T E C H N I C A L
R E P O R T

11019

Block-Threshold-Adapted Estimators via a maxiset approach

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

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

Block-Threshold-Adapted Estimators via a maxiset approach

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

June 30, 2011

Abstract

We study the performance of a large collection of block thresholding wavelet estimators, namely the *Horizontal Block Thresholding family.* In particular, we adopt a maxiset point of view, i.e. we are asking for the maximal functional space for a given estimator to converge in the L_2-sense with a chosen rate of convergence. We provide sufficient conditions on the choices of rates and threshold values to ensure large maxisets. By deriving maxiset embeddings, we identify the best estimator of such a family, that is the one associated with the largest maxiset. As a particularity of this paper we propose a refined maxiset approach that models method-dependent threshold values. By a series of simulation studies, we confirm the good performance of the best estimator when comparing to the other members of its family.

Keywords : Besov spaces, curve estimation, maxiset approach, rate of convergence, thresholding methods, wavelet-based estimation.

[*]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

Nonparametric estimation of functions by non-linear wavelet methods has proven to be a real success story, in particular for functions showing a high spatial variability. This is due to the fact that wavelets localize the information of a function in a few large coefficients for a wide range of function classes, the key property to the good performance of hard and soft thresholding estimators. Using empirical wavelet coefficients which are larger than a chosen threshold value, namely the Universal Threshold (U.T.), Donoho and Johnstone (1994), among others, showed that these estimators are not only near optimal over Besov spaces but also adaptive for the regularity parameter. Nevertheless these thresholding methods suffer from some criticisms over the last decade because of the presence of the suboptimal log-term in their minimax performance as shown in Cai (1999). Related to this, Autin (2004) emphasizes that such thresholding methods are *elitist*: they do not use small - but potentially important - empirical wavelet coefficients for reconstructing the function of interest.

In order to remedy the shortcomings of elitist procedures both from a theoretical and practical point of view, it has been shown in recent literature (Cai (1999), Hall et al. (1998, 1999), and Autin (2004, 2008), Autin et al. (2011), among others) that one can do better by using information from neighboring empirical wavelet coefficients. Cai (1997) proved that wavelet estimators based on *thresholding of empirical wavelet coefficients by blocks* (BT-methods) can be minimax optimal over Besov spaces (i.e. without the suboptimal log-term), such as the *Blockshrink estimator*. This estimator reconstructs functions using blocks of empirical wavelet coefficients for which the l_2-mean is larger than the noise level ϵ, up to a constant. From a different perspective, Autin (2008) proved that the set of functional spaces well estimated by wavelet estimators using BT-methods (the so-called maxiset) can be larger than the ones of hard and soft thresholding estimators. For instance, the *Maximum-block* estimator was proved to perform particularly well. This estimator uses blocks of empirical wavelet coefficients for which the l_∞-mean is larger than the U.T., up to a constant.

Indeed, these BT-estimators provide good visual reconstructions as shown in the Figures 1-4. This can be explained by the group-structure of the large true wavelet coefficients represented in the Figure 1 (the darker, the larger the coefficient magnitude). Note in particular the ability of the Blockshrink estimator to retrieve the local group-structure down to the finest scales where suggested by the presence of sharp local signal structure.

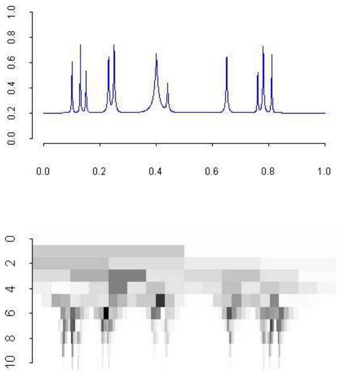

Figure 1: True function.

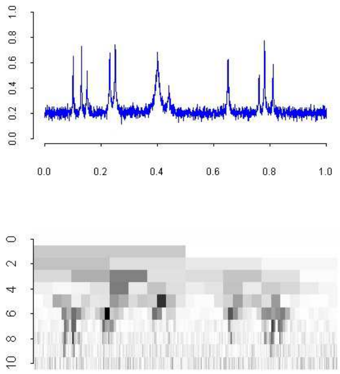

Figure 2: Noisy data.

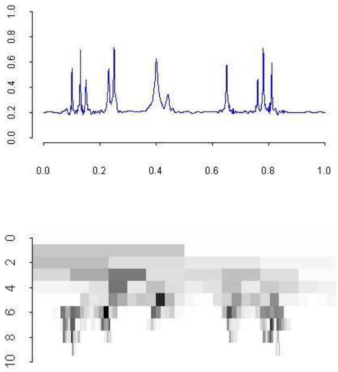

Figure 3: Blockshrink estimator.

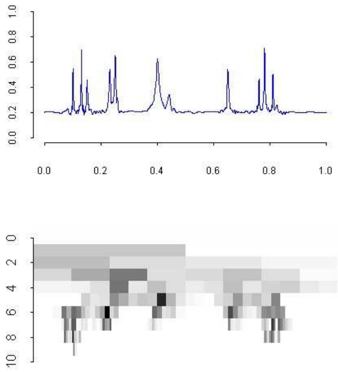

Figure 4: Maximum-block estimator.

In this paper we consider a wide range of wavelet estimators relying on BT-methods (BT-estimators), and we propose to study both their theoretical and numerical performances. We provide conditions to ensure that all the BT-estimators considered in the sequel perform well in maxiset sense in particular. These conditions deal with the *scores* of the blocks of empirical wavelet coefficients (their ℓ_p-mean) and the *threshold value* they are compared with.

The paper is organized as follows. After recalling in Section 2 the sequential version of the Gaussian white noise model, we introduce in Section 3 a general family of BT-estimators: the Horizontal Block Thresholding family. Estimators of such a family are associated with a thresholding rule on non overlapping blocks of coefficients. They are distinguished by the way how to compare scores of blocks - that are the l_p-means of blocks - to a threshold value that depends on the noise level $\epsilon > 0$ and may depend on the parameter p ($2 \leq p \leq \infty$). (We recall the usual fact that in this more abstract study the noise level ϵ is related to the sample size of the accompanying nonparametric curve estimation problem, and that all derived quantities such as rates of convergence and threshold values depending on ϵ do have the usual interpretation in terms of sample size.) For the specific case where the threshold value is of order $\epsilon(\log(\epsilon^{-1}))^{\frac{1}{2}-\frac{1}{p}}$ (using the convention $\frac{1}{\infty}=0$), the family under interest contains the Blockshrink and Maximum-block estimators.

In Section 4 we compute the set of all the functions *well estimated* by estimators belonging to the Horizontal Block Thresholding family. Precisely we identify all the functions for which the quadratic risk of these estimators does not exceed a given rate of convergence (see Theorem 4.1). We provide sufficient conditions on the possible choices of the rate of convergence and the threshold value to ensure that the maxisets of the estimators of the family contain Besov bodies (see Proposition 4.1), that is, we ensure that each estimator of the Horizontal Block Thresholding Family we are interested in performs well.

Further we show in Section 4 that for a wide range of threshold values, the family under study contains an estimator for which the maxiset at a given rate is the largest one. Hence it corresponds to the most performing estimator within the family according to the maxiset approach. Moreover, we point out that the best way to give a score to blocks may be different from one family to the other. Indeed, it depends on the threshold value under consideration (see Corollaries 4.1 and 4.2). This result is our most important contribution of this paper and it can be nicely interpreted through hypothesis testing ideas, in terms of false positives (erroneously active coefficients) and false negatives (erroneously deleted coefficients). The Corollary 4.1 shows that when using a conservative threshold value (such as the U.T.) that controls the false positives, one should apply the method that reduces the most the false negatives in order to get the largest maxiset. Choosing less conservative threshold values allows to get even larger maxisets but this choice leads to an important increase in false positives, and therefore, as expressed by the Corollary 4.2, a method that is able to control simultaneously false positives and negatives is required. Section 5 proposes numerical experiments to confirm the superiority of the best estimator using as a benchmark the informative results obtained by the keep-or-kill Oracle estimator (defined in, e.g., (9)). Finally after brief conclusive remarks in Section 6, Section 7 presents the proofs of our main results.

2 Wavelet setting and model

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 (1992). 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 $\theta = (\theta_{jk})_{jk}$ 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, i.e., of observations of f embedded into Gaussian white noise (see also equation (8) and its interpretation). I.e., we

dispose of observations of these coefficients which are assumed to be realizations of independent random variables:

$$\begin{aligned} \hat{\alpha} &= \alpha + \epsilon\xi, \\ \hat{\theta}_{jk} &= \theta_{jk} + \epsilon\xi_{jk}, \end{aligned} \tag{2}$$

where ξ, ξ_{jk} are i.i.d. $\mathcal{N}(0,1)$, $0 < \epsilon < 1$ is supposed to be the noise level, and where the sequence $(\theta_{jk})_{j,k}$ is sparse, meaning that only a small number of *large* coefficients contain nearly all the information about the signal. That motivates the use of keep-or-kill estimators, for which we recall the hard thresholding estimator:

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

where $\mathcal{S}_\epsilon = \left\{(j,k)\,; j \in \mathbb{N}, j \le j_{n_\epsilon}; 0 \le k < 2^j; \left|\hat{\theta}_{jk}\right| > t_\epsilon = m\epsilon\sqrt{\log(\epsilon^{-1})}\right\}$. If $\mathcal{S}_\epsilon$ is non empty, it forms an *unstructured* set of indices associated with *large empirical wavelet coefficients* (in the sequel, by *large empirical wavelet coefficients*, we understand those which belong to $\mathcal{S}_\epsilon$). Here,

- $0 < m < \infty$,
- n_ϵ is the integer such that $2^{-(n_\epsilon+1)} \le (mt_\epsilon)^2 < 2^{-n_\epsilon}$. For a general $t_\epsilon < 1$, n_ϵ is the finest scale up to which we can consider the empirical wavelet coefficients to reconstruct the signal f.

This term by term thresholding does not take into account the information that give us the clusters of wavelet coefficients that we observed in the Figure 1. This information allows to be more precise in the choice of the coefficients to keep. Indeed, we would not use in the reconstruction a large isolated wavelet coefficient because it is not likely to be part of the signal; while a small coefficient in the neighborhood of large coefficients would be kept. Under this model and considering the sequence $\{\theta_{jk}\}_{jk}$ to be in Besov spaces (see Definition 4.1), several impressive minimax results were obtained for such estimators (see in particular Cai (1999) and Cai and Zhou (2009)). Nevertheless, these results do not model the information given by the clusters of coefficients since the Besov norm is invariant under permutations within scale. Using the maxiset approach, for which the basics are recalled further down, we introduce a way to model these clusters of coefficients by introducing new functional spaces related to the methods (see Definition 4.2). This allows a more precise characterization of the performances of the estimators under consideration.

Let us consider an estimator $\tilde{f}^{(\lambda)}$ with a threshold value λ. The maxiset approach consists in computing the set of all the functions for which the rate of convergence of the quadratic-risk of the estimator $\tilde{f}^{(\lambda)}$ is at least as fast as a given rate of convergence ρ (with $\rho = \rho_\epsilon \to 0$ as $\epsilon \to 0$), i.e:

$$\sup_{0<\epsilon<1} \rho_\epsilon^{-1}\mathbb{E}\|\tilde{f}^{(\lambda)} - f\|_2^2 < \infty \iff f \in \mathcal{G}.$$

In this setting, the functional space $\mathcal{G}$ will be called maxiset of $\tilde{f}^{(\lambda)}$ for the rate of convergence ρ_ϵ. Obviously, the larger the maxiset, the better the procedure; and the slower the rate, the larger the maxiset (and conversely). In the existing maxiset literature, λ and ρ are generally chosen to be about the same order. Autin (2008) already gives results for more flexible choices of the rate and a given threshold value. In this paper, we go beyond this by further disconnecting the rate and the threshold value.

3 Horizontal Block Thresholding Estimators

For any $0 < m < \infty$ and any $2 \leq p \leq \infty$, let

- $(t_{\epsilon,p})_\epsilon$ and $(v_{\epsilon,p})_\epsilon$ be two sequences of positive real numbers continuously tending to 0 as ϵ goes to 0;
- $j_{v_{\epsilon,p}}$ be the integer such that $2^{-j_{v_{\epsilon,p}}} \leq (mv_{\epsilon,p})^2 < 2^{1-j_{v_{\epsilon,p}}}$ for any $0 < \epsilon < 1$;
- and let the length of the blocks be fixed in the order of $\log\left(\epsilon^{-1}\right)$ (this choice has been proven to be pertinent from both a minimax (see Cai (1999)) and a maxiset (see Autin (2008)) point of view.

In Section 4 we will use the tuning parameter m to link threshold values of the form $mt_{\epsilon,p}$ and rates of convergence of the L_2-risk of the form $\rho_\epsilon = (mv_{\epsilon,p})^\beta$. As usual, $0 < \beta < 1$ depends on the regularity of the given function space, e.g. we recall the form of the minimax-rate $\epsilon^{\frac{4s}{1+2s}}$ over Besov spaces $\mathcal{B}^\gamma_{2,\infty}$ (see Definition 4.1 below), which will amount to choosing $v_{\epsilon,p} = \epsilon$ - see also Example 2 below.
Let us now define a general BT-estimator $\tilde{f}^{(t)}_p$, $2 \leq p \leq \infty$ associated with a block-thresholding rule, namely the HBT$^{(t)}(p)$-estimator.

Definition 3.1. *[HBT$^{(t)}(p)$-estimator] Let $0 < \epsilon < 1$ and $2 \leq p \leq \infty$ and a given $m > 0$. Let us consider the following wavelet estimator $\tilde{f}^{(t)}_p$*

$$\begin{aligned}
\tilde{f}^{(t)}_p = \tilde{f}^{(t)\epsilon}_p &:= \hat{\alpha}\phi + \sum_{(j,k)\in H_\epsilon(t,p)} \hat{\theta}_{jk}\psi_{jk} \\
&= \hat{\alpha}\phi + \sum_{j\in\mathbb{N}, j<j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk}\ \mathbf{1}\left\{\|\hat{\theta}\ /\ B_{jk}(\epsilon)\|_p > m\ t_{\epsilon,p}\right\}\psi_{jk},
\end{aligned}$$

where, for any $0 < \epsilon < 1$, $B_{jk}(\epsilon)$ corresponds to the set of indices which contains k and such that $B_{jk}(\epsilon) \in \{B^{(1)}_j, B^{(2)}_j, \dots, B^{(u_{j,\epsilon})}_j\}$ with

$$B^{(u)}_j := \left\{k', 0 \leq k' < 2^j, (u-1)\lfloor\log(\epsilon^{-1})\rfloor \leq k' < u\lfloor\log(\epsilon^{-1})\rfloor\right\}, u \in \{1, 2, \dots, u_{j,\epsilon} = 2^j(\lfloor\log(\epsilon^{-1})\rfloor)^{-1}\}.$$

Notations used are the following

$$\begin{aligned}
\|\theta\ /\ B_{jk}(\epsilon)\|_p &:= \left(\frac{1}{\#B_{jk}(\epsilon)} \sum_{k'\in B_{jk}(\epsilon)} |\theta_{jk'}|^p\right)^{1/p} \quad \text{for } 2 \leq p < \infty, \\
\|\theta\ /\ B_{jk}(\epsilon)\|_\infty &:= \max\left\{|\theta_{jk'}|, k' \in B_{jk}(\epsilon)\right\}.
\end{aligned}$$

For any $2 \leq p \leq \infty$, $H_\epsilon(t,p)$ is the set of empirical wavelet coefficients used in the reconstruction based on a thresholding rule using the ℓ_p-norm and the threshold value $mt_{\epsilon,p}$.
Moreover, for any scale $j < j_{v_{\epsilon,p}}$, $\left(B^{(u)}_j, u \in \{1, 2, \dots, u_{j,\epsilon}\}\right)$ constitutes a set of non overlapping blocks of indices with size smaller or equal to $\log(\epsilon^{-1})$.

From now on, we will study the performances of these BT-estimators to address the following question: what is the best choice of ℓ_p-norm to consider ($2 \leq p \leq \infty$)? In the next section we use the maxiset approach to prove that among the different possibilities of choice of p the best one depends on the threshold value used. Comparing to Autin (2008) in which rules are based on the universal threshold value (up to a constant), we shall consider here a wide range of threshold values to refine existing maxiset results in the literature. Below are listed three examples of threshold values and rates we are interested in:

Example 1:

$v^{(1)}_{\epsilon,p} = t^{(1)}_{\epsilon,p} = \epsilon\sqrt{\log(\epsilon^{-1})}$
With such a choice, the threshold value corresponds to the universal threshold value (see Donoho and Johnstone (1994)) and the rate ρ_ϵ with $\beta = 4s(1+2s)^{-1}$ corresponds to the minimax rate over $\mathcal{B}^s_{2,\infty}$ up to a term $(\log\left(\epsilon^{-1}\right))^{1/2}$.

Example 2:

$v^{(2)}_{\epsilon,p} = (\log(\epsilon^{-1}))^{\frac{1}{p}-\frac{1}{2}} t^{(2)}_{\epsilon,p} = \epsilon$. With such a choice, the rate ρ_ϵ with $\beta = 4s(1+2s)^{-1}$ corresponds to the minimax rate over $\mathcal{B}^s_{2,\infty}$.

Example 3:

Choose $v^{(3)}_{\epsilon,p} = t^{(3)}_{\epsilon,p} = \epsilon(\log(\epsilon^{-1}))^{\frac{1}{2}-\frac{1}{p}}$. With such a choice, the threshold value corresponds to a lower threshold value than the universal threshold and the rate ρ_ϵ with $\beta = 4s(1+2s)^{-1}$ corresponds to the minimax rate over $\mathcal{B}^s_{2,\infty}$ up to a term $(\log\left(\epsilon^{-1}\right))^{1/2-1/p}$.

Remark 3.1. *For the case $t_{\epsilon,p} = \epsilon\left(\log(\epsilon^{-1})\right)^{\frac{1}{2}-\frac{1}{p}}$ (with $0<\epsilon<1$ and the convention $\frac{1}{\infty}=0$), the estimator $\tilde{f}^{(t)}_2$ is the BlockShrink estimator proposed by Cai (1999) and can be viewed as an hybrid version of NeighBlock estimator introduced by Cai and Silverman (2001), whereas $\tilde{f}^{(t)}_\infty$ is the Maximum Block estimator proposed by Autin (2008).*

Define the *Horizontal Block Thresholding* family, namely $(\mathrm{HBT}^{(t)}_\epsilon)$, as

$$\mathrm{HBT}^{(t)}_\epsilon = \left\{\tilde{f}^{(t)}_p,\ 2 \le p \le \infty\right\}.$$

At first glance, as $2 \le p \le \infty$ is real-valued, this family of estimators $(\mathrm{HBT}^{(t)}_\epsilon)$ seems to be uncountable. But it is not for a large choice of threshold value as we shall see in Propositions 7.2 and 7.3.

4 Main results

4.1 Functional spaces: definitions and embeddings

In this paragraph, we characterize the functional spaces which shall appear in the maxiset study of our estimators. Recall that, for later use of these functional spaces, we shall consider wavelet bases with V vanishing moments.

Definition 4.1. *Let* $0 < \gamma < V$. *We say that a function* $f \in L_2([0,1])$ *belongs to the Besov space* $\mathcal{B}^{\gamma}_{2,\infty}$ *if and only if:*

$$\sup_{J \in \mathbb{N}} 2^{2J\gamma} \sum_{j \geq J} \sum_{k=0}^{2^j-1} |\theta_{jk}|^2 < \infty.$$

Besov spaces naturally appear in estimation problems (see Autin (2004) and Cohen et al. (2001)). These spaces characterize the functions for which the energy of wavelet coefficients on scales larger than J ($J \in \mathbb{N}$) is decreasing exponentially in J. For an overview of these spaces, see Hardle et al. (1998).

We provide here the definition of a new function space which is the key to our results:

Definition 4.2. *Let* $m' > 0$, $0 < r < 2$ *and* $1 \leq p \leq \infty$. *We say that a function* f *belongs to the space* $\mathcal{W}^{(t,v)}_{r,m',p}$ *if and only if:*

$$\sup_{0<\lambda<1} (m' v_{\lambda,p})^{r-2} \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{ \|\theta \ / \ B_{jk}(\lambda)\|_p \leq m' t_{\lambda,p} \Big\} < \infty.$$

First, note that the larger r, the larger the functional space. Second, the spaces $\mathcal{W}^{(t,v)}_{r,m',p}$ ($m' > 0, 0 < r < 2$) are not invariant under permutations of wavelet coefficients within each scale. This makes them appear more interesting than weak Besov spaces (see Cohen et al. (2001) for an explicit definition) which are usually studied to derive maxiset results, in connection with near minimax rate over Besov spaces and for hard thresholding estimators as defined in equation (3). Indeed, it is precisely the non-invariance property that allows to distinguish functions according to the "block-neighborhood properties" of their wavelet coefficients.
Assuming some conditions on the choice of both rate of convergence and threshold value, these functional spaces are enlargements of classical Besov spaces as suggested by our following Proposition 4.1.

Proposition 4.1. *Let* $0 < s < V$ *and* $2 \leq p \leq \infty$. *Assume that* $(t_{\epsilon,p})_\epsilon$ *and* $(v_{\epsilon,p})_\epsilon$ *are such that*

$$\sup_{0<\epsilon<1} t_{\epsilon,p} \ {v_{\epsilon,p}}^{-1} < \infty. \tag{4}$$

Then, for any $m' > 0$

$$\mathcal{B}^{s}_{2,\infty} \subseteq \mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m',p}. \tag{5}$$

Our following Proposition 4.2 shows that, for the same parameters m' and r ($m' > 0, 0 < r < 2$), the functional spaces $\mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m',p}$ ($p \geq 2$) are embedded. The larger p the larger $\mathcal{W}^{(t,v)}_{r,m',p}$.

Proposition 4.2. *Let* $2 \leq p < q \leq \infty$. *Assume that for any* $0 < \epsilon < 1$, $t_{\epsilon,p}$ *and* $v_{\epsilon,p}$ *only depend on* ϵ. *Then, for any* $m' > 0$ *and any* $0 < r < 2$, *we have the following embeddings of spaces:*

$$\mathcal{W}^{(t,v)}_{r,m',p} \subseteq \mathcal{W}^{(t,v)}_{r,m',q}.$$

The assertion of this proposition changes however, if the scores over blocks $\{B_{j,k}\}$ appearing in Definition 4.2 are compared to a "threshold" value which depends in a particular way also on p, namely by rescaling with the length of the blocks:

Proposition 4.3. *Let $s > 0$ and $2 \le p < q \le \infty$. Assume that for any $0 < \epsilon < 1$, $\left(\log(\epsilon^{-1})\right)^{\frac{1}{p}} t_{\epsilon,p}$ and $v_{\epsilon,p}$ only depend on ϵ. Then, for any $m' > 0$ and any $0 < r < 2$, we have the following embeddings of spaces:*

$$\mathcal{W}^{(t,v)}_{r,m',q} \subseteq \mathcal{W}^{(t,v)}_{r,m',p}.$$

Further insights into these two crucial results can be found by investigating their proofs, but it is essential to note that these results prepare the ground to find the "maxiset-optimal" estimator $\tilde{f}^{(t)}_p$, i.e. the best p, according to the more refined specification of both rate (via $v_{\epsilon,p}$) and threshold value (via $t_{\epsilon,p}$). For this we refer to Corollaries 4.1 and 4.2 below.

4.2 Maxiset results

In this paragraph we provide the maximal space (maxiset) of any $\tilde{f}^{(t)}_p \in \mathrm{HBT}^{(t)}_\epsilon$ associated to a large collection of rates, that are $(mv_{\epsilon,p})^{\frac{4s}{1+2s}}$ $(s > 0)$ and of threshold values $t_{\epsilon,p}$. Some assumptions are necessary for sequences $(v_{\epsilon,p})_\epsilon$ and $(t_{\epsilon,p})_\epsilon$ $(2 \le p \le \infty)$ to ensure the validity of our next results. Firstly we suppose that, for any $c > 0$ there exists $m_c > 0$ - the bigger c the bigger m_c - such that

$$\sup_{0<\epsilon<1} v_{\epsilon,p}^{-c}\, \mathbb{P}\left(\|Z(\epsilon)\|_p > m_c\, t_{\epsilon,p}\right) < \infty, \tag{6}$$

where $Z(\epsilon) = (Z_1, \ldots, Z_{\lceil \log(\epsilon^{-1})\rceil})$ is a vector with i.i.d. $\mathcal{N}(0, \epsilon^2)$ entries. Secondly, assume that

$$\sup_{0<\epsilon<1} \frac{v_{\epsilon,p}}{\epsilon} > 0, \quad \sup_{0<\epsilon<1} \frac{\epsilon(\log(\epsilon^{-1}))^{\frac{1}{2}}}{v_{\epsilon,p}} > 0, \quad \text{and} \quad \sup_{0<\epsilon<1} \frac{t_{\epsilon,p}}{\epsilon(\log(\epsilon^{-1}))^{\frac{1}{2}-\frac{1}{p}}} > 0. \tag{7}$$

For any chosen $s > 0$, the reader can check that assumptions (6) and (7) are satisfied by the three examples given in the Section 3. We can now state the main theorem.

Theorem 4.1. *Let $s > 0$ and $2 \le p \le \infty$. We have the following equivalence:*

$$\text{For any } m \ge 2m_4,\ \sup_{0<\epsilon<1} (mv_{\epsilon,p})^{-\frac{4s}{1+2s}}\, \mathbb{E}\|\tilde{f}^{(t)}_p - f\|_2^2 < \infty \iff f \in \mathcal{B}^{\frac{s}{1+2s}}_{2,\infty} \cap \left(\bigcap_{m' \ge m_4} \mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m',p} \right).$$

Note that, if assumption (4) is satisfied, maxisets of estimators in $\mathrm{HBT}^{(t)}_\epsilon$ are quite large functional spaces since from (5) of Proposition 4.1 and Besov embedding properties, we deduce that the functional space $\mathcal{B}^{\frac{s}{1+2s}}_{2,\infty} \cap (\bigcap_{m' \ge m_4} \mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m',p})$ contains the space $\mathcal{B}^s_{2,\infty}$.

Remark that we got sufficient conditions (see assumptions (4), (6) and (7)) on choices of $(t_{\epsilon,p})_\epsilon$ and $(v_{\epsilon,p})_\epsilon$ to ensure that all the BT-estimators in our families are good estimators, in the sense that they outperform hard and soft thresholding estimators. Indeed, in such cases, our BT-estimators reconstruct sets of functions at least as large as the Besov spaces with a rate of convergence faster than or in the same order as the ones associated with $v_{\epsilon,p} = \epsilon\sqrt{\log(\epsilon^{-1})}$. Consequently, these methods are better than any diagonal procedures (see Theorem 2 in Cai (2008)).

We now state the following corollaries 4.1 and 4.2 that are a direct consequences of Theorem 4.1 and Propositions 4.2 and 4.3.

Corollary 4.1. *Under assumptions (6) and (7), $\tilde{f}_\infty^{(t)}$ is the best estimator in the maxiset sense among the $HBT_\epsilon^{(t)}$ family when choosing $t_{\epsilon,p}$ such that, for any $0 < \epsilon < 1$, $t_{\epsilon,p}$ and $v_{\epsilon,p}$ only depend on ϵ.*

Corollary 4.2. *Under assumptions (6) and (7), $\tilde{f}_2^{(t)}$ is the best estimator in the maxiset sense among the $HBT_\epsilon^{(t)}$ family when choosing $t_{\epsilon,p}$ such that, for any $0 < \epsilon < 1$, $\left(\log(\epsilon^{-1})\right)^{\frac{1}{p}} t_{\epsilon,p}$ and $v_{\epsilon,p}$ only depend on ϵ.*

The two previous corollaries allow to identify the BT-estimators within the family with the largest maxiset. They clearly indicate that the best way to get large maxisets is to choose threshold values that are of lower order than the universal threshold. This very interesting fact has a powerful interpretation in terms of false discoveries for which we refer to the Section 5.

Remark 4.1. *In Section 2 we emphasize the importance of taking into account the clusters among the wavelet coefficient sequences across scales in order to fully capture the behavior of these estimators. Once done through the maxiset approach, this reveals that the Blockshrink estimator is able to reconstruct at the optimal minimax rate a set of functions which is larger than the Besov space $B^s_{2,\infty}$.*

5 Numerical experiments

We first introduce the notations of the nonparametric model we are dealing with:

$$Y_i = f\left(i/N\right) + \sigma\zeta_i, \; 1 \le i \le N, \; \zeta_i \text{ are i.i.d. } \mathcal{N}\left(0,1\right). \qquad (8)$$

We refer the reader to the classical literature (e.g., Tsybakov (2008)) for details about the equivalence between this nonparametric regression model and the sequence model given by equation (2). We only recall that the noise level ϵ is such that $\epsilon = \frac{\sigma}{\sqrt{N}}$.
This section proposes numerical experiments designed to check whether our theoretical conclusions can be observed in a practical setting. The previous theory does not model all the complexity encountered in practice with the choice of the wavelet functions, of the primary resolution scale, etc. Therefore, we choose a classical setting for numerical experiments, using Daubechies Extremal Phase wavelets with 8 vanishing moments. To illustrate our theoretical results we choose to consider two examples of $\text{HBT}_\epsilon^{(t)}$ that have already been discussed (see Examples 1 and 2 in Section 4) and respectively correspond to the choice of threshold value $t_{\epsilon,p} = \epsilon\left(\log(\epsilon^{-1})\right)^{\frac{1}{2}}$ and $t_{\epsilon,p} = \epsilon\left(\log(\epsilon^{-1})\right)^{\frac{1}{2}-\frac{1}{p}}$. Let us respectively denote by $\text{HBT}_\epsilon^{(t),1}$ and $\text{HBT}_\epsilon^{(t),2}$ these two families of BT-estimators. Following Cai (1997), we set the threshold $\hat{\lambda} = \hat{\sigma}(5N^{-1}\log(N))^{\frac{1}{2}}$ (resp. $\hat{\lambda}_p = \hat{\sigma}(5N^{-1}\log(N))^{\frac{1}{2}-\frac{1}{p}}$) for all BT-methods associated with $\text{HBT}_\epsilon^{(t),1}$ (resp. $\text{HBT}_\epsilon^{(t),2}$). We follow a standard approach to estimate σ by the Median Absolute Deviation (MAD) divided by 0.6745 over the wavelet coefficients at the finest wavelet scale $J-1$ (see e.g., Vidakovic (1999)).

We generate the data sets from a large panel of functions often used in wavelet estimation studies (Antoniadis et al. (2001)) with various Signal to Noise Ratios $SNR = \{5, 10, 15, 20\}$ and sample sizes $N = \{512, 1024, 2048\}$. We define the SNR as the logarithmic decibel scale of the ratio of the standard deviation of the function values to the standard deviation of the noise. We compute the Integrated Squared Error of the estimators $\tilde{f}_p^{(t)}$, $p \in \{2, 3, 5, 10, \infty\}$ at the ℓ-th Monte Carlo replication ($ISE^{(l)}\left(\tilde{f}_p^{(t)}\right),\ 1 \le l \le M$) as follows:

$$ISE^{(l)}\left(\tilde{f}_p^{(t)}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{f}_p^{(l)}\left(\frac{i}{N}\right) - f\left(\frac{i}{N}\right)\right)^2 .$$

We generate 2000 Monte Carlo replications and compute the Mean ISE as follows $MISE\left(\tilde{f}_p^{(t)}\right) = \frac{1}{M}\sum_{l=1}^{M} ISE^{(l)}\left(\tilde{f}_p^{(t)}\right)$. In practice, we compute the ISE over a subinterval of $[0,1]$ in order to avoid that potential boundary effects mask the comparison of our methods.

In fact, there are numerous connections between keep-or-kill estimation and hypothesis testing, see e.g. Abramovich et al. (2006). In order to emphasize that, we report in the Tables 1 and 2 the number of false positives/negatives (i.e., type I/II errors) obtained by comparing the set of indices of wavelet coefficients $H_\epsilon(t,p)$ (as defined in Definition 3.1) kept by each estimators, with the set of indices of 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 $\mathcal{S}^{\mathcal{O}} = \left\{(j,k)\,;\, j\in\mathbb{N}, j < j_{v_{\lambda_{\frac{\sigma}{\sqrt{N}}},p}}\,; 0 \le k < 2^j; |\theta_{jk}| > \frac{\sigma}{\sqrt{N}}\right\}$.
The results suggest similar behavior for different values of N and SNR. To keep clear the presentation of the results, we only report those for $N = 2048$ and $SNR = 10$ in Tables 1 and 2.

The Figures 5 and 6 summarize the MISE results. We observe the optimality of the estimator $\tilde{f}_\infty^{(t)} \in \mathrm{HBT}_\epsilon^{(t),1} \cap \mathrm{HBT}_\epsilon^{(t),2}$ (resp. $\tilde{f}_2^{(t)} \in \mathrm{HBT}_\epsilon^{(t),2}$) for all the tested functions as it was suggested by the Corollary 4.1 (resp. Corollary 4.2). In addition, there is a gradual improvement of the MISE performances when p increases (resp. decreases) reflecting the embeddings of the maxisets of the BT-estimators considered (see Section 4).

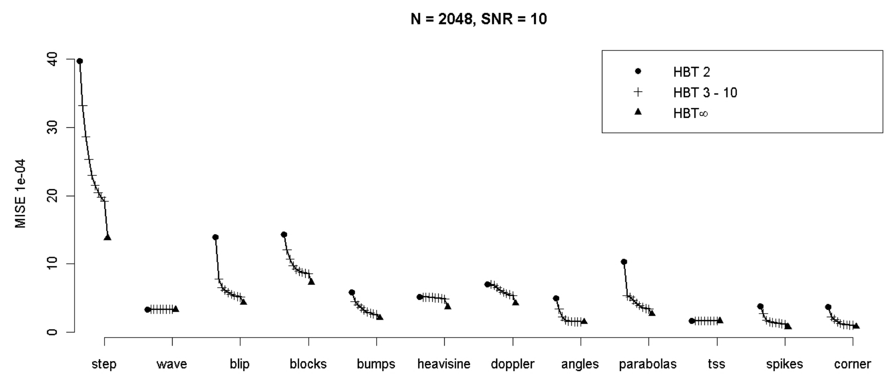

Figure 5: $MISE$ of the non overlapping BT-estimator in $\mathrm{HBT}_\epsilon^{(t),1}$ for different values of $2 \le p \le \infty$ for estimating various functions with a SNR equal to 10.

Looking at the number of false positives/negatives for the BT-estimators reported in the Tables 1 and 2, we can check that the best estimators in each family allow to reduce the percentage of false negatives with a comparatively small increase in the number of false positives yielding their good performances in terms of MISE. In the family $\mathrm{HBT}_\epsilon^{(t),1}$, the conservative universal threshold strongly controls the false positives but discards many small coefficients that would be useful for the reconstruction. Using the structure among the coefficients allows to reduce the number of false negatives. With such high threshold value, the method that appears to be the most powerful to reduce the false negatives is the estimator $\tilde{f}_\infty^{(t)}$.

Nevertheless, comparing Corollaries 4.1 and 4.2, our results point out that the $\mathrm{HBT}_\epsilon^{(t),1}$ family, based on a 'large' threshold, reaches a certain limit of detection of true discoveries that only a smaller order threshold would allow to overcome. That is confirmed when comparing the MISE of $\tilde{f}_\infty^{(t)}$ and $\tilde{f}_2^{(t)}$. The latter have lower MISE for all the tested functions with improvements ranging from 4 up to nearly 21 percent lower MISE. This emphasizes also that the use of a less conservative threshold gives rise to false positives, and thus we need a method that is able to treat simultaneously the false positives and the false negatives to control the overall risk (i.e., $p < \infty$). As we observe in the Figure 6, the best results are obtained for $p = 2$.

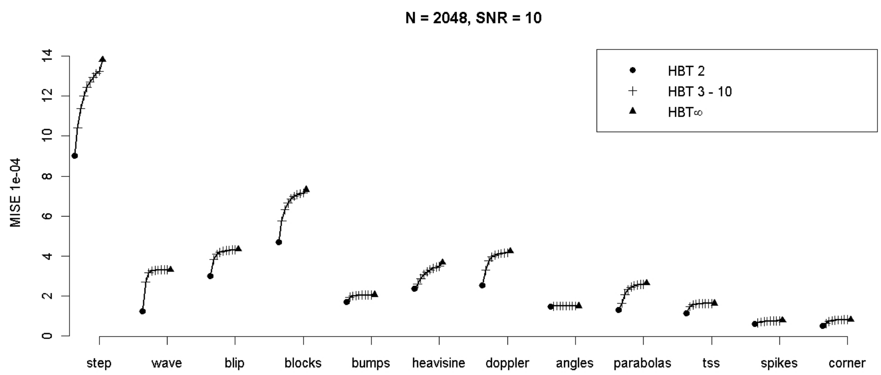

Figure 6: $MISE$ of the non overlapping BT-estimator in $\mathrm{HBT}_\epsilon^{(t),2}$ for different values of $2 \leq p \leq \infty$ for estimating various functions with a SNR equal to 10.

threshold	$\mathrm{HBT}_{\epsilon}^{(t),1}$				$\mathrm{HBT}_{\epsilon}^{(t),2}$				
method	$\hat{f}_2$	$\hat{f}_5$	$\hat{f}_{10}$	$\hat{f}_{\infty}$	$\hat{f}_2$	$\hat{f}_5$	$\hat{f}_{10}$	$\hat{f}_{\infty}$	$\hat{f}^{\mathcal{O}}$
	Function: Step								
MISE	39.8	25.3	19.2	13.8	9.0	12.0	13.2	13.8	3.4
False +	1.0	2.1	5.1	10.4	17.6	11.7	10.7	10.4	0.0
False −	50.8	42.9	37.9	31.6	24.2	29.2	30.9	31.6	0.0
size	19.2	28.2	36.2	47.7	62.4	51.6	48.8	47.7	69.0
	Function: Wave								
MISE	3.3	3.3	3.3	3.3	1.2	3.3	3.3	3.3	0.7
False +	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
False −	36.0	36.0	36.0	35.6	9.7	34.9	35.6	35.6	0.0
Size	24.0	24.0	24.0	24.4	50.3	25.1	24.4	24.4	59.0
	Function: Blip								
MISE	14.0	6.2	5.1	4.3	3.0	4.2	4.3	4.3	0.9
False +	2.1	6.9	8.9	12.8	22.3	13.7	13.0	12.8	0.0
False −	26.9	19.1	17.2	14.6	9.7	14.2	14.5	14.6	0.0
Size	16.3	28.8	32.6	39.2	53.6	40.5	39.5	39.2	41.0
	Function: Blocks								
MISE	14.3	9.8	8.5	7.3	4.7	6.6	7.2	7.3	1.8
False +	6.0	8.4	10.5	15.1	29.1	17.4	15.6	15.1	0.0
False −	152.3	137.6	130.4	119.8	92.2	114.6	118.6	119.8 0.0	
Size	47.7	64.7	74.1	89.3	130.9	96.8	91.0	89.3	194.0
	Function: Bumps								
MISE	5.9	3.6	2.6	2.1	1.7	2.0	2.1	2.1	0.7
False +	16.2	21.1	33.0	40.7	54.3	41.3	40.8	40.7	0.0
False −	110.5	92.3	81.4	74.5	62.9	73.9	74.4	74.5	0.0
Size	115.7	138.8	161.6	176.2	201.4	177.4	176.4	176.2	210.0
	Function: Heavisine								
MISE	5.1	5.1	4.9	3.7	2.4	3.1	3.5	3.7	1
False +	0.0	0.0	0.1	0.7	1.9	0.8	0.7	0.7	0
False −	26.0	26.0	25.3	22.3	18.8	20.9	21.9	22.3	0
Size	8.0	8.0	8.8	12.4	17.2	14.0	12.9	12.4	34

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

threshold	$HBT_\epsilon^{(t),1}$				$HBT_\epsilon^{(t),2}$				
method	$\hat{f}_2$	$\hat{f}_5$	$\hat{f}_{10}$	$\hat{f}_\infty$	$\hat{f}_2$	$\hat{f}_5$	$\hat{f}_{10}$	$\hat{f}_\infty$	$\hat{f}^{\mathcal{O}}$
	Function: Doppler								
MISE	7.0	6.5	5.4	4.2	2.5	4.0	4.2	4.2	1.2
False +	4	4.2	5.3	8.3	11.8	8.9	8.4	8.3	0.0
False −	37	35.5	32.0	27.9	17.5	26.7	27.6	27.9	0.0
Size	42	43.7	48.2	55.4	69.3	57.2	55.8	55.4	75.0
	Function: Angles								
MISE	5.0	1.8	1.6	1.5	1.5	1.5	1.5	1.5	0.8
False +	0.7	2.8	3.0	3.3	3.6	3.3	3.3	3.3	0.0
False −	14.8	10.4	10.0	9.9	9.8	9.9	9.9	9.9	0.0
Size	20.9	27.4	28.0	28.4	28.8	28.4	28.4	28.4	35.0
	Function: Parabolas								
MISE	10.3	4.7	3.4	2.7	1.3	2.3	2.6	2.7	0.8
False +	0.9	3.5	4.7	5.5	7.0	5.9	5.6	5.5	0.0
False −	12.8	9.5	8.2	6.8	3.0	5.9	6.6	6.8	0.0
Size	14.2	20.1	22.5	24.7	30.0	26.0	25.0	24.7	26.0
	Function: Time Shift sine								
MISE	1.7	1.7	1.7	1.6	1.1	1.6	1.6	1.6	0.6
False +	2.0	2.0	2.0	2.2	4.9	2.4	2.2	2.2	0.0
False −	5.0	5.0	5.0	4.9	2.8	4.7	4.8	4.9	0.0
Size	23.0	23.0	23.0	23.3	28.1	23.7	23.4	23.3	26.0
	Function: Spikes								
MISE	3.8	1.5	1.2	0.8	0.6	0.7	0.8	0.8	0.3
False +	3.0	14.0	16.3	20.4	23.7	20.9	20.5	20.4	0.0
False −	25.0	16.0	13.3	9.9	8.0	9.4	9.8	9.9	0.0
Size	42.1	62.1	67.0	74.5	79.7	75.5	74.8	74.5	64.0
	Function: Corner								
MISE	3.8	1.5	1.0	0.8	0.5	0.8	0.8	0.8	0.3
False +	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
False −	17.4	12.7	11.1	10.0	7.2	9.5	9.9	10.0	0.0
Size	9.6	14.3	15.9	17.1	19.9	17.5	17.1	17.1	27.0

Table 2: MISE (10^{-4}), average number of false positives/negatives and average number of non zero empirical wavelet coefficients in the estimator.

6 Summary of results and conclusion

In this paper we introduced the family of Horizontal Block Thresholding estimators. We studied the performances of the estimators of this family under the L_2-risk using the maxiset approach. We remark the good maxiset performances for a wide range of rates and threshold values, and we identified the best procedure in some cases, that is the one using the l_2-norm and a threshold value of order ϵ. This paper shows the importance of adapting the threshold to the method in order to enlarge the maxiset.

For a given (fixed) threshold value, there is the following interpretation of this family according to p: the situation of $p = \infty$ corresponds to a methodology that really focuses on the reduction of false negatives, all the coefficients in a block where one coefficient passes over the threshold are kept. This method, however, has to accommodate with really high threshold values in order that also the false positives are well controlled. Methods with lower p are meant to control simultaneously false positives and negatives. For those, one has to reduce the value of the threshold if one aims to really improve the estimation. But as a trade-off lowering the threshold calls for correcting on the side of the chosen method in order to not lose control of the false positives and to expect a higher potential for the method. Developing these fine tunings for the link between chosen threshold method and associated threshold value, is precisely the content of this paper using the maxiset approach.

Our numerical experiments for fixed threshold (as $\mathrm{HBT}_\epsilon^{(t),1}$) confirm the best procedure for $p = \infty$. On the contrary, for methods with a threshold value that depends on p (as $\mathrm{HBT}_\epsilon^{(t),2}$), the best procedure is the one associated with $p = 2$.

For future research, it would be interesting to model the improvements related to the James Stein correction used in Cai (2008). Another important point is about overlapping horizontal block thresholding theory. Indeed, the behavior of overlapping and non overlapping block thresholding procedures according to p and to the threshold value is similar - as it is shown in the Figures 7 and 8. Comparing non overlapping and overlapping $\mathrm{HBT}_\epsilon^{(t),2}$ we discover that the latter reduces the MISE for almost all the tested functions up to nearly 38 percent for the 'Doppler' function. Nevertheless, while the maxisets of the overlapping $\mathrm{HBT}_\epsilon^{(t),1}$ methods can be easily computed using a similar approach to ours in that paper, it appears to be non trivial for overlapping $\mathrm{HBT}_\epsilon^{(t),2}$.

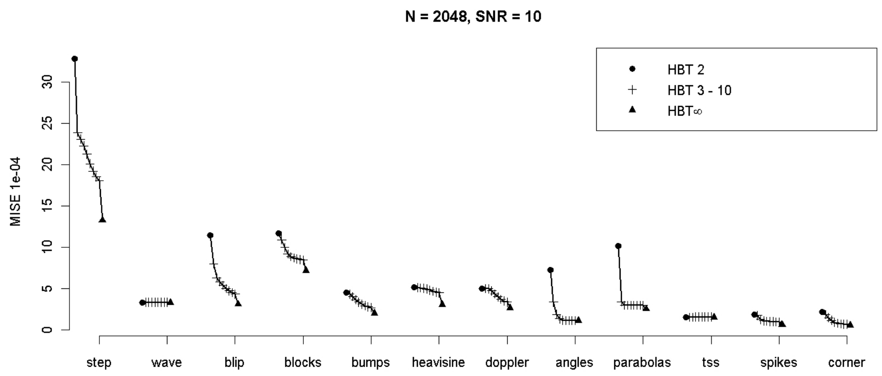

Figure 7: *MISE* of the overlapping $\mathrm{HBT}_\epsilon^{(t),1}$ estimators for different values of $2 \le p \le \infty$ for estimating various functions with a SNR equal to 10.

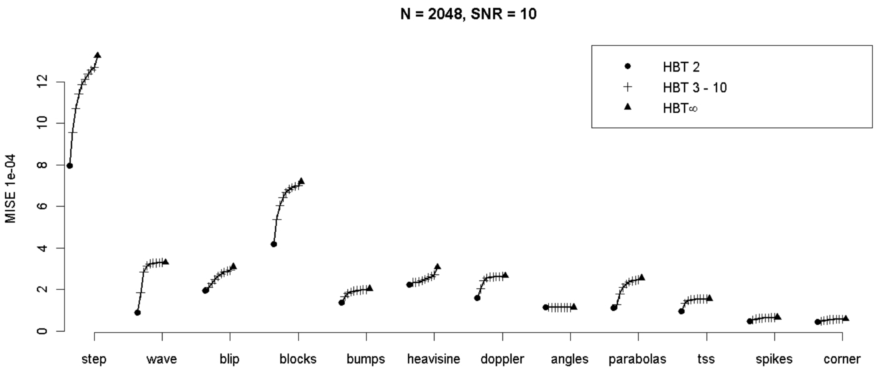

Figure 8: $MISE$ of the overlapping $\mathrm{HBT}_\epsilon^{(t),2}$ estimator for different values of $2 \le p \le \infty$ for estimating various functions with a SNR equal to 10.

7 Proofs of results

In this Section, C denotes a constant that does not depend on m and ϵ and that may be different from one line to an other line. Proofs of many results use the following lemma recalled from Cai and Silverman (2001).

Lemma 7.1. *Let $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ and $0 < p \le q \le \infty$. Then the following inequalities hold:*

$$\left(\sum_{i=1}^{d} |x_i|^q\right)^{\frac{1}{q}} \le \left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}} \le d^{\frac{1}{p}-\frac{1}{q}} \left(\sum_{i=1}^{d} |x_i|^q\right)^{\frac{1}{q}}.$$

7.1 Proof of Proposition 4.1

Proof. Fix $2 \le p \le \infty$ and let $f \in \mathcal{B}^s_{2,\infty}$. There exists $C > 0$ such that, for any $j \in \mathbb{N}$, the wavelet coefficients of f satisfy:

$$\sum_{k=0}^{2^j-1} \theta_{jk}^2 \le C\ 2^{-2js}.$$

Fix $m' \ge m_4$. For any $0 < \lambda < 1$, let $j^{(s)}_{v_{\lambda,p}}$ be the integer such that $2^{-j^{(s)}_{v_{\lambda,p}}} \le (m' v_{\lambda,p})^{\frac{2}{1+2s}} < 2^{1-j^{(s)}_{v_{\lambda,p}}}$.

$$\sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{ \|\theta\ /\ B_{jk}(\lambda)\|_p \le m' t_{\lambda,p} \right\} = A_0 + A_1$$

with, using Lemma 7.1

$$\begin{aligned}
A_0 &= \sum_{j<j^{(s)}_{v_{\lambda,p}}} \sum_{k=0}^{2^j-1} \theta^2_{jk} \mathbf{1}\Big\{ \|\theta \ / \ B_{jk}(\lambda)\|_p \leq m' t_{\lambda,p} \Big\} \\
&\leq \sum_{j<j^{(s)}_{v_{\lambda,p}}} \sum_{k=0}^{2^j-1} \theta^2_{jk} \mathbf{1}\Big\{ \|\theta \ / \ B_{jk}(\lambda)\|_2 \leq m' t_{\lambda,p} \Big\} \\
&\leq \lfloor \log(\lambda^{-1}) \rfloor (m' t_{\lambda,p})^2 \sum_{j<j^{(s)}_{v_{\lambda,p}}} \sum_{u=1}^{u_{j,\lambda}} \mathbf{1}\Big\{ \|\theta \ / \ B^{(u)}_j(\lambda)\|_2 \leq m' t_{\lambda,p} \Big\} \\
&\leq 2^{j^{(s)}_{v_{\lambda,p}}} (m' t_{\lambda,p})^2 \\
&\leq C \ (m' v_{_{\lambda,p}})^{\frac{4s}{1+2s}}.
\end{aligned}$$

The last inequality is obtained using $\sup\limits_{0<\lambda<1} t_{\lambda,p} \ {v_{\lambda,p}}^{-1} < \infty$. Since $f \in \mathcal{B}^s_{2,\infty}$,

$$\begin{aligned}
A_1 &= \sum_{j \geq j^{(s)}_{v_{\lambda,p}}} \sum_{k=0}^{2^j-1} \theta^2_{jk} \mathbf{1}\Big\{ \|\theta \ / \ B_{jk}(\lambda)\|_p \leq m' t_{\lambda,p} \Big\} \\
&\leq \sum_{j \geq j^{(s)}_{v_{\lambda,p}}} \sum_{k=0}^{2^j-1} \theta^2_{jk} \\
&\leq C \ 2^{-2sj^{(s)}_{v_{\lambda,p}}} \\
&\leq C \ (m' v_{_{\lambda,p}})^{\frac{4s}{1+2s}}.
\end{aligned}$$

Hence, for any $m' \geq m_4$,

$$\sup_{0<\lambda<1} (m' v_{_{\lambda,p}})^{-\frac{4s}{1+2s}} \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta^2_{jk} \mathbf{1}\Big\{ \|\theta \ / \ B_{jk}(\lambda)\|_p \leq m' t_{\lambda,p} \Big\} < \infty,$$

that is to say, $f \in \bigcap_{m' \geq m_4} \mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m',p}$.

□

7.2 Proof of Proposition 4.2 and of Proposition 4.3

For all $m' > 0$, $0 < r < 2$ and all $2 \leq p < q \leq \infty$, the definitions of $\mathcal{W}^{(t,v)}_{r,m',p}$ and $\mathcal{W}^{(t,v)}_{r,m',q}$ only differ by the indicator functions they are associated with. These spaces defined in Definition 4.2 control the energy of the coefficients that are not used in the reconstruction. That is exactly the complementary set to $H_\epsilon(t, m', p)$. Therefore, the embeddings given in Propositions 4.2 and 4.3, holding for a wide range of threshold values $(t_{\epsilon,p})_\epsilon$, can be deduced from the embedding properties of the sets of indices $H_\epsilon(t, m', p)$ stated in the following lemmas and proved further down.

Lemma 7.2. *Assume that, for any $2 \leq p \leq \infty$, $t_{\epsilon,p}$ only depends on ϵ. Then, for any $m > 0$, any $2 \leq p_1 \leq p_2 \leq \infty$ and any $0 < \epsilon < 1$*

- $H_\epsilon(t, m, p_1) \subseteq H_\epsilon(t, m, p_2)$,
- $H_\epsilon(t, m, \infty)$ *is the set of indices which contains all the empirical wavelet coefficients larger than the threshold value and their block-neighborhood, i.e. the empirical wavelet coefficients belonging to the same block.*

Lemma 7.3. *Assume that, for any* $2 \le p \le \infty$, $\big(\log(\epsilon^{-1})\big)^{\frac{1}{p}} t_{\epsilon,p}$ *only depends on* ϵ. *Then, for any* $m > 0$, *any* $2 \le p_1 \le p_2 \le \infty$ *and any* $0 < \epsilon < 1$

- $H_\epsilon(t, m, p_2) \subseteq H_\epsilon(t, m, p_1)$.

The two previous lemmas show that the embedding properties depend on the choice of the threshold $t_{\epsilon,p}$. More precisely, for any $0 < \epsilon < 1$, if $t_{\epsilon,p}$ only depends on ϵ then the larger p the bigger the set of indices $H_\epsilon(t, m, p)$ $(m > 0)$. If the quantity $(\log(\epsilon^{-1}))^{\frac{1}{p}} t_{\epsilon,p}$ only depends on ϵ then the larger p the smaller the set of indices $H_\epsilon(t, m, p)$.

Proof. Let $m > 0$ and $2 \le p_1 \le p_2 \le \infty$. Proofs of Lemmas 7.2 and 7.3 are a direct consequence of Lemma 7.1. Indeed, if for any $0 < \epsilon < 1$, $t_{\epsilon,p}$ only depends on ϵ then, for any (j, k):

$$\|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_{p_1} > m \ t_{\epsilon,p_1} \Longrightarrow \|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_{p_2} > m \ t_{\epsilon,p_2}.$$

Hence $H_\epsilon(t, m, p_1) \subseteq H_\epsilon(t, m, p_2)$.
In the same way, if for any $0 < \epsilon < 1$, $\big(\log(\epsilon^{-1})\big)^{\frac{1}{p}} t_{\epsilon,p}$ only depends on ϵ then, noting that the size of blocks is of order $\log(\epsilon^{-1})$, for any (j, k):

$$\|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_{p_2} > m \ t_{\epsilon,p_2} \Longrightarrow \|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_{p_1} > m \ t_{\epsilon,p_1}.$$

Hence $H_\epsilon(t, m, p_2) \subseteq H_\epsilon(t, m, p_1)$.

□

Remark 7.1. *For any* $2 \le p \le \infty$ *and any* $0 < \epsilon < 1$, *notice that the assumption of Lemma 7.2 is satisfied with the choice* $t_{\epsilon,p} = \epsilon$ *whereas the assumption of Lemma 7.3 is satisfied with the choice* $t_{\epsilon,p} = \epsilon(\log(\epsilon^{-1}))^{\frac{1}{2}-\frac{1}{p}}$.

7.3 Proof of Theorem 4.1

Proof. Notice that the result can be proven by replacing the supremum over ϵ in $[0, 1[$ by the supremum over ϵ in $]0, \epsilon_v[$, where ϵ_v is such that for any $0 < \epsilon \le \epsilon_v$ and such that $0 < mv_{\epsilon,p} < 1$.

$\Longrightarrow$

Suppose that, for any $m \ge 2m_4$ and any $0 < \epsilon < \epsilon_v$, $\mathbb{E}\|\tilde{f}_p^{(t)} - f\|_2^2 \le C \ (mv_{\epsilon,p})^{\frac{4s}{1+2s}}$. Then,

$$\begin{aligned}\sum_{j \ge j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 &\le \mathbb{E}\|\tilde{f}_p^{(t)} - f\|_2^2 \\ &\le C \ (mv_{\epsilon,p})^{\frac{4s}{1+2s}} \\ &\le C \, 2^{-\frac{s}{1+2s} \ 2j_{v_{\epsilon,p}}}.\end{aligned}$$

So, using the continuity of $v_{\epsilon,p}$ with respect to ϵ, $f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}}$. Let $m' \geq m_4$ and put $m := 2m'$,

$$
\begin{aligned}
(m'v_{\epsilon,p})^{\frac{-4s}{1+2s}} \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{\|\theta \ / \ B_{jk}(\epsilon)\|_p \leq m't_{\epsilon,p}\Big\} &= \Big(\frac{mv_{\epsilon,p}}{2}\Big)^{\frac{-4s}{1+2s}} \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{\|\theta \ / \ B_{jk}(\epsilon)\|_p \leq \frac{m}{2}t_{\epsilon,p}\Big\} \\
&\leq A_2 + A_3 + A_4,
\end{aligned}
$$

with

$$
\begin{aligned}
A_2 &= \Big(\frac{mv_{\epsilon,p}}{2}\Big)^{\frac{-4s}{1+2s}} \ \mathbb{E}\left[\sum_{j<j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{\|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_p \leq mt_{\epsilon,p}\Big\}\right] \\
&\leq \Big(\frac{mv_{\epsilon,p}}{2}\Big)^{\frac{-4s}{1+2s}} \ \mathbb{E}\|\tilde{f}_p^{(t)} - f\|_2^2 \\
&\leq C,
\end{aligned}
$$

$$
\begin{aligned}
A_3 &= \Big(\frac{mv_{\epsilon,p}}{2}\Big)^{\frac{-4s}{1+2s}} \ \mathbb{E}\left[\sum_{j<j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{\|\hat{\theta} - \theta \ / \ B_{jk}(\epsilon)\|_p > \frac{m}{2}t_{\epsilon,p}\Big\}\right] \\
&\leq C \ \Big(\frac{mv_{\epsilon,p}}{2}\Big)^{\frac{-4s}{1+2s}} \ \mathbb{P}\Big(\|Z(\epsilon)\|_p > \frac{m}{2} \ t_{\epsilon,p}\Big) \\
&\leq C.
\end{aligned}
$$

The last inequality uses assumption (6) - with $c = 2$ - and the fact that $m \geq 2m_4 \geq 2m_2$.

Now

$$
\begin{aligned}
A_4 &= \Big(\frac{mv_{\epsilon,p}}{2}\Big)^{\frac{-4s}{1+2s}} \sum_{j\geq j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\Big\{\|\theta \ / \ B_{jk}(\epsilon)\|_p \leq \frac{m}{2}t_{\epsilon,p}\Big\} \\
&\leq \Big(\frac{mv_{\epsilon,p}}{2}\Big)^{\frac{-4s}{1+2s}} \sum_{j\geq j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \\
&\leq C \ \Big(\frac{mv_{\epsilon,p}}{2}\Big)^{\frac{-4s}{1+2s}} \ 2^{-\frac{s}{1+2s} \ 2j_{v_{\epsilon,p}}} \\
&\leq C.
\end{aligned}
$$

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_2, A_3 and A_4 and when using the continuity on $v_{\epsilon,p}$ with respect to ϵ, one deduces that $f \in \mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m',p}$. Following the arbitrary choice of $m' \geq m_4$, one gets $f \in \cap_{m'\geq m_4}\mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m',p}$.

$\Longleftarrow$

Suppose that $f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap \left(\bigcap_{m' \geq m_4} \mathcal{W}_{\frac{2}{1+2s},m',p}^{(t,v)} \right)$. For any any $m \geq 2m_4$ and any $0 < \epsilon < \epsilon_v$, the quadratic risk of the estimator $\tilde{f}_p^{(t)}$ can be decomposed as follows:

$$\begin{aligned}
\mathbb{E}\|\tilde{f}_p^{(t)} - f\|_2^2 &= \mathbb{E}\left[\sum_{j < j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{ \|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_p \leq mt_{\epsilon,p} \right\} \right] \\
&\quad + \sum_{j < j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\hat{\theta}_{jk} - \theta_{jk})^2 \mathbf{1}\left\{ \|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_p > mt_{\epsilon,p} \right\} \right] \\
&\quad + \sum_{j \geq j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 + \epsilon^2 \\
&= A_5 + A_6 + A_7.
\end{aligned}$$

Since $f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap \mathcal{W}_{\frac{2}{1+2s},2m,p}^{(t,v)}$ and due to assumption (6)

$$\begin{aligned}
A_5 &= \mathbb{E}\left[\sum_{j < j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{ \|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_p \leq mt_{\epsilon,p} \right\} \right] \\
&\leq \sum_{j < j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{ \|\theta \ / \ B_{jk}(\epsilon)\|_p \leq 2mt_{\epsilon,p} \right\} \\
&\quad + C \ \mathbb{P}\left(\|Z(\epsilon)\|_p > mt_{\epsilon,p} \right) \\
&\leq C \left(mv_{\epsilon,p} \right)^{\frac{4s}{1+2s}}.
\end{aligned}$$

The last inequality arises from $m \geq m_2$.

Using the Cauchy-Schwarz inequality and assumption (6)

$$\begin{aligned}
A_6 &= \sum_{j < j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\hat{\theta}_{jk} - \theta_{jk})^2 \mathbf{1}\left\{ \|\hat{\theta} \ / \ B_{jk}(\epsilon)\|_p > mt_{\epsilon,p} \right\} \right] \\
&\leq \sum_{j < j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\hat{\theta}_{jk} - \theta_{jk})^2 \mathbf{1}\left\{ \|\theta \ / \ B_{jk}(\epsilon)\|_p > \frac{m}{2} t_{\epsilon,p} \right\} \right] \\
&\quad + C \, \epsilon^2 \ 2^{j_{v_{\epsilon,p}}} \mathbb{P}^{\frac{1}{2}} \left(\|Z(\epsilon)\|_p > \frac{m}{2} \ t_{\epsilon,p} \right) \\
&\leq C \left(\left(mv_{\epsilon,p} \right)^{\frac{4s}{1+2s}} + \mathbb{P}^{\frac{1}{2}} \left(\|Z(\epsilon)\|_p > \frac{m}{2} \ t_{\epsilon,p} \right) \right) \\
&\leq C \left(mv_{\epsilon,p} \right)^{\frac{4s}{1+2s}},
\end{aligned}$$

since $m \geq 2m_4$.

Since $f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}}$ and $v_{\epsilon,p}$ satisfies (7)

$$\begin{aligned} A_7 &= \epsilon^2 + \sum_{j \geq j_{v_{\epsilon,p}}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \\ &\leq \epsilon^2 + C\, 2^{-\frac{2s}{1+2s} j_{v_{\epsilon,p}}} \\ &\leq C\, (m v_{\epsilon,p})^{\frac{4s}{1+2s}}. \end{aligned}$$

When combining the bounds of A_5, A_6 and A_7 one deduces that

$$\sup_{0<\epsilon<1} (m v_{\epsilon,p})^{-\frac{4s}{1+2s}} \mathbb{E} \| \tilde{f}_p^{(t)} - f \|_2^2 < \infty.$$

This ends the proof. □

References

F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. *The Annals of Statistics*, 34(2):584–653, 2006.

A. Antoniadis, J. Bigot, and J. Sapatinas. Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, 6(6):1–83, 2001.

F. Autin. *Maxiset Point of View in Nonparametric Estimation.* PhD thesis, Ph.D thesis at university of Paris 7 - France, 2004.

F. Autin. Maxisets for μ thresholding rules. *Test*, 17:332–349, 2008.

F. Autin, J-M. Freyermuth, and R. von Sachs. Ideal denoising within a family of tree-structured wavelet estimators. *submitted*, 2011.

T. Cai. On adaptivity of blockshrink wavelet estimator over besov spaces. Technical report, Technical report 97-05, Purdue University, 1997.

T. Cai. Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Annals of Statistics*, 27(3):898–924, 1999.

T. Cai. On information pooling, adaptability and superefficiency in nonparametric function estimation. *Journal of Multivariate Analysis*, 99:421–436, 2008.

T. Cai and B. Silverman. Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya*, 63:127–148, 2001.

T. Cai and H. Zhou. A data-driven block thresholding approach to wavelet estimation. *Annals of Statistics*, 37:569–595, 2009.

A. Cohen, R. De Vore, G. Kerkyacharian, and D. Picard. Maximal spaces with given rate of convergence for thresholding algorithms. *Applied and Computational Harmonic Analysis*, 11, 2001.

I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia. 1992.

D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 1994.

P. Hall, G. Kerkycharian, and Picard D. Block threshold rules for curve estimation using kernel and wavelet methods. *Annals of Statistics*, 26(3):922–942, 1998.

P. Hall, G. Kerkycharian, and Picard D. On the minimax optimality for block thresholded wavelet estimators. *Statistica Sinica*, 9, 1999.

W. Hardle, G. Kerkycharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129 of *Lectures Notes in Statistics*. Springer-Verlag, Hamburg, 1998.

A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics, 214 pp. 2008.

B. Vidakovic. *Statistical Modelling by Wavelets*. John Wiley & Sons, Inc., New York, 384 pp. 1999.