T E C H N I C A L

R E P O R T

11023

Combining thresholding rules: a new way to improve the performance of wavelet estimators

AUTIN, F., FREYERMUTH, J.M. and R. von SACHS

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Combining thresholding rules: a new way to improve the performance of wavelet estimators

Florent Autin [*] Jean-Marc Freyermuth [†] Rainer von Sachs [‡]

September 12, 2011

Abstract

In this paper, we address the situation where we cannot differentiate wavelet-based threshold estimators because their sets of *well-estimated* functions (maxisets) are not nested. As a generic solution, we propose to proceed via a combination of these estimators in order to achieve new estimators which perform better in the sense that the involved maxisets contain the union of the previous ones. Throughout the paper we propose illuminating interpretations of the maxiset results and provide conditions to ensure that this combination generates larger maxisets. As an example, we propose to combine vertical- and horizontal-block thresholding estimators that are already known to perform well. We discuss the limitations of our method, and we confirm our theoretical results through numerical experiments.

Keywords : Besov spaces, curve estimation, wavelet methods, maximal spaces, rate of convergence, thresholding rules, white Gaussian noise.

[*]Université d'Aix-Marseille 1, CNRS UMR 6632,C.M.I., 39 rue F. Joliot Curie, 13453 Marseille Cedex 13. Email: autin@cmi.univ-mrs.fr. Research partly supported by A.N.R. project *Oracle*.

[†]Université catholique de Louvain, ISBA, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium. Email: Jean-Marc.Freyermuth@uclouvain.be. Financial support from the contract "Projet d'Actions de Recherche Concertées" nr. 07/12/002 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain", is gratefully acknowledged.

[‡]Université catholique de Louvain, ISBA, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium. Email: rvs@uclouvain.be. Financial support from the contract "Projet d'Actions de Recherche Concertées" nr. 07/12/002 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain", is gratefully acknowledged.

1 Introduction

The literature about wavelet-based nonparametric function estimation is very large. Many thresholding rules/procedures have been proposed and compared from both a practical and a theoretical point of view. In the last decade, a new theoretical way, dual to the minimax approach, has been proposed to assess and compare their theoretical performances. This approach consists in determining the maxiset of a thresholding rule that is the largest function space where the procedure reconstructs the target function at a given rate of convergence. As previously discussed in Cohen et al. [16], Kerkyacharian and Picard [21],[22], Autin [3], [6] and in Autin et al. [8], [9], this approach can be successful in order to differentiate minimax-equivalent procedures whenever their maxisets are nested. Without such embeddings, the comparison would be impossible. Even if it has often been viewed as a problem, this just reveals the fact that the procedures are well-suited to estimate different classes of functions. Hence, the best procedure within of a family of thresholding rules - that is the one with the largest maxiset - does not always exist.

In this paper we address this problem of the existence of *a best thresholding rule*. We generalize the framework of the μ-thresholding rules introduced by Autin [6] to work with and claim that: "if you cannot discriminate between thresholding rules then combine them". Indeed, we prove that we can construct a new thresholding rule borrowing strength from other well chosen ones (those with non nested maxisets) to yield better maxiset results and numerical performances.

Following the results of our numerical simulations given in detail in Section 6, the Figure 1 shows an example of our method that combines two block thresholding rules: horizontal- and vertical-block thresholding procedures that are, to the best of our knowledge, the ones with largest but not nested maxisets encountered in the literature. We recall that estimators induced by these rules are respectively the *Blockshrink estimator* studied by Cai [10] and Autin et al. [9] and the *Hard Tree estimator* studied by Autin [6] and Autin et al. [8]. Our numerical results clearly illustrate the need to use the combination of the previous estimators (called the Block Tree estimator) rather than the Blockshrink estimator or the Hard Tree estimator since it behaves well over all the twelve functions considered here. More information about these numerical experiments can be found in Section 6.

The rest of the paper is organized as follows: Section 2 and Section 3 describe our theoretical model and the maxiset approach, then, in Section 4, we define and give the maxiset properties of the thresholding rules. In Section 5 we describe our method to combine thresholding rules and its limitation. Finally, Section 7 ends the paper with detailed proofs of our theoretical results.

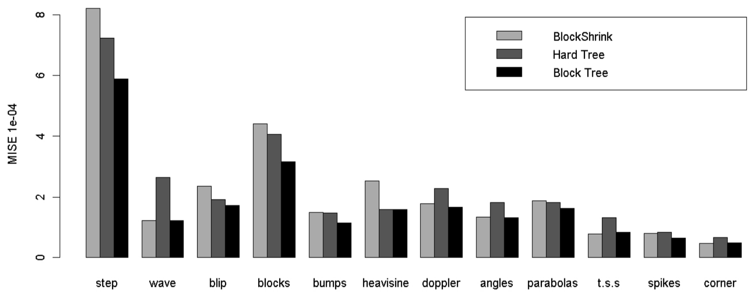

Figure 1: Quadratic risk of estimators Blockshrink, Hard Tree and their combination Block Tree.

2 Background of study

Let us consider a compactly supported wavelet basis of $L_2([0,1])$ with V vanishing moments ($V \in \mathbb{N}^*$) which has been previously periodized $\left\{\phi, \psi_{jk}, j \in \mathbb{N}, k \in \{0, \ldots, 2^j - 1\}\right\}$. Examples of such bases are given in Daubechies [17]. Any function $f \in L_2([0,1])$ can be written as follows:

$$f = \alpha\phi + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \theta_{jk}\psi_{jk}. \tag{1}$$

The coefficient α and the components of $\theta = (\theta_{jk})_{j,k}$ are respectively the scaling/wavelet coefficients of f. They correspond to the L_2-scalar products between f and the scaling/wavelet functions ϕ and ψ_{jk}.

We consider the sequential version of the Gaussian white noise model: we dispose of observations of the wavelet coefficients of the goal function f which are assumed to be realizations of independent random variables:

$$\hat{\alpha} \quad = \quad \alpha + \epsilon\xi \quad \text{ and } \quad \hat{\theta} = \left(\hat{\theta}_{jk}\right)_{j,k} = \left(\theta_{jk} + \epsilon\xi_{jk}\right)_{j,k}, \tag{2}$$

where $\xi, (\xi_{jk})_{j,k}$ are i.i.d. $\mathcal{N}(0,1)$, $0 < \epsilon < 1/e$ is the noise level.
We focus on performances of KK-estimators which are wavelet estimators relying on a *Keep-or-Kill* rule. Such a KK-estimator $\hat{f}$ can be written as follows:

$$\hat{f} = \hat{\alpha}\phi + \sum_{(j,k)\in\mathcal{K}_\epsilon} \hat{\theta}_{jk}\psi_{jk}, \tag{3}$$

where $\mathcal{K}_\epsilon$ is a finite set of indices that may be random or deterministic.

3 Maxiset approach

In order to assess the theoretical efficiency of estimators, Cohen et al. [16] suggested the maxiset point of view. This new setting offers a complementary approach to the minimax one and was successfully applied in order to differentiate between *minimax-optimal* estimators.

In the following, we consider families of threshold estimators $\hat{f}_{\mu,m}$ associated with a given thresholding rule μ (see Definition 4.1) and indexed by some index set M of positive real numbers $m \in M$, which will be specified later. Generally speaking, providing *maxiset performances of an estimator* $\hat{f}$ means determining the largest functional space (maxiset) $\mathcal{G}$ over which the quadratic risk of this estimator converges at a prespecified rate v, i.e.

$$\sup_{0<\epsilon<1/e} v_\epsilon^{-1}\mathbb{E}\|\hat{f}-f\|_2^2 < \infty \iff f \in \mathcal{G}.$$

More specifically, in our context, we shall say that the functional space $\mathcal{G}_{\mu,M}$ is the *maxiset of the thresholding rule* μ for the rate of convergence v and the L_2-loss function if and only if

$$\sup_{m\in M}\sup_{0<\epsilon<1/e} v_\epsilon^{-1}\mathbb{E}\|\hat{f}_{\mu,m}-f\|_2^2 < \infty \iff f \in \mathcal{G}_{\mu,M}.$$

In other words, the space $\mathcal{G}_{\mu,M}$ can be viewed as the intersection of the maxisets of estimators $\hat{f}_{\mu,m}$ $(m \in M)$.

Therefore, from the maxiset point of view, *the larger the maxiset the better the rule*. Obviously, the size of the maxiset depends on the chosen rate; the slower the rate the larger the maxiset. When comparing distinct rules of reconstruction, we say that one is better than the other if the maxiset of the one contains the maxiset of the other, for the same given rate.

The first maxiset results were provided by Cohen et al. [16] and Kerkyacharian and Picard [21], [22] who determined the maximal functional spaces for estimators based on Hard and Soft thresholding rules, respectively. They also proved that estimators built from the local bandwith selection rule of Lepski [23] were at least as efficient as the latter ones.

As discussed in Autin [3], thresholding rules with larger maxiset can be constructed from rules that are *not elitist* - i.e., rules that do not only keep all the 'large' empirical wavelet coefficients but equally consider some well chosen small ones. As examples, we cite estimators that rely on vertical-block thresholding rules (see Autin [5], Autin et al. [8])) or horizontal-block thresholding rules (see, among others, Hall et al. [19], [20], Cai [10], [11] [12], [13], [14] and Autin et al. [9]). When looking at these estimators, their maxisets are larger than those of estimators based on an elitist rules, including Hard, Soft Thresholding estimators, and also many Bayes estimators (see Autin et al. [4]).

Nevertheless the following open question arises from these previous works: in order to estimate a signal what is the best choice among vertical- and horizontal-block thresholding rules?

As emphasized in the Introduction the maxisets of vertical- and horizontal-block thresholding estimators are not embedded and thus these estimators can not be differentiated one from the other. Indeed, as shown by the quadratic risks of the estimators in the Figure 1 for several test functions, it seems to be difficult to identify a winning method. Hence, from both a theoretical and a practical point of view the answer to the question is not clear.

As a way out, in this paper we propose to combine existing thresholding rules so as to get a new well-performing rule which reconstructs at least as many functions as the ones generated by the vertical- and horizontal- block thresholding rules. To reach this goal we first introduce a large family of wavelet estimators built from thresholding rules.

4 Maxiset properties of thresholding rules

From now on, let us introduce some notations.

- $t_\epsilon = \epsilon\sqrt{\ln(\epsilon^{-1})}$ $(0 < \epsilon < 1/e)$,

- j_λ is the integer such that $2^{-j_\lambda} \le \lambda^2 < 2^{1-j_\lambda}$ $(\lambda > 0)$,

- $\theta = (\theta_{jk})_{j,k}$ (resp. $\hat{\theta} = (\hat{\theta}_{jk})_{j,k}$) is the set of all wavelet coefficients (resp. all empirical wavelet coefficients) of signal f.

4.1 Estimators built from thresholding rules

Let us now introduce a family of wavelet estimators built from thresholding rules. The following definition is a slight generalisation of the one given by Autin [6] in that condition (2.3) therein is not required here.

Definition 4.1. *Let $m > 0$ and consider the sequential model (2). An estimator $\hat{f}_{\mu,m}$ is called (μ, m)-thresholding estimator if, there exists a thresholding rule that is a sequence of positive functions $\left(\mu_{jk}\left(m, t_\epsilon, \hat{\theta}\right)\right)_{j,k}$ that are monotonically decreasing with respect to m and such that*

$$\hat{f}_{\mu,m} = \hat{\alpha}\phi + \sum_{j\in\mathbb{N},\ j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk}\mathbf{1}\left\{\mu_{jk}\left(m, t_\epsilon, \hat{\theta}\right) > mt_\epsilon\right\}\psi_{j,k}.$$

As remarked in Autin et al. [6], any (μ, m)-thresholding estimator $\hat{f}_{\mu,m}$ is a *limited estimator* in the sense that the reconstruction of f by such an estimator does not use empirical wavelet coefficients $\hat{\theta}_{jk}$ with $j \ge j_{mt_\epsilon}$.

Some examples of μ-thresholding rules and of the (μ, m)-thresholding estimators $(m > 0)$ associated with are the following:

- The Hard Thresholding rule μ^H: $\mu^H_{jk}\left(m, t_\epsilon, \hat{\theta}\right) := \left|\hat{\theta}_{jk}\right|$.

The (μ^H, m)-thresholding estimator relies on the basic rule to keep in the signal reconstruction only empirical wavelet coefficients larger than the threshold value mt_ϵ in absolute value. The other ones are killed.

- The Hard Tree Thresholding rule μ^T: $\mu^T_{jk}\left(m, t_\epsilon, \hat{\theta}\right) := \max_{(j',k')\in\mathcal{T}_{j,k}(mt_\epsilon)} \left|\hat{\theta}_{j'k'}\right|$.

The (μ^{T}, m)-thresholding estimator was already studied by Autin [5] and Autin et al. [8]. It relies on the rule to keep empirical wavelet coefficients with level strictly smaller than j_{mt_ϵ}, larger in absolute value than the threshold mt_ϵ and with their ancestors in the dyadic-tree rooted at $(j_0, k_0) := (0, 0)$. Here $\mathcal{T}_{j,k}(mt_\epsilon)$ corresponds to the dyadic-tree rooted à (j, k) and being reduced to indices with level strictly smaller than j_{mt_ϵ}. This estimator is tree structured (i.e. the empirical wavelet coefficients that have been kept for the signal reconstruction satisfy the hereditary constraint of Engel [18]). They can be viewed as both a hybrid wavelet version of Lepski's kernel method (see Autin [5]) and a vertical-block thresholding method (see Autin et al. [8]).

- The BlockShrink rule μ^{B}: $\mu^{B}_{jk}\left(m, t_\epsilon, \hat{\theta}\right) := \left(\sum_{k' \in \mathcal{P}_{j,k}(\epsilon)} \hat{\theta}^2_{jk'}\right)^{\frac{1}{2}}.$

The (μ^{B}, m)-thresholding estimator was studied by Cai [10] and Autin et al. [9]. It relies on the rule to keep empirical wavelet coefficients if the l_2-norm of the empirical wavelet coefficients with index k' belonging to its block $\mathcal{P}_{j,k}(\epsilon)$ is larger than the threshold value mt_ϵ. $\mathcal{P}_{j,k}(\epsilon)$ are non-overlapped blocks with common size $\lceil \ln(\epsilon^{-1}) \rceil = \lceil -\ln(F^{-1}(t_\epsilon)) \rceil$, where F^{-1} is the inverse function of $F : \epsilon \longrightarrow F(\epsilon) := t_\epsilon$ and $\lceil x \rceil$ denotes the smallest integer bigger than or equal to x. A precise description is given in Cai [10] and in Autin et al. [9].

More generally, for wise choices of μ, the resulting estimators show good theoretical and practical performances. In particular we recall in the next section that the sets of functions they are able to *well estimate* are quite large for 'classical' minimax rates (see Cohen et al. [16], Autin [3] and [6]).

4.2 Maxiset results

To begin, let us define the functional spaces that shall appear in our future maxiset results.

Definition 4.2. *Let $0 < u < V$. A function $f \in L_2([0, 1])$ belongs to the Besov space $\mathcal{B}^u_{2,\infty}$ if and only if:*

$$\sup_{J \geq 0} 2^{2Ju} \sum_{j \geq J} \sum_{k=0}^{2^j - 1} \theta^2_{jk} < \infty.$$

Following Autin [3], for any chosen rate v, Besov spaces $\mathcal{B}^u_{2,\infty}$ usually appear when studying the maxisets of wavelet estimators that kill any empirical wavelet coefficient with a level larger than or equal to a maximum resolution level $j_\epsilon = O\left(\ln(v_\epsilon^{-1})\right)$ $(0 < \epsilon < 1/e)$.

Definition 4.3. *Let $m' \geq 1$, $0 < r < 2$ and a thresholding rule μ be given. A function $f \in L_2([0, 1])$ belongs to the space $W_{\mu,m'}(r)$ if and only if:*

$$\sup_{m \geq m'} \sup_{0 < \lambda < \frac{1}{e}} (m\lambda)^{r-2} \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j - 1} \theta^2_{jk} \mathbf{1}\left\{\mu_{jk}\left(m, \lambda, \theta\right) \leq m\lambda\right\} < \infty.$$

The spaces $W_{\mu,m'}(r)$ contain functions for which we can control the energy of the coefficients that do not survive the μ-thresholding rule.

Definition 4.4. *A thresholding rule μ is said to satisfy the U-property if and only if, for any (j,k), any (m,ϵ) and any sequence of real numbers θ,*

$$\mu_{jk}\left(m, t_\epsilon, \theta\right) \text{ only depends on parameters } mt_\epsilon \text{ and } \theta, \tag{4}$$

In the sequel, we use $\tilde{\mu}_{jk}\left(mt_\epsilon, \theta\right) := \mu_{jk}\left(m, t_\epsilon, \theta\right)$ to denote a thresholding rule μ satisfying the U-property.

Proposition 4.1. *Consider a thresholding rule μ satisfying the U-property. Then, for any $0 < r < 2$ and any $m' \geq 1$*

$$W_{\mu,m'}(r) = W_\mu(r),$$

where $W_\mu(r)$ is the set of functions $f \in L_2([0,1])$ such that

$$\sup_{\lambda>0} \lambda^{r-2} \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{\tilde{\mu}_{jk}\left(\lambda, \theta\right) \leq \lambda\right\} < \infty.$$

Remark 4.1. *a) The proof of the previous proposition is obvious by considering the required change of variables.*
b) Notice that both μ^{H} and μ^{T} satisfy the U-property whereas μ^{B} does not. For the latter case, various values of m' generate distinct functional spaces $W_{\mu^B,m'}(r)$. Indeed, notice that $W_{\mu^B,m'}(r)$ can be rewritten as the space of functions f such that

$$\sup_{m\geq m'} \sup_{0<\lambda<\frac{m}{e}} \lambda^{r-2} \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{\left(\sum_{k'\in\mathcal{P}_{j,k}\left(F^{-1}\left(\frac{\lambda}{m}\right)\right)} \theta_{jk'}^2\right)^{\frac{1}{2}} \leq \lambda\right\} < \infty.$$

The aim of the following paragraph is to characterize the maxisets associated with thresholding rules (see also Autin [6]). As usual in the maxiset setting, we shall suppose that a *Large Deviation* property, namely LD-property, will hold to derive our results.

Definition 4.5. *We say that a thresholding rule μ satisfies the LD-property if and only if for any given $\nu > 0$, there exists $m_{\mu,\nu} \geq 1$ such that for any $m \geq m_{\mu,\nu}$ and any sequence of real numbers θ and Gaussian random variables $\hat{\theta}$ connected to θ via model (2),*

$$\sup_{0<\epsilon<1/e} \epsilon^{-\nu}\, \mathbb{P}\left(|\mu_{jk}(m, t_\epsilon, \hat{\theta}) - \mu_{jk}(m, t_\epsilon, \theta)| > m_{\mu,\nu} t_\epsilon\right) \leq \frac{1}{2}. \tag{5}$$

Remark 4.2. *Notice that for the examples of thresholding rules we gave, the LD-property (5) is satisfied for:*

- μ^H *when choosing* $m_{\mu^H,\nu} = \sqrt{2\nu + 4\ln(2)}$,
- μ^T *when choosing* $m_{\mu^T,\nu} = \sqrt{2(\nu + 2 + 2\ln(2))}$ *(due to the concentration inequality for standard Gaussian variables),*

- μ^B, *when choosing* $m_{\mu^B,\nu}$ *such that* $m^2_{\mu^B,\nu} - 2\ln(m_{\mu^B,\nu}) = 2\nu + 1$ *(obtained from inequality* (9.9) *given in Cai [11]).*

Definition 4.6. *We say that a thresholding rule* μ *satisfies the Sparsity-property if and only there exists* $C_\mu > 0$ *such that for any* $0 < \epsilon < 1/e$, *any* $m > 0$ *and any sequence of real numbers* θ

$$\begin{aligned} & \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \mathbf{1}\{\mu_{jk}(m, t_\epsilon, \theta) > \frac{mt_\epsilon}{2}\} \\ \leq \quad & C_\mu \ln(\epsilon^{-1}) \sum_{n\in\mathbb{N}} (m2^n t_\epsilon)^{-2} \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\{\mu_{jk}(m2^n, t_\epsilon, \theta) \leq m2^n t_\epsilon\}. \end{aligned} \tag{6}$$

Notice that rules μ^H, μ^T and μ^B satisfy the *Sparsity*-property (see also Autin [6]). Moreover, from Definition 4.3, the following lemma holds:

Lemma 4.1. *Let* $s > 0$ *and* $m' \geq 1$. *Consider a thresholding rule* μ *that satisfies the Sparsity-property. Then,*

$$f \in W_{\mu,m'}\left(\tfrac{2}{1+2s}\right)$$
$$\Downarrow$$
$$\sup_{m\geq m'} \sup_{0<\epsilon<1/e} (mt_\epsilon)^{\frac{2}{1+2s}} \left(\ln(\epsilon^{-1})\right)^{-1} \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \mathbf{1}\left\{\mu_{jk}(m, t_\epsilon, \theta) > \frac{m}{2} t_\epsilon\right\} < \infty,$$

where $\theta = (\theta_{jk})_{j,k}$ *is defined as in equation (1).*

When thinking of practical purposes, the choice of m is crucial. To avoid the gap between theoretical and practical settings, we provide our results for a large range of values m.

Theorem 4.1. *Let* $s > 0$. *Consider a thresholding rule* μ *such that the LD-property and the Sparsity-property hold. Then, for any* $m' \geq m_{\mu,4}$, *we have the following equivalence:*

$$\forall m \geq 2m', \sup_{0<\epsilon<1/e} (mt_\epsilon)^{-\frac{4s}{1+2s}} \mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 < \infty \iff f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap W_{\mu,m'}\left(\frac{2}{1+2s}\right).$$

There is a natural and interesting interpretation of this theorem. The *Sparsity*-property ensures that the functions to be estimated have a sufficient degree of sparsity to be able to control the variance of our thresholding rules at the level required by the prespecified rate. In such a situation, Theorem 4.1 shows us that one should enlarge as much as possible the space $W_{\mu,m'}$. To do so, the thresholding rules keep as much as possible well-chosen coefficients, in other words, they attempt to reduce as much as possible the estimation bias. This situation describes a wide range of thresholding rules, among them, we are naturally interested in those which at least outperform (in the maxiset sense) the Hard thresholding one: they are called *Cautious rules* and are defined as follows:

Definition 4.7. *We say that a thresholding rule μ is cautious if and only if, for any (j,k) and any sequence of real numbers θ, the following property holds:*

$$|\theta_{jk}| \geq mt_\epsilon \Longrightarrow \mu_{jk}(m, t_\epsilon, \theta) \geq mt_\epsilon, \quad \text{for any } 0 < \epsilon < 1/e.$$

Rules μ^{H}, μ^{T}, μ^{B} are clearly cautious. From Definition 4.3 and Proposition 4.1, one gets:

Proposition 4.2. *Let μ be a cautious rule. Then, for any $m' \geq 1$ and any $0 < r < 2$,*

$$W_{\mu,m'}(r) \supseteq W_{\mu^H}(r).$$

As a direct consequence of Theorem 4.1 and Proposition 4.2, we get the following corollary.

Corollary 4.1. *Let μ be a cautious rule such that the LD-property and the Sparsity-property hold. Consider $m' \geq m_{\mu,4}$. Then, the set of functions well estimated by (μ, m)-thresholding estimators ($m \geq 2m'$) are quite large. Indeed, for any $s > 0$*

$$f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap W_{\mu^H}\left(\frac{2}{1+2s}\right) \Longrightarrow \sup_{m \geq 2m'} \sup_{0<\epsilon<1/e} (mt_\epsilon)^{-\frac{4s}{1+2s}} \mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 < \infty.$$

Remark 4.3. *We recall that the set $\mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap W_{\mu^H}\left(\frac{2}{1+2s}\right)$ can be considered as a large functional space since it contains the space $\mathcal{B}_{2,\infty}^{s}$ (see among others Autin [3]).*

5 Combining thresholding rules to get larger maxisets

We show here how to construct more powerful thresholding procedures in the maxiset sense by combining many thresholding rules with possibly non nested maxisets, provided that the *LD*-property and the *Sparsity*-property are satisfied. We then show the limitation of that method of combination by constructing an upper bound, that is the largest maxiset that could possibly be attained by estimators built from our procedure. Finally we apply it to the introductory example concerning the Blockshrink and Hard Tree estimators.

Let us begin with the following lemma.

Lemma 5.1. *Let $\mu^{(1)}$ and $\mu^{(2)}$ be two thresholding rules which satisfy the LD-property. Consider, for any $m > 0$,*

$$\begin{aligned}
\hat{f}_{\mu^{(3)},m} &= \hat{\alpha}\phi + \sum_{j=0}^{j_{mt_\epsilon}-1} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk} \mathbf{1}\left\{\mu_{j,k}^{(3)}(m, t_\epsilon, \hat{\theta}) > mt_\epsilon\right\} \psi_{j,k} \\
&= \hat{\alpha}\phi + \sum_{j=0}^{j_{mt_\epsilon}-1} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk} \mathbf{1}\left\{\max\left(\mu_{j,k}^{(1)}(m, t_\epsilon, \hat{\theta}), \mu_{j,k}^{(2)}(m, t_\epsilon, \hat{\theta})\right) > mt_\epsilon\right\} \psi_{j,k}
\end{aligned}$$

Then $\mu^{(3)}$ is a rule satisfying the LD-property, with $m_{\mu^{(3)},\nu} = \max(m_{\mu^{(1)},\nu+1}, m_{\mu^{(2)},\nu+1})$, for any $\nu > 0$.

This Lemma reflects the key point of our method to get estimators with larger maxisets. Indeed the following theorem holds:

Theorem 5.1. *Let $s > 0$ and $\mu^{(1)}$ and $\mu^{(2)}$ be two thresholding rules which satisfy the LD-property and the Sparsity-property. Consider estimators $\hat{f}_{\mu^{(3)},m}$ $(m > 0)$ defined as in the previous lemma. If $\mu^{(3)}$ satisfies the Sparsity-property too, then for any $m' \geq m_{_{\mu^{(3)},4}}$*

$$a)\ \forall m \geq 2m',\ \sup_{0<\epsilon<1/e} (mt_\epsilon)^{-\frac{4s}{1+2s}} \mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 < \infty$$

$$\Updownarrow$$

$$f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap W_{\mu^{(3)},m'}\left(\tfrac{2}{1+2s}\right),$$

$$b)\ W_{\mu^{(3)},m'}\left(\tfrac{2}{1+2s}\right) \supset W_{\mu^{(1)},m'}\left(\tfrac{2}{1+2s}\right) \cup W_{\mu^{(2)},m'}\left(\tfrac{2}{1+2s}\right).$$

In the previous theorem, the equivalence given in a) means that considering the maximum of two thresholding rules generates a new thresholding rule for which maxiset have been determined, provided that the *LD*-property and the *Sparsity*-property are satisfied. The embedding property b) is quite interesting since it proves that from two chosen thresholding rules $\mu^{(1)}$ and $\mu^{(2)}$ with possibly non nested maxisets and satisfying the *LD*-property and the *Sparsity*-property, we are able to construct a new rule $\mu^{(3)}$ which is at least as efficient as the two previous ones in the maxiset sense, provided that $\mu^{(3)}$ satisfies the *Sparsity*-property.

5.1 Limitation of the method

In this paragraph, we point out the limitation of our method for enlarging maxisets. We deduce this limitation from the obvious fact that if at least one of the thresholding rules is cautious, then its combination with other rules following Lemma 5.1 is cautious too. We give this result in the Theorem 5.2 but we need first to define a new important functional space:

Definition 5.1. *Let $m' \geq 1$ and $0 < r < 2$, a function f belongs to the space $W^*_{m'}(r)$ if and only if:*

$$\sup_{m \geq 2m'} \sup_{0<\lambda<\frac{2m}{e}} m^{-2}\lambda^r \left[-\ln\left(F^{-1}\left(\frac{\lambda}{2m}\right)\right)\right]^{-1} \sum_{j \leq j_\lambda+1} \sum_{k=0}^{2^j-1} \mathbf{1}\{|\theta_{jk}| > \lambda\} < \infty.$$

Theorem 5.2. *(Limitation of the method for enlarging maxisets). Let $s > 0$ and μ be a cautious thresholding rule satisfying the LD-property. Then, for any $m' \geq m_{_{\mu,4}}$*

$$\forall m \geq 2m',\ \sup_{0<\epsilon<1/e} (mt_\epsilon)^{-\frac{4s}{1+2s}} \mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 < \infty \Longrightarrow f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap W^*_{m'}\left(\frac{2}{1+2s}\right).$$

As a conclusion, we proved first, that larger maxisets can be obtained by combining existing thresholding rules that satisfy both the *LD*- and the *Sparsity*-properties. Second, there exists a well defined limitation to this method, meaning that thresholding rules emerging from our procedure fail whenever we are dealing with functions that cannot be estimated with the prespecified rate by any cautious rule. It remains an open question to determine if the largest maxiset attainable by combining thresholding rules using our method is as large as the functional space given in Theorem 5.2.

5.2 Example: Combining Blocks thresholding rules

Here we propose to provide an example of our method. The largest maxisets of thresholding rules which have been determined up to now are those of μ^{T} and μ^{B} and are known to be not embedded.

As previously precised, these two thresholding rules satisfy the *LD*-property and the *Sparsity*-property. When combining these two rules, we get the following rule, called Block Tree rule,

$$\mu_{jk}^{BT}\left(m, t_\epsilon, \hat{\theta}\right) := \max\left(\max_{(j',k')\in\mathcal{T}_{j,k}(mt_\epsilon)}\left|\hat{\theta}_{j'k'}\right|, \left(\sum_{k'\in\mathcal{P}_{j,k}(\epsilon)}\hat{\theta}_{jk'}^2\right)^{\frac{1}{2}}\right),$$

that clearly satisfies the *Sparsity*-property. From Theorem 5.1, we get:

Corollary 5.1. *Let $s > 0$ and $m' \geq 4.1 \geq \max(m_{\mu^T,5}, m_{\mu^B,5})$. Then*

$$\forall m \geq 2m', \quad \sup_{0<\epsilon<1/e}(mt_\epsilon)^{-\frac{4s}{1+2s}}\,\mathbb{E}\|\hat{f}_{\mu^{BT},m} - f\|_2^2 < \infty \iff f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap W_{\mu^{BT},m'}\left(\frac{2}{1+2s}\right).$$

6 Numerical experiments

We propose to illustrate our theoretical results with numerical experiments. It is well known that the sequence model given by the equation (2) and the following nonparametric regression model are equivalent with the calibration $\epsilon = \frac{\sigma}{\sqrt{N}}$. Let us introduce the notations of the nonparametric model we are dealing with:

$$Y_i = f\left(i/N\right) + \sigma\zeta_i,\ 1 \leq i \leq N,\ \zeta_i \text{ are i.i.d. } \mathcal{N}\left(0,1\right). \tag{7}$$

We generate the data sets from a large panel of functions often used in wavelet estimation studies (see Antoniadis et al. [2]), the number of observations $N = 2048$, the Signal to Noise Ratio, defined as the logarithmic decibel scale of the ratio of the standard deviation of the function values to the standard deviation of the noise, is set to 10. We use the Daubechies least asymmetric wavelets with 8 vanishing moments. We use the universal threshold value $\hat{\sigma}\sqrt{2\ln(N)}$ for the Hard Tree estimator $\hat{f}_{HT}$ and $\hat{\sigma}\sqrt{5\ln(N)}$ for the Blockshrink estimator $\hat{f}_B$ as suggested in Cai [10]. We adopt the standard approach to estimate σ by computing the Median Absolute Deviation (MAD) over the thresholded wavelet coefficients at the finest wavelet scale $J-1$ (see e.g., Vidakovic [26]).
The Integrated Squared Error of the estimator $\hat{f}$ at the m-th Monte Carlo replication ($1 \leq m \leq M$) ($ISE^{(m)}\left(\hat{f}\right)$) is computed as follows:

$$ISE^{(m)}\left(\hat{f}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}^{(m)}\left(\frac{i}{N}\right) - f\left(\frac{i}{N}\right)\right)^2. \tag{8}$$

The Mean ISE (MISE) is computed over $M = 2000$ Monte Carlo replications:

$$MISE\left(\hat{f}\right) = \frac{1}{M}\sum_{m=1}^{M} ISE^{(m)}\left(\hat{f}\right).$$

We use the connections between keep-or-kill estimation and hypothesis testing (see Abramovich et al.[1]) in order to report in Table 1 the number of false positives/negatives (i.e., type I/II errors). This is obtained by comparing the set of indices of wavelet coefficients kept by each estimator with the set of indices kept by the keep-or-kill Oracle estimator

$$\hat{f}^{\mathcal{O}} = \hat{\alpha}\phi + \sum_{(j,k)\in\mathcal{S}^{\mathcal{O}}} \hat{\theta}_{jk}\psi_{jk}, \tag{9}$$

where $\mathcal{S}^{\mathcal{O}} = \left\{(j,k)\,;j \in \mathbb{N}, j < j_{\lambda_{\frac{\sigma}{\sqrt{N}}},p}; 0 \leq k < 2^j; |\theta_{jk}| > \frac{\sigma}{\sqrt{N}}\right\}.$

When comparing the MISE results of the Blockshrink and of the Hard Tree estimators in Table 1 we understand that in practical situations we would not be able to decide which one to use. Indeed, according to the test function, it could be either the Blockshrink or the Hard Tree that performs the best. When not optimal, their MISE can be larger up to 70 (resp. 115) percent compared to the other method. That is a potential huge loss for a practitioner that does not choose the method adapted to the target function we want to reconstruct. This observation is exactly what the maxiset approach suggests when the maxiset of these two methods are non nested. When looking at the results of the Block Tree estimator $\hat{f}_{HBT}$, it provides almost always the lowest MISE. If this is not the case, the deviation w.r.t the MISE of the Blockshrink or of the Hard Tree does not pass over a reasonable 8 percent. There is no doubt that the Block Tree estimator is to be preferred over the two others. Table 1 shows the impressive synergy when combining methods to increase the true discoveries at a comparatively low price in terms of false positives yielding these good performances of the Block Tree estimator.

Remark 6.1. *When comparing the behaviors of Blockshrink and Hard Tree estimators, they are quite sensitive to the choice of the wavelet family and regularity. Nevertheless, whatever the setting Block Tree estimator remains the estimator to be preferred.*

	$\hat{f}_B$	$\hat{f}_{HT}$	$\hat{f}_{BHT}$	$\hat{f}^{\mathcal{O}}$	$\hat{f}_B$	$\hat{f}_{HT}$	$\hat{f}_{BHT}$	$\hat{f}^{\mathcal{O}}$
	Function: Step				Function: Doppler			
MISE	8.21	7.23	5.89	2.56	1.78	2.28	1.66	1.02
False positives	23.3	0.9	24.0	0.0	15.0	8.0	18.5	0
False negatives	15.2	22.1	10.9	0.0	14.6	20.9	13.2	0
Size	58.1	28.8	63.1	50.0	62.4	49.2	67.3	62
	Function: Wave				Function: Angles			
MISE	1.23	2.65	1.23	0.75	1.35	1.81	1.33	0.76
False positives	5.0	4.7	9.6	0.0	2.3	0.7	2.9	0.0
False negatives	9.3	20.5	4.9	0.0	8.9	13.9	8.3	0.0
Size	49.7	38.2	58.7	53.0	28.4	21.8	29.6	35.0
	Function: Blip				Function: Parabolas			
MISE	2.37	1.92	1.73	0.78	1.88	1.82	1.63	0.83
False positives	18.5	1.0	19.4	0.0	10.9	1.0	11.9	0.0
False negatives	10.3	13.0	7.7	0.0	4.8	7.3	4.3	0.0
Size	45.1	24.9	48.6	37.0	30.1	17.8	31.6	24.0
	Function: Blocks				Function: time.shift.sine			
MISE	4.40	4.06	3.17	1.44	0.78	1.31	0.84	0.57
False positives	40.2	0.5	40.6	0.0	5.8	1.1	6.8	0.0
False negatives	57.7	80.8	48.7	0.0	1.3	5.5	1.2	0.0
Size	134.6	71.7	143.9	152.0	29.5	20.6	30.6	25.0
	Function: Bumps				Function: Spikes			
MISE	1.49	1.48	1.15	0.57	0.80	0.85	0.65	0.35
False positives	77.5	2.2	77.7	0.0	26.4	1.5	26.8	0.0
False negatives	40.0	73.0	33.9	0.0	11.8	21.6	10.2	0.0
Size	206.5	98.2	212.8	169.0	81.6	47.0	83.7	66.0
	Function: Heavisine				Function: Corner			
MISE	2.54	1.58	1.58	0.79	0.46	0.67	0.47	0.25
False positives	4.4	1.0	5.4	0.0	2.1	1.1	3.1	0.0
False negatives	16.3	15.6	13.0	0.0	4.1	7.4	3.6	0.0
Size	17.1	14.4	21.4	28.0	20.0	15.7	21.5	22.0

Table 1: MISE (10^{-4}), number of false positives/negatives and average size of the number of non zero empirical wavelet coefficients in the estimator.

7 Appendix

This section aims at proving the results provided in our study. In the sequel C denotes a generic constant which does not depend on ϵ and that may be different from one line to the other.

7.1 Proof of Lemma 4.3

Proof. Fix $s > 0$, $m \geq m' \geq 1$ and $0 < \epsilon < 1/e$. Let μ be a thresholding rule that satisfies the *Sparsity*-property and consider $f \in W_{\mu,m'}\left(\frac{2}{1+2s}\right)$.

Because of the *Sparsity*-property,

$$
\begin{aligned}
& \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \mathbf{1}\{\mu_{jk}(m,t_\epsilon,\theta) > \frac{mt_\epsilon}{2}\} \\
\leq \quad & C_\mu \ln(\epsilon^{-1}) \sum_{n\in\mathbb{N}} (m2^n t_\epsilon)^{-2} \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\{\mu_{jk}(m2^n, t_\epsilon, \theta) \leq m2^n t_\epsilon\} \\
\leq \quad & C \ln(\epsilon^{-1}) \sum_{n\in\mathbb{N}} (m2^n t_\epsilon)^{-2} (m2^n t_\epsilon)^{2-\frac{2}{1+2s}} \\
\leq \quad & C \ln(\epsilon^{-1}) (mt_\epsilon)^{-\frac{2}{1+2s}}.
\end{aligned}
$$

So

$$
\sup_{m\geq m'} \sup_{0<\epsilon<1/e} (mt_\epsilon)^{\frac{2}{1+2s}} \left(\ln(\epsilon^{-1})\right)^{-1} \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \mathbf{1}\left\{\mu_{jk}(m,t_\epsilon,\theta) > \frac{m}{2} t_\epsilon\right\} < \infty.
$$

□

7.2 Proof of Theorem 4.1

Proof. ($\Longrightarrow$) Let a thresholding rule μ satisfy the *LD*- and the *Sparsity*- properties and $m' \geq m_{\mu,4}$. Suppose that there exists $C > 0$ such that $\mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 \leq C\ (mt_\epsilon)^{\frac{4s}{1+2s}}$, for any $m \geq 2m'$ and any $0 < \epsilon < 1/e$.

Let $m \geq 2m'$. Then,

$$
\begin{aligned}
\sum_{j\geq j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \quad & \leq \quad \mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 \\
& \leq \quad C\ (mt_\epsilon)^{\frac{4s}{1+2s}} \\
& \leq \quad C\, 2^{-\frac{2s}{1+2s}\ j_{mt_\epsilon}}.
\end{aligned}
$$

Using the continuity of t_ϵ in ϵ, we deduce that $f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}}$. Moreover,

$$
\begin{aligned}
& \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{\mu_{jk}\left(\frac{m}{2}, t_\epsilon, \theta\right) \leq \frac{m}{2} t_\epsilon\right\} \\
= \quad & A_1 + A_2 + A_3,
\end{aligned}
$$

with,

$$
\begin{aligned}
A_1 &= \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{\mu_{jk}\left(\frac{m}{2}, t_\epsilon, \theta\right) \le \frac{m}{2} t_\epsilon\Big\} \mathbf{1}\Big\{\mu_{jk}(m, t_\epsilon, \hat{\theta}) \le mt_\epsilon\Big\}\right] \\
&\le \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{\mu_{jk}(m, t_\epsilon, \hat{\theta}) \le mt_\epsilon\Big\}\right] \\
&\le \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 \\
&\le C, \\
A_2 &= \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{\mu_{jk}\left(\frac{m}{2}, t_\epsilon, \theta\right) \le \frac{m}{2} t_\epsilon\Big\} \mathbf{1}\Big\{\mu_{jk}(m, t_\epsilon, \hat{\theta}) > mt_\epsilon\Big\}\right] \\
&\le \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{|\mu_{jk}(m, t_\epsilon, \hat{\theta}) - \mu_{jk}(m, t_\epsilon, \theta)| > \frac{m}{2} t_\epsilon\Big\}\right] \\
&= \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbb{P}\left(|\mu_{jk}(m, t_\epsilon, \hat{\theta}) - \mu_{jk}(m, t_\epsilon, \theta)| > \frac{m}{2} t_\epsilon\right) \\
&\le C\ (mt_\epsilon)^{-\frac{4s}{1+2s}} \epsilon^4 \\
&\le C.
\end{aligned}
$$

The last inequalities use the decay property of functions μ_{jk} with respect to the first variable, the LD-property and the fact that $m \ge 2m_{_{\mu,4}}$.

Now

$$
\begin{aligned}
A_3 &= \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \sum_{j\ge j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{\mu_{jk}\left(\frac{m}{2}, t_\epsilon, \theta\right) \le \frac{m}{2} t_\epsilon\right\} \\
&\le \left(\frac{mt_\epsilon}{2}\right)^{-\frac{4s}{1+2s}} \sum_{j\ge j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \\
&\le C\ (mt_\epsilon)^{-\frac{4s}{1+2s}} 2^{-\frac{2s}{1+2s} j_{mt_\epsilon}} \\
&\le C.
\end{aligned}
$$

The last inequality holds since we have already proved that $f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}}$. When combining the bounds of A_1, A_2 and A_3 and when using the continuity on t_ϵ in ϵ, one deduces that $f \in W_{\mu,m'}(\frac{2}{1+2s})$.

($\Longleftarrow$) Suppose that $f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap W_{\mu,m'}(\frac{2}{1+2s})$ with $m' \ge m_{_{\mu,4}}$. For any any $m \ge 2m'$ and any $0 < \epsilon < 1/e$, the quadratic risk of the estimator $\hat{f}_{\mu,m}$ can be decomposed as follows:

$$
\begin{aligned}
\mathbb{E}\|\hat{f}_{\mu,m}-f\|_2^2 &= \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\theta_{jk}^2\mathbf{1}\left\{\mu_{jk}\left(m,t_\epsilon,\hat{\theta}\right)\le mt_\epsilon\right\}\right]\\
&\quad+\mathbb{E}\left[\sum_{j<j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}(\hat{\theta}_{jk}-\theta_{jk})^2\mathbf{1}\left\{\mu_{jk}\left(m,t_\epsilon,\hat{\theta}\right)> mt_\epsilon\right\}\right]\\
&\quad+\sum_{j\ge j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\theta_{jk}^2+\epsilon^2\\
&= A_4+A_5+A_6.
\end{aligned}
$$

Since $f\in\mathcal{B}_{2,\infty}^{\frac{s}{1+2s}}\cap W_{\mu,m'}(\frac{2}{1+2s})$ and due to the LD-property

$$
\begin{aligned}
A_4 &= \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\theta_{jk}^2\mathbf{1}\left\{\mu_{jk}\left(m,t_\epsilon,\hat{\theta}\right)\le mt_\epsilon\right\}\right]\\
&\le \sum_{j<j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\theta_{jk}^2\mathbf{1}\left\{\mu_{jk}\left(2m,t_\epsilon,\theta\right)\le 2mt_\epsilon\right\}\\
&\quad+\sum_{j<j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\theta_{jk}^2\mathbb{P}\left(|\mu_{jk}(m,t_\epsilon,\hat{\theta})-\mu_{jk}(m,t_\epsilon,\theta)|>mt_\epsilon\right)\\
&\le C\left[(mt_\epsilon)^{\frac{4s}{1+2s}}+\epsilon^4\right]\\
&\le C\,(mt_\epsilon)^{\frac{4s}{1+2s}}.
\end{aligned}
$$

Using the Cauchy-Schwarz inequality, the LD-property and Lemma 4.3,

$$
\begin{aligned}
A_5 &= \sum_{j<j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\mathbb{E}\left[(\hat{\theta}_{jk}-\theta_{jk})^2\mathbf{1}\left\{\mu_{jk}\left(m,t_\epsilon,\hat{\theta}\right)> mt_\epsilon\right\}\right]\\
&\le \sum_{j<j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\mathbb{E}\left[(\hat{\theta}_{jk}-\theta_{jk})^2\mathbf{1}\left\{\mu_{jk}\left(m,t_\epsilon,\theta\right)> \frac{m}{2}t_\epsilon\right\}\right]\\
&\quad+C\,\epsilon^2\sum_{j<j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\mathbb{P}^{\frac{1}{2}}\left(|\mu_{jk}(m,t_\epsilon,\hat{\theta})-\mu_{jk}(m,t_\epsilon,\theta)|>\frac{m}{2}t_\epsilon\right)\\
&\le C\left((mt_\epsilon)^{\frac{4s}{1+2s}}+\epsilon^2\right)\\
&\le C\,(mt_\epsilon)^{\frac{4s}{1+2s}}.
\end{aligned}
$$

Since $f\in\mathcal{B}_{2,\infty}^{\frac{s}{1+2s}}$

$$
\begin{aligned}
A_6 &= \epsilon^2+\sum_{j\ge j_{mt_\epsilon}}\sum_{k=0}^{2^j-1}\theta_{jk}^2\\
&\le \epsilon^2+C\,2^{-\frac{2s}{1+2s}j_{mt_\epsilon}}\\
&\le C\,(mt_\epsilon)^{\frac{4s}{1+2s}}.
\end{aligned}
$$

When combining the bounds of A_4, A_5 and A_6 and using the continuity of t_ϵ in ϵ one deduces that

$$\sup_{0<\epsilon<\frac{1}{e}} (mt_\epsilon)^{-\frac{4s}{1+2s}} \mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 < \infty.$$

This ends the proof. $\square$

7.3 Proof of Lemma 5.1

Proof. It is obvious that $\mu^{(3)}$ is a thresholding rule that generates $(\mu^{(3)}, m)$-thresholding estimators. Suppose that $\mu^{(1)}$ and $\mu^{(2)}$ satisfy the LD-property and consider for any $\nu > 0$, $m_{_{\mu^{(3)},\nu}} = \max(m_{_{\mu^{(1)},\nu+1}}, m_{_{\mu^{(2)},\nu+1}})$. Then, for any $0 < \epsilon < 1/e$ and any $m \geq m_{_{\mu^{(3)},\nu}}$

$$\begin{aligned}
& \mathbb{P}\left(|\mu_{jk}^{(3)}(m, t_\epsilon, \hat{\theta}) - \mu_{jk}^{(3)}(m, t_\epsilon, \theta)| > m_{_{\mu^{(3)},\nu}} t_\epsilon\right) \\
\leq\ & \mathbb{P}\left(|\mu_{jk}^{(1)}(m, t_\epsilon, \hat{\theta}) - \mu_{jk}^{(1)}(m, t_\epsilon, \theta)| > m_{_{\mu^{(3)},\nu}} t_\epsilon\right) + \mathbb{P}\left(|\mu_{jk}^{(2)}(m, t_\epsilon, \hat{\theta}) - \mu_{jk}^{(2)}(m, t_\epsilon, \theta)| > m_{_{\mu^{(3)},\nu}} t_\epsilon\right) \\
\leq\ & \mathbb{P}\left(|\mu_{jk}^{(1)}(m, t_\epsilon, \hat{\theta}) - \mu_{jk}^{(1)}(m, t_\epsilon, \theta)| > m_{_{\mu^{(1)},\nu+1}} t_\epsilon\right) + \mathbb{P}\left(|\mu_{jk}^{(2)}(m, t_\epsilon, \hat{\theta}) - \mu_{jk}^{(2)}(m, t_\epsilon, \theta)| > m_{_{\mu^{(2)},\nu+1}} t_\epsilon\right) \\
\leq\ & \epsilon^{\nu+1} \\
\leq\ & \frac{\epsilon^\nu}{2}.
\end{aligned}$$

Hence $\mu^{(3)}$ satisfies the LD-property too. $\square$

7.4 Proof of Theorem 5.1

Proof. $a)$ is a direct consequence of Lemma 5.1 and Theorem 4.1.
$b)$ becomes obvious when looking at the definition of spaces $W_{\mu,m'}(r)$ (with $0 < r < 2$). Indeed, for any $m \geq m'$, any $0 < \epsilon < 1/e$ and any sequence of real numbers θ,

$$\mu_{j,k}^{(3)}(m, t_\epsilon, \theta) = \max\left(\mu_{j,k}^{(1)}(m, t_\epsilon, \theta), \mu_{j,k}^{(2)}(m, t_\epsilon, \theta)\right) \geq \mu_{j,k}^{(i)}(m, t_\epsilon, \theta), \quad \text{for } i \in \{1, 2\}.$$

$\square$

7.5 Proof of Theorem 5.2

Proof. Consider a cautious rule μ that satisfies the LD-property. Assume that there exists $C > 0$ such that, for any $0 < \epsilon < 1/e$, any $m' \geq m_{_{\mu,4}}$ and any $m \geq 2m'$

$$\mathbb{E}\|\hat{f}_{\mu,m} - f\|_2^2 \leq C\,(mt_\epsilon)^{\frac{4s}{1+2s}}.$$

Consider $m' \geq m_{_{\mu,4}}$. Then,

$$\begin{aligned}
\sum_{j \geq j_{2m't_\epsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 &\leq \mathbb{E}\|\hat{f}_{\mu,2m'} - f\|_2^2 \\
&\leq C\,(2m't_\epsilon)^{\frac{4s}{1+2s}} \\
&\leq C\, 2^{-\frac{2js}{1+2s}}.
\end{aligned}$$

Using the continuity of t_ϵ in ϵ, one gets $f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}}$.

Let us now prove that f necessarily belongs to $W^*_{m'}(\frac{2}{1+2s})$, i.e.

$$\sup_{m\geq 2m'} \sup_{0<\lambda<\frac{2m}{e}} m^{-2}\lambda^r \left[-\ln\left(F^{-1}\left(\frac{\lambda}{2m}\right)\right)\right]^{-1} \sum_{j\leq j_\lambda+1} \sum_{k=0}^{2^j-1} \mathbf{1}\{|\theta_{jk}|>\lambda\} < \infty.$$

When considering the change of variables $t_\epsilon = \lambda(2m)^{-1}$, ($m \geq 2m'$, $0<\lambda<\frac{2m}{e}$) one aims at proving that, for any $m \geq 2m'$:

$$\sup_{0<\epsilon<\frac{1}{e}} \epsilon^2 \left(mt_\epsilon\right)^{-\frac{4s}{1+2s}} \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \mathbf{1}\{|\theta_{jk}|>2mt_\epsilon\} < \infty.$$

Since μ is a cautious rule, for any $m \geq 2m'$ and $0<\epsilon<\frac{1}{e}$,

$$\begin{aligned}
\epsilon^2 \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \mathbf{1}\{|\theta_{jk}|>2mt_\epsilon\} &\leq \epsilon^2 \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \mathbf{1}\{\mu_{jk}(m,t_\epsilon,\theta)>2mt_\epsilon\} \\
&\leq \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} (\hat\theta_{jk}-\theta_{jk})^2 \mathbf{1}\{\mu_{jk}(m,t_\epsilon,\theta)>2mt_\epsilon\}\right] \\
&= B_1+B_2,
\end{aligned}$$

with

$$\begin{aligned}
B_1 &= \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} (\hat\theta_{jk}-\theta_{jk})^2 \mathbf{1}\{\mu_{jk}(m,t_\epsilon,\theta)>2mt_\epsilon \text{ and } \mu_{jk}(m,t_\epsilon,\hat\theta)>mt_\epsilon\}\right] \\
&\leq \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} (\hat\theta_{jk}-\theta_{jk})^2 \mathbf{1}\{\mu_{jk}(m,t_\epsilon,\hat\theta)>mt_\epsilon\}\right] \\
&\leq \mathbb{E}\|\hat f_{\mu,m}-f\|_2^2 \\
&\leq C\left(mt_\epsilon\right)^{\frac{4s}{1+2s}},
\end{aligned}$$

and, because of the LD-property and the Cauchy-Schwarz inequality

$$\begin{aligned}
B_2 &= \mathbb{E}\left[\sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} (\hat\theta_{jk}-\theta_{jk})^2 \mathbf{1}\{\mu_{jk}(m,t_\epsilon,\theta)>2mt_\epsilon \text{ and } \mu_{jk}(m,t_\epsilon,\hat\theta)\leq mt_\epsilon\}\right] \\
&\leq C\,\epsilon^2 \sum_{j<j_{mt_\epsilon}} \sum_{k=0}^{2^j-1} \mathbb{P}^{\frac{1}{2}}\left(|\mu_{jk}(m,t_\epsilon,\hat\theta)-\mu_{jk}(m,t_\epsilon,\theta)|>mt_\epsilon\}\right) \\
&\leq C\,\epsilon^2 \\
&\leq C\left(mt_\epsilon\right)^{\frac{4s}{1+2s}}.
\end{aligned}$$

The last inequality is obtained because of $m \geq m_{\mu,4}$.

Combining B_1 and B_2 and still using the continuity of t_ϵ in ϵ, one gets $f \in W^*_{m'}(\frac{2}{1+2s})$. This ends the proof. □

References

[1] Abramovich, F., Benjamini, Y., Donoho, D., and Johnstone, I. (2006). Adapting to Unknown Sparsity by Controlling the False Discovery Rate. *Annals of Statistics*, **34**(2), 584-653.

[2] Antoniadis, A., Bigot, J., and Sapatinas, T. (2001). Wavelet Estimators in Nonparametric Regression: a Comparative Simulation Study. *Journal of Statistical Software*, **6**(6), 1-83.

[3] Autin, F. (2004). Maxiset point of view in nonparametric estimation. *Ph.D. thesis* Université Paris 7 - Denis Diderot.

[4] Autin, F., Picard, D., Rivoirard, V. (2006). Large variance gaussian priors in Bayesian nonparametric estimation: a maxiset approach. *Mathematical Methods of Statistics*, **15**(4), 349-373.

[5] Autin, F. (2008). On the performances of a new thresholding procedure using tree structure. *Electron. Journ. of Statist.*,volume 2, 412-431.

[6] Autin, F. (2008). Maxisets for μ-thresholding rules. *Test*, **17**(2), 332-349.

[7] Autin, F., Le Pennec, E., Loubes, J.-M., and Rivoirard, V. (2010). Maxisets for model selection. *Constr. Approx.*, **31**(2), 195-229.

[8] Autin, F., Freyermuth, J.-M. and von Sachs, R. (2011). Ideal Denoising within a Family of Tree-Structured Wavelet Estimators. *Electron. Journ. of Statist.*, **5**, 829-855.

[9] Autin, F., Freyermuth, J.-M., and von Sachs, R. (2011). Block-Threshold-Adapted Estimators via a maxiset approach. Discussion Paper 2011/17, ISBA, Université Catholique de Louvain.

[10] Cai, T. (1997). On adaptivity of blockshrink wavelet estimator over Besov spaces. *Technical report 97-05*, Purdue University.

[11] Cai, T. (1999). Adaptive Wavelet Estimation: a Block Thresholding and Oracle Inequality Approach. *Annals of Statistics*, **27**(3), 898-924.

[12] Cai, T., and Silverman, B.W. (2001). Incorporating Information on Neighboring Coefficients into Wavelet Estimation. *Sankhya*, **63**, 127-148.

[13] Cai, T., and Silverman, B.W. (2008). On Information Pooling, Adaptability and Superefficiency in Nonparametric Function Estimation. *J. Multivariate Analysis*, **99**, 412-436.

[14] Cai, T., and Zhou, H. (2009). A data-driven block thresholding approach to wavelet estimation. *Ann. of Stat.*, **37**, 569-595.

[15] Cohen, A., Dahmen W., Daubechies I., and DeVore, R. (2001). Tree Approximation and Optimal Encoding. *Appl. Comput. Harmon. Anal.*, **11**(2), 192-226.

[16] Cohen, A., De Vore, R., Kerkyacharian, G., and Picard, D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, **11**, 167-191.

[17] Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

[18] Engel, J. (1994). A simple wavelet approach to nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.*, **49**(2), 242-254.

[19] Hall, P., Kerkycharian, G., and Picard D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. of Stat.*, **26**(3), 922-942.

[20] Hall, P., Kerkycharian, G., and Picard D. (1998). On the minimax optimality for block thresholded wavelet estimators. *Statist. Sinica.*, **9**, 33-49.

[21] Kerkyacharian, G., and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases. *Test*, **9**(2), 283-344.

[22] Kerkyacharian, G., and Picard, D. (2002). Minimax or maxisets? *Bernoulli*, **8**(2), 219-253.

[23] Lepski, O.V. (1991). Asymptotically minimax adaptive estimation I: Upperbounds. Optimally adaptive estimates. *Theory Probab. Appl.*, **36**, 682-697.

[24] Lorentz, G.G. (1950). Some new functional spaces. *Ann. of Math. Statist. Probab. Lett.*, **51**(2), 37-55.

[25] Massart, P. *Concentration inequalities and model selection* Lectures on probability theory and statistics (Saint-Flour, 2003), Lecture Notes in Math., 1896, Springer, Berlin, 2007.

[26] Vidakovic, B. (1999). Statistical Modelling by Wavelets, John Wiley & Sons, Inc., New York, 384 pp.