Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 41: 240–258, 2014
doi: 10.1111/sjos.12012

Block-threshold-adapted Estimators via a Maxiset Approach

FLORENT AUTIN

Université d'Aix-Marseille, CNRS UMR 6632

JEAN-MARC FREYERMUTH

Katholieke Universiteit Leuven

RAINER VON SACHS

Université Catholique de Louvain-La-Neuve

ABSTRACT. We study the maxiset performance of a large collection of block thresholding wavelet estimators, namely the *horizontal block thresholding family*. We provide sufficient conditions on the choices of rates and threshold values to ensure that the involved adaptive estimators obtain large maxisets. Moreover, we prove that any estimator of such a family reconstructs the Besov balls with a near-minimax optimal rate that can be faster than the one of any separable thresholding estimator. Then, we identify, in particular cases, 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 approach that models method-dependent threshold values. By a series of simulation studies, we confirm the good performance of the best estimator by comparing it with the other members of its family.

Key words: Besov spaces, curve estimation, minimax and maxiset approaches, rate of convergence, thresholding methods, wavelet-based estimation

1. Introduction

Non-parametric estimation of functions by non-linear wavelet methods has proven to be a real success story, in particular for functions showing locally varying regularity. Wavelets provide sparse representation of functions, that is, they localize the information of a function in a few large coefficients for a wide range of function classes. This property is the key to understanding the good performances of hard and soft thresholding estimators, which use only the empirical wavelet coefficients that are larger than a given threshold value, often chosen to be the universal threshold (UT) value (see among others Donoho & Johnstone, 1994). In this case, these estimators are near optimal over Besov balls, that is, they attain, up to a logarithmic factor in sample size, the optimal rate of convergence for a given risk (often the L_2-risk) for a relatively large class of functions of highly inhomogeneous spatial regularity. More than that, they are also adaptive for the regularity—or smoothness—parameter of these function classes, meaning that, even without its knowledge, they can reproduce this near-optimal rate of convergence.

The choice of the UT value has thus become very popular. Its second, more practical, motivation is to deliver asymptotic noise-free reconstructions. Being first of all proportional to the noise level of the data, it is also determined using the tail behaviour of the distribution of the maximum of standard Gaussian random variables. It ensures that asymptotically all the observed wavelet coefficients that are purely noise are removed. The UT value is known to be a large threshold value being often too conservative in practical applications and causing too many false negatives (i.e. suppressing too many true signal coefficients).

Hard and soft thresholding methods, in as much as they are 'separable' (or diagonal) rules that decide to keep a coefficient merely because of its individual magnitude, have been criticized over the last decade: their *minimax rate*, that is, the fastest rate of convergence of a given risk taken uniformly over all elements in the considered function class, suffers from a suboptimal log-term, as shown in Cai (2008). Related to this, Autin (2004) emphasized that such thresholding methods are too *elitist*: they have a tendency to suppress small—but important—empirical wavelet coefficients for reconstructing the function of interest. To remedy both the theoretical and practical shortcomings of elitist procedures, it has been shown in recent literature (see, among others, Cai, 1999; Hall *et al.*, 1998, 1999; Autin, 2004, 2008; Autin *et al.*, 2011) that one can use information from neighbouring empirical wavelet coefficients.

Cai (1997, 2002) proved that wavelet estimators based on *thresholding of empirical wavelet coefficients by blocks* (called 'block thresholding' or BT methods hereafter) can be *minimax optimal* over Besov balls, that is, they attain the L_2-minimax rate without the suboptimal log-term. There are many popular examples such as the so-called BlockShrink estimator of Cai (1997), the block James-Stein estimator of Cai (1999), or the NeighBlock/NeighCoef of Cai & Silverman (2001). The BlockShrink estimator, which is of particular interest in this paper, reconstructs functions using a non-separable thresholding rule. More precisely, it keeps an empirical wavelet coefficient if the block it belongs to (i.e. a well-defined set of empirical wavelet coefficients in a neighbourhood of its location on the same scale) has an l_2-mean norm larger than a given threshold value. Here, in contrast to using the UT, the threshold value is chosen to be purely proportional to the noise level, without the additional protection to values in the tail of standard Gaussian random variables.

To study the performance of the BlockShrink estimator, Autin (2008) and Chesneau (2008) adopted a different perspective, namely that of the *maxiset* approach. This approach aims at providing the largest set of functions that are 'well' estimated by a given estimator. Here, 'well' refers to a given minimal rate of convergence, usually chosen to be equal or close to the optimal minimax rate of convergence to allow for pertinent comparison of both the minimax and the maxiset approaches. Autin (2008) proved that the sets of functions well estimated by wavelet estimators using BT methods can be larger than the ones of (separable) hard and soft thresholding estimators. For instance, the *Maximum-Block* estimator was proven to perform particularly well (Autin, 2008). This estimator uses blocks of empirical wavelet coefficients for which the l_∞-mean norm, that is, the maximum element in the block of coefficients, is larger than a threshold value that is of the order of the UT value.

The BlockShrink and Maximum-Block estimators provide good visual reconstructions of function as shown in Figures 3 and 4, which started from the observation of the well-known function *Bumps* (Figure 1) in a noisy version (Figure 2). This can be explained by the group structure of the large true wavelet coefficients represented in Figure 1, lower panel (the darker the grey scale, the larger is 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.

Figures 1–4 show the true function, noisy function and two BT estimates associated with their 2^j wavelet coefficients at level j from 0 to 10. The darker the grey scale, the larger is the magnitude of the coefficients.

In this paper, we study more specifically a series of BT estimators relying on non-overlapping blocks, hereafter the horizontal block thresholding family. Instead of considering, as in Cai (1997) and Autin (2008), respectively, only ℓ_2-mean and ℓ_∞-mean norms as, what in the sequel we call, block 'scores' for the construction of the blocks, here we consider the whole

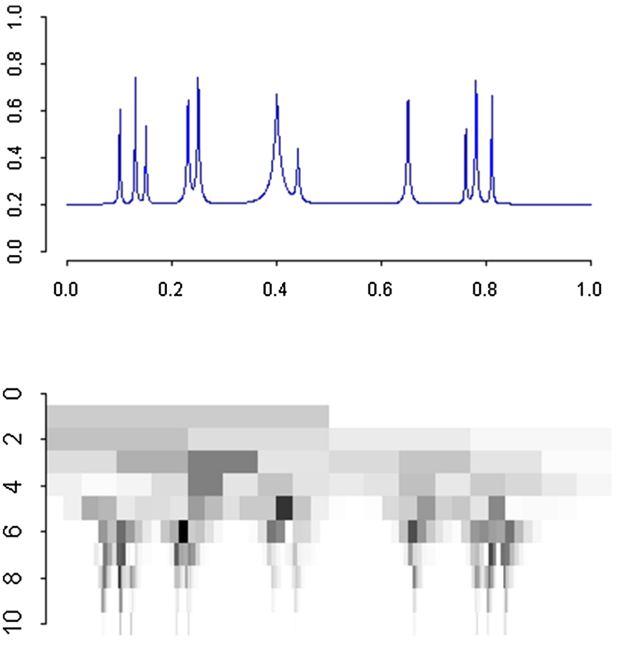

Fig. 1. Function *Bumps*: true version.

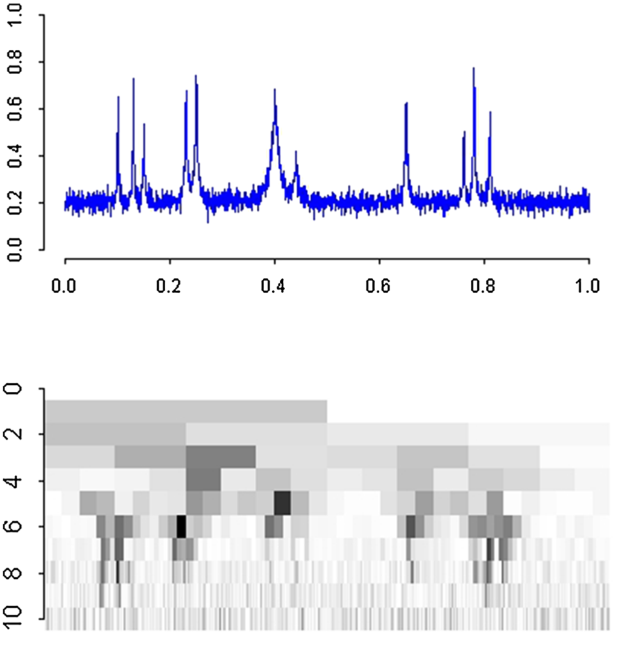

Fig. 2. Function *Bumps*: noisy version.

range of $2 \leq p \leq +\infty$ associated with threshold values that are functions not only of the noise level but possibly also of p. We compare those by studying both their theoretical performance via the maxiset approach and their numerical performance.

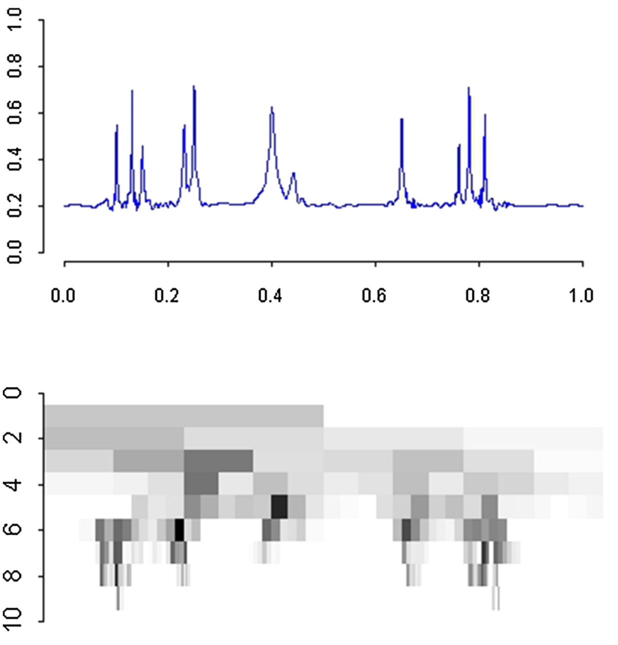

Fig. 3. BlockShrink estimator.

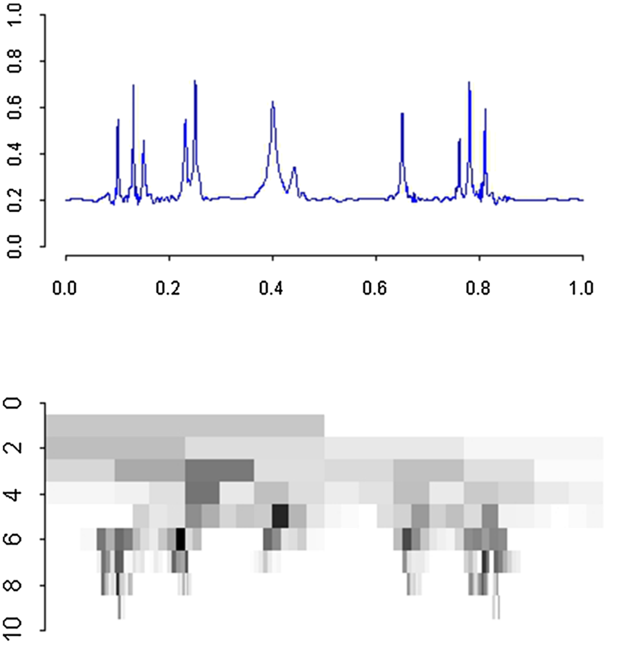

Fig. 4. Maximum-Block estimator.

The paper is organized as follows. After recalling in Section 2 the necessary essentials on abstract wavelet estimation in function spaces, we introduce in Section 3 our general horizontal block thresholding family. Our definition addresses a latent problem related to handling blocks at boundaries. In Section 4, we compute the set of all the functions well estimated by estimators belonging to that family. Precisely, we identify all the functions for which the quadratic risk of these estimators does not exceed a given rate of convergence (theorem 1). We provide sufficient conditions on the choices of the rate of convergence and the threshold values to ensure that the maxisets of the estimators of the family are not ‘degenerated’, that is, that they contain sufficiently interesting subsets of functions to guarantee that each estimator of the horizontal block thresholding family 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 best-performing estimator within the family according to the maxiset approach. Moreover, we point out that the best way to give a score to blocks indeed depends on the threshold value under consideration (corollaries 1 and 2). This result is an important contribution of this paper: it shows the ability of the maxiset approach to describe the behaviour of estimators with regard to the values of two parameters with interdependent effects (the score and the threshold value). This can be nicely interpreted in terms of a hypothesis testing language, through the control of the number of false positives (erroneously kept coefficients) and the number of false negatives (erroneously deleted coefficients). Moreover, it allows us to search for the best estimation procedure in the studied family. Finally, Section 5 proposes to check whether our theoretical results agreed with the practical behaviour of the studied estimators, and Section 6 gives brief conclusive remarks. All the proofs are given in the Supporting Information, which can be found in the online version of this paper.

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 $\left\{\phi, \psi_{jk}, j \in \mathbb{N}, k \in \{0,\dots,2^j-1\}\right\}$. Examples of such bases are given by 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 and the wavelet coefficients of f. They correspond to the L_2-scalar products between f and the scaling function ϕ and between f and the wavelet functions ψ_{jk}.

In the prominent denoising context of non-parametric regression, which we adopt in this work, we model our noisy data $Y_i, 1 \le i \le N$, to be observations of the equidistantly sampled true underlying signal f corrupted by the additive zero-mean Gaussian noise of variance σ^2,

$$Y_i = f\left(\frac{i}{N}\right) + \sigma\zeta_i,\ 1 \le i \le N;\ \zeta_i \text{ are independent and identically distributed } \mathcal{N}(0,1). \tag{2}$$

To study our estimation procedures of f, motivated from (1) and (2), we adopt the general abstract framework of the sequential Gaussian white noise. It is useful for the asymptotic study of model (2), whereas the level of noise ε in equation (3) is appropriately calibrated to be

$\varepsilon \approx \sigma/\sqrt{n}$. Under this model, we observe noisy (empirical) coefficients ($\hat{\alpha}$) and $\hat{\theta} = (\hat{\theta}_{jk})_{j,k}$ such that

$$\begin{aligned} \hat{\alpha} &= \alpha + \varepsilon\xi, \\ \hat{\theta}_{jk} &= \theta_{jk} + \varepsilon\xi_{jk}, \end{aligned} \tag{3}$$

where ξ and ξ_{jk} are independent and identically distributed $\mathcal{N}(0, 1)$ and ε is supposed to belong to $]0, \exp(-1)[$ for convenience.

The sequence $\theta = (\theta_{jk})_{j,k}$ is supposed *a priori* to be sparse, meaning that only a small number of *large* coefficients contain nearly all the information about the signal. This motivates us to use keep-or-kill estimators, for which we recall below the hard thresholding estimator $\hat{f}_h$. It suppresses all empirical wavelet coefficients below a threshold value λ_ε that is of the order of the UT value and keeps the surviving ones unchanged. The threshold value λ_ε foremost depends on the noise level ε and also on an additional proportionality parameter m, the role of which in our theoretical study is discussed in Section 4.1

$$\hat{f}_h = \hat{\alpha}\phi + \sum_{(j,k)\in\mathcal{S}_{\varepsilon,m}} \hat{\theta}_{jk}\psi_{jk}, \tag{4}$$

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

(i) the real number m belongs to $]0, +\infty[$, and
(ii) the integer j_{λ_ε} is such that $2^{-j_{\lambda_\varepsilon}} \le \lambda_\varepsilon^2 < 2^{1-j_{\lambda_\varepsilon}}$. If non-negative, $j_{\lambda_\varepsilon} - 1$ is the finest scale up to which the method considers the empirical wavelet coefficients to reconstruct the signal f.

This term-by-term thresholding does not take into account the information that gives us the clusters of wavelet coefficients of large magnitudes that we observed in Figures 1 and 2. This information will allow us to be more precise in the choice of coefficients to keep. Indeed, on the one hand, we should not use in the reconstruction a large *isolated* wavelet coefficient because its being isolated would likely make it not part of the signal. As a consequence, we should reduce the number of false positives (wrongly selected coefficients). On the other hand, a small coefficient, usually killed by the threshold if considered as is, should rather be not excluded if it occurs in the neighbourhood of large coefficients (caused by local structure in the function domain). Here, it is a matter of controlling the number of false negatives (wrongly discarded coefficients).

Under the aforementioned model, several impressive minimax results were obtained for such keep-or-kill estimators (e.g. Donoho *et al.*, 1995) by considering the function f associated to the sequence $\theta = (\theta_{jk})_{j,k}$ to be in a particularly large function space, a Besov space, allowing for spatially variable regularity of the underlying function (basically by a control in L_q of its generalized derivatives). For a more general overview of Besov spaces $\mathcal{B}^s_{q,q'}$, see Hardle *et al.* (1998). In the sequel of our paper, a less general class of Besov spaces, those with $q = 2$ and $q' = +\infty$, is needed to characterize the maxiset of our estimators.

Definition 1. Let $\theta = (\theta_{jk})_{j,k}$ be the sequence of coefficients of a function $f \in L_2([0, 1])$ projected onto a wavelet basis with V vanishing moments. We say that a function f belongs to the Besov space $\mathcal{B}^s_{2,\infty}$, with $0 < s < V$, if and only if

$$\sup_{J \in \mathbb{N}} 2^{2Js} \sum_{j \geq J} \sum_{k=0}^{2^j-1} \theta_{jk}^2 = \|f\|^2_{\mathcal{B}^s_{2,\infty}} < +\infty.$$

Interpreting this definition, we observe that Besov spaces, which naturally appear in estimation problems (see, among others, Cohen *et al.*, 2001; Autin, 2004), characterize functions for which the energy of wavelet coefficients on scales larger than J ($J \in \mathbb{N}$) decreases exponentially in J.

Interestingly enough, the aforementioned results on minimax rates over Besov balls do not model the information given by the clusters of coefficients because the Besov norm $\|.\|_{\mathcal{B}^s_{2,\infty}}$ is invariant under permutations within scale. Using the *maxiset approach*, for which the basic concepts are recalled now hereafter, we introduce a way to take into account these clusters of coefficients by introducing new function spaces related to the methods (definition 3). This allows a more precise characterization of the performances of the BT estimators under consideration.

Let us consider a keep-or-kill estimator $\tilde{f}$ built from the observations $\hat{\alpha}$ and $\hat{\theta} = (\hat{\theta}_{jk})_{j,k}$. Studying the *maxiset performance of* $\tilde{f}$ consists in computing the set of all functions for which the rate of convergence of the L_2-risk of the estimator $\tilde{f}$ is at least as fast as a given rate of convergence ρ (with $\rho = \rho_\varepsilon \to 0$ as $\varepsilon \to 0$). This set of radius $R > 0$ is denoted as $\mathrm{MS}(\tilde{f}, \rho)(R)$ and is defined by

$$\mathrm{MS}\left(\tilde{f}, \rho\right)(R) = \left\{ f \in L_2([0,1]) : \sup_{0<\varepsilon<\exp(-1)} \rho_\varepsilon^{-1} \mathbb{E}_f \|\tilde{f} - f\|_2^2 \leq R^2 \right\}. \quad (5)$$

In this setting, a function space $\mathcal{G}$ is said to be the maxiset of the estimator $\tilde{f}$ for the rate of convergence ρ_ε if the equivalence

$$\sup_{0<\varepsilon<\exp(-1)} \rho_\varepsilon^{-1} \mathbb{E}_f \|\tilde{f} - f\|_2^2 < \infty \iff f \in \mathcal{G} \quad (6)$$

holds, meaning that, for any $R > 0$, there exist some radii $R_1 > 0$ and $R_2 > 0$ of balls of $\mathcal{G}$ such that

$$\mathcal{G}(R_1) \subset \mathrm{MS}\left(\tilde{f}, \rho_\varepsilon\right)(R) \subset \mathcal{G}(R_2).$$

Obviously, the larger the maxiset, the better is the procedure; and the slower the rate, the larger is the maxiset (and conversely). It is important to recall that there is a connection between the minimax and maxiset approaches. Indeed, for any function space $\mathcal{F}$ with minimax rate ρ_ε, we necessarily obtain the following embedding property: $\mathcal{F} \subset \mathcal{G}$. As already emphasized by Kerkycharian & Picard (2002) and Autin (2004), the maxiset approach is also more optimistic than the minimax one because it characterizes the whole nature of the functions that are well estimated by a method. Therefore, although being interested in both approaches, we often pay more attention to the maxiset approach than the minimax approach.

As discussed by Autin (2004), estimators with large maxisets can be constructed from thresholding rules that are *not elitist*—that is, 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 (Autin, 2008; Autin *et al.*, 2011) or, indeed, on horizontal block thresholding rules that had been preliminarily studied by Autin (2008) and Chesneau (2008). When we look at these procedures, their maxisets are larger than those of procedures based on an elitist rule, including hard and soft thresholding estimators, and also many Bayes procedures (Autin *et al.*, 2006).

Motivated by the discussion started in Section 1, we propose to focus on the performance of a wide range of wavelet estimators based on a horizontal block thresholding rule (BT estimators).

3. Horizontal block thresholding estimators

We are now in the position to define our BT estimators using the abstract notation of the previous section. Prior to that, we explain their intuitive construction: we split each large enough level j of empirical wavelet coefficients into neighbouring disjoint groups (i.e. non-overlapping blocks) of the same length depending on the noise level ε and decide for some chosen levels to keep or kill each individual coefficient on the basis of the comparison of the value of the ℓ_p-mean norm of all the empirical coefficients in its block with a chosen threshold value. Interestingly, this threshold will now also depend on p, that is, $\lambda_{\varepsilon,p} = mt_{\varepsilon,p}$ ($0 < m < +\infty$ and $2 \leq p \leq +\infty$).

Let us now fix the (asymptotic) choice of the parameters of our method. For any $0 < m < +\infty$ and any $2 \leq p \leq +\infty$, let

(i) $t = (t_{\varepsilon,p})_\varepsilon$ and $v = (v_{\varepsilon,p})_\varepsilon$ be two sequences of positive real numbers continuously tending to 0 as ε goes to 0 (we recall that $t_{\varepsilon,p}$ determines the threshold value, whereas $v_{\varepsilon,p}$ will determine the rate of convergence of the L_2-risk of the associated BT estimator);

(ii) $j_{o,\varepsilon}$ be the fixed primary resolution scale, chosen to be the smallest integer such that $2^{j_{o,\varepsilon}} > \log \varepsilon^{-1}$;

(iii) $j_{mv_{\varepsilon,p}}$ be the finest considered resolution scale, chosen to be the smallest integer such that $2^{j_{mv_{\varepsilon,p}}} \geq (mv_{\varepsilon,p})^{-2}$; and

(iv) l_ε be the length of the blocks. It has been proven pertinent from both minimax (Cai, 1999, 2002) and maxiset (Autin, 2008) points of view to choose l_ε to be of the order of $\log \varepsilon^{-1}$. This specification remains too vague because the number of blocks at a scale j may not divide 2^j in an integer number. The treatment of this problem has often been neglected in the theoretical literature, whereas practical procedures employed some refinements to handle incomplete blocks at the boundaries. Here, we propose as a solution to calibrate the block length to be of the order $\log \varepsilon^{-1}$ but using the relation $l_\varepsilon = 2^{j_{o,\varepsilon}}$. This choice avoids all the subjectivity related to boundary handling and can be viewed as a way to recover the usual dyadic structure of the multiresolution analysis.

In Section 4, we will use the tuning parameter m to link threshold values of the form $mt_{\varepsilon,p}$ and rates of convergence of the L_2-risk of the form

$$\rho_\varepsilon = (mv_{\varepsilon,p})^\beta. \tag{7}$$

As usual and following the minimax rates of regular enough functions, we shall consider $\beta \in]0, 2[$. We recall that the minimax rates over the balls of the Besov spaces $\mathcal{B}^s_{2,\infty}$ (definition 1) with $s > 1/2$ are all of the order of $\varepsilon^{4s/(1+2s)}$. Considering such a rate means choosing $v_{\varepsilon,p} = \varepsilon$ (example 3 with $p = 2$).

Let us now define a general BT estimator $\tilde{f}^{(t,v)}_{m,p}$ associated with a block-thresholding rule depending on the noise level ε. We first define the involved blocks of translational parameters by

$$B^{(u)}_j(\varepsilon) = \{k \in \mathbb{N} : (u-1)l_\varepsilon \leq k < ul_\varepsilon\}, \text{ for any } u \in \left\{1, 2, \ldots, 2^j l_\varepsilon^{-1}\right\}.$$

For any sequence of wavelet coefficients θ and any sequence of empirical wavelet coefficients $\hat{\theta}$ associated with θ by (3), we denote, for any (j,k), by $\theta \,/\, B_{jk}(\varepsilon)$ the block of wavelet coefficients that contain θ_{jk} and $\hat{\theta} / B_{jk}(\varepsilon)$ the block of empirical wavelet coefficients that contain $\hat{\theta}_{jk}$:

$$\theta / B_{jk}(\varepsilon) = \left\{\theta_{jk'} : k' \in B_j^{(u)}(\varepsilon)\right\} \text{ with } k \in B_j^{(u)}(\varepsilon),$$
$$\hat{\theta} / B_{jk}(\varepsilon) = \left\{\hat{\theta}_{jk'} : k' \in B_j^{(u)}(\varepsilon)\right\} \text{ with } k \in B_j^{(u)}(\varepsilon).$$

Definition 2. Let $0 < \varepsilon < \exp(-1)$, $2 \le p \le +\infty$ and a given $m > 0$. We define the BT estimator $\tilde{f}_{m,p}^{(t,v)}$ as follows:

$$\tilde{f}_{m,p}^{(t,v)} = \hat{\alpha}\phi + \sum_{j=0}^{j_{o,\varepsilon}-1} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk}\psi_{jk} + \sum_{j=j_{o,\varepsilon}}^{j_{mv_{\varepsilon,p}}-1} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk}\, \mathbf{1}\left\{\|\hat{\theta} / B_{jk}(\varepsilon)\|_p > m\, t_{\varepsilon,p}\right\} \psi_{jk}, \tag{8}$$

where

$$\|\hat{\theta} / B_{jk}(\varepsilon)\|_p = \left(l_\varepsilon^{-1} \sum_{k' \in B_{jk}(\varepsilon)} |\hat{\theta}_{jk'}|^p \right)^{1/p} \quad \text{if } p < +\infty,$$
$$\|\hat{\theta} / B_{jk}(\varepsilon)\|_\infty = \max\left\{|\hat{\theta}_{jk'}| : k' \in B_{jk}(\varepsilon)\right\}.$$

According to the definition of the BT estimator $\tilde{f}_{m,p}^{(t,v)}$, the empirical wavelet coefficients selected by the method are those associated with the coarsest scales ($j < j_{o,\varepsilon}$) as well as those on finer scales ($j_{o,\varepsilon} \le j < j_{mv_{\varepsilon,p}}$) belonging to blocks of empirical wavelet coefficients with a large *p-score*, that is, a large ℓ_p-mean norm.

Remark 1. The *BlockShrink* estimator of Cai (1999) is analogous to $\tilde{f}_{m,2}^{(t,v)}$ with the choice $t_{\varepsilon,2} = v_{\varepsilon,2} = \varepsilon$. The *Maximum-Block* estimator of Autin (2008) is analogous to the estimator $\tilde{f}_{m,\infty}^{(t,v)}$ with the choice $t_{\varepsilon,\infty} = v_{\varepsilon,\infty} = \varepsilon\sqrt{\log\varepsilon^{-1}}$. For convenience, we shall keep the names *BlockShrink* and *Maximum-Block* to denote the estimators $\hat{f}_{m,2}^{(t,v)}$ and $\hat{f}_{m,\infty}^{(t,v)}$, respectively, whatever the choices of the sequences t and v.

From now on, we will study the performance of these BT estimators to address the following question: what is the best choice of ℓ_p-mean norm to consider ($2 \le p \le +\infty$)? 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. In the following are listed three examples of threshold values and rates we are particularly interested in.

Example 1. $t_{\varepsilon,p}^{(1)} = v_{\varepsilon,p}^{(1)} = \varepsilon\sqrt{\log\varepsilon^{-1}}$. With such a choice, the threshold value associated with $t_{\varepsilon,p}^{(1)}$ is in the order of the UT (Donoho & Johnstone, 1994), and the rate of convergence ρ_ε as in (7), with $\beta = 4s(1+2s)^{-1}$ and $s > 1/2$, corresponds to the minimax rate over any ball of the Besov space $\mathcal{B}_{2,\infty}^s$, up to a term of order $(\log\varepsilon^{-1})^{1/2}$.

Example 2. $t_{\varepsilon,p}^{(2)} = \varepsilon\sqrt{\log\varepsilon^{-1}}\, l_\varepsilon^{-(1/p)}$ *and* $v_{\varepsilon,p}^{(2)} = \varepsilon\sqrt{\log\varepsilon^{-1}}$, *with the convention* $1/{+\infty} = 0$ *for* $p = +\infty$. With such a choice, the order of the threshold value associated with $t_{\varepsilon,p}^{(2)}$ is lower than the UT for $p < +\infty$. The rate of convergence ρ_ε as in (7), with $\beta = 4s(1+2s)^{-1}$ and $s > 1/2$, corresponds to the minimax rate over any ball of the Besov space $\mathcal{B}_{2,\infty}^s$ up to a term of order $(\log\varepsilon^{-1})^{1/2}$.

Example 3. $t^{(3)}_{\varepsilon,p} = v^{(3)}_{\varepsilon,p} = \varepsilon\sqrt{\log\varepsilon^{-1}}\, l_\varepsilon^{-1/p}$, *with the convention* $1/+\infty = 0$ *for* $p = +\infty$. With such a choice, the order of the threshold value associated with $t^{(3)}_{\varepsilon,p}$ is lower than the UT for $p < +\infty$. The rate of convergence ρ_ε as in (7), with $\beta = 4s(1+2s)^{-1}$ and $s > 1/2$, corresponds to the minimax rate over any ball of the Besov space $\mathcal{B}^s_{2,\infty}$ up to a term of order $(\log\varepsilon^{-1})^{1/2-1/p}$ that is a constant for the particular case $p = 2$.

Notice that for a chosen $p \in [2, +\infty[$, the rate in example 3 is faster than the rate in examples 1 and 2. Actually, it corresponds to the fastest rate of convergence for which our asymptotical results in Section 4 hold. In spite of the fact that the faster is the rate of convergence, the smaller is the maxiset, we shall see in Section 4 that for the choice of the sequences t and v proposed in example 3, the maxiset of the BT estimator $\tilde{f}^{(t,v)}_{m,p}$ for the rate of convergence $(mv_{\varepsilon,p})^{4s/1+2s}$ is quite large because it contains at least the Besov space $\mathcal{B}^s_{2,\infty}$, provided m is large enough (theorem 1 and remark 4).

For any chosen $m > 0$, define the horizontal block thresholding family of wavelet estimators, namely $\mathrm{HBT}^{(m,t,v)}$, as

$$\mathrm{HBT}^{(m,t,v)} = \left\{\tilde{f}^{(t,v)}_{m,p} : 2 \le p \le +\infty\right\}.$$

At first glance, as $2 \le p \le +\infty$ is real valued, the family of estimators $\mathrm{HBT}^{(m,t,v)}$ seems to be uncountable. But it is not, whatever the choice of the threshold values $t_{\varepsilon,p}$, if the rates $v_{\varepsilon,p}$ are all greater than or equal to ε, as its elements only differ by the sets of the blocks that are kept by the related methods and the number of inspected blocks is finite. To prepare for our future results, we introduce the three following kinds of $\mathrm{HBT}^{(m,t,v)}$ families:

(i) $\mathrm{HBT}^{(m,t,v),1}$: when choosing t and v such that, as in example 1,
$t_{\varepsilon,p}$ and $v_{\varepsilon,p}$ do not depend on the parameter p. (*Hyp*-1)

(ii) $\mathrm{HBT}^{(m,t,v),2}$: when choosing t and v such that, as in example 2,
$l_\varepsilon^{\frac{1}{p}} t_{\varepsilon,p}$ and $v_{\varepsilon,p}$ do not depend on the parameter p. (*Hyp*-2)

(iii) $\mathrm{HBT}^{(m,t,v),3}$: when choosing t and v such that, as in example 3,
$l_\varepsilon^{\frac{1}{p}} t_{\varepsilon,p}$ and $l_\varepsilon^{\frac{1}{p}} v_{\varepsilon,p}$ do not depend on the parameter p. (*Hyp*-3)

We shall see in the next section that focusing on $\mathrm{HBT}^{(m,t,v),1}$ and $\mathrm{HBT}^{(m,t,v),2}$ families where the rate is the same one whatever the choice of p allows maxiset comparisons between the BT estimators within such a family to be derived (corollaries 1 and 2) whereas focusing on $\mathrm{HBT}^{(m,t,v),3}$ families allows nice minimax results for our BT estimators to be derived (theorem 2).

4. Main results

We first provide the definition of a new function space, which is the key to our results.

Definition 3. Let $m' > 0$, $0 < r < 2$ and $2 \le p \le +\infty$. We say that a function f belongs to the space $\mathcal{W}^{(t,v)}_{r,m',p}$ if and only if the sequence of its wavelet coefficients $\theta = (\theta_{jk})_{j,k}$ satisfies

$$\sup_{m\ge m'}\ \sup_{0<\varepsilon<\exp(-1)} (mv_{\varepsilon,p})^{r-2} \sum_{j\ge j_{o,\varepsilon}} \sum_{k=0}^{2^j-1} \theta_{jk}^2 \mathbf{1}\left\{\|\theta / B_{jk}(\varepsilon)\|_p \le m t_{\varepsilon,p}\right\} = \|f\|^2_{\mathcal{W}^{(t,v)}_{r,m',p}} < +\infty.$$

First, note that the larger the r, the larger is the function space. Second, in contrast to weak Besov spaces (see for an explicit definition Cohen *et al.*, 2001), which appear in the maxiset of elitist procedures, the function scores $\|.\|_{\mathcal{W}^{(t,v)}_{r,m',p}}$ of the spaces $\mathcal{W}^{(t,v)}_{r,m',p}$ ($m' > 0$ and $0 < r < 2$) are not invariant under permutations of wavelet coefficients within each scale. This is precisely the non-invariance property that allows us to distinguish functions according to the 'block-neighbourhood properties' of their wavelet coefficients. Nevertheless, the following occur:

(i) a translational-shift invariance of the blocks: if you shift one block $B_j^{(u)}(\varepsilon)$ on a level j by $u \to u + v (\text{mod } l_\varepsilon^{-1} 2^j)$, then the function score $\|.\|_{\mathcal{W}^{(t,v)}_{r,m',p}}$ is still the same.

(ii) a permutational-shift invariance of the coefficients within the same block: if you move the positions of the coefficients in a block $B_j^{(u)}(\varepsilon)$ on a level j, then the function score $\|.\|_{\mathcal{W}^{(t,v)}_{r,m',p}}$ is still the same.

The spaces $\mathcal{W}^{(t,v)}_{r,m',p}$ ($m' > 0$ and $0 < r < 2$) characterize sparse functions as highlighted in the following proposition.

Proposition 1. *Let $m' > 0$, $0 < r < 2$ and $2 \le p \le +\infty$. Then,*

$$f \in \mathcal{W}^{(t,v)}_{r,m',p}$$
$$\Downarrow$$
$$\sup_{m \ge m'} \sup_{0<\varepsilon<\exp(-1)} m^r v_{\varepsilon,p}^{r-2} t_{\varepsilon,p}^2 \left(\log \varepsilon^{-1}\right)^{\frac{2}{p}-1} \sum_{j=j_{o,\varepsilon}}^{j_{mv_{\varepsilon,p}}-1} \sum_{k=0}^{2^j-1} \mathbf{1}\left\{\|\theta / B_{jk}(\varepsilon)\|_p > \frac{m t_{\varepsilon,p}}{2}\right\} < +\infty,$$

where $\theta = (\theta_{jk})_{j,k}$ is the sequence of the wavelet coefficients of f.

Assuming some conditions on the choice of both the sequence of t and the sequence of v, these function spaces are enlargements of classical Besov spaces as suggested by proposition 2.

Proposition 2. *Let $2 \le p \le +\infty$. Assume that $t = (t_{\varepsilon,p})_\varepsilon$ and $v = (v_{\varepsilon,p})_\varepsilon$ are such that*

$$\lim_{\varepsilon \to 0} \frac{v_{\varepsilon,p}}{t_{\varepsilon,p}} > 0.$$

Then, for any $m' > 0$ and any $0 < s < V$,

$$\mathcal{B}^s_{2,\infty} \subset \mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m',p}. \tag{9}$$

Propositions 3 and 4 provide embedding properties that may exist between the function spaces $\mathcal{W}^{(t,v)}_{r,m',p}$, depending on the chosen sequences t and v.

Proposition 3. *Consider sequences t and v as in Hyp-1. Then, for any $m' > 0$, any $2 \le p < q \le +\infty$ and any $0 < r < 2$, the following embedding of spaces holds:*

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

The assertion of this proposition changes, however, if the p-scores over blocks appearing in definition 3 are compared with a threshold value that depends in a particular way also on p, namely by rescaling in some sense with the length of the blocks.

Proposition 4. *Consider sequences t and v as in Hyp-2. Then, for any $m' > 0$, any $2 \le p < q \le +\infty$ and any $0 < r < 2$, the following embedding of spaces holds:*

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

It is essential to note that these results prepare the ground to find the 'maxiset-optimal' estimator $\tilde{f}^{(t,v)}_{m,p}$, that is, the best p, according to the more refined specification of both threshold value (via $mt_{\varepsilon,p}$) and rate (via $v_{\varepsilon,p}$). For this, we refer to corollaries 1 and 2.

Remark 2. The proofs of propositions 3 and 4 are omitted because they are a direct consequence of lemma 1 given in the online Supporting Information.

4.1. Asymptotic results

When considering the maxiset approach for thresholding estimators $\tilde{f}_m$ with a parameter m, which calibrates the threshold value, the maxiset is usually sandwiched as

$$\mathcal{G}(R_{1,m}) \subset \mathrm{MS}\left(\tilde{f}_m, \rho_\varepsilon\right)(R) \subset \mathcal{G}(R_{2,m}), \tag{10}$$

where the involved radii $R_{1,m}$ and $R_{2,m}$ of $\mathcal{G}$ depend on m (see for instance Autin *et al.*, 2006).

The role of the parameter of calibration m is often ignored in asymptotic theory ($\varepsilon \to 0$). In both the minimax and maxiset settings, results are commonly established for any value of m provided it is large enough to guarantee the good performance of the studied thresholding method. Nevertheless, in our approach hereafter, we make sure to determine the maxisets of BT estimators $\tilde{f}^{(t,v)}_{m,p}$ independently of large enough m. This is like forcing the radii $R_{1,m}$ and $R_{2,m}$ not to depend on m but in return accepting that R is linked to m. Precisely, we propose in our study to look for the spaces $\mathcal{G}_p$ ($2 \le p \le +\infty$) satisfying the following embeddings for some $m_* > 0$: for any $C > 0$, there exist $R_1 > 0$ and $R_2 > 0$ such that

$$\mathcal{G}_p(R_1) \subset \bigcap_{m \ge 2m_*} \mathrm{MS}\left(\tilde{f}^{(t,v)}_{m,p}, v_{\varepsilon,p}^{\frac{4s}{1+2s}}\right)\left(Cm^{\frac{4s}{1+2s}}\right) \subset \mathcal{G}_p(R_2).$$

This can be rewritten as the following equivalence:

$$\sup_{m \ge 2m_*} \sup_{0<\varepsilon<\exp(-1)} (mv_{\varepsilon,p})^{-\frac{4s}{1+2s}} \mathbb{E}_f \| \tilde{f}^{(t,v)}_{m,p} - f \|_2^2 < +\infty \iff f \in \mathcal{G}_p.$$

This allows us to address (at least theoretically) the important problem of the choice of the best value for m. Indeed, in such a case, m only calibrates the rate of convergence; hence, the best choice of m will be the one that ensures the fastest reconstruction, that is, the smallest value of m considered. Finally, we would like to remark that controlling the maxiset results uniformly in m is also of primary importance when considering general thresholding rules for which the maxiset may not be embedded for different values of m as explained by Autin *et al.* (2012).

To present our maxiset results for the BT estimators within the horizontal block thresholding families, we propose to use a large collection of rates of convergence, which are $(mv_{\varepsilon,p})^{4s/1+2s}$ (with $s > 0$). The exponent terms are chosen to be the ones appearing in the minimax rates of function spaces with regularity s.

Some assumptions on the choices of both the sequence of threshold values t and the sequence of rates v are performed to ensure the validity of our forthcoming asymptotic results. Assume that, for any $2 \le p \le +\infty$,

$$\lim_{\varepsilon \to 0} \frac{\varepsilon(\log \varepsilon^{-1})^{\frac{1}{2}}}{v_{\varepsilon,p}} > 0, \quad \lim_{\varepsilon \to 0} \frac{v_{\varepsilon,p}}{t_{\varepsilon,p}} > 0 \quad \text{and} \quad \lim_{\varepsilon \to 0} \frac{t_{\varepsilon,p}}{\varepsilon(\log \varepsilon^{-1})^{\frac{1}{2}-\frac{1}{p}}} > 0. \tag{11}$$

The first limit on the left ensures that we focus on rates of convergence, which are faster than or of the same order as the best rate of convergence of separable rules to reconstruct Besov balls (Cai, 2008). The second limit will ensure that the proposed BT estimators have large maxisets, according to proposition 2. Finally, the third limit is a lower bound for our method-dependent threshold values. It ensures the existence of a large deviation property that will be useful in the proofs of our maxiset result.

For any chosen $s > 0$, the reader can check that assumption (11) is satisfied by the three examples given in Section 3.

Remark 3. Note that both the smallest threshold value and the fastest rate satisfying (11) are the ones given in example 3.

We now state the main theorem dealing with the maxiset performance of the estimators that belong to the horizontal block thresholding family. Firstly, we characterize in theorem 1 the well-reconstructed functions by such estimators for a wide range of parameters of calibration; secondly, we derive minimax results in theorem 2; thirdly, we exhibit the best estimator of $\mathrm{HBT}^{(m,t,v),1}$ and $\mathrm{HBT}^{(m,t,v),2}$ families thanks to corollaries 1 and 2.

Theorem 1 (Maxiset result). *Fix $s > 0$ and consider $2 \le p \le +\infty$. Let $t = (t_{\varepsilon,p})_\varepsilon$ and $v = (v_{\varepsilon,p})_\varepsilon$ satisfy (11). Define m_* as the real number such that $m_*^2 - 2\log m_* = 9$. Then,*

$$\sup_{m \ge 2m_*} \sup_{0<\varepsilon<\exp(-1)} (mv_{\varepsilon,p})^{-\frac{4s}{1+2s}} \mathbb{E}_f \| \tilde{f}^{(t,v)}_{m,p} - f \|_2^2 < +\infty \Longleftrightarrow f \in \mathcal{B}_{2,\infty}^{\frac{s}{1+2s}} \cap \mathcal{W}^{(t,v)}_{\frac{2}{1+2s},m_*,p}.$$

Remark 4. The maxisets of the estimators $\tilde{f}^{(t,v)}_{m,p}$ are quite large function spaces because from (9) of proposition 2 and the Besov embedding properties, the function spaces $\mathcal{B}_{2,\infty}^{s/1+2s} \cap \mathcal{W}^{(t,v)}_{2/1+2s,m_*,p}$ contain the Besov space $\mathcal{B}^s_{2,\infty}$.

From theorem 1 and remark 4, we immediately derive the following minimax result.

Theorem 2 (Minimax result). *Consider $0 < s < V$, $2 \le p \le +\infty$ and $R > 0$. Let $t = (t_{\varepsilon,p})_\varepsilon$ and $v = (v_{\varepsilon,p})_\varepsilon$ satisfy (11). Then, for any $m \ge 2m_*$,*

$$\sup_{0<\varepsilon<\exp(-1)} (mv_{\varepsilon,p})^{-\frac{4s}{1+2s}} \sup_{f: \|f\|_{\mathcal{B}^s_{2,\infty}} \le R} \mathbb{E}_f \| \tilde{f}^{t,v)}_{m,p} - f \|_2^2 < +\infty.$$

Therefore, we have obtained sufficient conditions (assumption (11)) on choices of t and v to ensure that all the BT estimators in our families are good estimators, in the sense that they are all adaptive and near minimax optimal over the balls of the Besov spaces $\mathcal{B}^s_{2,\infty}$, with $s > 1/2$. Furthermore, theorem 2 highlights the choices of t and v to be made to create BT estimators that strictly outperform the hard and soft thresholding estimators in the minimax sense. Consequently, these methods are better than any separable thresholding procedure, that is, one that decides to keep or kill a coefficient solely as a function of the magnitude of this individual coefficient (theorem 2 in Cai, 2008). A possible choice for t and v is given in example 3, whereas the minimax rates over the Besov balls are achieved by the BlockShrink estimator ($p = 2$), up to a constant.

Let us go back to the maxiset approach to go further in our interpretations. We state corollaries 1 and 2, which are direct consequences of theorem 1 and propositions 3 and 4.

Corollary 1. *Under assumption (11), $\tilde{f}_{m,\infty}^{(t,v)}$ is the best estimator in the maxiset sense among any HBT$^{(m,t,v),1}$ family.*

Corollary 2. *Under assumption (11), $\tilde{f}_{m,2}^{(t,v)}$ is the best estimator in the maxiset sense among any HBT$^{(m,t,v),2}$ family.*

Corollaries 1 and 2 give the best element of any HBT$^{(m,t,v),1}$ or HBT$^{(m,t,v),2}$ family of BT estimators. It is basic to remark that $\tilde{f}_{m,\infty}^{(t,v)}$ is the same procedure in both families. This allows us to compare them and see that the best way to obtain large maxisets is to choose threshold values that are of lower order than the UT and to choose small scores too. This very interesting fact has a powerful interpretation in terms of false discoveries, which we refer to in Section 5.

5. Numerical experiments

This section proposes numerical experiments designed to check whether our theoretical results can be observed in a practical setting, that is, in the context of non-parametric regression described in (2). Obviously, our theoretical approach cannot model all the complexity encountered in practice. Therefore, we choose a classical setting for numerical experiments, using the Daubechies least asymmetric wavelets with eight vanishing moments.

To illustrate our theoretical results, we consider two HBT$^{(m,t,v)}$ families associated with the choices $t_{\varepsilon,p} = \varepsilon(\log\varepsilon^{-1})^{1/2}$ and $t_{\varepsilon,p} = \varepsilon(\log\varepsilon^{-1})^{1/2} l_\varepsilon^{-1/p}$, and for both, we choose $m = 5$ to be the parameter of calibration. We shall respectively call these families the HBT$^{(m,t,v),1}$ family and the HBT$^{(m,t,v),2}$ family. Following Cai (1997), we set the threshold $\hat{\lambda} = \hat{\sigma}(5N^{-1}\log N)^{1/2}$ for all BT methods associated with the HBT$^{(m,t,v),1}$ family, and we choose $\hat{\lambda}_p = \hat{\sigma}(5N^{-1}\log N)^{1/2-1/p}$ for those associated with HBT$^{(m,t,v),2}$. We follow a standard approach to estimate σ by the median absolute deviation, divided by 0.6745, over the wavelet coefficients at the finest wavelet scale $J_N - 1$ such that $J_N = \lfloor \log_2 N \rfloor$ (e.g. Vidakovic, 1999). Such a choice of a finest scale means to consider, for any $2 \le p \le +\infty$, $v_{\varepsilon,p} = \varepsilon$, up to constant, in the sequential Gaussian white noise model. At last, we set the primary resolution scale to be $J_0 = \lceil \log_2 \log N \rceil$ and the length of the blocks to be 2^{J_0}.

We generate the data sets from a large panel of functions often used in wavelet estimation studies (see for instance 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, for integers p from 2 to 10 and for $p = +\infty$, the integrated squared error (ISE) of the BT estimators $\tilde{f}_{m,p}^{(t,v)}$ at the ℓth Monte Carlo replication (ISE$^{(l)}\left(\tilde{f}_{m,p}^{(t,v)}\right)$, $1 \le l \le M$) as follows:

$$\mathrm{ISE}^{(l)}\left(\tilde{f}_{m,p}^{(t,v)}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{f}_{m,p}^{(t,v)}\left(\frac{i}{N}\right) - f\left(\frac{i}{N}\right)\right)^2.$$

We generate $M = 2000$ Monte Carlo replications and compute the mean ISE (MISE) as follows

$$\mathrm{MISE}\left(\tilde{f}_{m,p}^{(t,v)}\right) = \frac{1}{M}\sum_{l=1}^{M}\mathrm{ISE}^{(l)}\left(\tilde{f}_{m,p}^{(t,v)}\right).$$

Because of numerous connections between keep-or-kill estimation and hypothesis testing (e.g. Abramovich *et al.*), 2006, we find very useful for interpreting our results reported in Tables 1 and 2 the number of false positives/negatives (i.e. type I/II errors). These are obtained

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

	Threshold								
	$\mathrm{HBT}^{(m,t,v),1}$				$\mathrm{HBT}^{(m,t,v),2}$				
Method	$\tilde{f}^{(t,v)}_{m,2}$	$\tilde{f}^{(t,v)}_{m,5}$	$\tilde{f}^{(t,v)}_{m,10}$	$\tilde{f}^{(t,v)}_{m,\infty}$	$\tilde{f}^{(t,v)}_{m,2}$	$\tilde{f}^{(t,v)}_{m,5}$	$\tilde{f}^{(t,v)}_{m,10}$	$\tilde{f}^{(t,v)}_{m,\infty}$	$\hat{f}^{\mathcal{O}}$
	Function: step								
MISE	23.13	10.27	9.05	8.33	7.86	8.33	8.33	8.33	2.54
False +	4.7	9.6	10.5	11.3	12.1	11.3	11.3	11.3	0.0
False −	25.9	20.4	19.5	18.7	17.7	18.7	18.7	18.7	0.0
Size	28.8	39.2	40.9	42.6	44.4	42.6	42.6	42.6	50.0
	Function: wave								
MISE	3.33	3.33	3.33	3.32	2.68	3.31	3.32	3.32	0.76
False +	0.0	0.0	0.0	0.0	1.6	0.1	0.0	0.0	0.0
False −	30.0	30.0	30.0	29.8	23.0	29.7	29.8	29.8	0.0
Size	24.0	24.0	24.0	24.2	32.6	24.4	24.2	24.2	54.0
	Function: blip								
MISE	5.99	3.69	3.26	2.80	2.39	2.72	2.78	2.8	0.78
False +	2.9	4.2	4.8	5.7	6.6	5.8	5.7	5.7	0.0
False −	16.4	13.0	12.2	11.1	9.8	10.9	11.0	11.1	0.0
Size	23.5	28.2	29.6	31.6	33.9	31.9	31.7	31.6	37.0
	Function: blocks								
MISE	13.29	8.86	7.29	5.90	4.75	5.52	5.76	5.90	1.41
False +	2.1	8.2	10.5	13.8	17.6	14.1	13.9	13.8	0.0
False −	104.2	89.3	82.0	74.5	67.2	72.6	73.9	74.5	0.0
Size	50.0	70.9	80.5	91.3	102.4	93.5	92.0	91.3	152.0
	Function: bumps								
MISE	6.71	3.49	2.57	1.96	1.55	1.86	1.93	1.96	0.56
False +	15.2	29.2	35.8	41.6	46.3	42.3	41.8	41.6	0.0
False −	94.4	70.7	61.2	53.2	46.5	52.1	52.9	53.2	0.0
Size	89.8	127.5	143.6	157.4	168.8	159.2	157.8	157.4	169.0
	Function: heavisine								
MISE	4.03	3.79	3.28	2.75	2.46	2.74	2.75	2.75	0.77
False +	0.0	0.3	0.9	1.9	2.6	1.9	1.9	1.9	0.0
False −	21.0	20.3	18.9	17.2	16.0	17.2	17.2	17.2	0.0
Size	8.0	8.9	11.0	13.7	15.6	13.7	13.7	13.7	29.0

MISE, mean integrated squared error.

by comparing the set of indices of wavelet coefficients kept by each estimators with the set of indices of the keep-or-kill oracle estimator

$$\hat{f}^{\mathcal{O}} = \hat{\alpha}\phi + \sum_{(j,k)\in\mathcal{S}^{\mathcal{O}}} \hat{\theta}_{jk}\psi_{jk}\,, \tag{12}$$

where $\mathcal{S}^{\mathcal{O}} = \left\{(j,k) \in \mathbb{N}^2 : j < J_N;\ k < 2^j;\ |\theta_{jk}| > \sigma/\sqrt{N}\right\}$.

The results suggest similar behaviour for different values of N and SNR. To keep clear the presentation of the results, we only report those for $N = 2048$ and $\mathrm{SNR} = 10$ in Tables 1 and 2.

Figures 5 and 6 summarize the MISE results. We observe the optimality of the estimator $\tilde{f}^{(t,v)}_{m,\infty} \in \mathrm{HBT}^{(m,t,v),1}$ ($\tilde{f}^{(t,v)}_{m,2} \in \mathrm{HBT}^{(m,t,v),2}$) for all the tested functions as suggested by corollary 1 (corollary 2). In addition, there is a gradual improvement of the MISE performance when p increases (decreases), reflecting the embeddings of the maxisets of the BT estimators considered (Section 4).

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

	Threshold								
	$\mathrm{HBT}^{(m,t,v),1}$				$\mathrm{HBT}^{(m,t,v),2}$				
Method	$\tilde{f}_{m,2}^{(t,v)}$	$\tilde{f}_{m,5}^{(t,v)}$	$\tilde{f}_{m,10}^{(t,v)}$	$\tilde{f}_{m,\infty}^{(t,v)}$	$\tilde{f}_{m,2}^{(t,v)}$	$\tilde{f}_{m,5}^{(t,v)}$	$\tilde{f}_{m,10}^{(t,v)}$	$\tilde{f}_{m,\infty}^{(t,v)}$	$\hat{f}^{\mathcal{O}}$
	Function: Doppler								
MISE	7.14	5.27	4.32	3.33	2.51	3.10	3.25	3.33	1.07
False +	2.0	3.1	4.5	6.2	7.7	6.2	6.2	6.2	0.0
False −	29.8	26.3	23.7	20.5	17.1	19.8	20.3	20.5	0.0
Size	34.2	38.8	42.8	47.6	52.6	48.5	47.9	47.6	62.0
	Function: angles								
MISE	1.57	1.57	1.57	1.57	1.56	1.57	1.57	1.57	0.75
False +	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
False −	11.0	11.0	11.0	11.0	10.8	11.0	11.0	11.0	0.0
Size	24.0	24.0	24.0	24.0	24.4	24.0	24.0	24.0	35.0
	Function: parabolas								
MISE	4.32	4.06	3.51	2.50	1.63	2.06	2.33	2.50	0.82
False +	0.0	0.1	0.5	1.2	1.9	1.5	1.3	1.2	0.0
False −	8.0	7.6	6.9	5.4	4.1	4.8	5.2	5.4	0.0
Size	16.0	16.5	17.6	19.9	21.9	20.7	20.2	19.9	24.0
	Function: time shift sine								
MISE	1.65	1.65	1.64	1.60	1.46	1.59	1.60	1.60	0.56
False +	3.0	3.0	3.0	3.2	3.8	3.2	3.2	3.2	0.0
False −	4.0	4.0	4.0	3.8	3.2	3.8	3.8	3.8	0.0
Size	24.0	24.0	24.0	24.5	25.7	24.5	24.5	24.5	25.0
	Function: spikes								
MISE	2.36	1.14	0.85	0.71	0.65	0.69	0.70	0.71	0.35
False +	5.3	8.2	9.3	10.2	10.8	10.3	10.2	10.2	0.0
False −	23.1	15.9	13.9	12.9	12.1	12.8	12.8	12.9	0.0
Size	49.2	59.2	62.4	64.3	65.7	64.5	64.4	64.3	67.0
	Function: corner								
MISE	1.91	0.93	0.70	0.53	0.45	0.52	0.53	0.53	0.24
False +	0.1	1.1	1.5	1.8	2.0	1.8	1.8	1.8	0.0
False −	9.8	6.7	5.5	4.5	4.0	4.5	4.5	4.5	0.0
Size	12.2	16.5	18.0	19.3	20.0	19.4	19.3	19.3	22.0

MISE, mean integrated squared error.

Looking at the number of false positives/negatives for the BT estimators reported in Tables 1 and 2, we can check that the best estimators in each family tend 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 $\mathrm{HBT}^{(m,t,v),1}$, the conservative UT strongly controls the false positives but discards many small coefficients that would be useful for the reconstruction. With such a high threshold value, the numerical experiments show that the estimator $\tilde{f}_{m,\infty}^{(t,v)}$ has the lowest MISE. This is indeed the ∞-score, which is the most powerful in reducing the false negatives using the structure among the coefficients.

By comparing corollaries 1 and 2, our results point out that the $\mathrm{HBT}^{(m,t,v),1}$ 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, estimators of the $\mathrm{HBT}^{(m,t,v),2}$ family. This is confirmed by our numerical experiments as can be observed in Figure 6; the best results are obtained for $\tilde{f}_{m,2}^{(t,v)}$. The latter have lower MISE than $\tilde{f}_{m,\infty}^{(t,v)}$ for all the tested functions with improvements of up to nearly 54% lower MISE. This emphasizes also that the use of

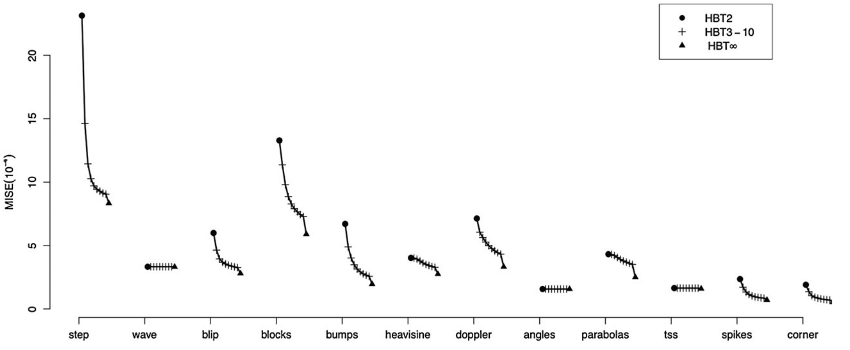

Fig. 5. Mean integrated squared error (MISE) of the non-overlapping block thresholding estimator in $\mathrm{HBT}^{(m,t,v),1}$ for different values of $2 \le p \le +\infty$ for estimating various functions with a signal-to-noise ratio equal to 10 and $N = 2048$.

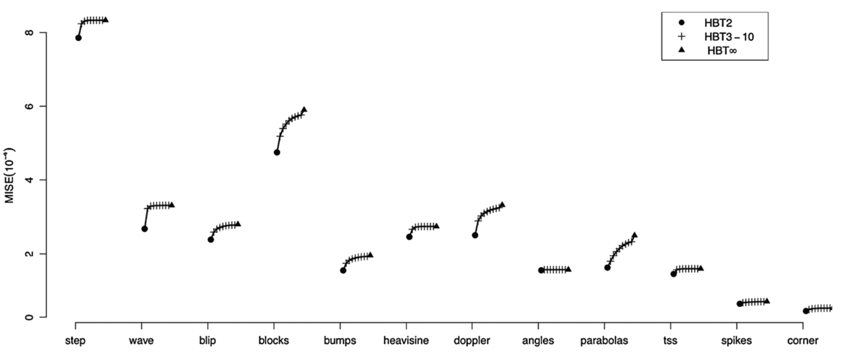

Fig. 6. Mean integrated squared error (MISE) of the non-overlapping block thresholding estimator in $\mathrm{HBT}^{(m,t,v),2}$ for different values of $2 \le p \le +\infty$ for estimating various functions with a signal-to-noise ratio equal to 10 and $N = 2048$.

less conservative thresholds increases the risk of false positives, and it needs to be controlled by using scores with smaller p (attributing a score with $p < +\infty$; coefficients larger than the threshold can also be discarded).

6. Summary of results and conclusion

In this paper, we introduced the family of non-overlapping horizontal block thresholding estimators. We studied the performance of the estimators of this family under the L_2-risk using the maxiset approach. We remark the good maxiset performance for a wide range of threshold values and rates, and we identified the best procedure in some cases, that is, the one using the l_2-norm and a threshold value in the order of the noise level ε. This paper shows the importance of adapting the threshold value and the score to enlarge the maxiset.

For a given threshold value (fixed with regard to p), there is the following interpretation of this family of BT estimators according to p: the score of $p = +\infty$ corresponds to a choice that really focuses on the reduction of false negatives; all the coefficients in a block are kept if only one coefficient passes over the threshold. This method, however, has to accommodate high threshold values to control the false positives. Scores with lower p are meant to simultaneously control false positives and negatives. For these, one has to reduce the value of the threshold if one aims to really improve the estimation. Another way to look at these results is to see the methods under study being parametrized by the score attributed to blocks and the threshold value. These parameters have some interdependence over which, using the maxiset approach, we were able to optimize to identify the procedure with the largest maxiset.

Our numerical experiments confirm our theoretical findings, that is, the best procedure in the $\mathrm{HBT}^{(m,t,v),1}$ family is obtained for $p = +\infty$. On the contrary, for methods with a threshold value that depends on p, such as the $\mathrm{HBT}^{(m,t,v),2}$ one, the best procedure is the one associated with $p = 2$. Because $\tilde{f}_{m,\infty}^{(t,v)}$ also belongs to $\mathrm{HBT}^{(m,t,v),2}$, the best estimator of both families is $\tilde{f}_{m,2}^{(t,v)}$.

It is worthwhile to mention that these BT estimators form spatially homogeneous partitions within scales. Among the various recent developments, those considering data-driven adaptive partitions by Evers & Heaton (2009) and Heaton (2009) have been proven extremely powerful in practice. Their theoretical study using the maxiset approach constitutes a very interesting challenge for future research.

Acknowledgement

The authors wish to thank the anonymous referees for their useful comments and suggestions.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Proofs of results.

References

Abramovich, F., Benjamini, Y., Donoho, D. & Johnstone, I. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34**, 584–653.

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

Autin, F. (2004). *Maxiset point of view in nonparametric estimation*, Ph.D. thesis, University of Paris 7, France.

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

Autin, F., Freyermuth, J.-M. & von Sachs, R. (2011). Ideal denoising within a family of tree-structured wavelet estimators. *Electron. J. Statist.* **5**, 829–855.

Autin, F., Freyermuth, J.-M. & von Sachs, R. (2012). Combining thresholding rules: a new way to improve the performance of wavelet estimators. *J. Nonparam. Statist.* **24**, 905–922.

Autin, F., Picard, D. & Rivoirard, V. (2006). Large variance Gaussian priors in Bayesian nonparametric estimation: a maxiset approach. *Math. Methods Statist.* **15**, 349–373.

Cai, T. (1997). On adaptivity of Blockshrink wavelet estimator over Besov spaces. Technical report 97-05, Purdue University. Available on (http://www.stat.purdue.edu/research/technical_reports/pdfs/1997/tr97-05.pdf).

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

Cai, T. (2002). On block thresholding in wavelet regression: adaptivity, block size, and threshold level. *Statist. Sinica* **12**, 1241–1273.

Cai, T. (2008). On information pooling, adaptability and superefficiency in nonparametric function estimation. *J. Multiv. Anal.* **99**, 421–436.

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

Chesneau, C. (2008). On the maxiset comparison between hard and block thresholding methods. *Statist. Probab. Let.* **18**, 675–681.

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

Daubechies, I. (1992). *Ten lectures on wavelets*, SIAM, Philadelphia.

Donoho, D. & Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.

Donoho, D., Johnstone, I., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc., Ser. B* **57**, 301–337.

Evers, L. & Heaton, T. (2009). Locally-adaptive tree-based thresholding. *J. Comput. Graph. Statist.* **18**, 961–977.

Hall, P., Kerkycharian, G. & Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.* **26**, 922–942.

Hall, P., Kerkycharian, G. & Picard, D. (1999). On the minimax optimality for block thresholded wavelet estimators. *Statist. Sinica* **9**, 33–49.

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

Heaton, T. (2009). Adaptive thresholding of sequences with locally variable strength. *Statist. and Comput.* **19**, 57–71.

Kerkycharian, G. & Picard, D. (2002). Minimax or maxisets? *Bernoulli* **8**, 219–253.

Vidakovic, B. (1999). *Statistical modelling by wavelets*, John Wiley & Sons Inc., New York.

Received July 2011, in final form January 2013

Florent Autin, Université d'Aix-Marseille 1, CNRS UMR 6632, C.M.I., 39 rue F. Joliot Curie, 13453 Marseille Cedex 13, France.
E-mail: autin@cmi.univ-mrs.fr