \mathrm{MS}\big(\hat{f}_{\mu^{(i)}}, \big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big), \quad \forall i \in \{3, 4\}, \tag{4.2}$$

$$\begin{aligned}\mathrm{MS}\big(\hat{f}_{\mu^{(2)}}, \big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big) &\subset \mathrm{MS}\big(\hat{f}_{\mu^{(4)}}, \big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big) \\ &\subset \mathrm{MS}\big(\hat{f}_{\mu^{(3)}}, \big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big).\end{aligned} \tag{4.3}$$

Proof Using Corollary 3.1, we have, for any $1 \le i \le 4$,

$$\mathrm{MS}\big(\hat{f}_{\mu^{(i)}}, \big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)}\big) = \mathcal{B}^{s/(1+2s)}_{p,\infty}(u) \cap W_{\mu^{(i)},u}\left(\frac{p}{1+2s}, p\right).$$

Now, for any f as in (2.1), we have

$$\max_{k' \in \mathcal{P}_j(k)} |\beta_{jk'}| \le \lambda \quad \Longrightarrow \quad |\beta_{jk}| \le \lambda,$$

$$\max\left(|\beta_{jk}|^p, \frac{1}{l_j} \sum_{k' \in \mathcal{P}_{jk}} |\beta_{jk'}|^p\right) \le \lambda^p \quad \Longrightarrow \quad |\beta_{jk}| \le \lambda.$$

So, using Proposition 4.1, the inclusion of spaces (4.2) holds. In the same way, the inclusion of spaces (4.3) also holds, since one has the both properties:

$$\max\left(|\beta_{jk}|^p, \frac{1}{l_j} \sum_{k' \in \mathcal{P}_{jk}} |\beta_{jk'}|^p\right) \le \lambda^p \quad \Longrightarrow \quad \frac{1}{l_j} \sum_{k' \in \mathcal{P}_{jk}} |\beta_{jk'}|^p \le \lambda^p,$$

$$\max_{k' \in \mathcal{P}_j(k)} |\beta_{jk'}| \le \lambda \quad \Longrightarrow \quad \max\left(|\beta_{jk}|^p, \frac{1}{l_j} \sum_{k' \in \mathcal{P}_{jk}} |\beta_{jk'}|^p\right) \le \lambda^p. \qquad \square$$

The previous proposition is important. Indeed, we see that block thresholding rules can outperform hard thresholding ones when conditions (3.3) and (3.4) are satisfied. In particular, condition (3.3) is satisfied if the lengths of the blocks are chosen small enough, as we can see in the following proposition:

Proposition 4.3 *Under the maxiset approach associated with the rate*

$$\big(u(\lambda_\varepsilon)\big)^{2sp/(1+2s)},$$

we have the following result:

[**Block thresholding rules**] *For any* $1 \le p < \infty$, *maximean- and maximum-block*(p) *thresholding rules such that the lengths* l_j *of the blocks* $\mathcal{P}_{jk}$ *do not exceed* $C(\log(\varepsilon^{-1}))^{p/2}$ *for some* $C > 0$ *outperform hard thresholding rules in the maxiset sense.*

Proof Note that it is just a consequence of the previous proposition, since the condition $l_j \leq C(\log(\varepsilon^{-1}))^{p/2}$ ensures that condition (3.3) of Theorem 3.2 is satisfied when dealing with block thresholding rules. □

This result is very interesting, since it allows one to give a theoretical explication about the better performances of block thresholding rules comparatively to hard thresholding rules which have been observed in the practical setting (see Hall et al. 1997 and Cai 1998, 1999, 2002) and which were not explained with the minimax approach.

4.2 On μ-thresholding rules based on hereditary constraints

In this last paragraph, we aim at underlining that, apart from the examples of block thresholding rules mentioned above, there exist other μ-thresholding rules able to outperform the hard thresholding one. Indeed, Autin (2004) has studied the maxiset performance of a new procedure called *hard tree procedure*, combining thresholding rules and hereditary constraints. By making use of the dyadic structure of the wavelet bases, the author has pointed out the relationship between this procedure and Lepski (1991)'s procedure and has proved that the maxisets of this procedure for usual rates are larger than the hard thresholding procedure ones.

There is no difficulty to prove that the *hard-tree procedure* belongs to the class of μ-thresholding rules. To be more explicit, let us use the same notation as in Autin (2004) and denote, for any j, k:

- I_{jk} the *dyadic interval* corresponding to the support of ψ_{jk}
- l_ψ the maximal size of the supports of the scaling function and the wavelet used for the reconstruction

Then the hard-tree rule corresponds to the μ-thresholding rule associated with the sequence of positive functions $(\mu_{jk}^{(5)})_{j,k}$ defined as follows:

$$\mu_{jk}^{(5)}(\lambda_\varepsilon, y_{\lambda_\varepsilon}) = \max\big(|y_{j'k'}|, I_{j'k'} \in \mathcal{T}_{jk}(\lambda_\varepsilon)\big),$$

where, for any $\varepsilon > 0$, $\mathcal{T}_{jk}(\lambda_\varepsilon)$ is the binary tree containing the set of *dyadic intervals* such that the following properties are satisfied:

- $I_{jk} \in \mathcal{T}_{jk}(\lambda_\varepsilon)$
- $I \in \mathcal{T}_{jk}(\lambda_\varepsilon) \Longrightarrow I \subset I_{jk}$ and $|I| > l_\psi \lambda_\varepsilon^2$
- Two distinct dyadic intervals of $\mathcal{T}_{jk}(\lambda_\varepsilon)$ with same length have their interiors disjointed
- The numbers of dyadic intervals of $\mathcal{T}_{jk}(\lambda_\varepsilon)$ of length $l_\psi 2^{-j'}$ $(j \leq j' < j_\varepsilon)$ is equal to $2^{j'-j}$
- Any set of all dyadic intervals of $\mathcal{T}_{jk}(\lambda_\varepsilon)$ with same length is forming a partition of I_{jk}

According to its definition, the *hard-tree rule* satisfies condition (3.3) of Theorem 3.2 and condition (4.1) of Proposition 4.1 (see Autin 2004). Hence, dealing with the maxiset approach associated with any rate satisfying condition (3.4) of Theorem 3.2, the hard-tree procedure outperforms the hard thresholding one in the maxiset sense.

As a conclusion, thanks to the maxiset approach, we prove in this paper that rules constructed with thresholding methods applied to *groups of empirical coefficients* are preferable to rules based on thresholding methods applied to *individual empirical coefficients*.

References

Autin F (2004) Maxiset point of view in non parametric estimation. PhD Thesis, Université Paris 7—Denis Diderot, France

Autin F, Picard D, Rivoirard V (2006) Large variance Gaussian priors in Bayesian nonparametric estimation: a maxiset point of view. Math Methods Stat 15(4)

Cai T (1998) Numerical comparisons of BlockJS estimator with conventional wavelet methods. Unpublished manuscript

Cai T (1999) Adaptive wavelet estimation: a block thresholding and oracle inequality approach. Ann Stat 27(3):898–924

Cai T (2002) On block thresholding in wavelet regression: adaptivity, block size and threshold level. Stat Sin 12(4):1241–1273

Cohen A, De Vore R, Kerkyacharian G, Picard D (2001) Maximal spaces with given rate of convergence for thresholding algorithms. Appl Comput Harmon Anal 11:167–191

Donoho D, Johnstone IM, Kerkyacharian G, Picard D (1995) Wavelet shrinkage: asymptotia? J Roy Stat Soc Ser B 57(2):301–369 (with discussion and reply by the authors)

Donoho D, Johnstone IM, Kerkyacharian G, Picard D (1996) Density estimation by wavelet thresholding. Ann Stat 24(2):508–539

Donoho D, Johnstone IM, Kerkyacharian G, Picard D (1997) Universal near minimaxity of wavelet shrinkage. In: Festschrift for Lucien Le Cam. Springer, New York, pp. 183–218

Hall P, Kerkyacharian G, Picard D (1999) On the minimax optimality of block thresholded wavelet estimators. Stat Sin 9(1):33–49

Hall P, Penev S, Kerkyacharian G, Picard D (1997) Numerical performance of block thresholded estimators. Stat Comput 7:115–124

Kerkyacharian G, Picard D (2000) Thresholding algorithms, maxisets and well concentrated bases. Test 9(2):283–344 (with comments, and a rejoinder by the authors)

Lepski O (1991) Asymptotically minimax adaptive estimation 1: upperbounds. Optimally adaptive estimates. Theory Probab Appl 36:682–697