Bernoulli **8**(2), 2002, 219–253

Minimax or maxisets?

GÉRARD KERKYACHARIAN[1,2] and DOMINIQUE PICARD[2]

[1]*Université Paris X-Nanterre, 200 avenue de la République, 92001 Nanterre Cedex, France.*
[2]*Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599, Université Paris VI and Université Paris VII, 16, rue de Clisson 75013, Paris, France.*

We discuss a new way of evaluating the performance of a statistical estimation procedure. This consists of investigating the maximal set where a given procedure has a given rate of convergence. Although the setting is not vastly different from the minimax context, it is in a sense less pessimistic and provides a functional set which is authentically connected to the procedure and the model. We also investigate more traditional concerns about procedures: oracle inequalities. Difficulties arise in the practical definition of this notion when the loss function is not the L_2 norm. We explain these difficulties and suggest a new definition in the case of L_p norms and pointwise estimation. We investigate the connections between maxisets and local oracle inequalities, and prove that verifying a local oracle inequality implies that the maxiset automatically contains a prescribed set linked with the oracle inequality. We have investigated the consequences of this statement on well-known efficient adaptive methods: wavelet thresholding and local bandwidth selection. We prove local oracle inequalities for these methods and draw conclusions about the maxisets associated with them.

Keywords: adaptive methods; oracle inequalities; saturation sets

1. Introduction

The recent appearance of nonparametric estimation methods offering a high degree of adaptivity has led to renewed interest in minimax theory.

During the 1970s and 1980s, minimax theory was essentially a jumble of results from worldwide researchers seeking solutions in situations where one specified the problem (density estimation, regression, spectral density, ...), the risk (L_2, L_p norms), and the functional class (Hölder, Sobolev, ...). Its impact on the statistical community was not uniformly positive. The main reasons were probably the lack of connection between the minimax paradigm and the actual situation where one is confronted with real data: minimax estimators either depended on smoothness assumptions which were mostly impossible to verify, or (for some procedures which were really new) were impracticable.

At the same time the statistical community was influenced by the practical need for spatial adaptation methods. During the 1990s, the development of wavelet methods and, in parallel, of local bandwidth selection led to some reduction in the gap between theory and practice.

The minimax paradigm has not disappeared. The general framework was enhanced with new spaces to better reflect spatial adaptivity (Besov, Triebel, ...). Moreover, the search for adaptive procedures by the minimax community is a useful approach to the question of tuning the smoothing parameters. Indeed, the easiest way to theoretically prove the high performance of these procedures for the analysis of functions with inhomogenous

1350–7265

smoothness, was still to establish that they allow minimax convergence rates close to optimal over large function classes. In parallel, a deep understanding of the minimax most striking evidence, the traditional trade-off between bias and stochastic terms has been an essential source of inspiration for the construction of these efficient methods.

However, part of the aversion and reluctance of the statistical community remained and some arguments are substantially difficult to deny. The tendency to expect the worst seems generally to be too pessimistic for practical purposes. Moreover, in the nonparametric context, the minimax theory investigates the rates of convergence for different sets of functions. Another drawback lies in the essential difficulty of making an appropriate deductive choice of these sets. Even in an adaptive context, this difficulty remains.

Our first aim will be to discuss a new way of evaluating the performance of a procedure. This approach, fairly standard in approximation theory (linking approximation procedures with saturation classes), is more unusual in statistics. It consists of investigating the maximal set where a procedure has a given rate of convergence. The setting is not vastly different from the minimax context, but it has the main advantage of providing a functional set which is authentically connected to the procedure and the model. In a sense it is also less pessimistic. When looking for minimax procedures over a fixed functional set, or adaptive procedures with respect to a range of sets indexed by a smoothing parameter (for example, Hölder spaces indexed by the smoothness parameter α) we are in fact seeking the most difficult functions in this set that can be estimated by a general procedure. But in fact this set of 'bad functions' is strongly dependent on the way smoothness is defined: most unfavourable a priori measures or sets of functions in Assouad's cube or Fano's pyramid do not look the same at all if we refer to Hölder classes or to Sobolev spaces, for instance. Moreover, they usually do not reflect what we expect to find in practical situations. As a consequence it is somewhat difficult to find the motivation to continue. When seeking maxisets, we look for functions which are the most difficult to estimate for a *given* procedure. Besides the fact that this is an interesting piece of information on the procedure itself, the main advantage of this approach is that the smoothness parameter will not come from an artificial external choice of spaces, but will be naturally connected to the procedure. We still are looking for the worst, but in a 'pragmatic' context, not in an imaginary one. Another incidental advantage of this approach is often to produce new classes of sets (or to rediscover forgotten ones, as here) which contain the classical Besov spaces, for instance. This provides an opportunity to enhance the minimax paradigm, since procedures automatically are minimax on their maxisets.

The second aim of this paper is to show that maxisets are connected not only with minimax theory, but also with another new and important way of evaluating the performance of statistical procedures, i.e. oracle inequalities. The concept of the oracle inequality was introduced in statistics by Donoho and Johnstone (1994) to reflect the idea of performing as if having an 'oracle' choosing the procedure (see also Donoho and Johnstone 1995). One of the major differences between oracle inequalities and minimax theory is that oracle inequalities are more oriented towards the function being estimated. This notion can prove very efficient in many contexts. However, it becomes more difficult to use when the loss function is not the L_2 norm. We will explain the difficulties that arise, and suggest a new definition in the case of L_p norms and pointwise estimation.

Surprisingly, the connections between maxisets and local oracle inequalities are in fact profoundly important and one of our goals in this paper will be to emphasize them. They will in particular be illustrated by Proposition 3, where it is stated that verifying a local oracle inequality implies that the maxiset automatically contains a prescribed set linked with the oracle inequality. We have investigated the consequences of this statement on efficient methods that are well known: wavelet threshold and local bandwidth selection. From the adaption or minimax point of view, all these procedures are equivalent. We can prove local oracle inequalities for both methods, as well as for a hybrid procedure which has turned out to be of particular interest in various contexts – especially for confidence interval purposes (see Picard and Tribouley 2000). We can precisely identify the maxisets of the thresholding procedure but not of the two other procedures. However, this allows us to formulate the following concluding remarks. As far as maxisets are concerned, local bandwidth selection and the hybrid procedure are at least as good as thresholding. Whether they are strictly better is an open question, as is the relation between the two procedures.

The paper is organized as follows. Section 2 is devoted to maxisets. In particular, we give the explicit maxisets for linear kernel methods as well as thresholding procedures. Section 3 concerns oracle inequalities. We provide definitions of such inequalities for L_p norms and in the local context. We investigate the consequences of oracle inequalities over the magnitude of the maxisets. Section 4 investigates the examples of adaptive procedures mentioned above. Finally, Sections 5 and 6 investigate the respective positions of the functional spaces appearing in the definition of the maxisets, and the consequences for comparisons of the procedures.

2. Maxisets

The study of the set of functions $f \in X$, for a family of operators U_n in some functional space X, such that

$$\|f - U_n(f)\|_X = \mathcal{O}(\epsilon_n),$$

where ϵ_n is a sequence of positive numbers decreasing to 0, is a classical topic in approximation theory. This family is known as the *saturation class* linked with the sequence U_n and the rate ϵ_n; see, for example, Butzer and Berens (1967), Butzer and Nessel (1971) and DeVore and Lorentz (1993).

We will now define maxisets and set out the motivation for their study. The definition is illustrated with nonparametric examples. We consider a sequence of models $\mathcal{E}_n = \{P_\theta^n, \theta \in \Theta\}$, where the P_θ^n are probability distributions on the measurable spaces Ω_n, and Θ is the set of parameters. We also consider a sequence of estimates $\hat{q}_n$ of a quantity $q(\theta)$ associated with this sequence of models, a loss function $\rho(\hat{q}_n, q(\theta))$, and a rate of convergence α_n tending to 0.

Definition 1. *The maxiset associated with the sequence $\hat{q}_n$, the loss function ρ, the rate α_n and the constant T is the set*

$$MS(\hat{q}_n, \rho, \alpha_n)(T) = \left\{ \theta \in \Theta, \sup_n \mathrm{E}_\theta^n \rho(\hat{q}_n, q(\theta))(\alpha_n)^{-1} \leqslant T \right\}.$$

In various parametric cases, we can easily prove in regular sequences of models that we have

$$MS(\hat{q}_n, \rho, n^{-1/2})(T) = \Theta$$

for various homogenous loss functions and sufficiently large constant T. Although it might be useful and interesting to further investigate those domains where the rate is different from $n^{-1/2}$ (domains of superefficiency, or underefficiency), we will focus, in this paper, on the nonparametric situation. Instead of a priori fixing a (functional) set such as a Hölder, Sobolev or Besov ball, as in the case of the minimax framework, we choose to situate the problem in a very wide context: the parameter set Θ can be very large, for example the set of bounded, measurable functions. Then the functional set (maxiset) is associated with the procedure in a genuine way. Let us start with two examples.

2.1. Density estimation: kernel methods

Let $X_1, \ldots, X_n$ be n independent and identically distributed random variables with density f. We wish to estimate f. Let us fix $2 \leqslant p < \infty$, and investigate the problem with L_p loss: that is, for a procedure $\hat{f}_n$,

$$\rho(\hat{f}_n, f) = \|\hat{f}_n - f\|_p^p.$$

We take as our set of parameters Θ the set of all densities included in a (large) L_p ball. This is reasonable, given our choice for the loss function. We will investigate the maxisets of the following sequence of kernel procedures:

$$\hat{E}_{j(n)}(x) := \frac{1}{n}\sum_{i=1}^{n} E_{j(n)}(x, X_i).$$

$E(u, v)$ is a kernel $E_j(u, v) = 2^j E(2^j u, 2^j v)$. Typically, E_j will be the projector onto the space V_j of a multiresolution analysis (i.e. $E(u, v) = \sum_{k\in\mathbb{Z}} \phi(u-k)\phi(v-k)$) or the convolution $E(u, v) = E(u - v)$. The sequence $j(n)$ is increasing, $2^{j(n)} = n^{(1-\alpha)}$, with $\alpha \in (0, 1)$ (see Kerkyacharian and Picard (1992).

Let $B_{s,p,q}$ denote a Besov space and $B_{s,p,q}(M)$ the associated ball of radius M (for the definition and properties of Besov spaces, see Meyer 1990; Nikol'skii 1975).

Then a consequence of Theorem 2.1 in Kerkyacharian and Picard (1993) is the following result (see also Härdle *et al*. 1998, Chapter 10):

Proposition 1. *Suppose that the following conditions hold:*

- *E is compactly supported.*
- *$\int (y-x)^k E(x, y)\mathrm{d}y = \delta_{0,k}$, for all $k = 0, 1, \ldots, N$.*

- $E \circ E_j = E_j \circ E$, *for all* $j \geqslant 0$ (*where* $E \circ E_j$ *stands for the composition of* E *and* E_j).
- $x \to \mathrm{E}(x, y)$ *is* N *times continuously differentiable.*
- $\alpha_n := (2^{j(n)}/n)^{p/2} = n^{-\alpha p/2}$.

Then

$$MS(\hat{E}_{j(n)}, f, \alpha_n) :=: \Theta \cap B_{s,p\infty} \text{ with } s = \frac{\alpha}{2(1-\alpha)} \text{ or } \alpha = \frac{2s}{1+2s}.$$

Here, and throughout the paper,

$$MS(\hat{E}_{j(n)}, f, \alpha_n) :=: \Theta \cap B_{s,p\infty} \tag{1}$$

means that:

(i) for any T, there exists M such that $MS(\hat{E}_{j(n)}, f, \alpha_n)(T) \subset \Theta \cap B_{s,p,\infty}(M)$;
(ii) for any M, there exists T such that $MS(\hat{E}_{j(n)}, f, \alpha_n)(T) \supset \Theta \cap B_{s,p,\infty}(M)$.

2.2. White noise model: wavelet thresholding

Let us examine the differential equation

$$\mathrm{d}Y_t^n = f(t)\mathrm{d}t + \frac{1}{\sqrt{n}}\mathrm{d}W_t, \qquad t \in [0, 1], \tag{2}$$

where W_t is a standard Brownian motion on [0, 1]. Our aim is again to estimate f. Let us fix $1 < p < \infty$, and investigate, as in Section 2.1, the problem with L_p loss. Let us fix, as above, $\alpha \in (0, 1)$. We take as our set of parameters Θ a ball in the space $B_{\alpha/2,p,\infty}$. This set corresponds to the idea of minimal regularity which is always necessary for nonlinear procedures. Notice that $\alpha/2$ is always smaller (and often much smaller) than s introduced in Proposition 1. In particular, if $\alpha < 2/p$, Θ contains discontinuous functions. For a scaling function ϕ and a wavelet ψ, let us define the following sequence of procedures:

$$\hat{f}^T(x) = \sum_{0 \leqslant j \leqslant J_n} \sum_k \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geqslant K t_n \phi\} g_{j-1,k}(x),$$

where

$$g_{jk} = \psi_{jk}, \qquad \text{for } j \geqslant 0, \qquad g_{-1,k} = \phi_{0k},$$

$$\hat{\beta}_{jk} = \int_{[0,1]} g_{j-1,k}(x)\mathrm{d}Y_n(x),$$

$$t_n = \sqrt{\frac{\log n}{n}}, \qquad 2^{J_n} \leqslant t_n^{-2} \leqslant 2^{J_n+1}.$$

Let us introduce the following sets of functions:

$$\mathcal{W}^*(p, q)(M) = \left\{ f \in L_p,\ \sup_{\lambda>0} \lambda^q \sum_{j\geqslant 0} 2^{j[(p/2)-1]} \#\{k, |\beta_{jk}| > \lambda\} \leqslant M^q \right\}. \tag{3}$$

Obviously, for $p = 2$, $\mathcal{W}^*(2, q)(M)$ selects the functions such that their total number of βs greater (in modulus) than λ for $j \geqslant 0$ is less than $(M\lambda^{-1})^q$. For $p > 2$, we also 'count' the βs greater than λ, but incur a penalty for large j. These spaces prove to have a special importance in approximation theory (Cohen *et al*. 2000), coding (Donoho 1996) and estimation (Donoho and Johnstone, 1996). Then a consequence of Theorem 7 in Cohen *et al*. (1999) is the following result:

Proposition 2. *For* $a_n = (t_n)^{\alpha p/2} = (\log n/n)^{\alpha p/2}$,

$$MS(\hat{f}^T, f, \alpha_n) :=: \Theta \cap \mathcal{W}^*(p, 1-\alpha)p).$$

3. Local oracle inequalities and maxisets

This section is divided into two parts. Section 3.1 is essentially concerned with oracle inequalities. In particular, we begin with the standard case of L_2 oracle inequalities. Then we explain how to overcome the difficulty of generalizing to other norms and to local inequalities. This part is a priori essentially unconnected to the previous section on maxisets. The relations between the two notions are clarified in Section 3.2, where the consequences of local oracle inequalities in term of maxisets are studied.

We again consider a sequence of models $\mathcal{E}_n$ in which we estimate a function f defined on $\mathcal{X} \mapsto \mathbb{R}$. $\mathcal{X}$ is a measurable space equipped with a measure μ, such that $\mu(\mathcal{X}) < \infty$. The most common example of $\mathcal{X}$ is $[0, 1]$ or $[0, 1]^d$ equipped with the Lebesgue measure. f is assumed to belong to some basic functional space $\mathcal{V}$ (e.g. $\mathcal{V} = L_p$). We consider a sequence of linear operators E_j, $j \geqslant 0$, associated with any measurable function f, defined on $(\mathcal{X}, \mu)$, a measurable function $E_j f$. Typically, as above, E_j will be the projector onto the space V_j of a multiresolution analysis $(E_j f(x) = \int \sum_{k\in\mathbb{Z}} 2^j \phi(2^j u - k)\phi(2^j x - k) f(u)\mathrm{d}u)$, or the convolution $E_j f(x) = \int 2^j E(2^j(u-x)) f(u)\mathrm{d}u$.

3.1. From L_2 oracle inequalities to local ones

3.1.1. L_2 oracle inequalities

Following Donoho and Johnstone (1994), we say that the estimate $\hat{f}$ satisfies an L_2 *oracle inequality* with the class $\mathcal{C}$ of estimators at rate c_n if, for all $n \geqslant 1$,

$$\mathrm{E}_n \|\hat{f} - f\|_2^2 \leqslant c_n \quad \inf\{\mathrm{E}_n \|\hat{\Phi} - f\|_2^2,\ \hat{\Phi} \in \mathcal{C}\}. \tag{4}$$

Inequality (4) says precisely that, up to rate c_n, $\hat{f}$ behaves as the oracle estimate of the class $\mathcal{C}$ – that is, the best estimate among the class $\mathcal{C}$ (as if an oracle was telling us for each function which estimator was to be chosen). c_n measures the loss of efficiency of $\hat{f}$ compared to the oracle estimator (which generally is not an estimator since the optimal choice may depend heavily on the function being estimated).

As a prototype example, it can easily be proved that the wavelet thresholding estimator in the white noise model satisfies an L_2 oracle inequality with the class $\{\hat{E}_j^n = \int E_j(x, t)\mathrm{d}Y_t^n, j \geqslant 0\}$ of estimators at the rate

$$c_n = 1 + \log n,$$

if the E_j are the projections on the V_j.

Hence, we immediately see that oracle inequalities may be a very useful property for a procedure. However, it seems that hitherto there has been no full agreement in the statistical community about the most suitable distance for reflecting the visual properties of estimation procedures. In particular, two functions may look very different although they are very close in L_2 norm. As a consequence, it is natural to ask whether we can also prove oracle inequalities for different norms (L_p, for instance), as well as oracle inequalities at a point. Let us first observe that an oracle inequality of type (4) gives us information about the quality of the procedure for L_p norms, with $1 \leqslant p \leqslant 2$, because of the finiteness of the measure μ. However, it does not tell us anything about the other norms. To be able to consider oracle inequalities for general L_p norms, it is more convenient to have a slightly different understanding of (4).

Let us evaluate in the prototype example above (still in the white noise model) the quantity

$$\inf\{\mathrm{E}_n\|\hat{E}_j^n - f\|_2^2, j \geqslant 0\}.$$

Standard calculations give

$$\mathrm{E}_n\|\hat{E}_j^n - f\|_2^2 = c\frac{2^j}{n} + \|E_j f - f\|_2^2.$$

Hence we observe the standard trade-off between an increasing and a decreasing quantity. This last quantity is decreasing in j because we use the L_2 norm and a family of projection operators on increasing subspaces. Precisely this will be the difficulty when we want to extend to other situations. Let us introduce

$$j_\lambda(f) := \inf\{j \in \mathbb{N}, 2^{-j/2}\|E_j f - f\|_2 \leqslant \lambda\}.$$

So, for $\lambda > 0$ we have, for $j \geqslant 1$,

$$j_\lambda(f) = j \Leftrightarrow 2^{-(j-1)/2}\|E_{j-1}f - f\|_2 > \lambda \geqslant 2^{-j/2}\|E_j f - f\|_2,$$

and, for $j = 0$,

$$j_\lambda(f) = 0 \Leftrightarrow \lambda \geqslant \|E_0 f - f\|_2.$$

Setting

$$\lambda_n = \left(\frac{c}{n}\right)^{1/2}, \qquad j_n^* = j_{\lambda_n}(f), \tag{5}$$

it is not difficult to prove that

$$\frac{c\,2^{j_n^*}}{2n} \leqslant \inf\{\mathrm{E}_n\|\hat{E}_j^n - f\|_2^2,\ j \geqslant 0\} \leqslant \frac{2c\,2^{j_n^*}}{n},$$

using the following lemma.

Lemma 1. *Let $\{a_j\}$ and $\{b_j\}$, $j \in \mathbb{N}$, be two sequences, the former non-increasing and the latter non-negative and non-decreasing. Let $j^* = \inf\{j \in \mathbb{N},\ a_j \leqslant b_j\}$. Then*

$$b_{j^*-1} \leqslant \inf\{j \in \mathbb{N},\ a_j + b_j\} \leqslant 2b_{j^*}.$$

(*By convention* $b_{-1} = b_0$.)

Proof. Clearly $b_{j^*-1} < a_{j^*-1}$, if $j^* > 0$, and $b_{j^*} \geqslant a_{j^*}$. So

$$\inf\{j \in \mathbb{N},\ a_j + b_j\} \leqslant a_{j^*} + b_{j^*} \leqslant 2b_{j^*}$$

On the other hand,

$$j \geqslant j^* \Rightarrow a_j + b_j \geqslant b_j \geqslant b_{j^*} \geqslant b_{j^*-1}.$$

Further,

$$0 \leqslant j < j^* \Rightarrow a_j + b_j \geqslant a_j \geqslant a_{j^*-1} > b_{j^*-1}. \qquad \square$$

We observe that obviously (as j_n^* strongly depends on f) $\hat{E}_{j_n^*}^n$ is not a true estimator. Hence, without loosing much with respect to (4), we define the following property:

Definition 2. *We say that $\hat{f}$ satisfies an oracle inequality for the L_2 norm, on the space $\mathcal{V}$, the class E_j of estimators and at the rate $c_n = 1 + \log n$ if*

$$\mathrm{E}_n\|\hat{f} - f\|_2^2 \leqslant Cc_n 2^{j_n^*}\lambda_n^2, \qquad \forall f \in \mathcal{V}, \tag{6}$$

where the sequence j_n^*, defined in (5), reflects the complexity of the function with respect to the sequence E_j.

3.1.2. L_p oracle inequalities associated with a sequence of operators E_j

We begin by defining the L_p analogue of $j_\lambda(f)$. Let $F(f)(j)$ be a non-negative, non-increasing functional defined on $\mathbb{N}$. An important example is

$$\tilde{F}(f)(j) := \sup_{j' \geqslant j} 2^{-j'/2}\|E_{j'}f - f\|_p. \tag{7}$$

Now let

$$j_\lambda^F(f) := \inf\{j \in \mathbb{N},\ F(f)(j) \leqslant \lambda\}.$$

So, for $j \geqslant 1$,

$$j_\lambda^F(f) = j \Leftrightarrow F(f)(j-1) > \lambda \geqslant F(f)(j),$$

and, for $j = 0$,

$$j_\lambda^F(f) = 0 \Leftrightarrow \lambda \geqslant F(f)(0).$$

Again, let us define

$$\lambda_n = \left(\frac{1}{n}\right)^{1/2}, \qquad j_n^F = j_{\lambda_n}^F(f). \tag{8}$$

This leads to the following definition:

Definition 3. *For $1 \leqslant p < \infty$, we say that $\hat{f}$ satisfies an L_p oracle equality on $\mathcal{V}$, associated with a sequence of operators E_j and the functional F, at rate $c_n = 1 + \log n$ if the following inequalities are true for all $n \geqslant 1$:*

$$\mathrm{E}_n \|\hat{f} - f\|_p^p \leqslant C c_n (2^{j_n^F/2} \lambda_n)^p, \qquad \forall f \in \mathcal{V}; \tag{9}$$

$$\|E_{j_\lambda^F(f)} f - f\|_p^p \leqslant C' (2^{j_\lambda^F(f)/2} \lambda)^p, \qquad \forall f \in \mathcal{V}, \forall \lambda > 0. \tag{10}$$

Remarks. First, this definition easily generalizes to the case $p = \infty$, with the usual modification consisting of ignoring all the pth powers in (9) and (10). The inequalities also are embedded: because of the finiteness of the mesure μ, satisfying an L_p oracle inequality implies satisfying an L_q oracle inequality for any $1 \leqslant q \leqslant p$.

Second, inequality (10) is obvious in the case where $F(f) = \tilde{F}(f)$ is defined by (7). In fact, if $F(f) \neq \tilde{F}(f)$, this inequality is needed to establish a relation between $F(f)$ and the approximation properties of the sequence $E_j f$.

Finally, if we compare (9) with (6), we notice that the two right-hand sides are equivalent. If we now compare (9) with (4), we cannot deny that there might be a loss, since the only thing that can be said is that there exists C with

$$\inf\{\mathrm{E}_n \|\hat{E}_j^n - f\|_p^p, j \geqslant 0\} \leqslant C (2^{j_n^F/2} \lambda_n)^p.$$

For $p = 2$, the two quantities were of the same order. For $p \neq 2$, as we are considering L_p norms, we can only hope that they do not differ much – and also observe that this is confirmed by the minimax rates for standard classes of functions.

3.1.3. Local oracle inequalities associated with a sequence of operators E_j

We will mimic locally what has been done above. Let $F(f)(j, x)$ be a non-negative functional defined on $\mathbb{N} \times \mathcal{X}$, such that, for μ-almost every x, $j \to F(f)(j, x)$ is non-increasing. We also suppose that, for μ-almost every x, $F(f)(0, x) < \infty$. An important example is

$$\tilde{F}(f)(j, x) = \sup_{j' \geqslant j} 2^{-j'/2} |\mathrm{E}_{j'} f(x) - f(x)|. \tag{11}$$

F is now a 'local' functional. Let

$$j_\lambda^F(f, \mathrm{x}) = \inf\{j \in \mathbb{N}, F(f)(j, x) \leqslant \lambda\}.$$

Now let

$$t_n = \left(\frac{\log n}{n}\right)^{1/2} \quad \text{and} \quad j_n^*(x) = j_{t_n}^F(f, x). \tag{12}$$

For practical reasons, it is generally necessary also to introduce a stopping sequence J_n tending to infinity, reflecting the fact that, in practice, a procedure will never be able to consider an infinite number of possible bandwidths.

Definition 4. *Let $p \geqslant 1$ be fixed. We say that the sequence of estimators $\hat{f}_n$ satisfies a local oracle inequality of order p on $\mathcal{V}$ associated with a sequence of operators E_j, the 'local' functional F and the stopping sequence J_n, if the following inequalities hold for all $n \geqslant 1$:*

$$\mathrm{E}_n|\hat{f}_n(x) - f(x)|^p \leqslant C\{(2^{j_n^*(x)/2}t_n)^p + |E_{j_n^*(x)}f(x) - f(x)|^p + |E_{J_n}f(x) - f(x)|^p\} \qquad \forall x \in \mathcal{X}, \forall f \in \mathcal{V}, \tag{13}$$

$$\left\| \sup_{j' \geqslant j} |E_{j'}f - f| I\{j_\lambda^F(f, \cdot) = j\} \right\|_p^p \leqslant C'(2^{j/2}\lambda)^p \mu\{x, j_\lambda^F(f, x) = j\} \qquad \forall \lambda > 0, \forall j \geqslant 0, \forall f \in \mathcal{V}, \tag{14}$$

where $I\{A\}$ denotes the indicator function of the set A.

Remarks. If we omit the terms depending on J_n, and again compare (13) with (9), besides the localization of the inequality, we notice two differences. The first is the presence of the term $|E_{j_n^*(x)}f(x) - f(x)|^p$, which was not in (9). However, we could have added a similar term to (9) without changing the rates of convergence, because of (10). The second difference is that a logarithmic factor now appears in the rate t_n, where t_n replaces λ_n, while the logarithmic term c_n has disappeared.

Also, if we now compare (14) with (10), we see that we require here a local comparison between $F(f)(j, x)$ and $\sup_{j' \geqslant j}|E_{j'}f(x) - f(x)|I\{x, j_\lambda^F(f, x) = j\}$ instead of a global one. However, this comparison is made after averaging, that is, in a somewhat mild way.

The following definition corresponds to letting p tend to infinity:

Definition 5. *We say that the sequence of estimators $\hat{f}_n$ satisfies an 'exponential' oracle inequality on $\mathcal{V}$ associated with a sequence of operators E_j, the stopping sequence J_n and the 'local' functional F if there exist C, C' v_0, λ_0, such that the following inequalities hold for all $n \geqslant 1$ and all $f \in \mathcal{V}$:*

$$P_n\left\{(2^{j/2}t_n)^{-1} \sup_{x, j_n^*(x) = j} |\hat{f}_n(x) - f(x)| \geqslant \lambda\right\} \leqslant C \exp\left\{-\frac{\lambda^2}{2v_0}\right\} \qquad \forall \lambda \geqslant \lambda_0, \forall J_n \geqslant j \geqslant 0, \tag{15}$$

$$\|\sup_{j' \geqslant j} |E_{j'}f - f| I\{j_\lambda^F(f, \cdot) = j\}\|_\infty \leqslant C'(2^{j/2}\lambda) \qquad \forall \lambda \geqslant 0, \forall j \geqslant 0. \tag{16}$$

This oracle condition is of course much stronger than the previous ones. Using the fact that a sub-Gaussian random variable has moments of any order, we deduce that satisfying an 'exponential' oracle condition implies satisfying a local oracle of order p for any $p \geqslant 1$, especially since $\mu(\mathcal{X}) < \infty$.

3.2. Local oracle inequalities and maxisets

Let us begin with some definitions of sets which will be connected later to maxisets:

3.2.1. Besov bodies

Let us write, for $\gamma > 0$, $r > 0$,

$$\mathcal{B}_{\gamma,r,\infty}(M) = \{f \in \mathcal{V}, \|E_j f - f\|_{L_r(d\mu)} \leqslant M2^{-j\gamma}, \forall j \geqslant 0\}.$$

Though obviously depending on the sequence of kernels E_j, $\mathcal{B}_{\gamma,r,\infty}$ is deliberately referred to as a 'Besov body'. The reason is that, in fact, these spaces coincide for a large variety of kernels E_j (for instance, projectors on a multiresolution analysis, or translation kernels, with standard cancellation of the first moments; see Meyer 1990). In these cases, the balls also coincide with the standard Besov balls. Of course, we can also generalize the definition above with

$$\mathcal{B}_{\gamma,r,m}(M) = \left\{ f \in \mathcal{V}, \sum_{j \geqslant 0} (2^{j\gamma} \|E_j f - f\|_{L_r(\mathrm{d}\mu)})^m \leqslant M^m \right\}$$

3.2.2. Weak Besov bodies

Recall the definition of Lorentz spaces (also called weak L_q spaces or Marcinkiewicz spaces), for $q > 0$ and ν a non-negative measure:

$$L_{q,\infty}(\nu) = \left\{ g; \sup_{\lambda > 0} \lambda^q \nu\{|g| \geqslant \lambda\} < \infty \right\}.$$

Let us introduce the following measure on $\mathbb{N} \times \mathcal{X}$:

$$\nu_p = \sum_{j \geqslant 0} 2^{jp/2} \delta_j \otimes \mu,$$

where δ is the Dirac measure. For F a non-negative functional defined on $\mathbb{N} \times \mathcal{X}$ (see Section 3.1.1), suppose, for $p > q > 0$, that

$$\begin{aligned} \mathcal{W}(F)(p, q) &= \{f \in \mathcal{V}, F(f) \in L_{q,\infty}(\nu_p)\} \\ &= \{f \in \mathcal{V}, \sup_{\lambda > 0} \lambda^q \sum_{j \geqslant 0} 2^{j(p/2)} \mu\{x, F(f)(j, x) > \lambda\} < \infty\}, \end{aligned}$$

with associated ball

$$\mathcal{W}(F)(p, q)(M) = \left\{ f \in \mathcal{V}, \sup_{\lambda > 0} \lambda^q \sum_{j \geq 0} 2^{j(p/2)} \mu\{x, F(f)(j, x) > \lambda\} \leq M^q \right\}$$

We investigate two examples. For the first we consider the functional

$$F^1(f)(j, x) = 2^{-j/2} |E_j f(x) - f(x)|, \tag{18}$$

not necessarily monotone, and its associated ball $\mathcal{W}(F^1)(p, q)(M)$. Let $\alpha \in (0, 1)$, $q = p(1-\alpha)$. The following lines prove, using Markov inequality, that if f belongs to $\mathcal{B}_{\gamma,q,q}(M)$, and $\gamma = \alpha/(2(1-\alpha))$, then f belongs to $\mathcal{W}(F^1)(p, q)(M)$. Hence, in this case, $\mathcal{W}(F^1)(q/(1-\alpha), q)(M)$ appears as a weak analogue of $\mathcal{B}_{\gamma,q,q}(M)$. We summarize this fact in the inclusion

$$\mathcal{B}_{\alpha/(2(1-\alpha)),q,q}(M) \subset \mathcal{W}(F^1)\left(\frac{q}{1-\alpha}, q\right)(M).$$

Indeed, we have, as $p - q = \alpha q/(1-\alpha)$,

$$\begin{aligned}
\sum_{j \geq 0} 2^{j(p/2)} \mu\{x, F(f)(j, x) > \lambda\} &= \sum_{j \geq 0} 2^{j(p/2)} \mu\{x, 2^{-j/2} |E_j f(x) - f(x)| > \lambda\} \\
&\leq \sum_{j \geq 0} 2^{j(p/2} (\lambda 2^{j/2})^{-q} \|E_j f - f\|^q_{L_q(\mathrm{d}\mu)} \\
&\leq \lambda^{-q} M^q.
\end{aligned}$$

If ψ is a wavelet, and β_{jk} denotes the wavelet coefficient of f ($\beta_{jk} = \int f \psi_{jk}$), and $\chi_{jk}(x) = 2^{j/2} I\{2^j x - k \in [0, 1]\}$ is the Haar scaling function, let us consider, by way of a second example, the case of the functional

$$F^2(f)(j, x) = 2^{-j/2} \sum_k |\beta_{jk}| \chi_{jk}(x),$$

also not necessarily non-increasing, and its associated ball $\mathcal{W}(F^2)(p, q)(M)$. We notice that $\mathcal{W}(F^2)(p, q)(M)$ coincides with the set $\mathcal{W}^*(p, q)(M)$ introduced in (3).

In Section 6, we investigate in greater depth the weak Besov bodies for some classes of local functionals F. In particular, we establish that they happen to coincide rather often. For instance, we prove that if the E_j of (18) are projections on the spaces V_j, then $\mathcal{W}(F^1)(p, q)$ and $\mathcal{W}(F^2)(p, q)$ are equal.

3.2.3. *Local oracle inequalities and maxisets*

We consider a sequence of estimates $\hat{f}_n$ associated with a sequence of models $\mathcal{E}_n$. Let us, as above, define the maxiset associated with the sequence $\hat{f}_n$, the L_p loss and the rate $(t_n)^{\alpha p}$ (recall that $t_n = (\log n/n)^{1/2}$):

$$MS(\hat{f}_n,\ p,\ \alpha)(T) = \left\{ f \in \mathcal{V},\ \sup_n \mathrm{E}_n \|\hat{f}_n - f\|^p_{L_{p(\mathrm{d}\mu)}} (t_n)^{-\alpha p} \leqslant T \right\}.$$

The following proposition establishes a natural correspondence between the previous local oracle inequalities and maxisets:

Proposition 3. *Let β be a positive constant. Suppose that the sequence of estimates $\hat{f}_n$ satisfies a local oracle inequality of order p, associated with the sequence of operators E_j, the sequence J_n and the local functional F on the space $\mathcal{V} = \mathcal{B}_{\alpha/\beta, p, \infty}(M)$. Then, if J_n is such that $2^{J_n} < t_n^{-\beta} \leqslant 2^{J_n+1}$, then there exists M' such that*

$$\mathcal{W}(F)(p,\ p(1-\alpha))(M') \subset MS(\hat{f}_n,\ p,\ \alpha)(T).$$

Remarks. The constant M' may be chosen such that

$$T = C[2(1 + 2M'^{\,p(1-\alpha)}) + M^{p(1-\alpha)}],$$

with C is from inequality (13).

Membership of the set $\mathcal{W}(F)(p,\ q)(M)$, with $q = p(1-\alpha)$, may also be written in the following way. Let ν_0 be the measure on $\mathbb{N} \times \mathcal{X}$ defined, as above, by the formula

$$\nu_0 = \left(\sum_{j \in \mathbb{N}} \delta_j \right) \otimes \mu.$$

Then f belongs to the set $\mathcal{W}(F)(p,\ p(1-\alpha))(M)$ if and only if

$$\sup_{\lambda > 0} \lambda^{(1-\alpha)} \| 2^{(j/2)} I\{F(f)(j,\ x) > \lambda\} \|_{L_p(\nu_0)} \leqslant M^{(1-\alpha)}.$$

In this way, it is easier to let p tend to infinity. One can prove that we obtain as a limit, when $F(f) = \tilde{F}(f)$, $\mathcal{B}_{\alpha/(2(1-\alpha)), \infty, \infty}(M)$. This introduces the following proposition, corresponding to the case $p = \infty$, $F(f) = \tilde{F}(f)$.

Proposition 4. *Suppose that the sequence of estimates $\hat{f}_n$ satisfies the exponential oracle inequality associated with the sequence J_n and the local functional $\tilde{F}$* on *$\mathcal{V} = \mathcal{B}_{\alpha/\beta, \infty, \infty}(M)$. Then, if $2^{J_n} < t_n^{-\beta} \leqslant 2^{J_n+1}$ and*

$$MS(\hat{f}_n,\ \infty,\ \alpha)(T) = \left\{ f \in \mathcal{V},\ \sup_n \mathrm{E}_n \|\hat{f}_n(x) - f(x)\|_\infty (t_n)^{-\alpha} \leqslant T \right\}$$

for $\alpha < 1$, there exists M', such that

$$\mathcal{B}_{\alpha/(2(1-\alpha)), \infty, \infty}(M') \subset MS(\hat{f}_n,\ \infty,\ \alpha)(T).$$

Remark. The constant M' may be chosen such that

$$T = C(2 + \sqrt{2}) M'^{(1-\alpha)},$$

with C from inequality (15).

Proof of Proposition 3. Let f be arbitrary in $\mathcal{V} = \mathcal{B}_{\alpha/\beta,p,\infty}(M)$, $q = p(1-\alpha)$. Then

$$\begin{aligned}
\mathrm{E}_n\int|\hat{f}_n(x) - f(x)|^p \mathrm{d}\mu &= \sum_{j\geqslant 0}\mathrm{E}_n\int|\hat{f}_n(x) - f(x)|^p I\{j_n^*(x) = j\}\mathrm{d}\mu \\
&\leqslant C\sum_{j\geqslant 0}\int\{(2^{j/2}t_n)^p + |E_j f(x) - f(x)|^p \\
&\quad + |E_{J_n}f(x) - f(x)|^p\}I\{j_n^*(x) = j\}\mathrm{d}\mu \\
&\leqslant Ct_n^p\left\{2\sum_{j\geqslant 0}2^{jp/2}\mu\{j_n^*(x) = j\} + \|E_{J_n}f - f\|_p^p\right\} \\
&\leqslant 2C(1 + 2M'^q)t_n^{p-q} + CM^p 2^{-(\alpha/\beta)J_n p} \leqslant C(M)t_n^{\alpha p}.
\end{aligned}$$

We have used the definition of $\mathcal{V} = \mathcal{B}_{\alpha/\beta,p,\infty}(M)$ and the decomposition

$$\begin{aligned}
\sum_{j=0}^{\infty}\mu\{j_n^*(x) = j\}2^{jp/2} &= \mu\{j_n^*(x) = 0\} + \sum_{j=1}^{\infty}\mu\{j_n^*(x) = j\}2^{jp/2} \\
&\leqslant \mu\{F(f)(0, x) \leqslant t_n\} + \sum_{j=1}^{\infty}2^{jp/2}\mu\{F(f)(j-1, x) > t_n\} \\
&\leqslant \mu\{\mathcal{X}\} + 2\sum_{j=0}^{\infty}2^{jp/2}\mu\{F(f)(j, x) > t_n\} \\
&\leqslant \mu\{\mathcal{X}\} + 2\nu_p\{F(f) > t_n\} \\
&\leqslant \mu\{\mathcal{X}\} + 2\|F(f)\|^q_{L_{q,\infty}(\nu_p)}t_n^{-q} \\
&\leqslant \mu\{\mathcal{X}\} + 2M'^q t_n^{-q} \leqslant M'' t_n^{-q},
\end{aligned}$$

as $t_n \leqslant 1$ and $\mu\{\mathcal{X}\}$ is finite. □

Proof of Proposition 4. Because of the definition of $\mathcal{B}_{\alpha/(2(1-\alpha)),\infty,\infty}(M)$, we obtain that, for all j', $2^{-j'/2}|E_{j'}f(x) - f(x)| \leqslant M2^{-j'[1/2(1-\alpha)]}$. Hence, if we recall the definition of $j_n^*(x)$ (see (12)) and the fact that we use the functional $\tilde{F}$ to define $j_n^*(x)$, we find that we must have $j_n^*(x) \leqslant j_0$ such that $2^{j_0} \leqslant Ct_n^{-2(1-\alpha)}$.

Using the same argument, we see that the condition $f \in \mathcal{V}$ ensures that $J_n \leqslant j_0$.

By the two previous paragraphs, we obtain:

$$
\begin{aligned}
E_n\|\hat{f}_n(x) - f(x)\|_\infty &= E_n \sup_{j\geqslant 0}\sup_x |\hat{f}_n(x) - f(x)| I\{j_n^*(x) = j\} \\
&= E_n \sup_{0\leqslant j\leqslant j_0}\sup_x |\hat{f}_n(x) - f(x)| I\{j_n^*(x) = j\} \\
&\leqslant \sum_{0\leqslant j\leqslant j_0} E_n \sup_x |\hat{f}_n(x) - f(x)| I\{j_n^*(x) = j\} \\
&\leqslant \sum_{0\leqslant j\leqslant j_0} C2^{j/2} t_n \leqslant Ct_n^\alpha. \qquad \square
\end{aligned}
$$

4. Applications to wavelet thresholding and Lepski's procedures

In this section, we prove a local oracle inequality for wavelet thresholding, as well as for adaptive local bandwidth selection. We first state the assumptions on the sequence of models (which will be roughly the same for the different procedures). We exhibit standard examples where these conditions are fulfilled, and then prove the local oracle inequalities. It is interesting to notice that the associated functionals F^T and F^L are different. F^L is very easy to understand since it is, up to a constant, our standard example $\tilde{F}$ introduced in (7). Surprisingly, F^T is much less intuitive since it requires the introduction of the maximal function (see (29)). We also investigate the behaviour of a hybrid procedure, intermediate between the two previous ones, which turns out to be very efficient for the construction of confidence intervals (see Picard and Tribouley 2000).

4.1. Assumptions on the sequence of models

Our assumptions on the sequence of models will only be relevant with respect to its ability to estimate the $E_j f$. Let $p \geqslant 1$ be fixed. Let us also fix a constant $K \geqslant 4$ and an increasing sequence of integers J_n. The latter quantities will appear as tuning quantities for both procedures.

Moreover, we assume that there exist a sequence of estimates $\hat{E}_j^n$, and a sequence of class of distributions $\mathcal{C}_j$ such that $\int(\hat{E}_j^n - E_j f)\mathrm{d}\delta$ is defined for any $\delta \in \mathcal{C}_j$. In what follows, $\mathcal{C}_j$ will be:

- either the class $\mathcal{C}^D$ of all the Dirac masses of $\mathcal{X}$, for any j, and then $\int(\hat{E}_j^n - E_j f)\mathrm{d}\delta_x = \hat{E}_j^n(x) - E_j f(x)$. This case will be concerned with local bandwidth selection.
- or the class $\mathcal{C}_j^W$ of measures δ_{jk} with density functions $2^{j/2} g_{j-1,k}$, $k \in \mathbb{Z}$, associated with a pair (ϕ, ψ) of father and mother wavelets in the way defined above: $g_{jk} = \psi_{j,k}$, for $j \geqslant 0$, $g_{-1k} = \phi_{0k}$. Here the E_j are projection kernels on the space V_{j-1}, for $j \geqslant 1$, $E_1 = E_0$, and $\int E_j f \mathrm{d}\delta_{jk} = 2^{j/2}\beta_{jk}$. We define $\int \hat{E}_j^n \mathrm{d}\delta_{jk} = 2^{j/2}\hat{\beta}_{jk}$, and then $\int(\hat{E}_j^n - E_j f)\mathrm{d}\delta_{jk} = 2^{j/2}(\hat{\beta}_{jk} - \beta_{jk})$. This case will obviously be concerned with thresholding.

In either case, we assume there to exist: some constant C_1 such that, for all $n \geqslant 1$, for all $\delta \in \mathcal{C}_j$, for all j, $0 \leqslant j \leqslant J_n$,

$$\mathrm{E}_n \left| \int (\hat{E}_j^n - E_j f) \mathrm{d}\delta \right|^{2p} \leqslant C_1 \left(\frac{2^j}{n} \right)^p; \tag{19}$$

and some $\gamma > 0$ and some C_2 such that, for all $n \geqslant 1$, for all $\delta \in \mathcal{C}_j$, for all j, $0 \leqslant j \leqslant J_n$,

$$P_n \left(\left| \int (\hat{E}_j^n - E_j f) \mathrm{d}\delta \right| \geqslant \frac{K 2^{j/2}}{4} \left(\frac{\log n}{n} \right)^{1/2} \right) \leqslant C_2 \left(\frac{1}{n} \right)^{\gamma}. \tag{20}$$

Remarks. The expectation (19) and probability (20) are taken when f is the true parameter. Notice also that if condition (19) holds for one value of p, it automatically holds for any $p' \leqslant p$.

It is worth noting that neither (19) nor (20) implies the other. However, it is easy to verify that the following condition implies both for any $p \geqslant 1$: there exists C_3, $v_0 > 0$, $\lambda_0 \geqslant 0$ such that, for all $n \geqslant 1$, for all $\delta \in \mathcal{C}_j$, for all j, $0 \leqslant j \leqslant J_n$, for all $\lambda \geqslant \lambda_0$,

$$P_n \left(\left| \int \hat{E}_j^n - E_j f \,\mathrm{d}\delta \right| \geqslant \frac{\lambda 2^{j/2}}{n^{1/2}} \right) \leqslant C_3 \exp\left\{ \frac{-\lambda^2}{2 v_0} \right\}. \tag{21}$$

4.1.1. Examples of models where such conditions are satisfied

Let us take the two examples where E_j is either a projection on V_j or a convolutor with bandwidth 2^{-j}. It is well known that in the following basic models, in which the classical kernel estimator and the Bernstein and Rosenthal inequalities are used for $\hat{E}_j^n$, conditions (19) and (20) are satisfied:

- white noise model (see section 2.2),

$$\mathrm{d}Y_t^n = f(t)\mathrm{d}t + \frac{1}{\sqrt{n}} \mathrm{d}W_t, \qquad t \in [0, 1]; \tag{22}$$

- equispaced regression model, with Gaussian errors,

$$Y_i = f\left(\frac{i}{n} \right) + \epsilon_i, \qquad i = 1, \ldots, n; \tag{23}$$

- density model (see Section 2.1),

$$Y_1, \ldots, Y_n \text{ i.i.d., with density } f. \tag{24}$$

However, with more elaborate arguments one can also prove that they are satisfied for stationary processes of spectral density f, evolutionary spectra (Neumann and von Sachs 1997), locally stationary processes (Mallat *et al.* (1998), partially observed diffusion models (Hoffmann 1999a; 1999b and 1999c) multivariate extenstions ($t \in [0, 1]^d$) (Donoho, 1997; Neumann, 1998).

4.2. Local bandwidth selection

The following procedure has been introduced by Lepski (1991) and can be found presented in this local version in Lepski *et al*. (1997). It is associated with a general sequence of operators E_j.

Let t_n be $(\log n/n)^{1/2}$ as before. We define the index $\hat{j}(x)$ as the minimum of the set of admissible js at the point x, where $j \in \{0, \ldots, J_n\}$ is admissible at the point x if

$$|\hat{E}^n_{j'}(x) - \hat{E}^n_{j''}(x)| \leqslant K2^{j''/2}t_n, \qquad \forall j', j''; \, j \leqslant j' \leqslant j'' \leqslant J_n.$$

We also define the estimate

$$\hat{f}^L(x) = \hat{E}^n_{\hat{j}(x)}(x).$$

The sequence J_n will again be chosen in such a way that $2^{J_n} < t_n^{-\beta} \leqslant 2^{J_n+1}$ for some positive constant β. Let

$$\mathcal{M} = \left\{ f, \limsup_j |E_j f(x) - f(x)| = 0 \ \mu\text{-a.e.} \right\}.$$

Proposition 5. *If conditions* (19) *and* (20) *are satisfied for some order* p^* *and* $\gamma > p^*\beta/2$ *and for the class* $\mathcal{C}^D$, *then* $\hat{f}^L$ *satisfies, for any* $1 \leqslant p \leqslant p^*$, *the local oracle conditions or order p of Definition* 4 *on the space* $\mathcal{V} = \mathcal{M}$, *associated with the sequence of operators* E_j *and the functional*

$$F^L(f)(j, x) := \frac{4}{K} \sup_{j' \geqslant j} 2^{-j'/2} |E_{j'} f(x) - f(x)|.$$

Notice, in particular, that the conditions of Proposition 3 for any $1 \leqslant p \leqslant p^*$ are fulfilled, since $\mathcal{B}_{\alpha/\beta, p, \infty}(M) \subset \mathcal{M}$. Moreover, the result of Proposition 5 holds for any $p \geqslant 1$ if (21) holds.

Proof. First, we observe that, as mentioned above, because of the precise form of $F^L(f)$, we only have to establish (13), as (14) is naturally fulfilled in this case. We recall (see Section 3.1.3) that $j_n^*(x) = j_{t_n}^{F^L}(x)$. Notice that $j_n^*(x)$ is finite since $f \in \mathcal{M}$.

We begin the proof with the following lemma.

Lemma 2. *Under the conditions above,*

$$P\{\hat{j}(x) > j_n^*(x)\} \leqslant \frac{CJ_n^2}{n^\gamma}.$$

Proof. Observe that, by the definition of $\hat{j}(x)$, when $\hat{j}(x) \geqslant 1$, $\hat{j}(x) - 1$ is not admissible (when $\hat{j}(x)$ is admissible), so there exists j', $\hat{j}(x) - 1 < j' \leqslant J_n$, such that $|\hat{E}^n_{j'}(x) - \hat{E}^n_{\hat{j}(x)-1}(x)| \geqslant Kt_n 2^{j'/2}$. In addition, if $\hat{j}(x) - 1 \geqslant j_n^*(x)$, we have

$$|E_{j'}f(x) - E_{\hat{j}(x)-1}f(x)| \leqslant |E_{j'}f(x) - f(x)| + |E_{\hat{j}(x)-1}f(x) - f(x)| \leqslant \frac{K}{2}t_n 2^{j'/2}.$$

So, on the set $\hat{j}(x) > j_n^*(x)$, we have that there exists j', $\hat{j}(x) - 1 < j' \leqslant J_n$, with

$$|\hat{E}_{j'}^n(x) - \hat{E}_{\hat{j}(x)-1}^n(x)| \geqslant Kt_n 2^{j'/2} \quad \text{and} \quad |E_{j'}f(x) - E_{\hat{j}(x)-1}f(x)| \leqslant \frac{K}{2}t_n 2^{j'/2}.$$

Hence, using (20),

$$\begin{aligned}
&P\{\hat{j}(x) > j_n^*(x)\}\\
&\quad = P\left\{\bigcup_{j_n^*(x)\leqslant j\leqslant j'\leqslant J_n}\left\{|\hat{E}_{j'}^n(x) - \hat{E}_j^n(x)| \geqslant Kt_n 2^{j'/2};\ |E_{j'}f(x) - E_jf(x)| \leqslant \frac{K}{2}t_n 2^{j'}\right\}\right\}\\
&\quad \leqslant \sum_{j,j'\leqslant J_n} P\left\{|\hat{E}_{j'}^n(x) - E_{j'}f(x) - \hat{E}_j^n(x) + E_jf(x)| \geqslant \frac{K}{2}t_n 2^{j'/2}\right\}\\
&\quad \leqslant \sum_{j,j'\leqslant J_n} P\left\{|\hat{E}_{j'}^n(x) - E_{j'}f(x)| \geqslant \frac{K}{4}t_n 2^{j'/2}\right\} + P\left\{|\hat{E}_j^n(x) - E_jf(x)| \geqslant \frac{K}{4}t_n 2^{j/2}\right\}\\
&\quad \leqslant \frac{CJ_n^2}{n^\gamma}
\end{aligned}$$

Let us now investigate the two different cases. In the case $j_n^*(x) \leqslant J_n$, we can divide $\mathrm{E}_n|\hat{f}^L(x) - f(x)|^p$ into two terms:

$$T_1 = \sum_{0\leqslant j'\leqslant j_n^*(x)} \mathrm{E}_n|\hat{E}_{j'}^n(x) - f(x)|^p I\{\hat{j}(x) = j'\}$$

$$T_2 = \sum_{j_n^*(x)<j'\leqslant J_n} \mathrm{E}_n|\hat{E}_{j'}^n(x) - f(x)|^p I\{\hat{j}(x) = j'\}$$

To bound T_1, we observe that

$$\begin{aligned}
&|\hat{E}_{j'}^n(x) - f(x)|^p\\
&\qquad \leqslant 3^{p-1}(|\hat{E}_{j'}^n(x) - \hat{E}_{j_n^*(x)}^n(x)|^p + |\hat{E}_{j_n^*(x)}^n(x) - E_{j_n^*(x)}f(x)|^p + |E_{j_n^*(x)}f(x) - f(x)|^p).
\end{aligned}$$

On the set $\{\hat{j}(x) = j'\}$,

$$|\hat{E}_{j'}^n(x) - \hat{E}_{j_n^*(x)}^n(x)| \leqslant Kt_n 2^{j_n^*(x)/2}.$$

For the second term, we use (19) and the Cauchy–Schwartz inequality. We obtain

$$T_1 \leqslant 3^{p-1}\Bigg((Kt_n 2^{j_n^*(x)/2})^p P\{\hat{j}(x) \leqslant j_n^*(x)\}$$

$$+ \sum_{0\leqslant j' \leqslant j_n^*(x)} \left(\frac{2^{j_n^*(x)}}{n}\right)^{p/2} (P(\hat{j}(x) = j'))^{1/2} + |E_{j_n^*(x)} f(x) - f(x)|^p\Bigg)$$

$$\leqslant 3^{p-1}\left((Kt_n 2^{j_n^*(x)/2})^p + \left(\frac{2^{j_n^*(x)}}{n}\right)^{p/2} j_n^*(x)^{1/2} P\{\hat{j}(x) \leqslant j_n^*(x)\}^{1/2} + |E_{j_n^*(x)} f(x) - f(x)|^p\right)$$

$$\leqslant 3^{p-1}\left((Kt_n 2^{j_n^*(x)/2})^p + 2^{j_n^*(x)p/2}\left(\frac{t_n}{\log 2}\right)^p + |E_{j_n^*(x)} f(x) - f(x)|^p\right).$$

For the last inequality, we used the fact that $j_n^*(x) \leqslant J_n \leqslant \beta \log n/\log 2$, and $p \geqslant 1$.

For the term T_2, using Lemma 2, the Cauchy–Schwarz inequality, (19) and the definition of $\hat{j}(x)$, we obtain:

$$T_2 \leqslant \sum_{j_n^*(x)+1\leqslant j' \leqslant J_n} 2^{p-1}\Big\{[\mathrm{E}_n|\hat{E}_{j'}^n(x) - E_{j'} f(x)|^{2p}]^{1/2} P\{\hat{j}(x) = j'\}^{1/2} + |E_{j'} f(x) - f(x)|^p P\{\hat{j}(x) = j'\}\Big\}$$

$$\leqslant 2^{p-1}\Bigg\{\left[\sum_{j_n^*(x)+1\leqslant j' \leqslant J_n} \mathrm{E}_n|\hat{E}_{j'}^n(x) - E_{j'} f(x)|^{2p}\right]^{1/2} [P\{j_n^*(x) + 1\hat{j}(x) \leqslant J_n\}]^{1/2}$$

$$+ \sum_{j_n^*(x)+1\leqslant j' \leqslant J_n} \left(\frac{K2^{j'} t_n}{4}\right)^p P\{\hat{j}(x) = j'\}\Bigg\}$$

$$\leqslant 2^{p-1} C\left\{2^{J_n p/2} n^{-p/2}\left[\frac{CJ_n^2}{n^\gamma}\right]^{1/2} + \left(\frac{K2^{J_n} t_n}{4}\right)^p \frac{CJ_n^2}{n^\gamma}\right\}.$$

This concludes this case since, for $\gamma > p\beta/2$, the right-hand side of the last inequality is easily bounded by $Ct_n^p \leqslant Ct_n^p 2^{j^*_n(x)p/2}$.

The case $j_n^*(x) > J_n$ is parallel to the previous one except that the term T_2 has disappeared. Again we observe that

$$|\hat{E}_{j'}^n(x) - f(x)|^p \leqslant 3^{p-1}\Big(|\hat{E}_{j'}^n(x) - \hat{E}_{J_n}^n(x)|^p + |\hat{E}_{J_n(x)}^n - E_{J_n} f(x)|^p + |E_{J_n} f(x) - f(x)|^p\Big)$$

On the set $\{\hat{j}(x) = j'\}$,

$$|\hat{E}_{j'}^n(x) - \hat{E}_{J_n}^n(x)| \leqslant Kt_n 2^{J_n/2}.$$

For the second term, we again use (19) and the Cauchy–Schwarz inequality. Thus

$$\sum_{0\leqslant j'\leqslant J_n} \mathrm{E}_n|\hat{E}^n_{j'}(x) - f(x)|^p I\{\hat{j}(x) = j'\}$$

$$\leqslant 3^{p-1}\left((Kt_n 2^{J_n/2})^p + \sum_{0\leqslant j'\leqslant J_n}\left(\frac{2^{J_n}}{n}\right)^{p/2}(P(\hat{j}(x)=j'))^{1/2} + |E_{J_n}f(x) - f(x)|^p\right)$$

$$\leqslant 3^{p-1}\left((Kt_n 2^{J_n/2})^p + \left(\frac{2^{j_n}}{n}\right)^{p/2} J_n^{1/2} + |E_{J_n}f(x) - f(x)|^p\right)$$

$$\leqslant 3^{p-1}\left((Kt_n 2^{j_n^*(x)/2})^p + 2^{j_n^*(x)p/2}\left(\frac{t_n}{\log 2}\right)^p + |E_{J_n}f(x) - f(x)|^p\right). \qquad \square$$

4.3. Thresholding wavelet coefficients

Various descriptions of this procedure can be found, for instance in Donoho *et al*. (1994, 1996), in different frameworks (white noise, equispaced regression, density). As above, we consider $t_n = (\log n/n)^{1/2}$, and fix J_n such that $2^{J_n} < (n/\log n)^\beta \leqslant 2^{J_n+1}$. Note that in the references just cited we generally have $\beta = 2$.

Here the space $\mathcal{X}$ is [0, 1], equipped with the Lebesgue measure, and the E_j are the projection kernels on the spaces V_{j-1} (for $j \leqslant 1$, $E_0 = E_1$) of a multiresolution analysis generated by a pair of mother and father wavelets ϕ and ψ. We assume that ϕ and ψ are compactly supported and regular (at least bounded). We again assume conditions (19) and (20) on the sequence of models for these particular E_j, and recall the following thresholding procedure:

$$\hat{f}^T(x) = \sum_{0\leqslant j\leqslant J_n}\sum_k \hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geqslant Kt_n\} g_{j-1,k}(x),$$

where $g_{jk} = \psi_{jk}$, for $j > 0$, $g_{0k} = \phi_{0k}$, $\hat{\beta}_{jk} = \int_{[0,1]} \hat{E}_j(x) g_{j-1,k}(x)\mathrm{d}x$.

We prove an analogue of Proposition 5. Here, the difficulty lies in the definition of $F(f)(j, x)$. We have to be a little careful and, in particular, we introduce the following tools.

As usual, we denote by W_j the 'innovation' space defined by $V_{j+1} = V_j \oplus W_j$. W_j is spanned by the collection $\{\psi_{jk}, k \in \mathbb{N}\}$. If f is a function of L_2, $\Delta_j f = \sum_k \beta_{jk}\psi_{j,k}$ denotes its projection on W_j, and if χ_{jk} denotes the Haar wavelet, we define

$$\tilde{\Delta}_j f := \sum_k \beta_{jk}\chi_{jk}.$$

Notice that $\tilde{\Delta}_j f$ is in general a slightly modified version of $\Delta_j f$ and enjoys the following nice property: $|\tilde{\Delta}_j f(x)| = |\beta_{jk}|2^{j/2}$ when $\chi_{jk}(x) \neq 0$.

Now, for g locally integrable in $\mathbb{R}$, $r > 0$ let us define the maximal function

$$M_r(g)(x) = \sup_{\{B, x\in B\}} \left(|B|^{-1} \int_B |g|^r \right)^{1/r}. \tag{25}$$

The supremum is taken over all the balls B containing x, and $|B|$ denotes the volume of the ball. The function $M_r(g)$ obviously satisfies $M_r(g) \geqslant |g|$ a.e., and enjoys the following nice properties.

Lemma 3. *For any $r > 0$, there exist universal constants C_r, C'_r such that, for any $j \geqslant 0$, k,*

$$|\Delta_j f(x)| \leqslant C_r M_r(\tilde{\Delta}_j f)(x), \qquad \forall x \text{ a.e.}, \tag{26}$$

$$|\tilde{\Delta}_j f(x)| \leqslant C'_r M_r(\Delta_j f)(x), \qquad \forall x \text{ a.e.} \tag{27}$$

There also exists, for any f locally integrable, for any $q > r$, a constant C_{qr} such that

$$\int |M_r(f)|^q \leqslant C_{qr} \int |f|^q \leqslant C_{qr} \int |M_r(f)|^q. \tag{28}$$

Remark. Inequality (28) states the equivalence of the L_q norm of f and $M_r(f)$. This is a classical result in harmonic analysis. Its proof can be found in Stein (1993). The proof of (26) and (27) is given in Lemma 8.

Proposition 6. *If the inequalities* (19) *and* (20) *are satisfied for some order p^* and $\gamma > p^*$ and for the class $\mathcal{C}_j^W$, $\hat{f}^T$ satisfies a local oracle inequality of order p on the space $L_\infty(M)$ of functions bounded by M on* $[0, 1]$, *associated with the sequence J_n and the local functional*

$$F^T(f)(j, x) := \frac{K}{2C'_r} 2^{-j/2} \sum_{j' \geqslant j} M_r(\Delta_{j'} f)(x) \tag{29}$$

for any choice of $r < p$, $1 \leqslant p \leqslant p^*$.

Remark. Again, here the conditions of Proposition 3 are fulfilled for instance if $\alpha/\beta > 1/p$, since in this case $\mathcal{B}_{\alpha/\beta, p, \infty}(M) \subset L_\infty(M)$. □

Proof. The proof will follow, with some differences the proof of Proposition 5.

First, we need to prove inequality (14).

$$\int \sup_{j' \geqslant j} |E_{j'} f(x) - f(x)|^p I\{x, j_\lambda(f, x) = j\} \mathrm{d}\mu(x) \leqslant \int I\{x, j_\lambda(f, x) = j\} \left[\sum_{j' \geqslant j} |\Delta_{j'} f(x)| \right]^p \mathrm{d}\mu(x)$$

$$\leqslant \int I\{x, j_\lambda(f, x) = j\} \left[\sum_{j' \geqslant j} M_r(\Delta_{j'} f)(x)| \right]^p \mathrm{d}\mu(x)$$

$$\leqslant \int I\{x, j_\lambda(f, x) = j\} (2^{j/2}\lambda)^p \mathrm{d}\mu(x)$$

$$\leqslant C'(2^{j/2}\lambda)^p \mu\{x, j_\lambda(f, x) = j\}.$$

To establish (13) we have, as above, to distinguish two cases: $j_n^*(x) \leqslant J_n$ and $j_n^*(x) > J_n$. We only investigate the first case. The second can easily be treated using the same technique as above. In this particular case, we can divide $E_n|\hat{f}^T(x) - f(x)|^p$ into three terms:

$$e_1 = 3^{p-1}\mathrm{E}_n \left| \sum_{-1 \leqslant j \leqslant j_n^*(x)} \sum_k (\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geqslant Kt_n\} - \beta_{jk}) g_{jk}(x) \right|^p,$$

$$e_2 = 3^{p-1}\mathrm{E}_n \left| \sum_{j_n^*(x)+1 \leqslant j \leqslant J_n} \sum_k (\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geqslant Kt_n\} - \beta_{jk}) g_{jk}(x) \right|^p,$$

$$e_3 = 3^{p-1}|E_{J_n} f(x) - f(x)|^p.$$

Notice that we already have e_3 on the right-hand side of (13) so we only need to bound e_1 and e_2.

We prove the following lemma.

Lemma 4. *Under the conditions above, for all $n \geqslant 1$, for all j, such that $-1 \leqslant j \leqslant J_n$, for all k,*

$$\mathrm{E}_n \left| \hat{\beta}_{jk} - \beta_{jk} \right|^{2p} \leqslant Cn^{-p}, \tag{30}$$

$$P\left(\left| \hat{\beta}_{jk} - \beta_{jk} \right| \geqslant \frac{Kt_n}{4} \right) \leqslant C_2 n^{-\gamma}, \tag{31}$$

$$\mathrm{E}_n \left| \hat{\beta}_{jk} I\left\{ \left| \hat{\beta}_{jk} \right| \geqslant Kt(n) \right\} - \beta_{jk} \right|^p \leqslant C\left\{ t_n^p + n^{-\gamma'} \right\} \qquad \text{for } \gamma' = \frac{\gamma}{2} + \frac{p}{2}. \tag{32}$$

Proof. Inequality (30) follows directly from condition (19), and (31) follows from (20).

For (32), we have to investigate separately the different cases $|\beta_{jk}| > 2Kt_n$, $|\beta_{jk}| \leqslant Kt_n/2$, $Kt_n/2 \leqslant |\beta_{jk}| \leqslant 2Kt_n$. For the first case, we write

$$\begin{aligned}\mathrm{E}_n|\hat\beta_{jk}I\{|\hat\beta_{jk}|\geqslant Kt_n\}-\beta_{jk}|^p &= \mathrm{E}_n|\hat\beta_{jk}-\beta_{jk}|^p I\{|\hat\beta_{jk}|\geqslant Kt_n\}+|\beta_{jk}|^p P|\hat\beta_{jk}|<Kt_n\}\\ &\leqslant \mathrm{E}_n|\hat\beta_{jk}-\beta_{jk}|^p+|\beta_{jk}|^p P\{|\hat\beta_{jk}-\beta_{jk}|\geqslant Kt_n\}\\ &\leqslant C\{n^{-p/2}+n^{-\gamma}\}.\end{aligned}$$

Here we have used (30), (31) and the fact that boundedness of f implies that its wavelet coefficients are also bounded for $j\geqslant 0$.

For the second case, using the Cauchy–Schwarz inequality, we have

$$\begin{aligned}\mathrm{E}_n|\hat\beta_{jk}I\{|\hat\beta_{jk}|\geqslant Kt_n\}-\beta_{jk}|^p &\leqslant (\mathrm{E}_n|\hat\beta_{jk}-\beta_{jk}|^{2p})^{1/2}P\left\{|\hat\beta_{jk}-\beta_{jk}|\geqslant\frac{Kt_n}{2}\right\}^{1/2}+\left(\frac{Kt_n}{2}\right)^p\\ &\leqslant Cn^{-(p+\gamma)/2}+\left(\frac{Kt_n}{2}\right)^p.\end{aligned}$$

The third case uses the arguments of both previous cases:

$$\mathrm{E}_n|\hat\beta_{jk}I\{|\hat\beta_{jk}|\geqslant Kt_n\}-\beta_{jk}|^p\leqslant C\left\{n^{-p/2}+\left(\frac{Kt_n}{2}\right)^p\right\}.$$

This ends the proof of the lemma. □

To bound e_1 and e_2, we use the following triangular inequality, true for $p\geqslant 1$:

$$\left(\mathrm{E}\left|\sum_I X_i\right|^p\right)^{1/p}\leqslant\sum_I\left(\mathrm{E}|X_i|^p\right)^{1/p} \tag{33}$$

To bound e_1 we use (33), (32), the fact that as g is compactly supported only a finite number of k (N, say) at each level j are such that $g_{jk}(x)\neq 0$:

$$\begin{aligned}e_1 &\leqslant 3^{p-1}\left[\sum_{-1\leqslant j\leqslant j_n^*(x)}\sum_k\left\{\mathrm{E}_n|(\hat\beta_{jk}I\{|\hat\beta_{jk}|\geqslant Kt_n\}-\beta_{jk})g_{jk}(x)|^p\right\}^{1/p}\right]^p\\ &\leqslant 3^{p-1}\left[\sum_{-1\leqslant j\leqslant j_n^*(x)}\sum_k I\{g_{jk}(x)\neq 0\}\left(2^{jp/2}C\left\{t_n^p+\frac{1}{n^{\gamma'}}\right\}\right)^{1/p}\right]^p\\ &\leqslant 3^{p-1}\left[\sum_{-1\leqslant j\leqslant j_n^*(x)}NC'2^{j/2}\left\{t_n+\frac{1}{n^{\gamma'}/p}\right\}\right]^p\\ &\leqslant C''2^{j_n^*(x)p/2}\left\{t_n^p+\frac{1}{n^{\gamma'}}\right\}.\end{aligned}$$

To bound e_2, we observe that we can write:

$$\sum_{j_n^*(x)+1\leqslant j\leqslant J_n}\sum_k(\hat\beta_{jk}I\{|\hat\beta_{jk}|\geqslant Kt_n\}-\beta_{jk})g_{jk}(x)$$

$$=(E_{j_n^*(x)}f(x)-f(x)-E_{J_n}f(x)+f(x))$$

$$+\sum_{j_n^*(x)+1\leqslant j\leqslant J_n}\sum_k(\hat\beta_{jk}-\beta_{jk})I\{|\hat\beta_{jk}|\geqslant Kt_n\}I\left\{|\beta_{jk}|\leqslant\frac{Kt_n}{2}\right\}g_{jk}(x).$$

We are permitted to use the indicator function $I\{|\beta_{jk}|\leqslant Kt_n/2\}$ here because we are dealing with js larger than $j_n^*(x)$. Because $F(f)(j,x)$ is non-decreasing in j, we have, for $j\geqslant j_n^*(x)$, $F(f)(j,x)\leqslant Kt_n/2C'_r$, hence $M_r(\Delta_j f)\leqslant 2^{j/2}Kt_n/2C'_r$. Therefore, using (27), we obtain that necessarily $|\beta_{jk}|\leqslant Kt_n/2$ if $g_{jk}(x)\neq 0$. Now it remains to write:

$$e_2\leqslant C'\big\{|E_{j_n^*(x)}f(x)-f(x)|^p+|E_{J_n}f(x)-f(x)|^p$$

$$+\left[\sum_{j_n^*(x)+1\leqslant j'\leqslant J_n}\sum_k\left(\mathrm{E}|\hat\beta_{jk}-\beta_{jk}|^{2p}P\left\{|\hat\beta_{jk}-\beta_{jk}|\geqslant\frac{Kt_n}{2}\right\}\right)^{1/2p}2^{j/2}\|g\|_\infty I\{g_{jk}(x)\neq 0\}\right]^p\Bigg\}$$

$$\leqslant C''\left\{|\mathrm{E}_{j_n^*(x)}f(x)-f(x)|^p+|E_{J_n}f(x)-f(x)|^p+\frac{1}{n^{\gamma/2}}\right\}.\qquad\square$$

4.4. Local bandwidth selection using wavelet coefficients

We also consider the following procedure, which can be considered as a hybrid version between thresholding and Lepski's procedure. We define the index $\hat{j}^\beta(x)$ as the minimum of the set of β-admissible js at the point x, where $j\in\{0,\ldots,J_n\}$ is β-admissible at the point x if

$$|\hat\beta_{j'k}\chi_{j'k}(x)|\leqslant K2^{j'/2}t_n,\ \forall k,j';\ j\leqslant j'\leqslant J_n.$$

We also define the estimate

$$\hat f^H(x)=\hat E^n_{\hat j^\beta(x)}(x).$$

The sequence J_n is again chosen (for the sake of simplicity) in such a way that $2^{J_n}<t_n^{-\beta}\leqslant 2^{J_n+1}$. Notice that $\hat f^H$ looks very much like $\hat f^T$, except that it somehow 'fills the holes'. If j is such that $|\hat\beta_{jk}|\geqslant Kt_n$ and $g_{jk}(x)\neq 0$, and if for instance $|\hat\beta_{j-1,k'}|<Kt_n$ for all k' with $g_{j-1,k'}(x)\neq 0$, then $\hat f^H$ restores all that $|\hat\beta_{j-1,k'}|<Kt_n$ which were killed in the $\hat f^T$ expansion. This estimator has similar minimax and adaptation properties to $\hat f^T$, but turns out to be strictly more efficient for the construction of confidence intervals (see Picard and Tribouley 2000). We also can prove the following proposition:

Proposition 7. *If the inequalities* (19) *and* (20) *are satisfied for some order* p^* *and* $\gamma > p^*$ *and for the class* $\mathcal{C}^{W_j}$, $\hat{f}^H$ *satisfies a local oracle inequality of order p on the space* $L_\infty(M)$ *of functions bounded by M on* [0, 1], *associated with the sequence* J_n *and the local functional*

$$F^H(f)(j, x) := \left(\frac{K}{2C'r}\right)2^{-j/2}\sum_{j'\geqslant j}|\tilde{\Delta}_{j'}f(x)|,$$

for any choice of $r < p$, $1 \leqslant p \leqslant p^*$.

Proof. The proof combines the arguments of thresholding and bandwidth selection. We have, as above, to distinguish two cases: $j_n^*(x) \leqslant J_n$ and $j_n^*(x) > J_n$. We only investigate the first case. The following lemma summarizes the essential properties of $\hat{f}^H$.

Lemma 5. *Under the above conditions,*

(*i*) $P\{\hat{j}(x) > j_n^*(x)\} \leqslant CJ_n/n^\gamma$;
(*ii*) $\mathrm{E}_n|\hat{E}_j^n(x) - E_jf(x)|^{2p} \leqslant C_1(2^j/n)^p$, for all $n \geqslant 1$, for all $x \in \mathcal{X}$, for all j, $0 \leqslant j \leqslant J_n$.

Proof. (i) We observe that, by the definition of $\hat{j}^\beta(x)$, when $\hat{j}^\beta(x) \geqslant 1$, $j^* = \hat{j}^\beta(x)$ is not β-admissible (when $\hat{j}^\beta(x)$ is admissible), then there exists k, $/g_{j^*k}(x) \neq /|\hat{\beta}_{j^*k}| \geqslant Kt_n$. If, in addition, we suppose $j^* \geqslant j_n^*(x)$, then $|\beta_{j^*k}| \leqslant Kt_n/2$. Hence,

$$\begin{aligned}
P\{\hat{j}^\beta(x) > j_n^*(x)\} &= P\left\{\bigcup_{j_n^*(x)\leqslant j^*\leqslant J_n\ k,/\ g_{j^*k}(x)\neq 0}|\hat{\beta}_{j^*k} - \beta_{j^*k}| \geqslant \frac{Kt_n}{2}\right\}\\
&\leqslant \sum_{j\leqslant J_n}\ \sum_{k,/g_{j^*k}(x)\neq 0} P\left\{\{|\hat{\beta}_{j^*k} - \beta_{j^*k}| \geqslant \frac{Kt_n}{2}\right\}\\
&\leqslant \frac{CJ_n}{n^\gamma}.
\end{aligned}$$

(ii) Using (33), we easily obtain

$$\begin{aligned}
\mathrm{E}_n|\hat{E}_j^n(x) - E_jf(x)|^{2p} &\leqslant \sum_{0\leqslant j'\leqslant j}\ \sum_{k,/g_{j'k}(x)\neq 0}(\mathrm{E}_n|\hat{\beta}_{j'k} - \beta_{j'k}|^{2p}2^{jp}\|g\|_\infty)^{1/2p}]^{2p}\\
&\leqslant C\left(\frac{2^j}{n}\right)^p.
\end{aligned}$$

□

In the case $j_n^*(x) \leqslant J_n$, we can divide $\mathrm{E}_n|\hat{f}^H(x) - f(x)|^p$ into two terms:

$$T_1 = \sum_{0 \leqslant j' \leqslant j_n^*(x)} \mathrm{E}_n |\hat{E}_{j'}^n(x) - f(x)|^p I\{\hat{j}^\beta(x) = j'\},$$

$$T_2 = \sum_{j_n^*(x)+1 \leqslant j' \leqslant J_n} \mathrm{E}_n |\hat{E}_{j'}^n(x) - f(x)|^p I\{\hat{j}^\beta(x) = j'\}.$$

To bound T_1, we again observe that

$$|\hat{E}_{j'}^n(x) - f(x)|^p \leqslant 3^{p-1}(|\hat{E}_{j'}^n(x) - \hat{E}_{j_n^*(x)}^n(x)|^p + |\hat{E}_{j_n^*(x)}^n(x) - E_{j_n^*(x)} f(x)|^p + |\mathrm{E}_{j_n^*(x)} f(x) - f(x)|^p),$$

and on the set $\{\hat{j}^\beta(x) = j'\}$,

$$|\hat{E}_{n'}^n(x) - \hat{E}_{j_n^*(x)}^n(x)| = \left| \sum_{j' \leqslant j \leqslant j_n^*(x)} \sum_{k,/\ g_{j^*k}(x) \neq 0} \hat{\beta}_{jk} g_{jk}(x) \right|$$

$$\leqslant CKt_n 2^{j_n^*(x)/2}.$$

At this stage, we can bound T_1 and T_2 just as in the proof of Proposition 5.

To end our proof, we need to establish (14):

$$\int \sup_{j' \geqslant j} |E_{j'} f(x) - f(x)|^p I\{x, j_\lambda(f, x) = j\} \mathrm{d}\mu(x)$$

$$\leqslant \int I\{x, j_\lambda(f, x) = j\} \left[\sum_{j' \geqslant j} |\Delta_{j'} f(x)| \right]^p \mathrm{d}(x)$$

$$\leqslant \int \|g\|_\infty I\{x, j_\lambda(f, x) = j\} \left[\sum_{j' \geqslant j} |\tilde{\Delta}_{j'} f(x)| \right]^p \mathrm{d}\mu(x)$$

$$\leqslant \int I\{x, j_\lambda(f, x) = j\} (2^{j/2}\lambda)^p \mathrm{d}\mu(x)$$

$$\leqslant C'(2^{j/2}\lambda)^p \mu\{x, j_\lambda(f, x) = j\}.$$

5. Comparison among various adaptive procedures

We now compare the methods investigated above. Let us restrict ourselves to the case where the E_j are the projection kernels onto the spaces V_j of a multiresolution analysis generated by a pair of mother and father wavelets ϕ and ψ having the properties mentioned above. To simplify, let us also take the most common stopping sequence J_n corresponding to the case $\beta = 2$. From Propositions 3, 5, 6 and 7, we know that the maxisets $MS(\hat{f}_n, p, \alpha)(T)$ associated with the rate $t_n^{\alpha p}$ of $\hat{f}^L$ ($\hat{f}^T$, $\hat{f}^H$) contains the set $\mathcal{B}_{\alpha/2,p,\infty}(M) \cap \mathcal{W}(F^L)(p, q)(M)$,

$\mathcal{B}_{\alpha/2,p,\infty}(M) \cap \mathcal{W}(F^T)(p, q)(M)$, $(\mathcal{B}_{\alpha/2,p,\infty}(M) \cap \mathcal{W}(F^H)(p, q)(M))$, where $q = p(1-\alpha)$ and, if we omit the constants,

$$F^L(f)(j, x) = \sup_{j' \geqslant j} 2^{-j'/2} |E_{j'} f(x) - f(x)|$$

$$F^T(f)(j, x) = 2^{-j/2} \sum_{j' \geqslant j} M_r(\Delta_{j'} f)(x)$$

$$F^H(f)(j, x) = 2^{-j/2} \sum_{j' \geqslant j} |\tilde{\Delta}_{j'} f(x)|.$$

It is thus natural to ask how far is the inclusion from equality, and whether the above-mentioned spaces can be compared.

By virtue of Theorem 7 in Cohen *et al*. (1999) (see also the example in Section 2.2), we have equality between the spaces $\mathcal{B}_{\alpha/2,p,\infty} \cap \mathcal{W}^*(p, q)$ and the maxiset associated with the thresholding procedure $MS(\hat{f}_n^T, p, \alpha)$, in the sense described in (1).

Recalling that

$$F^2(f)(j, x) = 2^{-j/2} |\tilde{\Delta}_j f(x)|,$$

we already have that $\mathcal{W}^*(p, q) = \mathcal{W}(F^2)(p, q)$.

A consequence of Theorem 1 in Section 6 is that

$$\mathcal{W}(F^H)(p, q) = \mathcal{W}(F^2)(p, q) = \mathcal{W}^*(p, q).$$

If we introduce the auxiliary function

$$F^3(f)(j, x) = 2^{-j/2} |\Delta_j f(x)|,$$

and refer to Theorem 1 for the definition of T, we observe that

$$\begin{aligned} F^3(f)(j, x) &\leqslant 2^{-j/2}\{|E_{j+1}f - f| + |E_j f - f|\} \\ &\leqslant F^L(f)(j, x), \\ F^L(f)(j, x) &\leqslant T(F^3)(f)(j, x). \end{aligned}$$

We deduce, using the first inequality, that

$$\mathcal{W}(F^L)(p, q) \subset \mathcal{W}(F^3)(p, q),$$

and, from the second one, that

$$\mathcal{W}(F^3)(p, q) \subset \mathcal{W}(T(F^L))(p, q).$$

Now, using Theorems 1 and 3,

$$\mathcal{W}(F^L)(p, q) = \mathcal{W}(T(F^L))(p, q),$$

$$\mathcal{W}(F^3)(p, q) = \mathcal{W}(F^2)(p, q).$$

Hence,

$$\mathcal{W}(F^L)(p, q) = \mathcal{W}(F^2)(p, q) = \mathcal{W}(F^H)(p, q) = \mathcal{W}^*(p, q).$$

Now, writing

$$F^4(f)(j, x) = 2^{-j/2}|M_r(\Delta_j f)(x)|,$$

we have, from Theorem 2 below,

$$\mathcal{W}(F^4)(p, q) = \mathcal{W}(F^3)(p, q),$$

and from Theorem 1,

$$\mathcal{W}(F^4)(p, q) = \mathcal{W}(T(F^4))(p, q) = \mathcal{W}(F^H)(p, q),$$

Hence,

$$\mathcal{W}(F^L)(p, q) = \mathcal{W}(F^T)(p, q) = \mathcal{W}(F^H)(p, q) = \mathcal{W}^*(p, q).$$

As a consequence of the foregoing remarks, we can easily state that as far as maxisets are concerned, $\hat{f}_n^L$ and $\hat{f}_n^H$ are at least good as $\hat{f}_n^T$. Whether they are strictly better is an open question, as is the relationship between them.

6. Comparison of weak Besov bodies associated with different functionals

6.1. Weak Besov bodies associated with max or sum functionals

Let χ be a measure space with a σ-finite measure μ. Let $0 < p < \infty$, and let us define on $\mathbb{N} \times \mathcal{X}$ the measure $\nu_p = (\sum_{j \in \mathbb{N}} 2^{j(p/2)} \delta_j) \otimes \mu$. Let $G(j, x)$ be a measurable function defined on $\mathbb{N} \times \mathcal{X}$. Define

$$G^*(j, x) = \sup_{j' \geqslant j} |G(j', x)|,$$

$$T(G)(j, x) = 2^{-j/2} \sum_{j' \geqslant j} 2^{j'/2} |G(j', x)|.$$

Theorem 1. *Let* $0 < q < p$. *Then*

$$\mathcal{W}(F)(p, q) = \mathcal{W}(F^*)(p, q) = \mathcal{W}(T(F))(p, q).$$

Because of the definition of $\mathcal{W}(F)(p, q)$ (see (17)), this theorem is a consequence of the following lemma:

Lemma 6. *For* $0 < q < p < \infty$,

$$G \in L_{q,\infty}(\nu_p) \Leftrightarrow G^* \in L_{q,\infty}(\nu_p) \Leftrightarrow T(G) \in L_{q,\infty}(\nu_p).$$

Proof. Since $|G| \leqslant G^* \leqslant T(G)$, we only have to prove that

$$G \in L_{q,\infty}(\nu_p) \Rightarrow T(G) \in L_{q,\infty}(\nu_p).$$

We begin by proving that, for all $0 < q < p$, there exists some $C_q < \infty$, such that for all G,

$$\|T(G)\|^q_{L_q(\nu_p)} \leqslant C_q \|G\|^q_{L_q(\nu_p)}.$$

We observe that

$$\|G\|^q_{L_q(\nu_p)} = \sum_{j \geqslant 0} 2^{jp/2} \|G(j, \cdot)\|^q_{L_q(\mu)},$$

$$\|T(G)\|^q_{L_q(\nu_p)} = \sum_{j \geqslant 0} 2^{j(p-q)/2} \left\| \sum_{j' \geqslant j} 2^{j'/2} |G(j', \cdot)| \right\|^q_{L_q(\mu)}.$$

Hence, for $q \leqslant 1$,

$$\begin{aligned} \|T(G)\|^q_{L_q(\nu_p)} &\leqslant \sum_{j \geqslant 0} 2^{j(p-q)/2} \sum_{j' \geqslant j} 2^{j'q/2} \|G(j', \cdot)\|^q_{L_q(\mu)} \\ &= \sum_{j' \geqslant 0} 2^{j'q/2} \|G(j', \cdot)\|^q_{L_q(\mu)} \sum_{j \leqslant j'} 2^{j(p-q)/2} \\ &= c \sum_{j' \geqslant 0} 2^{j'q/2} 2^{j'(p-q)/2} \|G(j', \cdot)\|^q_{L_q(\mu)} \\ &= c \|G\|^q_{L_q(\nu_p)}. \end{aligned}$$

For $q > 1$, we observe that

$$\|G\|^q_{L_q(v_p)} < \infty \Leftrightarrow 2^{jp/2q} \|G(j, \cdot)\|_{L_q(\mu)} = \epsilon_j \in l_q(\mathbb{N}).$$

Hence

$$\begin{aligned} \|T(G)\|^q_{L_q(\nu_p)} &\leqslant \sum_{j \geqslant 0} 2^{j(p-q)/2} \left(\sum_{j' \geqslant j} 2^{j'/2} \|F(j', \cdot)\|_{L_q(\mu)} \right)^q \\ &= \sum_{j \geqslant 0} 2^{j(p-q)/2} \left(\sum_{j' \geqslant j} 2^{j'(1/2 - p/2q)} \epsilon_{j'} \right)^q \\ &= \sum_{j \geqslant 0} \left(\sum_{j' \geqslant j} 2^{-(j'-j)(p/2q - 1/2)} \epsilon_{j'} \right)^q. \end{aligned}$$

As $\alpha = p/2q - 1/2 > 0$,

$$\sum_{j' \geqslant j} 2^{-(j'-j)(p/2q-1/2)} \epsilon_{j'} = (\epsilon. * b.)_j,$$

where, $b_j = 1_{j \leqslant 0} 2^{-\alpha j}$. Now, using

$$\|\epsilon. * b.\|_q \leqslant \|b.\|_1 \|\epsilon.\|_q,$$

we have

$$\|T(G)\|^q_{L_q(\nu_p)} \leqslant c\|\epsilon_j\|^q_{l_q} = c\|F\|^q_{L_q(\nu_p)}. \tag{34}$$

All we now need to prove is that (34) can be extended to the associated weak spaces. This is a consequence of the following interpolation theorem, whose proof is given for the sake of completeness. □

Proposition 8. *Let* (Y, ν) *be a measure space, let* $0 < p_1 < q < p_2$, *and let* T *be a mapping from* $L_{p_1} + L_{p_2}$ *to the space of measurable functions satisfying* $|T(f_1 + f_2)| \leqslant |T(f_1)| + |T(f_2)|$ *a.e. We suppose that for all* $i \in \{1, 2\}$ *there exists a constant depending only on* T *and* p_i, *denoted by* $\|T\|_{p_i}$, *such that* $0 < \|T\|_{p_i} < \infty$ *and, for all* $f \in L_{p_i}(\nu)$,

$$\|T(f)\|_{L_{p_i(\nu)}} \leqslant \|T\|_{p_i} \|f\|_{L_{p_i(\nu)}}.$$

Then

$$\|T(f)\|_{L_{q,\infty}(\nu)} \leqslant C(p_1, p_2, q, \|T\|_{p_1}, \|T\|_{p_2}) \|f\|_{L_{q,\infty}(\nu)}.$$

Proof. Let $f \in L_{q,\infty}(Y, \nu)$, $0 < p_1 < q < p_2$, $0 < \lambda < \infty$. We have the following inequalities:

$$\begin{aligned}\int_Y |f|^{p_2} 1_{|f| \leqslant \lambda} \, d\nu &\leqslant \int_Y (|f| \wedge \lambda)^{p_2} \, d\nu = \int_0^\lambda p_2 x^{p_2 - 1} \nu(|f| > x) dx \\ &\leqslant \int_0^\lambda p_2 x^{p_2-1} \left(\frac{\|f\|_{L_{q,\infty}}}{x} \right)^q dx = \frac{p_2}{p_2 - q} \|f\|^q_{L_{q,\infty}} \lambda^{p_2 - q};\end{aligned}$$

$$\begin{aligned}\int_Y |f|^{p_1} 1_{|f| > \lambda} d\nu &= \int_0^\lambda p_1 x^{p_1 - 1} \nu(|f| > \lambda) dx + \int_\lambda^\infty p_1 x^{p_1-1} \nu(|f| > x) dx \\ &\leqslant \left(\frac{\|f\|_{L_{q,\infty}}}{\lambda} \right)^q \lambda^{p_1} + \int_\lambda^\infty p_1 x^{p_1 - 1} \left(\frac{\|f\|_{L_{q,\infty}}}{x} \right)^q dx = \frac{q}{q - p_1} \|f\|^q_{L_{q,\infty}} \lambda^{p_1 - q}.\end{aligned}$$

For a fixed $0 < \lambda < \infty$, let us decompose $f \in L_{q,\infty}$ as follows:

$$f = f 1_{|f| > \lambda} + f 1_{|f| \leqslant \lambda} = f_1 + f_2.$$

Using the previous inequalities, we have

$$\int_Y |f_1|^{p_1}\, \mathrm{d}\nu \leqslant \frac{q}{q-p_1}\|f\|^q_{L_{q,\infty}}\lambda^{p_1-q},$$

$$\int_Y |f_2|^{p_2}\, \mathrm{d}\nu \leqslant \frac{p_2}{p_2-q}\|f\|^q_{L_{q,\infty}}\lambda^{p_2-q}.$$

So

$$\begin{aligned}\nu(|T(f)| > 2\lambda) \leqslant \nu(|T(f_1)| > \lambda) + \nu(|T(f_2)| > \lambda) &\leqslant \left(\frac{\|T(f_1)\|_{p_1}}{\lambda}\right)^{p_1} + \left(\frac{\|T(f_2)\|_{p_2}}{\lambda}\right)^{p_2} \\ &\leqslant \left(\frac{\|T\|_{p_1}}{\lambda}\right)^{p_1}\|f_1\|^{p_1}_{p_1} + \left(\frac{\|T\|_{p_2}}{\lambda}\right)^{p_2}\|f_2\|^{p_2}_{p_2} \\ &\leqslant \left(\frac{q}{q-p_1}\|T\|^{p_1}_{p_1} + \frac{p_2}{p_2-q}\|T\|^{p_2}_{p_2}\right)\left(\frac{\|f\|_{L_{q,\infty}}}{\lambda}\right)^q.\end{aligned}$$

□

6.2. Weak Besov bodies associated with maximal functions

Let $\mathcal{X}$ be either of the spaces $\mathbb{R}^d$ or $[0, 1]^d$ and μ the Lebesgue measure. Let $0 < r < \infty$. For all measurable functions on $\mathcal{X}$, we recall that $M_r(g)(x) = \sup_{\{B, x\in B\subset\mathcal{X}\}}(|B|^{-1}\int_B |g|^r)^{1/r}$, where B denotes a ball of $\mathcal{X}$. For $F(j, x) = F_j(x)$ a non-negative functional defined on $\mathbb{N}\times\mathcal{X}$, let us extend the definition:

$$M_r(F)(j, x) = M_r(F_j)(x).$$

Theorem 2. *For* $0 < r < q < p$, $\mathcal{W}(F)(p, q) = \mathcal{W}(M_r(F))(p, q)$.

As above, this is a consequence of the following lemma:

Lemma 7. *Let* $F(j, x)$ *be a measurable function defined on* $\mathbb{N}\times\mathcal{X}$. *Then, in the previous notation, for all* $0 < r < q$,

$$F \in L_{q,\infty}(\nu_p) \Leftrightarrow M_r(F) \in L_{q,\infty}(\nu_p).$$

Proof. We begin with the classical result (see Stein 1993, Theorem 1, p. 13), that for any measurable function g on $\mathcal{X}$, $|g(x)| \leqslant M_1(g)(x)$ a.e. and, for all $q > 1$, there exists some C_q, depending only on $\mathcal{X}$, such that

$$\left(\int_{\mathcal{X}} M_1(g)^q\, \mathrm{d}\mu\right)^{1/q} \leqslant C_q\left(\int_{\mathcal{X}} |g|^q\, \mathrm{d}\mu\right)^{1/q}.$$

One can obviously deduce, as $M_r(g) = (M_1(|g|^r))^{1/r}$, that for all $0 < r < q < \infty$,

$$\left(\int_{\mathcal{X}} M_r(g)^q \,\mathrm{d}\mu\right)^{1/q} \leqslant C_{q/r}^{1/r}\left(\int_{\mathcal{X}} |g|^q \,\mathrm{d}\mu\right)^{1/q} \leqslant C_{q/r}^{1/r}\left(\int_{\mathcal{X}} M_r(g)^q \,\mathrm{d}\mu\right)^{1/q},$$

with an obvious extension for $q = \infty$.

Now let $0 < r < q < \infty$. As $M_r(F)(j, x) \geqslant |F(j, s)|$ ν_p a.e., we need only prove the lemma in one direction: $\|M_r(F)\|_{L_{q,\infty}(\nu_p)} \leqslant C\|F\|_{L_q,\infty(\nu_p)}$. But

$$\|M_r(F)\|^q_{L_{q,\infty}(\nu_p)} = \sum_{j\in\mathbb{N}} 2^{j(p/2)}\|M_r(F_j)\|^q_{L_q(\mu)} \leqslant C^{q/r}_{q/r}\sum_{j\in\mathbb{N}} 2^{j(p/2)}\|F_j\|^q_{L_q(\mu)} = C^{q/r}_{q/r}\|F\|^q_{L_q(\nu_p)}.$$

The lemma follows using Proposition 8. □

6.3. Weak bodies associated with wavelet coefficients

Let $\mathcal{X}$ again be either of the spaces $\mathbb{R}^d$ or $[0, 1]^d$ and μ the Lebesgue measure. Let $0 < p < \infty$, and let us define on $\mathbb{N} \times \mathcal{X}$ the measure $\nu_p = (\sum_{j\in\mathbb{N}} 2^{j(p/2)}\delta_j) \otimes \mu$. Suppose we have a compactly supported wavelet basis $\psi_{j,k}$. We associate the corresponding Haar wavelet $\chi_{j,k}(x) = 2^{j/2} 1_{[0,1]}(2^j x - k)$. For $f \in L_p(\mathcal{X})$, write the wavelet decomposition as

$$f = \sum_j \sum_k \lambda_{j,k}\psi_{j,k}.$$

We define

$$\sum_k \lambda_{j,k}\psi_{j,k}(x) = \Delta_j(f)(x),$$

$$\sum_k \lambda_{j,k}\chi_{j,k}(x) = \widetilde{\Delta_j(f)}(x).$$

Let us associate the two following functionals:

$$F_f(j, x) = 2^{-j/2}\Delta_j(f)(x);$$

$$\widetilde{F_f}(j, x) = \widetilde{\Delta_j(f)}(x).$$

Then we have the following theorem:

Theorem 3. *For all* $0 < r < q$,

$$\mathcal{W}(F)(p, q) = \mathcal{W}(\tilde{F})(p, q).$$

This is the consequence of Theorem 2 and the following lemma:

Lemma 8. *In the previous notation, for all* x,

$$|\widetilde{\Delta_j}(x)| \leqslant C'_r \|\psi\|_\infty N^{1/r} M_r(\Delta_j)(x) \ a.e.,$$

$$|\Delta_j(x)| \leqslant \|\psi\|_\infty N^{(1/r \vee 1)} M_r(\widetilde{\Delta_j})(x) \ a.e.$$

Recall that

$$\mathrm{supp}(\psi_{j,k}) = \left[\frac{k}{2^j}, \frac{k+N}{2^j}\right]; \ \mathrm{supp}(\chi_{j,k}) = \left[\frac{k}{2^j}, \frac{k+1}{2^j}\right].$$

Proof. A finite-dimensional argument leads straightforwardly, for all $0 < r < \infty$, to the existence of C_r, C'_r such that

$$C_r\left(\int_{[0,N]}\left|\sum_l \alpha_l \psi(u-l)\right|^r \mathrm{d}u\right)^{1/r} \leqslant \int_{[0,N]}\left|\sum_l \alpha_l \psi(u-l)\right|\mathrm{d}u$$

$$\leqslant C'_r\left(\int_{[0,N]}\left|\sum_l \alpha_l \psi(u-l)\right|^r \mathrm{d}u\right)^{1/r}.$$

Moreover, for all $k, j \in \mathbb{Z}$,

$$C_r\left(2^j\int_{[k/2^j,(k+N)/2^j]}\left|\sum_l \alpha_l \psi_{j,l}(u)\right|^r \mathrm{d}u\right)^{1/r} \leqslant 2^j\int_{[k/2^j,(k+N)/2^j]}\left|\sum_l \alpha_l \psi_{j,l}(u)\right|\mathrm{d}u$$

$$\leqslant C'_r\left(2^j\int_{[k/2^j,(k+N)/2^j]}\left|\sum_l \alpha_l \psi_{j,l}(u)\right|^r \mathrm{d}u\right)^{1/r}.$$

We first prove, for all x, that

$$|\widetilde{\Delta_j}(x)| \leqslant C'_r \|\psi\|_\infty N^{1/r} M_r(\Delta_j)(x):$$

Let $x \in [k/2^j, (k+1)/2^j]$. Then

$$|\widetilde{\Delta_j}(x)| = \left|\sum_l \lambda_{j,l}\chi_{j,l}(x)\right| = 2^{j/2}|\lambda_{j,k}| = 2^{j/2}\left|\int \Delta_j \psi_{j,k}\right| \leqslant 2^j \|\psi\|_\infty \int_{[k/2^j,(k+N)/2^j]}\left|\sum_l \lambda_{j,l}\psi_{j,l}\right|.$$

But, using the preceding remarks, this can be bounded by

$$C'_r \|\psi\|_\infty N^{1/r}\left(\frac{2^j}{N}\int_{[k/2^j,(k+N)/2^j]}\left|\sum_l \lambda_{j,l}\psi_{j,l}\right|^r \mathrm{d}u\right)^{1/r} \leqslant C'_r \|\psi\|_\infty N^{1/r} M_r(\Delta_j)(x).$$

Next, we prove that, for all x,

$$|\Delta_j(x)| \leqslant \|\psi\|_\infty N^{(1/r \vee 1)} M_r(\widetilde{\Delta_j})(x).$$

Let $x \in [k/2^j, (k+1)/2^j]$ as before. Then

$$|\Delta_j(x)| = \left| \sum_{k'=k-N+1}^{k} \lambda_{j,k'} \psi_{j,k'}(x) \right| \leqslant 2^{j/2} \|\psi\|_\infty \sum_{k'=k-N+1}^{k} |\lambda_{j,k'}|$$

$$\leqslant 2^{j/2} \|\psi\|_\infty N^{(1-1/r)} + \left(\sum_{k'=k-N+1}^{k} |\lambda_{j,k'}|^r \right)^{1/r}.$$

But we observe that

$$M_r(\widetilde{\Delta_j})(x) \geqslant \left(\frac{2^j}{N} \int_{[(k-N+1)/2^j,(k+1)/2^j]} \left| \sum_l \lambda_{j,l} \chi_{j,l} \right|^r \mathrm{d}u \right)^{1/r} = N^{-1/r} 2^{j/2} \left(\sum_{k'=k-N+1}^{k} |\lambda_{j,k'}|^r \right)^{1/r}.$$

□

Acknowledgements

We are very grateful to the two referees and an associate editor for their very positive and helpful comments.

References

Butzer, P.L. and Berens, H. (1967) *Semi-groups of Operators and Approximation*. Berlin: Springer-Verlag.

Butzer, P.L. and Nessel, R.J. (1971) *Fourier Analysis and Approximation*. New York: Academic Press.

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

Cohen, A., DeVore, R. and Hochmuth, R. (2000) Restricted nonlinear approximation. *Constr. Approx.*, **16**, 85–113.

DeVore, R. and Lorentz, G. (1993) *Constructive Approximation*. Berlin: Springer-Verlag.

Donoho, D. (1996) Unconditional bases and bit-level compression. *Appl. Comput. Harmon. Anal.*, **3**, 388–392.

Donoho, D. (1997) CART and best-ortho-basis, a connection. *Ann. Statist.*, **25**, 1870–1911.

Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D.L. and Johnstone, I.M. (1995) Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, **90**, 1200-1224.

Donoho, D.L. and Johnstone, I.M. (1996) Neoclassical minimax problems, thresholding and adaptive function estimation. *Bernoulli*, **2**, 39–62.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1994) Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B*, **57**, 301–369.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996) Density estimation by wavelet thresholding. *Ann. Statist.*, **24**, 508–539.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998) *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Statist. 129. New York: Springer-Verlag.

Hoffmann, M. (1999a) L_p estimation of the diffusion coefficient. *Bernoulli*, **5**, 447–481.

Hoffmann, M. (1999b) Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, **79**, 135–163.

Hoffmann, M. (1999c) On nonparametric estimation in nonlinear AR(1)-models. *Statist. Probab. Lett.*, **44**, 29–45.

Kerkyacharian, G. and Picard, D. (1992) Density estimation in Besov spaces. *Statist. Probab. Lett.*, **13**, 15–24.

Kerkyacharian, G. and Picard, D. (1993) Density estimation by kernel and wavelets methods: Optimality of Besov spaces. *Stat. Probab. Lett.*, **18**, 327–36.

Lepski, O.V. (1991) Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.*, **36**, 682–697.

Lepski, O.V., Mammen, E. and Spokoiny, V.G. (1997) Optimal spatial adaptation to inhomogenous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, **25**, 929–947.

Mallat, S., Papanicolaou, G. and Zhang, Z. (1998) Adaptive covariance estimation of locally stationary processes. *Ann. Statist.*, **26**, 1–47.

Meyer, Y. (1990) *Ondelettes et Opérateurs*. Paris: Hermann.

Neumann, M.H. (1998) Multivariate wavelet thresholding: A remedy against the curse of dimensionality? Technical report, Humboldt Universität, Berlin.

Neumann, M.H. and von Sachs, R. (1997) Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.*, **25**, 38–76.

Nikol'skii, S.M. (1975) *Approximation of Functions of Several Variables and Imbedding Theorems*. Berlin: Springer-Verlag.

Picard, D. and Tribouley, K. (2000) Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.*, **28**, 298–335.

Stein, E.M. (1993) *Harmonic Analysis*. Princeton, NJ: Princeton University Press.

Received December 1999 and revised August 2001