Minimax or Maxisets ?

Gérard Kerkyacharian [1] and Dominique Picard [2]

October 2, 2001

Abstract

In this paper, we discuss a new way of evaluating the performances of a statistical estimation procedure. This point of view consists in investigating the maximal set where a given procedure has a given rate of convergence. Although the setting is not extremely 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. This notion becomes more difficult even to be practically defined when the loss function is not the $_2$-norm. We explain the difficulties arising there, and suggest a new definition, in the cases of $_p$-norms and point-wise estimation. The connections between maxisets and local oracle inequalities are investigated: we 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 the previous statement on well known efficient adaptive methods: Wavelet thresholding and local bandwidth selection. We can prove local oracle inequalities for these methods and draw the conclusions about there associated maxisets.

Key words and phrases :Adaptive methods, oracle inequalities, saturation sets.

[1] G. Kerkyacharian, Université Paris X-Nanterre, 200 Avenue de la République F 92001 Nanterre cedex et Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599, Université Paris VI et Université Paris VII, 16, rue de Clisson, F-75013 Paris .

[2] D. Picard, Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599, Université Paris VI et Université Paris VII, 16, rue de Clisson, F-75013 Paris.

1 Introduction

The recent appearance of nonparametric estimation methods offering a high degree of adaptivity has given a renewal of interest into minimax theory.

Back to the seventies or eighties, the minimax context was essentially a forest of results from worldwide researchers seeking for solutions to problems where one specified the problem (density estimation, regression, spectral density,...), the risk ($_2$, $_p$ norms), and the functional class (Hölder, Sobolev,...). At this time the impact on the statistical community was not uniformly enthusiastic. The main reasons probably were the disconnection between minimax paradigm and the actual situation when one is confronted with real data: Either minimax estimators where depending on smoothness assumptions which were mostly impossible to verify, or (for some procedures which were really new) were impracticable. At the same time, a real interest in practice for methods developing spatial adaptation had a considerable influence on the statistical community. In the nineties, the development of wavelet methods and in parallel of local bandwidth selection has partially reduced the gap between theory and practice. The minimax paradigm has not disappeared. The general framework was enhanced with new spaces to better reflect the spatial adaptivity (Besov, Triebel,...). Moreover, the general concern of the minimax community for seeking for adaptive procedures is a serious answer to the question of tuning the smoothing parameters. Indeed, the easiest way to theoretically prove the high performances of these procedures for the analysis of functions with inhomogeneous 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 term has been an essential source of inspiration for the construction of these efficient methods.

However, part of the aversions and reluctance of the statistical community remained and some arguments are substantially difficult to deny. The inclination to expect the worst reveals to be generally too pessimistic to accurately reflect the 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 to a priori choose these sets, in an appropriate way. Even in a adaptive context, this difficulty remains.

Our first aim will be to discuss a new way of evaluating the performances of a procedure. This point of view, rather standard in approximation theory (Saturation class linked with an approximation procedure) is more unusual is Statistics. It consists in investigating the maximal set where a procedure has a given rate of convergence. The setting is not extremely 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 also it is less pessimistic. When looking for minimax procedures over a fixed functional set, or adaptive procedures with respect to a range of sets indiced by a smoothing parameter (like Hölder spaces indiced by the smoothness parameter α) we are seeking in fact for the most difficult functions in this set to be estimated by a general procedure. But in fact this set of "bad functions" is strongly depending on the definition of the smoothness : Most defavorable a priori measures or sets of functions in the 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 feel motivated. When seeking for maxisets, we look for functions which are the most difficult to estimate for a *given* procedure. Besides the fact that it is an interesting information for the procedure itself, it has the main advantage 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 side advantage of this point of view 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 prove that maxisets are not only connected to minimax theory, but also with another new and important way of evaluating the performances of statistical procedures e.g. oracle inequalities. The concept of oracle inequality has been introduced in statistics by Donoho and Johnstone (1994) to reflect the idea of performing, as if having an "oracle", "nearly" as the best of a whole class of procedures (see also Donoho and Johnstone 1995). One of the major differences between oracle inequalities and minimax theory is that oracle inequalities are more oriented toward the function to estimate. This notion proves to be very efficient, in many contexts. However it becomes more difficult to use when the loss function is not the $_2$-norm. We are going to explain the difficulties arising there, and suggest a new definition, in the cases of $_p$-norms and point-wise estimation.

Surprisingly, the connections between maxisets and local oracle inequalities are in fact deep important and one of our goal in this paper will be to emphasize them. Especially, they will 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 the previous statement on well known efficient methods:

Wavelet thresholding and local bandwidth selection. From adaptation or minimax point of view, all these procedures are equivalent. We can prove local oracle inequalities for both methods as well as for an hybrid procedure, which proved in different contexts to be of particular interest (especially for confidence interval, purposes, see Picard, Tribouley (2000). The maxisets of the thresholding procedure can be precisely identified. For the two other ones, we do not obtain a complete identification. However, this allows us to formulate the following concluding remarks: As far as maxisets are concerned, the local bandwidth selection and the hybrid procedure are at least as good as the thresholding. Whether they are strictly better is an opened question, as well as the comparison between them.

The paper is organized as follows: Section 2 is devoted to maxisets. Especially, the explicit maxisets for linear kernel methods as well as thresholding procedures are given.
Section 3 concerns oracle inequalities. We provide definitions of such inequalities for $_p$-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.
Sections 5 and 6 investigates the respective positions of the functional spaces appearing in the definition of the maxisets, and the consequences on the comparisons on the procedures.

2 Definition and examples of Maxisets, comparisons of procedures

It is a classical topic, in Approximation Theory, to study, for an a family of operators U_n in some functional space X, the set of functions $f \in X$, such that

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

were ϵ_n is a sequence of positive numbers, decreasing to 0. This is known as the saturation class linked with the sequence U_n, and the rate ϵ_n. See for example Butzer, Berens (1967), Butzer, Nessel (1971) , DeVore, Lorentz (1993).

We will, now, give the definition as well as a motivation of the maxisets. This 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's 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 to 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 *Let us define the* **maxiset** *associated with the sequence $\hat{q}_n$, the loss function ρ,*

the rate α_n and the constant T as the following set:

$$MS(\hat{q}_n, \rho, \alpha_n)(T) = \{\theta \in \Theta,\ \sup_n \ {}^n_\theta\rho(\hat{q}_n, q(\theta))(\alpha_n)^{-1} \leq T\}$$

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 homogeneous loss functions and large enough constant T. Although it might be useful and interesting to investigate more precisely the domains where the rate is precisely not $n^{-1/2}$ (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 it is the case in a minimax framework, we choose to settle the problem in a very wide context: The parameter set Θ can be very large, such as the set of bounded, measurable functions. Then, the functional set (maxiset) is associated with the procedure in a genuine way. Let us start with 2 examples :

2.1 Density Estimation: Kernel methods.

Let $X_1 \ldots, X_n$ be n i.i.d. random variables, having the density f. We want to estimate the density f. Let us fix $2 \leq p < \infty$, and investigate the problem with the $_p$-loss: i.e. for a procedure $\hat{f}_n$,

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

We take as set of parameters Θ, the set of all densities included in a (large) $_p$-ball. This is reasonable, given our choice for the loss function. We are going to 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. $E(u, v) = \sum_{k\in} \ \phi(u-k)\phi(v-k)$), or the following convolution : $E(u, v) = E(u-v)$. $j(n)$ is an increasing sequence : $2^{j(n)} = n^{(1-\alpha)}$, with $\alpha \in (0, 1)$ (see Kerkyacharian, Picard 1992).

Let us denote by $B_{s,p,q}$ the Besov space and $B_{s,p,q}(M)$ the associated ball of radius M (for the definition and properties of Besov spaces, we refer to Meyer(1990), Nikolskii(1975)).

Then a consequence of Theorem 2.1 in Kerkyacharian, Picard (1993) is the following result: (see also Härdle, Kerkyacharian, Picard, Tsybakov (1998) ch 10.)

Proposition 1 *Under the following conditions:*

- E *is compactly supported.*
- $\int (y-x)^k E(x,y)dy = \delta_{0,k}$, *for all* $k = 0, 1, \ldots, N$.
- $E \circ E_j = E_j \circ E$, *for all* $j \geq 0$. *(Here* $E \circ E_j$ *stands for the composition of* E *and* E_j*.)*
- $x \to E(x,y)$ *is* N *times continuously differentiable.*
- $\alpha_n := (\frac{2^{j(n)}}{n})^{p/2} = n^{-\alpha p/2}$.

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

Here and throughout the paper

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

means:

(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 observe the following differential equation,

$$dY_t^n = f(t)dt + \frac{1}{\sqrt{n}} dW_t, \ t \in [0,1], \tag{2}$$

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 the previous example, the problem with the $_p$-loss. Let us fix, as above α in $(0,1)$. We take as set of parameters Θ, a ball of the space $B_{\alpha/2,p,\infty}$. This set corresponds to the idea of a minimal regularity which is always necessary for the non linear procedures. Notice that $\alpha/2$ is always smaller (and often much smaller) than s introduced in the proposition above. In particular if $\alpha < 2/p$, Θ contains discontinuous functions. For a pair of scaling function and wavelet ϕ, ψ, let us define the following sequence of procedures:

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

$$\text{where} \quad g_{jk} = \psi_{jk}, \quad \text{for} \quad j \geq 0, \ g_{-1,k} = \phi_{0k},$$

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

$$t_n = \sqrt{\tfrac{\log n}{n}}, \ 2^{J_n} \leq t_n^{-2} \leq 2^{J_n+1}$$

Let us introduce the following sets of functions:

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

Obviously, for $p = 2$, $\mathcal{W}^*(2,q)(M)$ selects the functions such that their total number of β's greater (in modulus) than λ in the positive scales in less than $(M\lambda^{-1})^q$. For the cases $p > 2$, we also "count" the β's greater than λ, but with a penalization in the counting for the large scales. These spaces prove to have a special importance in approximation theory (Cohen, DeVore, Hochmuth, 1997), coding (Donoho, 1996) and estimation (Donoho and Johnstone, 1996). Then a consequence of Theorem 7 in Cohen, DeVore, Kerkyacharian and Picard (1999) is the following result:

Proposition 2 *For* $\alpha_n = (t_n) = (\frac{\log n}{n})^{(1/2)}$,

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

3 Local Oracle inequalities and Maxisets

This section is divided into 2 separated parts. The first part is essentially concerned with oracle inequalities. Particularly, we begin with the standard case of $_2$ oracle inequalities. Then, we explain a way to overcome the difficulty to generalize to other norms, and to local inequalities. This part is a priori essentially disconnected with the previous section about maxisets. The relations between the 2 notions are clarified in the second part of this 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 $\mathcal{X} \mapsto \ .$ $\mathcal{X}$ is a measurable space. It is 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 $\mathcal{V}$ (for instance $\mathcal{V} = \ _p$). We consider a sequence of linear operators $E_j, j \geq 0$ associating to 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 ($E_j f(x) = \int \sum_{k\in} \ 2^j \phi(2^j u - k)\phi(2^j x - k) f(u) du.$), or the following convolution : $E_j f(x) = \int 2^j E(2^j(u-x)) f(u) du$.

3.1 From $_2$-oracle inequalities to local ones.

3.1.1 $_2$-oracle inequalities.

Following Donoho and Johnstone (1994), we say that the estimate $\hat{f}$ satisfies a $_2$-**oracle inequality** with the class $\mathcal{C}$ of estimators at the rate c_n, if for all $n \geq 1$,

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

(4) is exactly expressing that, up to rate c_n, $\hat{f}$ is behaving as the oracle estimate of the class $\mathcal{C}$ -i.e. the best estimate among the class $\mathcal{C}$ (as if an oracle was telling 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 to estimate).

As a prototype example, it can easily be proved that the wavelet thresholding estimator in the white noise model satisfies a $_2$-oracle inequality with the class $\{\hat{E}_j^n = \int E_j(x,t)dY_t^n, j \geq 0\}$ of estimators at the rate

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

if the E_j's are the projections on the V_j's.

Hence we immediately see that oracle inequalities may be a very useful property for a procedure. However, it seems that up to now, there is no full agreement in the statistical community about the most suitable distance to reflect well the visual properties of estimation procedures. In particular, two functions may look very differently although they are very close in $_2$ norm. As a consequence, a natural question is : Can we also prove oracle inequalities for different norms ($_p$ for instance), and also 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 $_p$norms, with $1 \leq p \leq 2$, because of the finiteness of the measure μ. However, it does not tell anything about the other norms. To be able to consider oracle inequalities for general $_p$ norms, it is more convenient to have first a slightly different understanding of (4).
Let us evaluate in the prototype example above (still in the white noise model), the quantity :

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

We have, by standard calculations,

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

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 $_2$ norm and a family of projection operators

on increasing subspaces. This precisely will be the difficulty when we want to extend to other situations. Let us introduce :

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

So for $\lambda > 0$:

$$\text{for} \ \ j \geq 1 : \{j_\lambda(f) = j\} \iff 2^{-(j-1)/2}\|E_{j-1}f - f\|_2 > \lambda \geq 2^{-j/2}\|E_j f - f\|_2,$$

$$\text{for} \ \ j = 0 : \{j_\lambda(f) = 0\} \iff \{\lambda \geq \|E_0 f - f\|_2\}$$

and, for

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

it is not difficult, using the following lemma, to prove that:

$$\frac{c\ 2^{j_n^*}}{2n} \ \leq \inf\{ \ _n\|\hat{E}_j^n - f\|_2^2, \ j \geq 0\} \ \leq \frac{2c\ 2^{j_n^*}}{n}.$$

Lemma 1 *Let* a_j *and* $b_j \geq 0$, $j \in$ *be respectively non increasing and non decreasing sequences. Let* $j^* = \inf\{j \in \quad , \ a_j \leq b_j\}$. *Then :*

$$b_{j^*-1} \leq \inf\{j \in \quad , \ a_j + b_j\} \leq 2b_{j^*}.$$

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

Proof of the lemma: Clearly $b_{j^*-1} < a_{j^*-1}$, if $j^* > 0$, and $b_{j^*} \geq a_{j^*}$. So :

$$\inf\{j \in \quad , \ a_j + b_j\} \leq a_{j^*} + b_{j^*} \leq 2b_{j^*}$$

On the other side : $j \geq j^* \Rightarrow a_j + b_j \ \geq b_j \geq b_{j^*} \geq b_{j^*-1}$.

And $\quad 0 \leq j < j^* \Rightarrow \quad a_j + b_j \geq \ a_j \ \ \geq a_{j^*-1} > b_{j^*-1}$.

∎

Let us 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** $_2$ **norm,** *on the space* $\mathcal{V}$, *the class* E_j *of estimators and at the rate* $c_n = 1 + \log n$ *if :*

$$_n\|\hat{f} - f\|_2^2 \leq C c_n 2^{j_n^*} \lambda_n^2 \quad \forall\ f \in \mathcal{V} \tag{6}$$

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

3.1.2 $_p$ oracle inequalities associated to a sequence of operators E_j.

Let us begin by defining the $_p$ analogue of $j_\lambda(f)$.

Let $F(f)(j)$ be a non negative, non increasing functional defined on .

An important example is:

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

Now, let

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

So ;

$$\text{for } \ j \geq 1 : \{j_\lambda^F(f) = j\} \iff \{F(f)(j-1) > \lambda \geq F(f)(j)\},$$

$$\text{for } \ j = 0 : \{j_\lambda^F(f) = 0\} \iff \{\lambda \geq F(f)(0)\}$$

Again, let us define:

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

This leads to the following definition:

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

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

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

Remarks :

- This definition easily generalizes to the case $p = \infty$, with the usual modification consisting of ignoring all the p-th powers in (9) and (10). They also are embedded: Because of the finiteness of the measure μ, satisfying an $_p$ oracle inequality implies satisfying an $_q$ oracle inequality for any $1 \leq q \leq p$.

- We remark that (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 $E_j f$.

- If we compare (9) with (6), we notice that the 2 right hand sides are equivalent. If we now compare (9) with (4), we can't deny that there might be a loss, since the only thing that can be said is, there exists C with

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

For $p = 2$, the 2 quantities where of the same order. For $p \neq 2$, as we are considering $_p$ norms we can only hope that they do not differ a lot -and also observe that this is confirmed by the minimax rates for standard classes of functions-. $\diamondsuit$

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

We'll mimic locally what has been done above.

1. Let $F(f)(j,x)$ be a non negative functional defined on $\times \mathcal{X}$, 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

$$\tilde{F}(f)(j,x) = \sup_{j' \geq j} 2^{-j'/2} |E_{j'} f(x) - f(x)| \tag{11}$$

   F is now a "local" functional.

2. Let

$$j_\lambda^F(f,x) = \inf\{j \in \ , F(f)(j,x) \leq \lambda\}.$$

3. Let now

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

4. For practical reasons, it is generally necessary to introduce in addition 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 $p \geq 1$ 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 2 following inequalities are true for all $n \geq 1$.*

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

$$\| \sup_{j' \geq j} |E_{j'} f - f| I\{\ j_\lambda^F(f,.) = j\}\|_p^p \leq C'(2^{j/2}\lambda)^p \mu\{x,\ j_\lambda^F(f,x) = j\} \quad \forall\, \lambda > 0,\ \forall\, j \geq 0,\ \forall\, f \in \mathcal{V} \tag{14}$$

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 2 differences. The first one is the presence of 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 in (9) without changing the rates of convergence, because of (10). The second difference is that a logarithmic factor is now appearing in the rate t_n whereas c_n has now disappeared. - However, not completely if we notice that now t_n is replacing λ_n.-

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

$\diamondsuit$

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 are true for all $n \geq 1$ and all $f \in \mathcal{V}$.*

$$P_n\{(2^{j/2}t_n)^{-1} \sup_{x,\ j_n^*(x)=j} |\hat{f}_n(x) - f(x)| \geq \lambda\} \leq C \exp -\lambda^2/(2v_0) \quad \forall\, \lambda \geq \lambda_0,\ \forall\, J_n \geq j \geq 0 \qquad (15)$$

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

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 "exponential" oracle condition implies satisfying a local oracle of order p for any $p \geq 1$, since especially $\mu(\mathcal{X}) < \infty$.

3.2 Local-oracle inequalities and maxisets.

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

3.2.1 "Besov bodies".

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

$$\mathcal{B}_{\gamma,r,\infty}(M) = \{f \in \mathcal{V}, \|E_j f - f\|_{\ r(d\mu)} \leq M 2^{-j\gamma},\ \forall\ j \geq 0\}$$

Though, obviously depending on the sequence of kernels E_j, $\mathcal{B}_{\gamma,r,\infty}$ is intentionally denoted 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:

$$\mathcal{B}_{\gamma,r,m}(M) = \{f \in \mathcal{V}, \sum_{j\geq 0} (2^{j\gamma} \| E_j f - f \|_{\ r(d\mu)})^m \leq M^m\}$$

3.2.2 Weak "Besov bodies".

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

$$_{q,\infty}(\nu) = \{g;\ \sup_{\lambda>0} \lambda^q \nu\{|g| \geq \lambda\} < \infty\}. \tag{17}$$

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

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

where δ is the Dirac measure.

For F a non negative functional defined on $\times \mathcal{X}$, (see §3.1.1), let us put, for $p > q > 0$,

$$\begin{aligned} \mathcal{W}(F)(p,q) &= \{f \in \mathcal{V},\ F(f) \in \ _{q,\infty}(\nu_p)\} \\ &= \{f \in \mathcal{V}, \sup_{\lambda>0} \lambda^q \sum_{j\geq 0} 2^{j(p/2)} \mu\{x, F(f)(j,x) > \lambda\} < \infty\}, \end{aligned} \tag{18}$$

and the associated ball,

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

Let us investigate 2 examples :

1. The first one is associated with the following functional (notice that this one is not necessarily monotone)

   $$F^1(f)(j,x) = 2^{-j/2} |E_j f(x) - f(x)|, \text{ and associated } \mathcal{W}(F^1)(p,q)(M).$$

   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 $\gamma = \frac{\alpha}{2(1-\alpha)}$ then f belongs to $\mathcal{W}(F^1)(p,q)(M)$. Hence, in this

case, $\mathcal{W}(F^1)(\frac{q}{1-\alpha},q)(M)$ appears as a weak analogue of $\mathcal{B}_{\gamma,q,q}(M)$. We summarize this fact in the following inclusion :

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

Indeed, we have, as $p-q=\frac{\alpha q}{1-\alpha}$,

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

2. If ψ is a wavelet, and β_{jk} denotes the wavelet coefficient of f ($\beta_{jk}=\int f\psi_{jk}$), and $\chi_{jk}(x)=2^{j/2}I\{2^jx-k\in[0,1]\}$ is the Haar scaling function, let us consider the case of the following functional (also not necessarily non increasing)

$$F^2(f)(j,x)=2^{-j/2}\sum_k|\beta_{jk}|\chi_{jk}(x) \text{ and associated } \mathcal{W}(F^2)(p,q)(M).$$

We notice that $\mathcal{W}(F^2)(p,q)(M)$ coincides with the set $\mathcal{W}^*(p,q)(M)$ introduced in section 2, (3).

In section §6, we'll investigate more precisely the weak besov bodies for some classes of local functionals F. Especially, we'll establish that they happen to coincide rather often. For instance, we'll prove that if the E_j of the first example are the projection 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 $_p$-loss and the rate $(t_n)^{\alpha p}$. (We recall that $t_n=(\frac{\log n}{n})^{1/2}$).

$$MS(\hat{f}_n,p,\alpha)(T)=\{f\in\mathcal{V},\ \sup_n\ {}_n\|\hat{f}_n-f\|^p_{p(d\mu)}(t_n)^{-\alpha p}\leq T\}$$

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

Proposition 3 *Let β be a positive constant. If 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 $2^{J_n} < t_n^{-\beta} \leq 2^{J_n+1}$, 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 : (the constant C is coming from inequality (13))

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

- Belonging to 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 $\quad \times \mathcal{X}$ defined as above by the formula

$$\nu_0 = (\sum_{j\in} \delta_j) \otimes \mu.$$

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

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

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

$\diamondsuit$

Proposition 4 *If 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 $2^{J_n} < t_n^{-\beta} \leq 2^{J_n+1}$ and*

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

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

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

Remark :

The constant M' may be chosen such that : (the constant C here is coming from inequality (15))

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

$\diamondsuit$

3.2.4 Proof of the Proposition 3

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

$$\begin{aligned}
{}_n\int |\hat{f}_n(x) - f(x)|^p d\mu &= \textstyle\sum_{j\geq 0} {}_n\int |\hat{f}_n(x) - f(x)|^p I\{j_n^*(x) = j\} d\mu \\
&\leq C \textstyle\sum_{j\geq 0} \int \{(2^{j/2}t_n)^p + |E_j f(x) - f(x)|^p \\
&\quad + |E_{J_n} f(x) - f(x)|^p\} I\{j_n^*(x) = j\} d\mu \\
&\leq C t_n^p \{2 \textstyle\sum_{j\geq 0} 2^{jp/2} \mu\{j_n^*(x) = j\} + \|E_{J_n} f - f\|_p^p\} \\
&\leq 2C(1 + 2M'^q) t_n^{p-q} + CM^p 2^{-\frac{\alpha}{\beta} J_n p} \leq C(M) t_n^{\alpha p}
\end{aligned}$$

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

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

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

3.2.5 Proof of the Proposition 4

- Because of the definition of $\mathcal{B}_{\frac{\alpha}{2(1-\alpha)},\infty,\infty}(M)$, we get : For all j', $2^{-j'/2}|E_{j'} f(x) - f(x)| \leq M 2^{-j'[1/2(1-\alpha)]}$. 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 $j_n^*(x) \leq j_0$ such that $2^{j_0} \leq C t_n^{-2(1-\alpha)}$

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

- Using the previous remarks, we get:

$$\begin{aligned}
{}_n\|\hat{f}_n(x) - f(x)\|_\infty &= {}_n\sup_{j\geq 0} \sup_x |\hat{f}_n(x) - f(x)| I\{j_n^*(x) = j\} \\
&= {}_n\sup_{0\leq j\leq j_0} \sup_x |\hat{f}_n(x) - f(x)| I\{j_n^*(x) = j\} \\
&\leq \textstyle\sum_{0\leq j\leq j_0} {}_n\sup_x |\hat{f}_n(x) - f(x)| I\{j_n^*(x) = j\} \\
&\leq \textstyle\sum_{0\leq j\leq j_0} C 2^{j/2} t_n \leq C t_n^\alpha
\end{aligned}$$

4 Applications to wavelets thresholding and Lepski's procedures

In this section, we'll prove that for wavelet thresholding, as well as for adaptive local bandwidth selection, we are able to prove a local oracle inequality. We first state the assumptions on the sequence of models (which will be roughly the same for the different procedures). We

produce standard examples where these conditions are fulfilled, and prove then 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 (7). Surprisingly, F^T is much less intuitive since it requires the introduction of the maximal function (see (29)). We also investigate the behavior of an hybrid procedure, intermediate between the 2 previous ones, which proves to be very efficient for the construction of confidence intervals (see Picard, Tribouley 2000).

4.1 Assumptions on the sequence of models.

Our assumptions on the sequence of models will only take place with respect to its ability of estimating the $E_j f$. Let $p \geq 1$ be fixed.

Let us also fix a constant $K \geq 4$ and an increasing sequence of integers J_n. These last 2 quantities will appear as tuning quantities for both procedures.

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

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

In either case, we'll assume that there exist some constants C_1 and C_2 such that :

1. Moment Inequality of order p:

   There exists C_1 such that for all $n \geq 1$, for all $\delta \in \mathcal{C}_j$, for all j, $0 \leq j \leq J_n$,

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

2. Concentration Inequality of order γ:

   There exist $\gamma > 0$ and C_2 such that for all $n \geq 1$, for all $\delta \in \mathcal{C}_j$, for all j, $0 \leq j \leq J_n$,

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

Remarks :

- The expectation and probability considered above, are taken when f is the true parameter. Notice also, that if the condition (19) above is fulfilled for one p, it is automatically satisfied for any $p' \leq p$.

- It is worthwhile to notice that neither of the 2 conditions above implies the other one. However, it is easy to verify that the following one implies both of them for any $p \geq 1$:

  There exist $C_3,\ v_0 > 0,\ \lambda_0 \geq 0$ such that for all $n \geq 1$, for all $\delta \in \mathcal{C}_j$, for all $j,\ 0 \leq j \leq J_n$, for all $\lambda \geq \lambda_0$,

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

  $\diamondsuit$

4.1.1 Examples of Models where such conditions are satisfied

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

1. White Noise Model, see section 2.2

$$dY_t^n = f(t)dt + \frac{1}{\sqrt{n}} dW_t,\ t \in [0,1] \tag{22}$$

2. Equispaced regression model, with gaussian errors,

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

3. Density Model, see section 2.1

$$Y_1, \ldots, Y_n \text{ independent, identically distributed, with density } f \tag{24}$$

But with more elaborate arguments one can also prove that they are satisfied for stationary processes of spectral density f, or evolutionary spectra, Neumann and von Sachs (1997) , locally stationary processes Mallat, Papanicolaou, Zhang, (1998), partially observed diffusion models (Hoffmann, 1999 a,b,c) multivariate extensions ($t \in [0,1]^d$) (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, Mammen and Spokoiny (1997). It is associated with a general sequence of operators E_j.

Let t_n still be $(\frac{\log n}{n})^{1/2}$ and let us define:

1. The index $\hat{j}(x)$ as the minimum of admissible j's at the point x, where

   $j \in \{0, \ldots, J_n\}$ is admissible at the point x if

$$|\hat{E}^n_{j'}(x) - \hat{E}^n_{j"}(x)| \leq K 2^{j"/2} t_n, \ \forall j', \ j"; \ j \leq j' \leq j" \leq J_n$$

2. The following 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 $2^{J_n} < t_n^{-\beta} \leq 2^{J_n+1}$ for some positive constant β.

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

Proposition 5 *As soon, as the inequalities (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 $1 \leq p \leq p^*$ the local oracle conditions of order p of definition 4 on the space $\mathcal{V} = \mathcal{M}$, associated with the sequence of operators E_j and the functional*

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

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

4.2.1 Proof of the Proposition 5.

First, let us remark, that, as mentioned above, because of the precise form of $F^L(f)$, we have just to establish (13), as (14) is naturally fulfilled in this case. Let us recall, (see section 3.1.3) $j_n^*(x) = j_{t_n}^{F^L}(x)$. Notice that $j_n^*(x)$ is finite since $f \in \mathcal{M}$.
We'll begin this proof by the following lemma.

Lemma 2 *Under the conditions above,*

$$P\{\hat{j}(x) > j_n^*(x)\} \leq \frac{C J_n^2}{n^\gamma}.$$

Proof of the Lemma:

We remark that, by definition of $\hat{j}(x)$, when $\hat{j}(x) \geq 1$, $\hat{j}(x) - 1$ is not admissible, (when $\hat{j}(x)$ is admissible) so there exists j', $\hat{j}(x) - 1 < j' \leq J_n$ such that $|\hat{E}^n_{j'}(x) - \hat{E}^n_{\hat{j}(x)-1}(x)| \geq K t_n 2^{j'/2}$. In addition, if $\hat{j}(x) - 1 \geq j^*_n(x)$, we have:

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

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

$$|\hat{E}^n_{j'}(x) - \hat{E}^n_{\hat{j}(x)-1}(x)| \geq K t_n 2^{j'/2} \quad \text{and } |E_{j'}f(x) - E_{\hat{j}(x)-1}f(x)| \leq \frac{K}{2} t_n 2^{j'/2}.$$

Hence, using (20)

$$\begin{aligned}
&P\{\hat{j}(x) > j^*_n(x)\} = \\
&\quad P\left\{\cup_{j^*_n(x) \leq j < j' \leq J_n} \{|\hat{E}^n_{j'}(x) - \hat{E}^n_j(x)| \geq K t_n 2^{j'/2}; |E_{j'}f(x) - E_j f(x)| \leq \tfrac{K}{2} t_n 2^{j'}\}\right\} \\
&\quad \leq \textstyle\sum_{j,j' \leq J_n} P\left\{|\hat{E}^n_{j'}(x) - E_{j'}f(x) - \hat{E}^n_j(x) + E_j f(x)| \geq \tfrac{K}{2} t_n 2^{j'/2}\right\} \\
&\quad \leq \textstyle\sum_{j,j' \leq J_n} P\left\{|\hat{E}^n_{j'}(x) - E_{j'}f(x)| \geq \tfrac{K}{4} t_n 2^{j'/2}\right\} + P\left\{|\hat{E}^n_j(x) - E_j f(x)| \geq \tfrac{K}{4} t_n 2^{j/2}\right\} \\
&\quad \leq C J_n^2 n^{-\gamma}
\end{aligned}$$

Let us now investigate the 2 different cases:

Case 1: $j^*_n(x) \leq J_n$**.** In this particular case, we can divide $_n|\hat{f}^L(x) - f(x)|^p$ into 2 terms:

$$\begin{aligned}
T_1 &= \textstyle\sum_{0 \leq j' \leq j^*_n(x)} \ _n|\hat{E}^n_{j'}(x) - f(x)|^p I\{\hat{j}(x) = j'\}, \\
T_2 &= \textstyle\sum_{j^*_n(x) < j' \leq J_n} \ _n|\hat{E}^n_{j'}(x) - f(x)|^p I\{\hat{j}(x) = j'\}
\end{aligned}$$

1. To bound T_1 let us remark that :

$$\begin{aligned}
|\hat{E}^n_{j'}(x) - f(x)|^p \leq \ & 3^{p-1}\Big(|\hat{E}^n_{j'}(x) - \hat{E}^n_{j^*_n(x)}(x)|^p \\
& + |\hat{E}^n_{j^*_n(x)}(x) - E_{j^*_n(x)}f(x)|^p + |E_{j^*_n(x)}f(x) - f(x)|^p\Big)
\end{aligned}$$

On the set $\{\hat{j}(x) = j'\}$, $|\hat{E}^n_{j'}(x) - \hat{E}^n_{j^*_n(x)}(x)| \leq K t_n 2^{j^*_n(x)/2}$.

For the second term, we'll use (19) and the Cauchy-Schwartz inequality.

More precisely,

$$\begin{aligned}
T_1 \ &\leq 3^{p-1}\Big((K t_n 2^{j^*_n(x)/2})^p P\{\hat{j}(x) \leq j^*_n(x)\} + \\
&\quad \textstyle\sum_{0 \leq j' \leq j^*_n(x)} (\frac{2^{j^*_n(x)}}{n})^{p/2} (P(\hat{j}(x) = j'))^{1/2} + |E_{j^*_n(x)}f(x) - f(x)|^p\Big) \\
&\leq 3^{p-1}\Big((K t_n 2^{j^*_n(x)/2})^p + (\tfrac{2^{j^*_n(x)}}{n})^{p/2} j^*_n(x)^{1/2} P\{\hat{j}(x) \leq j^*_n(x)\}^{1/2} + |E_{j^*_n(x)}f(x) - f(x)|^p\Big) \\
&\leq 3^{p-1}\Big((K t_n 2^{j^*_n(x)/2})^p + 2^{j^*_n(x)p/2}(t_n/\log 2)^p + |E_{j^*_n(x)}f(x) - f(x)|^p\Big)
\end{aligned}$$

For the last inequality, we used that $j^*_n(x) \leq J_n \leq \beta \frac{\log n}{\log 2}$, and $p \geq 1$.

2. For the term T_2, lemma 2 will be the key point. Using, in addition, Cauchy-Schwartz inequality, (19) and the definition of $\hat{j}(x)$ we obtain:

$$\begin{aligned} T_2 &\leq \sum_{j_n^*(x)+1\leq j'\leq J_n} 2^{p-1}\Big\{[\ _n|\hat{E}^n_{j'}(x) - E_{j'}f(x)|^{2p}]^{1/2} P\{\hat{j}(x)=j'\}^{1/2} \\ &\quad + |E_{j'}f(x) - f(x)|^p P\{\hat{j}(x)=j'\}\Big\} \\ &\leq 2^{p-1}\Big\{[\textstyle\sum_{j_n^*(x)+1\leq j'\leq J_n}\ _n|\hat{E}^n_{j'}(x) - E_{j'}f(x)|^{2p}]^{1/2}[P\{j_n^*(x)+1\leq \hat{j}(x)\leq J_n\}]^{1/2} \\ &\quad + \textstyle\sum_{j_n^*(x)+1\leq j'\leq J_n}(K2^{j'}t_n/4)^p P\{\hat{j}(x)=j'\}\Big\} \\ &\leq 2^{p-1}C\left\{2^{J_n p/2}n^{-p/2}[\tfrac{CJ_n^2}{n^\gamma}]^{1/2} + (K2^{J_n}t_n/4)^p\tfrac{CJ_n^2}{n^\gamma}\right\} \end{aligned}$$

This concludes this case since, for $\gamma > p\beta/2$, the RHS of the last inequality is easily bounded by $Ct_n^p \leq Ct_n^p 2^{j_n^*(x)p/2}$.

Case 2: $j_n^*(x) > J_n$**.** This case is parallel to the previous one except that the term T_2 has disappeared. Again, we remark:

$$\begin{aligned} |\hat{E}^n_{j'}(x) - f(x)|^p \leq\ & 3^{p-1}\Big(|\hat{E}^n_{j'}(x) - \hat{E}^n_{J_n}(x)|^p \\ & + |\hat{E}^n_{J_n}(x) - E_{J_n}f(x)|^p + |E_{J_n}f(x) - f(x)|^p\Big) \end{aligned}$$

On the set $\{\hat{j}(x)=j'\}$, $|\hat{E}^n_{j'}(x) - \hat{E}^n_{J_n}(x)| \leq Kt_n 2^{J_n/2}$.

For the second term, we again use (19) and the Cauchy-Schwartz inequality.

$$\begin{aligned} \textstyle\sum_{0\leq j'\leq J_n}\ _n|\hat{E}^n_{j'}(x) - f(x)|^p I\{\hat{j}(x)=j'\} &\leq 3^{p-1}\Big((Kt_n2^{J_n/2})^p + \\ &\quad \textstyle\sum_{0\leq j'\leq J_n}(\frac{2^{J_n}}{n})^{p/2}(P(\hat{j}(x)=j'))^{1/2} + |E_{J_n}f(x) - f(x)|^p\Big) \\ &\leq 3^{p-1}\Big((Kt_n2^{J_n/2})^p + (\tfrac{2^{J_n}}{n})^{p/2}J_n^{1/2} + |E_{J_n}f(x) - f(x)|^p\Big) \\ &\leq 3^{p-1}\Big((Kt_n2^{j_n^*(x)/2})^p + 2^{j_n^*(x)p/2}(t_n/\log 2)^p + \\ &\quad |E_{J_n}f(x) - f(x)|^p) \end{aligned}$$

4.3 Thresholding Wavelet Coefficients.

Various descriptions of this procedure can be found, for instance in Donoho, Johnstone, Kerkyacharian, Picard (1994), (1996), in different frameworks, (white noise, equispaced regression, density). As above, we'll consider $t_n = (\frac{\log n}{n})^{1/2}$, and fix J_n such that $2^{J_n} < (\frac{n}{\log n})^\beta \leq 2^{J_n+1}$. Notice that in the papers above generally we have $\beta = 2$.

The space $\mathcal{X}$ is here $[0,1]$, equipped with the Lebesgue measure, the E_j's here are the projection kernels on the spaces V_{j-1} (for $j \geq 1$, $E_0 = E_1$) 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 the conditions (19) and

(20) on the sequence of models, for this particular E_j's and recall the following thresholding procedure:

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

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

We'll prove an analogue of Proposition 5. Here, the difficulty lies in the definition of $F(f)(j,x)$. We have to be a little more 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 $\{\psi_{jk},\ k \in \ \}$.If f in a function of $\ _2$, $\Delta_j f = \sum_k \beta_{jk} \psi_{j,k}$ 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 $\ $, $r > 0$, let us define the following 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 $M_r(g) \ge |g|$ *a.e.*, and enjoys the following nice properties.

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

$$|\Delta_j f(x)| \ \le \ C_r M_r(\tilde{\Delta}_j f)(x) \quad \forall\ x\ a.e. \tag{26}$$

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

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

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

Remark :

The last inequality (28) states the equivalence of the $\ _q$ norm of f and $M_r(f)$. This result is classical in Harmonic Analysis. Its proof can be found in Stein, E. (1993). The proof of (26) and (27) is given in section 6.3, lemma 8. $\diamondsuit$

Proposition 6 *As soon, as 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 $_\infty(M)$ of functions bounded by M on $[0,1]$, associated with the sequence J_n and the local functional*

$$F^T(f)(j,x) := (K/2C_r')2^{-j/2} \sum_{j' \geq j} M_r(\Delta_{j'} f)(x). \tag{29}$$

For any choice of $r < p$, $1 \leq p \leq p^*$.

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

4.3.1 Proof of the Proposition 6.

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

First, here, there is a need for proving (16).

$$\begin{array}{ll} \int \sup_{j' \geq j} |E_{j'} f(x) - f(x)|^p I\{x,\ j_\lambda(f,x) = j\} d\mu(x) & \leq \int I\{x,\ j_\lambda(f,x) = j\} [\sum_{j' \geq j} |\Delta_{j'} f(x)|]^p d\mu(x) \\ & \leq \int I\{x,\ j_\lambda(f,x) = j\} [\sum_{j' \geq j} M_r(\Delta_{j'} f)(x)|]^p d\mu(x) \\ & \leq \int I\{x,\ j_\lambda(f,x) = j\} (2^{j/2}\lambda)^p d\mu(x) \\ & \leq C'(2^{j/2}\lambda)^p \mu\{x,\ j_\lambda(f,x) = j\} \end{array}$$

We have, as above to distinguish the 2 different cases: $j_n^*(x) \leq J_n$ and $j_n^*(x) > J_n$. We'll only investigate the first case. The second can easily be treated with the same trick as above. In this particular case, we can divide $E_n|\hat{f}^T(x) - f(x)|^p$ into 3 terms:

$$\begin{array}{ll} e_1 & = 3^{p-1} \ _n|\sum_{-1 \leq j \leq j_n^*(x)} \sum_k (\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq Kt_n\} - \beta_{jk}) g_{jk}(x)|^p, \\ e_2 & = 3^{p-1} \ _n|\sum_{j_n^*(x)+1 \leq j \leq J_n} \sum_k (\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq Kt_n\} - \beta_{jk}) g_{jk}(x)|^p \\ e_3 & = 3^{p-1} |E_{J_n} f(x) - f(x)|^p \end{array}$$

We'll prove the following lemma.

Lemma 4 *Under the conditions above, $\forall n \geq 1,\ \forall j,\ -1 \leq j \leq J_n,\ \forall k$,*

$$_n|\hat{\beta}_{jk} - \beta_{jk}|^{2p} \leq Cn^{-p} \tag{30}$$

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

$$_n|\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq Kt(n)\} - \beta_{jk}|^p \leq C\{t_n^p + n^{-\gamma'}\} \text{ for } \gamma' = \gamma/2 + p/2. \tag{32}$$

Proof of the Lemma:

1. (30) directly follows from condition (19), as well as (31) follows from (20).

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

(a) In the first case, we write,

$$\begin{aligned} {}_n|\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq Kt_n\} - \beta_{jk}|^p &= {}_n|\hat{\beta}_{jk} - \beta_{jk}|^p I\{|\hat{\beta}_{jk}| \geq Kt_n\} + |\beta_{jk}|^p P\{|\hat{\beta}_{jk}| < Kt_n\} \\ &\leq {}_n|\hat{\beta}_{jk} - \beta_{jk}|^p + |\beta_{jk}|^p P\{|\hat{\beta}_{jk} - \beta_{jk}| \geq Kt_n\} \\ &\leq C\{n^{-p/2} + n^{-\gamma}\} \end{aligned}$$

Here, we used (30), (31) and the fact that f bounded implies that its wavelets coefficients are also bounded for $j \geq 0$.

(b) In the second case, using Cauchy-Schwarz inequality, we have,

$$\begin{aligned} &{}_n|\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq Kt_n\} - \beta_{jk}|^p \\ &\quad \leq ({}_n|\hat{\beta}_{jk} - \beta_{jk}|^{2p})^{1/2} P\{|\hat{\beta}_{jk} - \beta_{jk}| \geq Kt_n/2\}^{1/2} + (Kt_n/2)^p \\ &\quad \leq Cn^{-(p+\gamma)/2} + (Kt_n/2)^p \end{aligned}$$

(c) The third case uses the arguments of both previous cases:

$${}_n|\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq Kt_n\} - \beta_{jk}|^p \leq C\{n^{-p/2} + (Kt_n/2)^p\}$$

This ends the proof of the lemma. ∎

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

$$(E|\sum_I X_i|^p)^{1/p} \leq \sum_I (E|X_i|^p)^{1/p} \tag{33}$$

1. 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 $g_{jk}(x) \neq 0$:

$$\begin{aligned} e_1 &\leq 3^{p-1}[\textstyle\sum_{-1\leq j\leq j_n^*(x)} \sum_k \{ {}_n|(\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq Kt_n\} - \beta_{jk}) g_{jk}(x)|^p\}^{1/p}]^p \\ &\leq 3^{p-1}[\textstyle\sum_{-1\leq j\leq j_n^*(x)} \sum_k I\{g_{jk}(x) \neq 0\} (2^{jp/2} C\{t_n^p + \frac{1}{n^{\gamma'}}\})^{1/p}]^p \\ &\leq 3^{p-1}[\textstyle\sum_{-1\leq j\leq j_n^*(x)} NC' 2^{j/2}\{t_n + \frac{1}{n^{\gamma'/p}}\}]^p \\ &\leq C'' 2^{j_n^*(x)p/2}\{t_n^p + \frac{1}{n^{\gamma'}}\} \end{aligned}$$

2. To bound e_2, let us remark that we can write:

$$\begin{aligned} &\sum_{j_n^*(x)+1\leq j\leq J_n} \sum_k (\hat{\beta}_{jk} I\{|\hat{\beta}_{jk}| \geq Kt_n\} - \beta_{jk}) g_{jk}(x) \\ &\quad = (E_{j_n^*(x)} f(x) - f(x) - E_{J_n} f(x) + f(x)) I\{|\hat{\beta}_{jk}| < Kt_n\} \\ &\quad + \textstyle\sum_{j_n^*(x)+1\leq j\leq J_n} \sum_k (\hat{\beta}_{jk} - \beta_{jk}) I\{|\hat{\beta}_{jk}| \geq Kt_n\} I\{|\beta_{jk}| \leq Kt_n/2\} g_{jk}(x) \end{aligned}$$

We are allowed to put here the indicator function $I\{|\beta_{jk}| \leq Kt_n/2\}$ because we are dealing with j's larger than $j_n^*(x)$. Because $F(f)(j,x)$ is non decreasing in j, we have for $j \geq j_n^*(x)$, $F(f)(j,x) \leq Kt_n/2C_r'$, hence $M_r(\Delta_j f) \leq 2^{j/2} Kt_n/2C_r'$. Therefore, using lemma (3), (27), we get that necessarily $|\beta_{jk}| \leq Kt_n/2$ if $g_{jk}(x) \neq 0$. Now it remains to write:

$$\begin{aligned} e_2 &\leq C'\{|E_{j_n^*(x)} f(x) - f(x)|^p + |E_{J_n} f(x) - f(x)|^p + \\ &\quad \Big[\textstyle\sum_{j_n^*(x)+1\leq j'\leq J_n} \sum_k (E|\hat{\beta}_{jk} - \beta_{jk}|^{2p} P\{|\hat{\beta}_{jk} - \beta_{jk}| \geq Kt_n/2\})^{1/2p} 2^{j/2} \|g\|_\infty I\{g_{jk}(x) \neq 0\}\Big]^p\} \\ &\leq C''\{|E_{j_n^*(x)} f(x) - f(x)|^p + |E_{J_n} f(x) - f(x)|^p + \tfrac{1}{n^{\gamma/2}}\} \end{aligned}$$

4.4 Local bandwidth selection using wavelet coefficients.

We'll also consider the following procedure, which can be considered as a hybrid version between thresholding and Lepski's procedure.
Let us define:

1. The index $\hat{j}^\beta(x)$ as the minimum of β-admissible j's at the point x, where

   $j \in \{0, \ldots, J_n\}$ is β-admissible at the point x if

   $$|\hat{\beta}_{j'k}\chi_{j'k}(x)| \leq K2^{j'/2}t_n, \ \forall\, k,\ j';\ j \leq j' \leq J_n$$

2. The following estimate:

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

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

Proposition 7 *As soon, as 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 $_\infty(M)$ of functions bounded by M on $[0,1]$, associated to the sequence J_n and the local functional*

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

For any choice of $r < p,\ 1 \leq p \leq p^*$.

4.4.1 Proof of the Proposition 7

The proof combines the arguments of thresholding and bandwidth selection. We have, as above to distinguish the 2 different cases: $j_n^*(x) \leq J_n$ and $j_n^*(x) > J_n$. We'll only investigate the first case. The following lemma summarizes the essential properties of $\hat{f}^H$.

Lemma 5 *Under the conditions above,*

$$\begin{array}{lll} (i) & P\{\hat{j}(x) > j_n^*(x)\} & \leq \frac{CJ_n}{n^\gamma} \\ (ii) & {}_n|\hat{E}^n_j(x) - E_jf(x)|^{2p} & \leq C_1\left(\frac{2^j}{n}\right)^p,\ \forall n \geq 1,\ \forall x \in \mathcal{X},\ \forall j,\ 0 \leq j \leq J_n. \end{array}$$

Proof of the Lemma:

(i)- We remark that, by definition of $\hat{j}^\beta(x)$, when $\hat{j}^\beta(x) \geq 1$, $j^* = \hat{j}^\beta(x) - 1$ is not β-admissible, (when $\hat{j}^\beta(x)$ is, then there exists k, / $g_{j^*k}(x) \neq 0$ / $|\hat{\beta}_{j^*k}| \geq Kt_n$. If in addition, we suppose $j^* \geq j_n^*(x)$, then : $|\beta_{j^*k}| \leq Kt_n/2$. Hence,

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

(ii)- Using (33), we easily get

$$\begin{aligned} {}_n|\hat{E}_j^n(x) - E_j f(x)|^{2p} & \leq \left[\textstyle\sum_{0\leq j'\leq j}\sum_{k,\ /\ g_{j^*k}(x)\neq 0}\left({}_n|\hat{\beta}_{j'k} - \beta_{j'k}|^{2p}2^{jp}\|g\|_\infty\right)^{1/2p}\right]^{2p} \\ & \leq C(\tfrac{2^j}{n})^p \end{aligned}$$

∎

In the case $j_n^*(x) \leq J_n$, we can divide ${}_n|\hat{f}^H(x) - f(x)|^p$ into 2 terms:

$$\begin{aligned} T_1 &= \textstyle\sum_{0\leq j'\leq j_n^*(x)} {}_n|\hat{E}_{j'}^n(x) - f(x)|^p I\{\hat{j}^\beta(x) = j'\}, \\ T_2 &= \textstyle\sum_{j_n^*(x)+1\leq j'\leq J_n} {}_n|\hat{E}_{j'}^n(x) - f(x)|^p I\{\hat{j}^\beta(x) = j'\} \end{aligned}$$

To bound T_1 let us again remark that :

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

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

$$\begin{aligned} |\hat{E}_{j'}^n(x) - \hat{E}_{j_n^*(x)}^n(x)| &= |\textstyle\sum_{j'\leq j\leq j_n^*(x)}\sum_{k,\ /\ g_{j^*k}(x)\neq 0}\hat{\beta}_{jk}g_{jk}(x)| \\ &\leq CKt_n 2^{j_n^*(x)/2}. \end{aligned}$$

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 establishing (16).

$$\begin{aligned} \textstyle\int \sup_{j'\geq j}|E_{j'}f(x) - f(x)|^p I\{x,\ j_\lambda(f,x) = j\}d\mu(x) &\leq \textstyle\int I\{x,\ j_\lambda(f,x) = j\}[\sum_{j'\geq j}|\Delta_{j'}f(x)|]^p d\mu(x) \\ &\leq \textstyle\int \|g\|_\infty I\{x,\ j_\lambda(f,x) = j\}[\sum_{j'\geq j}|\tilde{\Delta}_{j'}f(x)|]^p d\mu(x) \\ &\leq \textstyle\int I\{x,\ j_\lambda(f,x) = j\}(2^{j/2}\lambda)^p d\mu(x) \\ &\leq C'(2^{j/2}\lambda)^p \mu\{x,\ j_\lambda(f,x) = j\} \end{aligned}$$

5 Comparison among various adaptive procedures.

We are now able to start a comparison between the methods investigated above. Let us restrict to the case where 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 $\beta = 2$. Using the result of propositions 3, 5, 6 and 7 we know that the respective maxisets $MS(\hat{f}_n, p, \alpha)(T)$ associated to the rate $t_n^{\alpha p}$ of $\hat{f}^L$ (resp. $\hat{f}^T$, $\hat{f}^H$) **contains** the set

$$\begin{array}{ll} & \mathcal{B}_{\alpha/2,p,\infty}(M) \cap \mathcal{W}(F^L)(p,q)(M) \\ \text{resp.} & \mathcal{B}_{\alpha/2,p,\infty}(M) \cap \mathcal{W}(F^T)(p,q)(M), \\ \text{resp.} & \mathcal{B}_{\alpha/2,p,\infty}(M) \cap \mathcal{W}(F^H)(p,q)(M) \end{array}$$

where $q = p(1-\alpha)$ and if we omit the constants,

$$\begin{array}{l} F^L(f)(j,x) = \sup_{j' \geq j} 2^{-j'/2} |E_{j'} f(x) - f(x)| \\ F^T(f)(j,x) = 2^{-j/2} \sum_{j' \geq j} M_r(\Delta_{j'} f)(x) \\ F^H(f)(j,x) = 2^{-j/2} \sum_{j' \geq j} |\tilde{\Delta}_{j'} f(x)| \end{array}$$

It is thus natural to ask the following questions:

1. How far from equality is the inclusion ?
2. Can the spaces mentioned above be compared ?

We'll precisely give the following answers :

1. Because of Theorem 7 in Cohen *et al* (1999) (see also the second example in §2), we have equality between the spaces $\mathcal{B}_{\alpha/2,p,\infty} \cap \mathcal{W}^*(p,q)$ and the maxiset associated to the thresholding procedure $MS(\hat{f}_n^T, p, \alpha)$, in the sense described in (1).
2. If we recall,
$$F^2(f)(j,x) = 2^{-j/2} |\tilde{\Delta}_j f(x)|,$$
we have already that $\mathcal{W}^*(p,q) = \mathcal{W}(F^2)(p,q)$
3. A consequence of the forthcoming theorem 1 is that
$$\mathcal{W}(F^H)(p,q) = \mathcal{W}(F^2)(p,q) = \mathcal{W}^*(p,q).$$
4. If we introduce the following auxiliary function, and refer to the forthcoming theorem 1 for the definition of T,
$$F^3(f)(j,x) = 2^{-j/2} |\Delta_j f(x)|,$$
we observe that
$$\begin{array}{ll} F^3(f)(j,x) & \leq 2^{-j/2} \{ |E_{j+1} f - f| + |E_j f - f| \} \\ & \leq F^L(f)(j,x), \\ F^L(f)(j,x) & \leq T(F^3(f)(j,x)) \end{array}$$
We deduce, using the first inequality:
$$\mathcal{W}(F^L)(p,q) \subset \mathcal{W}(F^3)(p,q),$$

from the second one:

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

Now, using theorem 1, and theorem 3

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

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

5. Now, if we introduce

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

using the forthcoming theorem 2:

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

using theorem 1:

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

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

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

6 Comparison of weak besov bodies associated to different functionals.

6.1 Weak Besov bodies associated to max or sum functionals

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

$$\begin{array}{lcl} G^*(j,x) & = & \sup_{j'\geq j} |G(j',x)|. \\ and & & \\ T(G)(j,x) & = & 2^{-\frac{j}{2}} \sum_{j'\geq j} 2^{\frac{j'}{2}} |G(j',x)|. \end{array}$$

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 (18), this theorem is a consequence of the following lemma:

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

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

Proof of the lemma 6:

Since $|G| \leq G^* \leq T(G)$, we only have to prove :

$$G \in \ _{q,\infty}(\nu_p) \Longrightarrow T(G) \in \ _{q,\infty}(\nu_p).$$

In fact let us first prove :

$$\forall\ 0 < q < p,\ \ \exists C_q < \infty,\ \text{ such that } \quad \forall\ G, \quad \|T(G)\|^q_{\ q(\nu_p)} \leq C_q \|G\|^q_{\ q(\nu_p)}.$$

- Let us observe that

  $\|G\|^q_{\ q(\nu_p)} = \sum_{j\geq 0} 2^{jp/2} \|G(j,.)\|^q_{\ q(\mu)}$ and

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

- Hence for $q \leq 1$:

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

- For $q > 1$, we remark that $\|G\|^q_{\ q(\nu_p)} < \infty \iff 2^{jp/2q} \|G(j,.)\|_{\ q(\mu)} = \epsilon_j \in l_q(\ \ )$, hence:

$$\begin{aligned} \|T(G)\|^q_{\ q(\nu_p)} &\leq \sum_{j\geq 0} 2^{j(p-q)/2} (\sum_{j'\geq j} 2^{j'/2} \|F(j',.)\|_{\ q(\mu)})^q \\ &= \sum_{j\geq 0} 2^{j(p-q)/2} (\sum_{j'\geq j} 2^{j'(\frac{1}{2}-\frac{p}{2q})} \epsilon_{j'})^q \\ &= \sum_{j\geq 0} (\sum_{j'\geq j} 2^{-(j'-j)(\frac{p}{2q}-\frac{1}{2})} \epsilon_{j'})^q \end{aligned}$$

  As $\alpha = \frac{p}{2q} - \frac{1}{2} > 0$, $\sum_{j'\geq j} 2^{-(j'-j)(\frac{p}{2q}-\frac{1}{2})} \epsilon_{j'} = (\epsilon_. \star b.)_j$, where $b_j = 1_{j\leq 0} 2^{-\alpha j}$. Now, using

$$\|\epsilon_. \star b.\|_q \leq \|b.\|_1 \|\epsilon_.\|_q,$$

  we have

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

We just need to prove that (34), can be extended to the associated weak spaces. It is a consequence of the following interpolation theorem, whose proof is given for sake of completeness.

∎

Proposition 8 *Let (Y,ν) be a measured space, $0<p_1<q<p_2$, T defined on $_{p_1}+\ _{p_2}$ with values in the space of measurable functions , verifying, $|T(f_1+f_2)|\le|T(f_1)|+|T(f_2)|a.e.$. We suppose that for all $i\in\{1,2\}$ there exists a constant only depending on T_i and p_i, denoted by $\|T\|_{p_i}$, such that $0<\|T\|_{p_i}<\infty$ and*

$$\forall f\in\ _{p_i}(\nu),\quad \|T(f)\|_{\ _{p_i}(\nu)}\le\|T\|_{p_i}\|f\|_{\ _{p_i}(\nu)}.$$

Then

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

Proof of Proposition 8:

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

   - $$\int_Y|f|^{p_2}1_{|f|\le\lambda}d\nu\le\int_Y(|f|\wedge\lambda)^{p_2}d\nu=\int_0^\lambda p_2x^{p_2-1}\nu(|f|>x)dx$$
   $$\le\int_0^\lambda p_2x^{p_2-1}\left(\frac{\|f\|_{\ _{q,\infty}}}{x}\right)^q dx=\frac{p_2}{p_2-q}\|f\|^q_{\ _{q,\infty}}\lambda^{p_2-q}.$$

   - $$\int_Y|f|^{p_1}1_{|f|>\lambda}d\nu=\int_0^\lambda p_1x^{p_1-1}\nu(|f|>\lambda)dx+\int_\lambda^\infty p_1x^{p_1-1}\nu(|f|>x)dx$$
   $$\le\left(\frac{\|f\|_{\ _{q,\infty}}}{\lambda}\right)^q\lambda^{p_1}+\int_\lambda^\infty p_1x^{p_1-1}\left(\frac{\|f\|_{\ _{q,\infty}}}{x}\right)^q dx=\frac{q}{q-p_1}\|f\|^q_{\ _{q,\infty}}\lambda^{p_1-q}.$$

2. For a fixed $0<\lambda<\infty$, let us decompose $f\in\ _{q,\infty}$, in the following way :

   $$f=f1_{|f|>\lambda}+f1_{|f|\le\lambda}=f_1+f_2.$$

   Using the previous inequalities we have :

   $$\int_Y|f_1|^{p_1}d\nu\le\frac{q}{q-p_1}\|f\|^q_{\ _{q,\infty}}\lambda^{p_1-q}\ ;\quad\int_Y|f_2|^{p_2}d\nu\le\frac{p_2}{p_2-q}\|f\|^q_{\ _{q,\infty}}\lambda^{p_2-q}.$$

   So

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

6.2 Weak Besov bodies associated with maximal functions

Let $\mathcal{X}$ be one of the spaces ^{d} or $[0,1]^d$ and μ the Lebesgue measure. Let $0 < r < \infty$. For all measurable function on $\mathcal{X}$, we recall that, $M_r(g)(x) = \sup_{\{B, x\in B\subset\mathcal{X}\}} \left(|B|^{-1} \int_B |g|^r\right)^{1/r}$ where B denotes a ball of $\mathcal{X}$.

For $F(j,x) = F_j(x)$ a non negative functional defined on $\times\mathcal{X}$, Let us extend the definition :

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

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

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

Lemma 7 *Let $F(j,x)$ be a measurable function defined on $\times\mathcal{X}$. Then, with the previous notations :*

$$\forall 0 < r < q \ ; \ F \in \ _{q,\infty}(\nu_p) \Longleftrightarrow M_r(F) \in \ _{q,\infty}(\nu_p).$$

Proof of the lemma : It is classical (see Stein 1993), theorem 1, p 13, that for any measurable function g on $\mathcal{X}$, $|g(x)| \leq M_1(g)(x) a.e.$ and $\forall q > 1$, $\exists C_q$, depending only on $\mathcal{X}$,such that

$$\left(\int_{\mathcal{X}} M_1(g)^q d\mu\right)^{\frac{1}{q}} \leq C_q \left(\int_{\mathcal{X}} |g|^q d\mu\right)^{\frac{1}{q}}.$$

One can obviously induce, as $M_r(g) = (M_1(|g|^r))^{\frac{1}{r}}$, that $\forall 0 < r < q < \infty$,

$$\left(\int_{\mathcal{X}} M_r(g)^q d\mu\right)^{\frac{1}{q}} \leq C_{\frac{q}{r}}^{\frac{1}{r}} \left(\int_{\mathcal{X}} |g|^q d\mu\right)^{\frac{1}{q}} \leq C_{\frac{q}{r}}^{\frac{1}{r}} \left(\int_{\mathcal{X}} M_r(g)^q d\mu\right)^{\frac{1}{q}}.$$

There is an obvious modification for $q = \infty$.

Let now $0 < r < q < \infty$. As $M_r(F)(j,x) \geq |F(j,s)|, \nu_p-$almost everywhere we have just to prove the lemma in one direction : $\|M_r(F)\|_{\ _{q,\infty}(\nu_p)} \leq C\|F\|_{\ _{q,\infty}(\nu_p)}$ But,

$$\|M_r(F)\|^q_{\ _{q}(\nu_p)} = \sum_{j\in} 2^{j(p/2)} \|M_r(F_j)\|^q_{\ _q(\mu)} \leq C_{\frac{q}{r}}^{\frac{q}{r}} \sum_{j\in} 2^{j(p/2)} \|F_j\|^q_{\ _q(\mu)} = C_{\frac{q}{r}}^{\frac{q}{r}} \|F\|^q_{\ _{q}(\nu_p)}.$$

We can conclude using proposition 8.

6.3 Weak bodies associated with wavelet coefficients

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

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

Let us 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 :

$$F_f(j,x) = 2^{-\frac{j}{2}}\Delta_j(f)(x); \quad \widetilde{F_f}(j,x) = \widetilde{\Delta_j(f)}(x).$$

then we have the following theorem :

Theorem 3 $\forall 0 < r < q$,

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

this is the consequence of theorem 2 and the following lemma :

Lemma 8 *With the previous notations*

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

$$|\Delta_j(x)| \leq \|\psi\|_\infty N^{(\frac{1}{r}\vee 1)} M_r(\widetilde{\Delta_j})(x) \quad \forall x \quad a.e.$$

Let us recall

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

Proof of the lemma : The following remarks are obvious, due to a finite dimensional argument : *Remarks* :

- $\forall 0 < r < \infty, \exists\, C_r \;,\; C'_r$ such that :

$$C_r\left(\int_{[0,N]} |\sum_l \alpha_l\psi(u-l)|^r du\right)^{\frac{1}{r}} \leq \int_{[0,N]} |\sum_l \alpha_l\psi(u-l)|du \leq C'_r\left(\int_{[0,N]} |\sum_l \alpha_l\psi(u-l)|^r du\right)^{\frac{1}{r}}.$$

- Moreover : $\forall k \in \quad , \quad \forall j \in$

$$C_r\left(2^j\int_{[\frac{k}{2^j},\frac{k+N}{2^j}]} |\sum_l \alpha_l\psi_{j,l}(u)|^r du\right)^{\frac{1}{r}} \leq 2^j\int_{[\frac{k}{2^j},\frac{k+N}{2^j}]} |\sum_l \alpha_l\psi_{j,l}(u)|du$$

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

$\diamondsuit$

Let us now prove :

1.

$$\forall x,\ |\widetilde{\Delta_j}(x)| \leq C'_r \|\psi\|_\infty N^{\frac{1}{r}} M_r(\Delta_j)(x) :$$

Let $x \in [\frac{k}{2^j}, \frac{k+1}{2^j}]$.

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

But using the preceding remarks this can be bounded by :

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

2.

$$\forall x,\quad |\Delta_j(x)| \leq \|\psi\|_\infty N^{(\frac{1}{r} \vee 1)} M_r(\widetilde{\Delta_j})(x) :$$

Let $x \in [\frac{k}{2^j}, \frac{k+1}{2^j}]$.

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

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

But let us observe that :

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

We are very grateful to the very positive and helpful comments of two referees and the Case-Editor.

References

[1] Butzer, P.L., Berens, H. (1967) *Semi-groups of operators and approximation* Springer. Berlin.

[2] Butzer, P.L.,Nessel, R.J. (1971) *Fourier analysis and approximation* Academic Press. New York.

[3] Cohen, A., DeVore, R. and Hochmuth, R. (1997) Restricted nonlinear approximation, preprint Laboratoire d'Analyse Numérique, Université Paris VI, to appear in Constructive Apporximation.

[4] Cohen, A., DeVore, R., Kerkyacharian, G. and Picard, D.(1999) Maximal sets with given rates of convergence for thresholding algorithms, preprint Laboratoire d'Analyse Numérique, Université Paris VI.

[5] DeVore, R.,and Lorentz G. (1993) *Constructive Approximation* Springer-Verlag.

[6] DeVore, R., Jawerth, B. and Popov, V. (1991) Compression of wavelet decomposition. *American Journal of Mathematics*, **114**, 737-785 .

[7] Donoho, D. (1996) Unconditional bases and bit-level compression, *Applied Computational Harmonic Analysis* **3**, 388-392.

[8] Donoho, D. (1997) CART, and Best Orthogonal Basis, a connection, *Ann. Statist.,* **25**, 1870-1911.

[9] Donoho, D.L. and Johnstone, I. M (1994) Ideal spatial adaptation via Wavelet shrinkage.*Biometrika* **81**, 425-455

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

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

[12] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1994) Wavelet Shrinkage: Asymptopia? .*Journal of the Royal Statistical Society, Series B* **57**, 301-369. With discussion.

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

[14] Donoho, D.L., Johnstone. (1996) Neoclassical minimax problems, thresholding and adaptive function estimation Bernoulli **1**, 39-62.

[15] Härdle W., Kerkyacharian, G., Picard, D., Tsybakov, A. (1998) *Wavelet, Approximation and Statistical Applications* Lecture Notes in Statistics, 129 Springer Verlag, New York.

[16] Hoffmann, M. (1999a). L_p estimation of the diffusion coefficient. *Bernoulli*, **5**, 447-481.

[17] Hoffmann, M. (1999b). Adaptive estimation in diffusion processes. *Stochastic processes and their Applications*, **79**, 135-163.

[18] Hoffmann, M. (1999c). On nonparametric estimation in nonlinear AR(1)-models. *Statistics and Probability Letters*, **44**, 29-45.

[19] Kerkyacharian, G. and Picard, D. (1992) *Density estimation in Besov Spaces.* Statistics and Probability Letters **13** 15-24

[20] Kerkyacharian, G. and Picard, D. (1993) *Density Estimation by Kernel and Wavelets methods - Optimality of Besov spaces.* . Statistics and Probability Letters **18** 327-336

[21] Lepskii, O.V. (1991) *Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates.* Theory Probab. Appl. (36) pp 682-697

[22] Lepskii, O.V., E. Mammen and V.G. Spokoiny (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Annals of Statistics*, **25** , 929-947.

[23] Mallat, S., Papanicolaou, G., Zhang, Z. (1998) Adaptive covariance estimation of locally stationary processes. *Ann. Statis.* **26**

[24] Neumann M.B. (1998) Multivariate wavelet thresholding: A remedy against the curse of dimensionality ? Technical report, Humboldt Universitat. Berlin

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

[26] Meyer, Y. (1990) *Ondelettes et Opérateurs*, Hermann, Paris.

[27] Nikolskii S.M.(1975) *Approximation of functions of several variables and imbedding theorems* (Russian). Sec. ed., Moskva, Nauka 1977 English translation of the first ed., Berlin 1975.

[28] Picard, D., Tribouley, K. (2000). Adaptive Confidence Interval for Pointwise Curve Estimation. *Ann. Statist.* **28** pp 298-335.

[29] Stein, E. (1993). *Harmonic Analysis.* Princeton Math. series.