The maxiset point of view for estimating integrated quadratic functionals

Vincent Rivoirard and Karine Tribouley

CNRS UMR 8628, Paris XI-Orsay and Modal'X, Paris X-Nanterre

Abstract: This paper explores the estimation of $\int f^2$ where f is a functional parameter in the white noise model. To compare different estimation procedures, we adopt the maxiset point of view, i.e., we point out the entire set of functions on which a given procedure achieves a given target rate. Quadratic and soft (local and global) thresholding wavelet procedures are considered. We compute the maxisets for these procedures and prove that, most of the time, thresholding procedures outperform the quadratic one. The comparison of performances in the maxiset setting of local and global thresholding depends on the target rate; none of them is always preferable.

Key words and phrases: Besov spaces, integrated functional, maxiset theory, minimax theory, thresholding, wavelet.

1. Introduction.

Our aim, in this paper, is to investigate the estimation of $\theta(f) = \int f^2$, where f is the functional parameter of the classical white noise model

$$Y_\epsilon(t) = \int_0^t f(u)du + \epsilon B(t), \quad t \in [0,1], \tag{1.1}$$

$B(t)$ is a standard Brownian motion on $[0,1]$, $Y_\epsilon = (Y_\epsilon(t), 0 \leq t \leq 1)$ is the observed variable, and $\epsilon \to 0$.

In a general way, for the non-parametric framework there are three steps to take when estimating $\theta(f)$: the choice of the method (kernel, series, wavelet,...); the determination of parameters of the method (the bandwidth h, the number N, the level j,...); the evaluation of the quality of the procedure $\hat{\theta}_\epsilon$ (the word "procedure" fixes the method/parameter) by computing its rate. It is well known that the rate has to be associated with a function space. More precisely, for the procedure $\hat{\theta}_\epsilon$ and function space $\mathcal{F} \subset \mathbb{L}_2$, we point out the associated (quadratic) rate ρ_ϵ that results from the computation of $\sup_{g\in\mathcal{F}} \mathbb{E}[(\hat{\theta}_\epsilon - \theta(g))^2]$.

When non-parametric problems are explored, the minimax theory is the most popular point of view: it consists in ensuring that the procedure $\hat{\theta}_\epsilon$ to be used achieves the best rate on a given function space $\mathcal{F}$. But the rate might be unknown (in an adaptive framework) and the choice of $\mathcal{F}$ is arbitrary (what kind of spaces has to be considered: Sobolev spaces? Besov spaces? why?). Moreover, $\mathcal{F}$ could contain very bad functions g (in the sense that $\theta(g)$ is difficult to estimate). Since $\theta(f)$ might be easier to estimate, the procedure could be too pessimistic and not adapted to the data. More embarrassing in practice, several minimax procedures may be proposed and the practitioner has no way to decide between them. To answer these questions, another point of view has recently appeared: the maxiset point of view (see for instance Kerkyacharian, and Picard (2000)). It consists in deciding the accuracy of the estimate by fixing a prescribed rate ρ_ϵ and pointing out all functions g such that $\theta(g)$ can be estimated by the procedure $\hat{\theta}_\epsilon$ at the target rate ρ_ϵ. Roughly speaking, the maxiset of the procedure $\hat{\theta}_\epsilon$ for the rate ρ_ϵ is the set of all such functions. The maxiset point of view brings answers to the previous questions. Indeed, there is no a priori functional assumption and then, the practitioner does not need to restrict his study to an arbitrary function space. The practitioner states the desired accuracy and fixes the quality of the used procedure. Obviously, he chooses the procedure with the largest maxiset.

Let us come back to the problem of estimating $\theta(f) = \int f^2$. This problem has been intensively studied in the minimax theory and is now completely solved. Generally, f is assumed to belong to the Besov space $\mathcal{B}^\alpha_{p,\infty}$ for $\alpha > 0$, $p \geq 1$. One gets different rates according to the regularity α of the function f. If f is regular, it is possible to estimate $\theta(f)$ with the parametric rate. Otherwise, the (non-parametric) rates depend on α and on p when $p < 2$. Moreover, as in the problem of estimating the entire function f, two forms of rates have been pointed out when f is *dense* ($p \geq 2$) or f is *sparse* ($p < 2$). Procedures have been proposed to achieve the minimax rate in each case. Under some conditions, in the case where $p \geq 2$, quadratic methods or global thresholding methods are shown to be minimax or adaptive minimax (see Tribouley (2000)); in the case $p < 2$, Cai and Low (2005) prove that a local thresholding method is minimax.

In this paper, we study wavelet estimation methods and focus on thresholding methods. We consider soft local thresholding, soft global thresholding, and no thresholding (the resulting method is then called the quadratic procedure). Our aim is to answer the following questions.

- Is the use of Besov spaces arbitrary in the minimax point of view?
- If f is supposed to be regular, the usual quadratic estimate is optimal in the minimax sense. But why not use a non-quadratic procedure? Could it be better?
- The soft local thresholding procedure has been proposed by Cai and Low (2005) to obtain minimax procedures in the non-regular sparse case. The global thresholding procedure proposed by Tribouley (2000) solves the problem of adaptation in the dense case. Is it judicious to use these procedures instead of the quadratic one in a more general context?
- If the practitioner is convinced about the performances of the thresholding methods, is it preferable to use global or local thresholding procedures?

For each procedure, we compute the maxiset associated with a target rate. We prove that the classical regularity assumptions of the minimax theoreticians make sense in that the maxiset of the quadratic procedures for polynomial rates is exactly a Besov space (see Theorem 1). Next, Theorem 2 states that the maxisets of the thresholding procedures are weak Besov spaces; they are at least as large as classical Besov spaces. These spaces are directly connected to the thresholding methods, and we define a weak local version and a weak global version. Actually, they appear in a quite natural way in the context of the functional estimation using wavelet methods, see for instance Donoho (1993) or Johnstone (1994). See also Kerkyacharian, and Picard (2000) for a precise study of the links between wavelet thresholding methods and weak Besov spaces in the functional estimation framework. Given a target rate and estimation procedures, we compute their maxisets as a basis for comparisons. For polynomial rates, and with an optimal choice of parameters, we establish that the local thresholding procedure is always best in the sense that it achieves the given target rate on the largest set of functions. We also prove that the maxiset of the global thresholding procedure is the maxiset of the quadratic procedure and we deduce that both have the same performance. With a different choice for the parameters, the global thresholding procedure outperforms the quadratic one and is not comparable with the local thresholding procedure since neither maxiset is included in the other. In this case, we point out the shape of the functions (sparse or dense) for which each procedure is adapted. Finally, we study the case where the target rate is more general than the polynomial one. In particular, we focus on rates of the form $\epsilon^{2r}|\log(\epsilon)|^{2r'}$; such rates appear in the minimax adaptive framework. With these rates, the maxiset conclusions are different according to the value of r'. This proves the influence of the target rate when we compare procedures in the maxiset framework.

The paper is organized as follows. In Section 2, we present the model and introduce the maxiset setting. In Section 3, we define the function spaces that are candidates for maxisets of procedures presented in Section 4. All the results are given in Section 5. Section 6 is devoted to the connections with the minimax results. The proofs concerning the properties of the function spaces are postponed to Section 7, and those concerning maxisets are postponed to Section 8.

2. Model and problem.

As usual, we translate the original functional model (1.1) into the sequence space model. For this purpose, let us take a wavelet function ψ and an associated scaling function denoted ψ_{-1}. We assume that these functions are compactly supported, see for example the Daubechies wavelets (Daubechies (1992)). By translations and dilations, we obtain a $\mathbb{L}_2$-orthonormal wavelet basis denoted $(\psi_{jk})_{j\geq -1, k\in\mathbb{Z}}$, which enables us to translate (1.1) into the following sequence space model:

$$y_{jk} = \beta_{jk} + \epsilon\, z_{jk},\ j \in \{-1\} \cup \mathbb{N}, \quad k \in \mathbb{Z},$$

where (z_{jk}) is a sequence of i.i.d. standard Gaussian variables, (y_{jk}) is the sequence of observed variables, and $(\beta_{-1,k})$ (respectively $(\beta_{j,k})$ for $j \neq -1$) are the coefficients of f on the scaling function (respectively on the wavelet function): $f = \sum_{j=-1}^{+\infty} \sum_k \beta_{jk}\psi_{jk}$. Let us note that at each level $j \geq -1$, the number of non-zero wavelet coefficients is smaller or equal to $[\max(2^j, 1) + l_\psi - 1]$, where l_ψ is the maximal size of the supports of ψ and ψ_{-1}. Since the wavelet basis is an orthonormal basis of $\mathbb{L}_2$, the parameter to be estimated is $\theta = \theta(f) = \sum_{j=-1}^{\infty} \sum_k \beta_{jk}^2$.

Definition 1. *Let $R>0$ and let $\rho_\epsilon>0$ be the target rate. If $\hat\theta$ denotes an estimator of θ, the maxiset of $\hat\theta$ of radius R for the rate ρ_ϵ is denoted $MS(\hat\theta,\rho_\epsilon)(R)$, and is defined by*

$$MS(\hat\theta,\rho_\epsilon)(R)=\left\{f\in\mathbb{L}_2([0,1]):\quad \sup_\epsilon \rho_\epsilon^{-1}\mathbb{E}\left[(\hat\theta-\theta)^2\right]\le R^2\right\}.$$

We write $MS(\hat\theta,\rho_\epsilon)=\mathcal{A}$ to mean that $\forall R,\ \exists R',\ MS(\hat\theta,\rho_\epsilon)(R)\subset\mathcal{A}(R')$ and $\forall R',\ \exists R,\ \mathcal{A}(R')\subset MS(\hat\theta,\rho_\epsilon)(R)$, where $R,R'>0$ are the radii of balls of $MS(\hat\theta,\rho_\epsilon)$ and $\mathcal{A}$ respectively.

3. Function spaces.

In Section 3.1, we recall the definitions of the function spaces that play an important role in the sequel. Note that, here, they appear with definitions depending on the wavelet basis. However, as noted by Meyer (1990) and Cohen, DeVore, and Hochmuth (2000), most of them also have different definitions, so this dependence on the basis is not crucial. Next, in Section 3.2, we explore the links between the spaces that we have introduced.

3.1. Definitions.

Recall the definition of the Besov spaces in terms of wavelet coefficients.

Definition 2. *Let $s>0$ and $R>0$. A function $f=\sum_{j=-1}^{+\infty}\sum_k\beta_{jk}\psi_{jk}\in\mathbb{L}_2([0,1])$ belongs to the* **Besov ball** *$\mathcal{B}^s_{p,\infty}(R)$ if and only if*

$$\left[\sup_{j\ge -1}2^{j(s+\frac{1}{2}-\frac{1}{p})p}\sum_k|\beta_{jk}|^p\right]^{1/p}\le R.$$

Since we focus on the estimation of $\theta=\sum_{j\ge -1}\sum_k\beta_{jk}^2$, the space $\mathcal{B}^s_{2,\infty}$ has specific interest. Note that, when $p=2$, f belongs to $\mathcal{B}^s_{2,\infty}$ if and only if

$$\sup_{J\ge -1}2^{2Js}\sum_{j\ge J}\sum_k\beta_{jk}^2<+\infty.$$

This characterization is often used in the sequel. We now introduce spaces in the class of Lorentz spaces that are directly connected to the estimation procedures considered in this paper.

Definition 3. *Let $0<r<2$ and $R>0$. A function $f=\sum_{j=-1}^{+\infty}\sum_k\beta_{jk}\psi_{jk}\in$*

$\mathbb{L}_2([0,1])$ *belongs to the* **weak local Besov ball** $W^L_{r,\gamma}(R)$ *if and only if*

$$\left[\sup_{\lambda>0} \lambda^{r-2} \sum_{j \geq -2\gamma \log_2(\lambda)} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq \lambda\sqrt{j}}\right]^{1/2} \leq R,$$

and to the **weak global Besov ball** $W^G_{r,\gamma}(R)$ *if and only if*

$$\left[\sup_{\lambda>0} \lambda^{r-2} \sum_{j \geq -2\gamma \log_2(\lambda)} \sum_k \beta_{jk}^2 1_{\sum_k \beta_{jk}^2 \leq \lambda^2 2^{j/2} \sqrt{j}}\right]^{1/2} \leq R.$$

Actually, to check that a function decomposed on the wavelet basis belongs to one of these weak Besov spaces, it is enough to verify that $f \in \mathbb{L}_2([0,1])$ and to evaluate the supremum for $\lambda \leq 1$. For weak local Besov spaces, we focus on the number of the wavelet coefficients that are smaller than a prescribed threshold. For weak global Besov spaces, we do the same job, but level by level and for a function of the wavelet coefficients. In fact, weak Besov spaces have already been introduced in the maxiset context in statistics (see Cohen, DeVore, Kerkyacharian, and Picard (2001), Kerkyacharian, and Picard (2000, 2002), Rivoirard (2004), or Autin, Picard, and Rivoirard (2004)) and in approximation theory (see Cohen, DeVore, and Hochmuth (2000) for instance). The main difference lies in the level j's we consider: we do not care about wavelet coefficients when $j < -2\gamma \log_2(\lambda)$, and this difference is crucial in the sequel. These spaces play an important role in approximation theory (see DeVore and Lorentz (1993)). We show in the next section that they are strongly connected to classical Besov spaces and they appear as weak versions of Besov spaces.

3.2. Links between function spaces.

Recall the inclusions between the Besov spaces.

Property 1. *Let* $s, s' > 0$ *and* $p \geq 1$. *Then, we have*

$$\begin{aligned} &\mathcal{B}^s_{p,\infty} \subset \mathcal{B}^s_{2,\infty} \quad \text{if} \quad p \geq 2, \\ &\mathcal{B}^s_{p,\infty} \subset \mathcal{B}^{s'}_{2,\infty} \quad \text{if} \quad p \leq 2, \quad s - 1/p \geq s' - 1/2. \end{aligned}$$

We establish links between the weak Besov spaces.

Property 2. *Let $0 < r, r_1, r_2 < 2$ and $\gamma, \gamma_1, \gamma_2 > 0$. Suppose W^* is either the weak local Besov space or the weak global Besov space. Then*

$$W^*_{r_1,\gamma} \subset W^*_{r_2,\gamma} \text{ if } 0 < r_1 < r_2 < 2 \quad \text{and} \quad W^*_{r,\gamma_1} \subset W^*_{r,\gamma_2} \text{ if } 0 < \gamma_1 < \gamma_2. \tag{3.1}$$

Let $0 < r < 2$. If $0 < \gamma < 2 - r$, we have

$$W^L_{2-r,\gamma} \not\subset W^G_{2-r,\gamma}, \quad \text{and} \quad W^G_{2-r,\gamma} \not\subset W^L_{2-r,\gamma}. \tag{3.2}$$

If $\gamma = 2 - r$

$$W^G_{2-r,2-r} \subsetneq W^L_{2-r,2-r}. \tag{3.3}$$

We now establish links between weak Besov spaces and Besov spaces.

Property 3. *If $0 < r < 2$ and $\gamma > 0$,*

$$\mathcal{B}^{r/(4\gamma)}_{2,\infty} \subset W^*_{2-r,\gamma}, \tag{3.4}$$

where W^* is either the weak local Besov space W^L or the weak global Besov space W^G. We also have, for $s > 0$ and $p \geq 1$,

$$\mathcal{B}^s_{p,\infty} \subset W^L_{1,1} \quad \text{if} \quad p < 2, \text{ and } s > 1/(2p), \text{ or if } p \geq 2, \text{ and } s \geq 1/4;$$
$$\mathcal{B}^s_{p,\infty} \subset W^G_{1,1} \quad \text{if} \quad p < 2, \text{ and } s \geq 1/p - 1/4, \text{ or if } p \geq 2, \text{ and } s \geq 1/4.$$

To compare the estimation procedures from the maxiset point of view, it is crucial to know whether the inclusion (3.4) is strict or not.

Property 4. *When $0 < r < 2$, and $\gamma > 0$,*

$$\mathcal{B}^{r/(4\gamma)}_{2,\infty} \subsetneq W^L_{2-r,\gamma}, \text{ if } \gamma \leq 2 - r \quad \text{and} \quad \mathcal{B}^{r/(4\gamma)}_{2,\infty} \subsetneq W^G_{2-r,\gamma}, \text{ if } \gamma < 2 - r, \tag{3.5}$$

but

$$\mathcal{B}^{r/(4(2-r))}_{2,\infty} = W^G_{2-r,2-r}. \tag{3.6}$$

Proofs are postponed to Section 7. The proofs of the strict inclusions or the non-inclusions are interesting because they are constructive: we build explicit functions belonging to a specified space, but not belonging to another.

4. Procedures of estimation.

We present different estimation procedures that are essentially thresholding procedures. For local thresholding, we refer to Cai and Low (2005), and for global thresholding to Tribouley (2000). Let j_0, j_1 be levels such that $j_0 \leq j_1$, and let τ be a threshold to be chosen (eventually depending on j, ϵ). We consider estimates

$$\hat{\theta} = \sum_{j=-1}^{j_1-1} \sum_k \hat{\theta}_{jk},$$

where for $j = -1, \ldots, j_0 - 1$, and all k, $\hat{\theta}_{jk} = y_{jk}^2 - \epsilon^2$, and for all $j \geq j_0$, and all k,

$$\hat{\theta}_{jk} = \hat{\theta}_{jk}^L = \left(y_{jk}^2 - \mu\epsilon^2\right) 1_{|y_{jk}|>\epsilon\sqrt{\tau}} - \epsilon^2 \mathbb{E}\left(z_{jk}^2 - \mu\right) 1_{|z_{jk}|>\sqrt{\tau}}$$

or

$$\hat{\theta}_{jk} = \hat{\theta}_{jk}^G = 2^{-j} \sum_k \left(y_{jk}^2 - \lambda\epsilon^2\right) 1_{\sum_k \left(y_{jk}^2 - \epsilon^2\right) > \epsilon^2 \sqrt{2^j \tau}},$$

where μ, λ are real parameters (eventually depending on ϵ or j). We recall their minimax properties.

Remark 1. In the sequel, 2^{j_0} and 2^{j_1} are powers of ϵ but, since j_0 and j_1 are integers, integer parts should be used. To avoid tedious notation, and without loss of generality, we ignore this point.

4.1. The quadratic procedures.

If $j_0 = j_1$, $\hat{\theta}$ is the *classical quadratic estimator*. In this case, we note $\hat{\theta} = \hat{\theta}^Q$. If f is in $B_{p,\infty}^{\alpha}$ for $p \geq 2$, $\alpha > 0$ or $p \leq 2$, $\alpha \geq 1/p - 1/4$, it achieves the optimal minimax rate

$$\rho_\epsilon = \epsilon^{2r} = \begin{cases} \epsilon^2 & \text{if} \quad \alpha \geq 1/4, p \geq 2 \text{ or } \alpha \geq 1/p - 1/4, p \leq 2, \\ \epsilon^{16\alpha/(1+4\alpha)} & \text{if} \quad \alpha \leq 1/4, p \geq 2, \end{cases}$$

as soon as the smoothing parameter j_0 is chosen such that

$$2^{j_0} = \begin{cases} \epsilon^{-2} & \text{if} \quad \alpha \geq 1/4, p \geq 2 \text{ or } \alpha \geq 1/p - 1/4, p \leq 2, \\ \epsilon^{-4/(1+4\alpha)} & \text{if} \quad \alpha \leq 1/4, p \geq 2. \end{cases}$$

4.2. The local thresholding procedures.

When $j_1 > j_0$, the estimate $\hat{\theta}$ built with the sequence $\hat{\theta}_{jk}^L$ is a *local thresholding estimator*. In this case, we note $\hat{\theta} = \hat{\theta}^L$. The choice $\mu = 0$ is associated with hard thresholding procedures, while $\mu = \tau$ is associated with soft thresholding procedures. Cai and Low (2005) proved that the soft local procedure is minimax on $B^{\alpha}_{p,\infty}$ when $p \le 2$. More precisely, set $s = \alpha + 1/2 - 1/p$. The minimax rates

$$\rho_\epsilon = \epsilon^{2r} \quad = \quad \begin{cases} \epsilon^2 & \text{if} \quad \alpha > 1/(2p), p \le 2, \\ \epsilon^{4-2p/(1+2ps)} & \text{if} \quad \alpha \le 1/(2p), p \le 2, \end{cases}$$

are achieved if j_0 satisfies

$$2^{j_0} = \begin{cases} \epsilon^{-2} & \text{if} \quad \alpha > 1/(2p), p \le 2, \\ \epsilon^{-2p/(1+2ps)} & \text{if} \quad \alpha \le 1/(2p), p \le 2, \end{cases}$$

and $2^{j_1} = |\log \epsilon| \epsilon^{-1/2s}$, $\tau = \kappa(j - j_0)$, κ a constant. If non-limited procedures are allowed, we take $j_1 = +\infty$ and $\tau = \kappa j$, and the procedure is adaptive minimax on $B^{\alpha}_{p,\infty}$ for $\alpha > 1/(2p)$, $p \le 2$.

4.3. The global thresholding procedures.

When $j_1 > j_0$, the estimate $\hat\theta$ built with the sequence $\hat\theta^G_{jk}$ is a *global thresholding estimator*. In this case, we note $\hat\theta = \hat\theta^G$. The choice $\lambda = 1$ leads to a hard thresholding procedure, and $\lambda = 1 + 2^{-j/2}\sqrt{\tau}$ is associated with a soft thresholding procedure. For $\lambda = 1$, Tribouley (2000) proved that if $2^{j_0} = \epsilon^{-2}$, $2^{j_1} = \epsilon^{-4}$, and $\tau = \kappa j$, the procedure is adaptive minimax on the space $B^{\alpha}_{p,\infty}$ for $p \ge 2$, $\alpha > 0$. This means that the adaptive minimax rate

$$\rho_\epsilon = \begin{cases} \epsilon^2 & \text{if} \quad \alpha \ge 1/4, p \ge 2, \\ (|\log \epsilon|^{1/4}\epsilon)^{16\alpha/(1+4\alpha)} & \text{if} \quad \alpha < 1/4, p \ge 2, \end{cases}$$

is achieved. Note that the logarithmic term is the price to pay for adaptation.

5. Main results.

In this section, we apply the maxiset theory for the procedures defined previously. As mentioned earlier, use of a "procedure" means that we fix the method (quadratic, local thresholding, global thresholding) and the parameters of the method (j_0, j_1, τ). Let ρ_ϵ be the target rate of convergence. First we focus on polynomial rates $\rho_\epsilon = \epsilon^{2r}$, $0 < r \le 1$. In Section 5.1, we determine the maxisets associated with the procedures described earlier. In Section 5.2, for the same rates, we compare procedures by comparing maxisets. Finally, we study the case where general rates are considered: the maxiset computations are in Section 5.3 and comparisons between procedures are in Section 5.4.

5.1. Maxisets when the target rate is polynomial.

Theorem 1 deals with quadratic procedures, which means that we consider the case $j_0 = j_1$. In Theorem 2, we establish results about the thresholding procedures for non-limited procedures and for limited procedures. The smoothing parameter j_0 is proportional to $|\log_2 \epsilon|$. We always consider the case $\tau = \kappa j$ for some κ large enough.

Theorem 1. *Let $0 < r \leq 1$ and $0 < \gamma \leq 2 - r$. If $\hat{\theta}^Q$ is the quadratic estimate with $2^{j_0} = 2^{j_1} = \epsilon^{-2\gamma}$, then $MS(\hat{\theta}^Q, \epsilon^{2r}) = \mathcal{B}_{2,\infty}^{r/(4\gamma)}$.*

Next, we focus on the soft thresholding procedures: we take $\mu = \tau$ for the local thresholding, and $\lambda - 1 = m 2^{-j/2}\sqrt{\tau}$ for some constant $m > 0$ for the global thresholding.

Theorem 2. *Let $0 < r \leq 1$ and $0 < \gamma \leq 2 - r$. If $\hat{\theta}^L$ is the soft local thresholding estimate and $\hat{\theta}^G$ is the soft global thresholding estimate, with $2^{j_0} = \epsilon^{-2\gamma}$, $2^{j_1} = +\infty$ and $\tau = \kappa j$ for κ great enough, $MS(\hat{\theta}^L, \epsilon^{2r}) = W^L_{2-r,\gamma}$, and $MS(\hat{\theta}^G, \epsilon^{2r}) = W^G_{2-r,\gamma}$. If $2^{j_1} = \epsilon^{-2\gamma'}$ for some $\gamma' > \gamma$, the maxisets are $MS(\hat{\theta}^L, \epsilon^{2r}) = W^L_{2-r,\gamma} \cap \mathcal{B}_{2,\infty}^{r/(4\gamma')}$ and $MS(\hat{\theta}^G, \epsilon^{2r}) = W^G_{2-r,\gamma} \cap \mathcal{B}_{2,\infty}^{r/(4\gamma')}$.*

Note that Theorems 1 and 2 are special cases of Theorems 4 and 5. Note also that each procedure depends only on the choice of γ (and γ' for limited procedures), but the larger γ, the larger the maxiset.

5.2. Maxiset comparisons of procedures when the target rate is polynomial.

We compare our estimation procedures for polynomial rates of convergence. For this purpose, assume that the rate is of the form $\rho_\epsilon = \epsilon^{2r}$ with $0 < r \leq 1$. We still consider the quadratic, local thresholding and global thresholding procedures (denoted $\hat{\theta}^Q$, $\hat{\theta}^L$, $\hat{\theta}^G$) with thresholds introduced in Theorem 2. For each of them we take $2^{j_0} = \epsilon^{-2\gamma}$, $\gamma \leq 2 - r$ and $2^{j_1} = \epsilon^{-2\gamma'}$, where $\gamma' \in [\gamma, +\infty]$ is assumed to be a large enough constant (see Remark 3). Using the properties of Section 3.2 and Theorems 1 and 2, we immediately deduce:

Theorem 3. *When $\gamma < 2 - r$, the quadratic procedure is outperformed by the local and global thresholding ones since $MS(\hat{\theta}^Q, \epsilon^{2r}) \subsetneq MS(\hat{\theta}^L, \epsilon^{2r})$ and $MS(\hat{\theta}^Q, \epsilon^{2r}) \subsetneq MS(\hat{\theta}^G, \epsilon^{2r})$. Moreover, local and global thresholding are not comparable since*

$MS(\hat\theta^L, \epsilon^{2r}) \not\subset MS(\hat\theta^G, \epsilon^{2r})$ and $MS(\hat\theta^G, \epsilon^{2r}) \not\subset MS(\hat\theta^L, \epsilon^{2r})$. When $\gamma = 2 - r$, the quadratic procedure and global thresholding achieve the same performance since $MS(\hat\theta^Q, \epsilon^{2r}) = MS(\hat\theta^G, \epsilon^{2r})$. Moreover, the local thresholding procedure outperforms global thresholding and the quadratic procedure since $MS(\hat\theta^Q, \epsilon^{2r}) = MS(\hat\theta^G, \epsilon^{2r}) \subsetneq MS(\hat\theta^L, \epsilon^{2r})$.

Non-comparability of local and global thresholding when $\gamma < 2 - r$ could appear as an illustration of a drawback of the maxiset setting where the order is not total. However, we can draw interesting conclusions from these maxiset results in the lights of counter-examples of Section 7. Indeed, in Section 7, we point out what are the functions that belong to the maxiset of one procedure and not to the maxiset of the other one, according to their sparsity. And as a conclusion, roughly speaking, local thresholding is convenient when estimating sparse functions, global thresholding for dense ones. The last point shows that, from the maxiset point of view, local thresholding is the best choice for an appropriate choice of γ, and global thresholding should be avoided when γ is taken as large as possible.

5.3. Extensions of previous results for general rates.

To generalize results of Section 5.1 for quadratic and thresholding procedures, consider a continuous function $u : [0,1] \longrightarrow \mathbb{R}_+$ such that

$$\exists\, \delta > 0, \exists\, M > 0, \forall\, x \in [0,1], \forall\, y \in [x,1], \quad u(y)y^{\delta-2} \le M u(x) x^{\delta-2}. \qquad (5.1)$$

The following theorem is a generalization of Theorem 1.

Theorem 4. *Let $\gamma > 0$. If $\hat\theta^Q$ is the quadratic estimate with $2^{j_0} = 2^{j_1} = \epsilon^{-2\gamma}$, and if (5.1) is satisfied for some $\delta \ge \max(\gamma, 1)$, then $MS(\hat\theta^Q, u^2(\epsilon)) = \mathcal{B}_{2,\gamma,\infty}(u)$, where*

$$\mathcal{B}_{2,\gamma,\infty}(u)(R) := \left\{ f : \quad \sup_{\lambda > 0} u(\lambda)^{-1} \sum_{j \ge -2\gamma \log_2(\lambda)} \sum_k \beta_{jk}^2 \le R^2 \right\}.$$

Theorem 2 is generalized as follows.

Theorem 5. *Let $\gamma > 0$. Let $\hat\theta^L$ be the soft local thresholding estimate, and $\hat\theta^G$ the soft global thresholding estimate, with $2^{j_0} = \epsilon^{-2\gamma}, 2^{j_1} = +\infty$, and $\tau = \kappa j$ for*

κ large. If (5.1) is satisfied for $\delta \geq \max(\gamma, 1)$, then $MS(\hat{\theta}^L, u^2(\epsilon)) = W^L_\gamma(u)$, and $MS(\hat{\theta}^G, u^2(\epsilon)) = W^G_\gamma(u)$, where

$$W^L_\gamma(u)(R) := \left\{ f : \quad \sup_{\lambda>0} u(\lambda)^{-1} \sum_{j\geq -2\gamma\log_2(\lambda)} \sum_k \beta^2_{jk} 1_{|\beta_{jk}|\leq\lambda\sqrt{j}} \leq R^2 \right\},$$

$$W^G_\gamma(u)(R) := \left\{ f : \quad \sup_{\lambda>0} u(\lambda)^{-1} \sum_{j\geq -2\gamma\log_2(\lambda)} \sum_k \beta^2_{jk} 1_{\sum_k \beta^2_{jk}\leq\lambda^2 2^{j/2}\sqrt{j}} \leq R^2 \right\}.$$

If $2^{j_1} = \epsilon^{-2\gamma'}$ for some $\gamma' > \gamma$, the maxisets are $MS(\hat{\theta}^L, u^2(\epsilon)) = W^L_\gamma(u) \cap \mathcal{B}_{2,\gamma',\infty}(u)$ and $MS(\hat{\theta}^G, u^2(\epsilon)) = W^G_\gamma(u) \cap \mathcal{B}_{2,\gamma',\infty}(u)$.

Theorems 4 and 5 are proved in Section 8. Note that $u(\epsilon) = \epsilon^r$ for $r > 0$ corresponds to the particular case studied before with $M = 1$ and $\delta = 2 - r$. Theorem 5 is especially interesting when the target rate is non-parametric. In this case, rates like $u^2(\epsilon) = \epsilon^{2r}|\log(\epsilon)|^{2r'}$ for $r' \geq 0$ are of interest because they appear in the minimax adaptive framework. Here (5.1) is still satisfied with $M = 1$ and $\delta = 2 - r$. For instance, to study the adaptive global thresholding procedure, $u(\epsilon) = \epsilon^{8\alpha/(1+4\alpha)}|\log(\epsilon)|^{2\alpha/(1+4\alpha)}$ $(0 < \alpha < 1/4)$. In the lights of these results, it is of particular interest to compare thresholding and quadratic procedures when the rate is of this form. That is the goal of the following section.

5.4. Maxiset comparison.

Here we compare thresholding (with thresholds introduced in Theorem 5) and quadratic procedures for the rate $u^2(\epsilon)$, where $u(\epsilon) = \epsilon^r|\log(\epsilon)|^{r'}$. We still consider $\hat{\theta}^Q$, $\hat{\theta}^L$ and $\hat{\theta}^G$: the quadratic, local thresholding and global thresholding procedures. First, we have to state properties on the links between the function spaces (that are the maxisets of the procedures). Using similar arguments as for Property 2, it is easy to state the following result.

Property 5. *Let $0 < r < 2$, $r' \geq 0$. If $0 < \gamma < 2 - r$, then $W^L_\gamma(u) \not\subset W^G_\gamma(u)$ and $W^G_\gamma(u) \not\subset W^L_\gamma(u)$.*

We now establish the links between generalized Besov spaces and generalized weak Besov spaces. Note that the power of the logarithmic term plays a role for weak global Besov spaces.

Property 6. *Let $0 < r < 2$ and $r' \geq 0$. Then, $\mathcal{B}_{2,\gamma,\infty}(u) \subsetneq W^L_\gamma(u)$ if $0 < \gamma \leq 2 - r$. If $0 < \gamma < 2 - r$ or if $\gamma = 2 - r$, $r' > 1/2$,*

$$\mathcal{B}_{2,\gamma,\infty}(u) \subsetneq W^G_\gamma(u), \quad W^G_\gamma(u) \not\subset W^L_\gamma(u). \tag{5.2}$$

If $\gamma = 2 - r$, $r' < 1/2$,

$$\mathcal{B}_{2,2-r,\infty}(u) = W^G_{2-r}(u). \tag{5.3}$$

If $\gamma = 2 - r$, $r' = 1/2$, for any $R \leq (4-2r)^{1/4}$ there exists R' such that

$$\mathcal{B}_{2,2-r,\infty}(u)(R) \subset W^G_{2-r}(u)(R) \subset \mathcal{B}_{2,2-r,\infty}(u)(R').$$

Theorem 6. *When $\gamma < 2 - r$, the first two conclusions of Theorem 3 remain valid when ϵ^{2r} is replaced with $u^2(\epsilon)$. When $\gamma = 2-r$, if $r' > 1/2$, the thresholding procedure outperforms the quadratic one and local and global thresholding are not comparable since*

$$MS(\hat\theta^Q, u^2(\epsilon)) \subsetneq MS(\hat\theta^L, u^2(\epsilon)) \quad \text{and} \quad MS(\hat\theta^Q, u^2(\epsilon)) \subsetneq MS(\hat\theta^G, u^2(\epsilon)),$$

and

$$MS(\hat\theta^L, u^2(\epsilon)) \not\subset MS(\hat\theta^G, u^2(\epsilon)) \quad \text{and} \quad MS(\hat\theta^G, u^2(\epsilon)) \not\subset MS(\hat\theta^L, u^2(\epsilon)).$$

If $r' < 1/2$, the quadratic procedure and the global thresholding one achieve the same performance and the local thresholding procedure outperforms the other ones since $MS(\hat\theta^Q, u^2(\epsilon)) = MS(\hat\theta^G, u^2(\epsilon)) \subsetneq MS(\hat\theta^L, u^2(\epsilon))$. If $r' = 1/2$, this last result remains valid if $MS(\hat\theta^G, u^2(\epsilon))(R') \subset W^G_{2-r}(u)((4-2r)^{1/4})$.

Remark 2. When $r' = 1/2$, we need an accurate control of constants involved in the proofs of Property 6 and Theorem 6 that cannot be reached by using thresholding procedures. So, whether the last result of Theorem 6 remains true for any value of R' remains an open question; we conjecture that the answer is yes.

6. Connections with minimax results.

In this section, our goal is to establish connections between maxiset and minimax results. Indeed, in the first part, we show how to deduce minimax properties of a given procedure from maxiset results. To prove that a procedure is minimax on

$\mathcal{F}$, we point out the minimax rate ρ_ϵ associated with $\mathcal{F}$. Then we compute the maxiset of the procedure for the rate ρ_ϵ by using theorems of the previous section and prove that $\mathcal{F}$ is included in the maxiset. Note that many of the minimax results established in Section 6.1 are already known. In the second part, we focus on procedures that are optimal on Besov spaces $\mathcal{B}^\alpha_{p,\infty}$ from a minimax point of view, and we compare these procedures from a maxiset point of view.

6.1. Minimax properties of procedures deduced from maxiset results. Recall that the minimax rate on $\mathcal{B}^\alpha_{p,\infty}$ is ϵ^2 if $p \geq 2$, $\alpha \geq 1/4$, or $p < 2$, $\alpha > 1/(2p)$. It is also the adaptive minimax rate. When $p \geq 2$, $\alpha < 1/4$, the minimax rate is $\epsilon^{16\alpha/(1+4\alpha)}$, but the adaptive minimax rate is $(|\log \epsilon|^{1/4}\epsilon)^{16\alpha/(1+4\alpha)}$.

We begin with quadratic procedures, which means that $j_1 = j_0$ is the only parameter to fix (equivalently γ). Applying Theorem 1 for the minimax rates and using the inclusions between the Besov spaces given in Property 1, we obtain the following result.

Result 1. *The quadratic procedure built with $\gamma = 1$ is minimax on $\mathcal{B}^\alpha_{p,\infty}$ if $p \geq 2$, $\alpha \geq 1/4$ or $p < 2$, $\alpha \geq 1/p - 1/4$. The quadratic procedure built with $\gamma = 2/(1+4\alpha)$ is minimax on $\mathcal{B}^\alpha_{p,\infty}$ if $p \geq 2$, $\alpha \leq 1/4$.*

Let us focus now on local thresholding procedures. The soft procedures ($\mu = \tau$) for which we take γ as large as possible are considered, and the threshold is chosen as in Theorem 2. We apply Theorem 2 when $j_1 = \infty$ for the minimax rates, and we use Property 3 giving results about the inclusions of the Besov spaces in the weak local Besov spaces.

Result 2. *The local soft thresholding procedure built with $\gamma = 1$ is minimax on $\mathcal{B}^\alpha_{p,\infty}$ if $p \geq 2$, $\alpha \geq 1/4$ or $p < 2$, $\alpha > 1/(2p)$. The local soft thresholding procedure built with $\gamma = 2/(1+4\alpha)$ is minimax on $\mathcal{B}^\alpha_{p,\infty}$ if $p \geq 2$, $\alpha \leq 1/4$.*

Now we study the global thresholding procedures. The soft procedures are considered ($\lambda = 1 + m2^{-j/2}\sqrt{\tau}$ for some constant $m > 0$). We consider non-limited procedures, which means that $j_1 = \infty$ and the threshold is chosen as in Theorem 2. We apply Theorem 2 and use Property 3 giving results on the inclusions of the Besov spaces in the weak global Besov spaces.

Result 3. *The global soft thresholding procedure built with $\gamma = 1$ is minimax* *on* $\mathcal{B}^{\alpha}_{p,\infty}$ *if* $p \geq 2,\ \alpha \geq 1/4$ *or* $p < 2,\ \alpha \geq 1/p - 1/4$*. The global soft thresholding procedure built with* $\gamma = 2/(1+4\alpha)$ *is minimax on* $\mathcal{B}^{\alpha}_{p,\infty}$ *if* $p \geq 2,\ \alpha \leq 1/4$*.*

Lastly, we study the adaptive thresholding procedures (we take $\gamma = 1$ and the thresholds as before) and adaptive minimax rates on $\mathcal{B}^{\alpha}_{p,\infty}$, $p \geq 2$. In this case, the target rate is generalized: take $u^2(\epsilon) = \epsilon^2$ if $\alpha \geq 1/4$ and $u^2(\epsilon) = (\sqrt{|\log(\epsilon)|}\epsilon^2)^{8\alpha/(1+4\alpha)}$ if $\alpha < 1/4$. Note that the following property (proved in Section 7) holds.

Property 7. *Let* $p \geq 2$ *and* $\alpha > 0$*. Then* $\mathcal{B}^{\alpha}_{p,\infty} \subset \mathcal{B}^{\alpha}_{2,\infty} \subset W^G_1(u)$*, and* $\mathcal{B}^{\alpha}_{p,\infty} \not\subset W^L_1(u)$ *if* $\alpha < 1/4$*.*

Now, using Theorem 6 for $r = 8\alpha/(1+4\alpha)$, $r' = 2\alpha/(1+4\alpha) < 1/2$ and $\gamma = 1$, we have the following.

Result 4. *The adaptive soft local procedure is not adaptive minimax on* $\mathcal{B}^{\alpha}_{p,\infty}$ *for* $p \geq 2,\ \alpha < 1/4$*. The adaptive soft global procedure is adaptive minimax on* $\mathcal{B}^{\alpha}_{p,\infty}$ *for* $p \geq 2,\ \alpha > 0$*.*

6.2. Comparisons between procedures.

The parameters of all procedures are chosen to have good minimax properties (see Sections 4.1, 4.2 and 4.3). Depending on the rate, they are non-adaptive. Applying Theorem 3, we obtain the following result.

Result 5. *If the target rate is* ϵ^2 *or* $\epsilon^{16\alpha/(1+4\alpha)}$ *for some* $0 < \alpha < 1/4$*, the quadratic procedure is as good as the (non-adaptive) soft global procedure for the maxiset criterion. The soft local procedure outperforms the quadratic one (and then also the non-adaptive soft global thresholding procedure) from the maxiset point of view.*

Now, let $u^2(\epsilon) = (\sqrt{|\log(\epsilon)|}\epsilon^2)^{8\alpha/(1+4\alpha)}$ for $\alpha > 0$ be the best rate achievable by adaptive procedures on $\mathcal{B}^{\alpha}_{p,\infty}$ when $\alpha < 1/4, p \geq 2$. We focus on non-adaptive procedures and choose $\gamma = 2/(1+4\alpha)$. Applying Theorem 6 with $r = 8\alpha/(1+4\alpha) = 2 - \gamma$ and $r' = 2\alpha/(1+4\alpha) < 1/2$, we obtain the following result.

Result 6. *If the target rate is* $u^2(\epsilon)$*, the non-adaptive soft local procedure outperforms the quadratic one and the non-adaptive soft global thresholding procedure (that achieves the same performance as the quadratic one) from the* *maxiset point of view.*

Choosing now $\gamma = 1$, we consider adaptive procedures. Applying Theorem 6 with $r = 8\alpha/(1+4\alpha) > 2-\gamma$, we obtain the following result.

Result 7. *If the target rate is $u^2(\epsilon)$, the quadratic procedure is the worst method from the maxiset point of view.*

7. Proofs for the results on function spaces.

In the sequel, c denotes a positive constant that may change from line to line. For the sake of simplicity, and without loss of generality, we assume that the number of non-zero wavelet coefficients of any signal at each level j is exactly 2^j. Detailed proofs of Properties 2, 4 and 6 are available on http://www3.stat.sinica.edu.tw/statistica

7.1. Proof of Property 2.

Inclusions (3.1) are obvious. Inclusion (3.3) is a direct consequence of (3.6) and of the first part of (3.5). To establish the first part of (3.2), we consider a **sparse** sequence (β_{jk}) such that at each level j, only one wavelet coefficient takes the value $j^{1/4}2^{-j\beta}$, with $0 < \beta < r/(4(2-r)) = 1/(4-2r) - 1/4$, while the others are 0. Using straightforward computations, we easily prove that the function $f = \sum_{j=-1}^{+\infty}\sum_k \beta_{jk}\psi_{jk}$ belongs to $W^L_{2-r,\gamma}$ but not to $W^G_{2-r,\gamma}$. To establish that the second part of (3.2) holds, we build a **dense** sequence such that all its wavelet coefficients at the level j take the value $\beta_{jk} = 2^{-j\beta}\sqrt{j}$, with $1/(2(2-r)) + 1/4 < \beta \leq \min(1/2 + r/(4\gamma); 1/(2-r))$, which is possible as soon as $\gamma < 2-r$. Using straightforward computations, we easily prove that the function $f = \sum_{j=-1}^{+\infty}\sum_k \beta_{jk}\psi_{jk}$ belongs to $W^G_{2-r,\gamma}$ but not to $W^L_{2-r,\gamma}$.

7.2. Proof of Property 3.

Inclusion (3.4) is very simple to obtain since we just omit the terms $1_{|\beta_{jk}|\leq\lambda\sqrt{j}}$ and $1_{\sum_k \beta_{jk}^2 \leq \lambda^2 2^{j/2}\sqrt{j}}$ in the definition of weak Besov spaces. To prove the last points, note that if $p \geq 2$ and $s \geq 1/4$, or if $p < 2$ and $s \geq 1/p - 1/4$, we have $\mathcal{B}^s_{p,\infty} \subset \mathcal{B}^{1/4}_{2,\infty} \subset W^*_{1,1}$, where $W^*_{1,1}$ is either the weak local Besov space $W^L_{1,1}$ or the weak global one $W^G_{1,1}$. Moreover, when $p < 2$ and $1/(2p) < s < 1/p - 1/4$, we have

$$\begin{aligned}\lambda^{-1}\sum_{j\geq -2\log_2(\lambda)}\sum_k \beta_{jk}^2 1_{|\beta_{jk}|\leq\lambda\sqrt{j}} &= \lambda^{-1}\sum_{j\geq -2\log_2(\lambda)}\sum_k |\beta_{jk}|^p|\beta_{jk}|^{2-p}1_{|\beta_{jk}|\leq\lambda\sqrt{j}}\\ &\leq \lambda^{-1}\sum_{j\geq -2\log_2(\lambda)}\sum_k |\beta_{jk}|^p(\lambda\sqrt{j})^{2-p}.\end{aligned}$$

If f is assumed to belong to $\mathcal{B}^s_{p,\infty}$, we get for any $0<\lambda<1$,

$$\begin{aligned}\lambda^{-1}\sum_{j\geq -2\log_2(\lambda)}\sum_k \beta_{jk}^2 1_{|\beta_{jk}|\leq\lambda\sqrt{j}} &\leq c\lambda^{-1}\sum_{j\geq -2\log_2(\lambda)} 2^{-jp(s+1/2-1/p)}(\lambda\sqrt{j})^{2-p}\\ &\leq c[\log_2(\lambda^{-1})]^{1-p/2}\lambda^{2ps-1}\leq c',\end{aligned}$$

where c' is a constant. So, $f\in W^L_{1,1}$ and $\mathcal{B}^s_{p,\infty}\subset W^L_{1,1}$.

7.3. Proof of Property 4.

Property 4 is a particular case of Property 6. The proof of the strict inclusion of the Besov space in the weak local Besov space is similar for the polynomial rate and for the generalized rate $u^2(\epsilon)$. We prove strict inclusion for the polynomial rate. The proof of the links between the Besov space and the weak global Besov space is more complicated in the case of the generalized rate $u^2(\epsilon)$, because there is a question on the power of the logarithmic term. In the next subsection we prove Property 6, which implies the second statement of (3.5) and the equality (3.6). Let us prove the first part of (3.5). Using Property 3, the inclusion is valid for any $0<r<2$ and any $0<\gamma\leq 2-r$. To prove strict inclusion, consider the sequence (β_{jk}) such that at each level j, $n_j=\lfloor 2^{jm}\rfloor$ wavelet coefficients take the value $\sqrt{j}2^{-j(r/(4\gamma)+m/2)}$ for some constant $m\in]0,(2-r)/(2\gamma)[$ with the others equal to 0 (the notation $\lfloor 2^{jm}\rfloor$ denotes the integer part of 2^{jm}). The function $f=\sum_{j=-1}^{+\infty}\sum_k\beta_{jk}\psi_{jk}$ does not belong to $\mathcal{B}^{r/(4\gamma)}_{2,\infty}$, but belongs to $W^L_{2-r,\gamma}$.

7.4. Proof of Property 6.

The first statement is proved with the same argument as was Property 4. Now consider the case of the weak global Besov space. The following inclusions are obvious: $\mathcal{B}_{2,\gamma,\infty}(u)\subset W^G_\gamma(u)$, and $\mathcal{B}_{2,\gamma,\infty}(u)\subset W^L_\gamma(u)$, if $\gamma\leq 2-r$. Using Property 5, we have $\mathcal{B}_{2,\gamma,\infty}(u)\subsetneq W^G_\gamma(u)$, if $\gamma<2-r$. This proves (5.2) when $\gamma<2-r$. Still considering $\mathcal{B}_{2,2-r,\infty}(u)\subset W^G_{2-r}(u)$, we want to prove that inclusion is strict when $r'>1/2$. Consider the **dense** sequence (β_{jk}) such that at each level j, and for any k, $\beta_{jk} = 2^{-j\beta} j^{\alpha}$, where

$$\beta = \frac{1}{4} + \frac{1}{2(2-r)} > \frac{1}{2} \quad \text{and} \quad \frac{r'}{2} < \alpha \leq \frac{1}{2-r}\left(r' - \frac{r}{4}\right),$$

which is possible if and only if $r' > 1/2$. The function built with the sequence (β_{jk}) belongs to $W^G_{2-r}(u)$, but not to $\mathcal{B}_{2,2-r,\infty}(u)$ and not to $W^L_{2-r}(u)$. Let us now prove (5.3). We assume that the function $f = \sum_j \sum_k \beta_{jk}\psi_{jk}$ belongs to $W^G_\gamma(u)(R)$ with $\gamma = 2 - r$ and $u(\lambda) = \lambda^r[\log(\lambda^{-1})]^{r'}$. Set $B_j = \sum_k \beta^2_{jk}$. Using the definition of the weak global Besov space, we have that

$$\forall\, \lambda > 0, \quad \forall\, j \geq -2(2-r)\log_2(\lambda), \quad B_j 1_{B_j \leq \lambda^2 2^{j/2}\sqrt{j}} \leq R^2\, u(\lambda). \tag{7.1}$$

Then, for any given λ, we study the behavior of B_j when $B_j > \lambda^2 2^{j/2}\sqrt{j}$. Let us set, if $r' < 1/2$, $m(r, r', R)$ such that

$$R^2\left(\frac{1}{4-2r}\right)^{r'} = \left[-2(2-r)\log_2(m(r,r',R))\right]^{1/2-r'}.$$

Then determine, if there exist, indexes j such that

$$\exists\, \lambda \in]0; m(r,r',R)], \text{ such that } \left\{\begin{array}{lll} j & \geq & -2(2-r)\log_2(\lambda) \\ B_j & > & \lambda^2 2^{j/2}\sqrt{j}, \end{array}\right. \tag{7.2}$$

or, equivalently,

$$\exists\, \lambda \in]0; m(r,r',R)], \text{ such that } \left\{\begin{array}{lll} 2^{-\frac{1}{(4-2r)}j} & \leq & \lambda \\ \lambda & < & (B_j 2^{-j/2} j^{-1/2})^{1/2}. \end{array}\right.$$

We deduce that these indexes j must verify

$$B_j > 2^{(\frac{1}{2} - \frac{1}{2-r})j}\sqrt{j} = 2^{-\frac{r}{4-2r}j}\sqrt{j}. \tag{7.3}$$

Let j_0 denote such an index and take $\lambda_0 = (B_{j_0} 2^{-j_0/2} j_0^{-1/2})^{1/2}$. Using (7.3), we have

$$\log_2(\lambda_0) \;=\; \frac{1}{2}\log_2(B_{j_0} 2^{-j_0/2} j_0^{-1/2}) > -\frac{j_0}{2(2-r)}.$$

Since we have $j_0 \geq -2(2-r)\log_2(\lambda_0)$, Assumption (7.1) has to be satisfied for λ_0 and j_0. Then

$$\begin{aligned} B_{j_0} 1_{B_{j_0} \leq \lambda_0^2 2^{j_0/2}\sqrt{j_0}} \leq R^2 u(\lambda_0) \quad &\Longleftrightarrow \quad B_{j_0} \leq R^2 u(\lambda_0) \\ &\Longleftrightarrow \quad B_{j_0} \leq R^2 \lambda_0^r \left(\log_2(\lambda_0^{-1})\right)^{r'}, \end{aligned}$$

which yields that

$$B_{j_0} \leq \left[R^2 \left(\frac{1}{4-2r} \right)^{r'} \right]^{2/(2-r)} 2^{-j_0 \frac{r}{4-2r}} j_0^{(4r'-r)/(4-2r)},$$

which contradicts (7.3). Note that for any $m(r, r', R)$ we also obtain a contradiction if

$$r' = \frac{1}{2} \quad \text{and} \quad R^2 \left(\frac{1}{4-2r} \right)^{1/2} \leq 1.$$

We deduce that there exists no index j such that (7.2) is true. It means that for any $0 < \lambda \leq m(r, r', R)$, for any $j \geq -2(2-r)\log_2(\lambda)$, $B_j \leq \lambda^2 2^{j/2}\sqrt{j}$. It follows that, with $\gamma = 2 - r$,

$$\sup_{\lambda > 0} u(\lambda)^{-1} \sum_{j \geq -2\gamma \log_2(\lambda)} \sum_k \beta_{jk}^2$$
$$\leq c \left(\sup_{0 < \lambda \leq m(r,r',R)} u(\lambda)^{-1} \sum_{j \geq -2(2-r)\log_2(\lambda)} \sum_k \beta_{jk}^2 1_{B_j \leq \lambda^2 2^{j/2}\sqrt{j}} + \|f\|_2^2 \right).$$

This completes the proof since f is supposed to belong to $W^G_{2-r,2-r}$.

7.5. Proof of Property 7.

We consider the **dense** function introduced in Section 7.1. with $\gamma = 1$. Exactly as in Section 7.1, we prove that f does not belong to $W_1^L(u)$. Take $\beta = 1/2 + r/4$ with $r = 8\alpha/(1+4\alpha)$. Then, if $\alpha < 1/4$,

$$\sup_j 2^{pj(\alpha+1/2-1/p)} \sum_k |\beta_{jk}|^p \;=\; \sup_j 2^{pj(\alpha+1/2-1/p)} 2^j 2^{-pj\beta} j^{p/2} < +\infty$$

because $\alpha + 1/2 - \beta < 0$. We conclude that $f \in \mathcal{B}^\alpha_{p,\infty}$ and $\mathcal{B}^\alpha_{p,\infty} \not\subset W_1^L(u)$.
Let us prove now that $\mathcal{B}^\alpha_{2,\infty} \subset W_1^G(u)$. Assume $f \in \mathcal{B}^\alpha_{2,\infty}$. If $\alpha \geq 1/4$ then, for any $\lambda > 0$,

$$u(\lambda)^{-1} \sum_{j=-2\log_2(\lambda)}^{+\infty} \sum_k \beta_{jk}^2 1_{\sum_k \beta_{jk}^2 \leq 2^{j/2}\lambda^2\sqrt{j}}$$
$$\leq u(\lambda)^{-1} \sum_{j=-2\log_2(\lambda)}^{+\infty} \sum_k \beta_{jk}^2$$
$$\leq u(\lambda)^{-1} \sum_{j=-2\log_2(\lambda)}^{+\infty} 2^{-2j\alpha} \leq c u(\lambda)^{-1} \lambda^{4\alpha}.$$

If $\alpha < 1/4$ then, with $2^{j_\alpha} = (|\log(\lambda)|\lambda^4)^{-\frac{1}{1+4\alpha}}$ we get, for any $\lambda > 0$,

$$\begin{aligned} u(\lambda)^{-1} \sum_{j=-2\log_2(\lambda)}^{+\infty} \sum_k \beta_{jk}^2 1_{\sum_k \beta_{jk}^2 \leq 2^{j/2}\lambda^2\sqrt{j}} &\leq u(\lambda)^{-1} \left(\sum_{j=-2\log_2(\lambda)}^{j_\alpha - 1} 2^{j/2}\lambda^2\sqrt{j} + \sum_{j=j_\alpha}^{+\infty} \sum_k \beta_{jk}^2 \right) \\ &\leq cu(\lambda)^{-1} \left(2^{j_\alpha/2}\lambda^2\sqrt{|\log(\lambda)|} + 2^{-2j_\alpha\alpha} \right). \end{aligned}$$

Taking the supremum in $\lambda > 0$, we conclude that $f \in W_1^G(u)$ and then $\mathcal{B}_{2,\infty}^\alpha \subset W_1^G(u)$.

Remark 3. Note that it can easily be proved that any function considered in Section 7 belongs to $\mathcal{B}_{2,\gamma',\infty}(u) \subset \mathbb{L}_2([0,1])$, when γ' is large enough.

8. Proofs for the results on the statistical procedures.

The results for the maxiset theory are based on a sharp study of the bias of the estimation procedures. In Section 8.1, we give an upper bound and a lower bound for the expected quadratic errors due to the procedures. In Section 8.2, we deduce the proof of the main results stated in Theorem 4 and Theorem 5. In the following sections, we prove the preliminary results. Arguments for global thresholding being similar to those of local thresholding, proofs are available on http://www3.stat.sinica.edu.tw/statistica

8.1. Preliminary results.

We give bounds for the quadratic error of our procedures. It is worth noting that we do not make any regularity assumption on the function f. First, we state the results concerning the local thresholding procedure and next we deal with the global one. The maxisets of the local thresholding estimates are determined in the following propositions.

Proposition 1. *Let $\tau = \kappa j$ and let a constant $K_2 < 1$. Then, for κ large* *enough, there exists some constant* $c_2 > 0$ *such that*

$$\mathbb{E}(\hat{\theta}^L - \theta)^2 \leq c_2 \left[2^{j_0} \epsilon^4 + \theta\, \epsilon^2 + \left(\sum_{j=j_0}^{j_1-1} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq \epsilon\sqrt{\tau}} \right)^2 + \left(\sum_{j=j_0}^{j_1-1} \sum_k \epsilon^2 (\tau + |\mu - \tau|) 1_{|\beta_{jk}| > K_2 \epsilon \sqrt{\tau}} \right)^2 + \left(\sum_{j=j_1}^{+\infty} \sum_k \beta_{jk}^2 \right)^2 \right].$$

If $\mu/\tau > c(1-K_2^2)$, *where* $c > 1$ *is a constant, there exists some constant* $c_1 > 0$ *with*

$$\mathbb{E}(\hat{\theta}^L - \theta)^2 \geq c_1 \left[\left(\sum_{j=j_0}^{j_1-1} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq K_2 \epsilon\sqrt{\tau}} + \sum_{j=j_1}^{+\infty} \sum_k \beta_{jk}^2 \right)^2 \right].$$

Proposition 2. *Consider a continuous function* $u : [0,1] \longrightarrow \mathbb{R}_+$ *such that (5.1) is satisfied. Then, if* $K_2 \geq 1/2$ *and* $2^{j_0} = \epsilon^{-2\gamma}$,

$$\sup_{\epsilon > 0} u(\epsilon)^{-1} \sum_{j \geq j_0} \sum_k \epsilon^2 \tau 1_{|\beta_{jk}| > K_2 \epsilon \sqrt{\tau}} \leq \frac{4M}{1 - 2^{-\delta}} \sup_{\epsilon > 0} u(\epsilon)^{-1} \sum_{j \geq j_0} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq \epsilon\sqrt{\tau}}.$$

We deal next with the global thresholding estimate.

Proposition 3. *Let* $\tau = \kappa j$ *and take constants* $K_2 < 1 < K_1$. *Then, for* κ *large enough, there exists some constant* $c_2 > 0$ *such that if* $\lambda - 1 = m 2^{-j/2}\sqrt{\tau}$, *for* $m \geq 0$,

$$\mathbb{E}(\hat{\theta}^G - \theta)^2 \leq c_2 \left[\epsilon^2 \theta + \left(\sum_{j=j_0}^{j_1-1} 2^j \epsilon^2 [(\lambda - 1) + 2^{-j/2}\sqrt{\tau}] 1_{\sum_k \beta_{jk}^2 > K_2 \epsilon^2 \sqrt{2^j \tau}} \right)^2 + \epsilon^2 \theta + \left(\sum_{j=j_0}^{j_1-1} \sum_k \beta_{jk}^2 1_{\sum_k \beta_{jk}^2 < K_1 \epsilon^2 \sqrt{2^j \tau}} \right)^2 + \left(\sum_{j=j_1}^{+\infty} \sum_k \beta_{jk}^2 \right)^2 \right].$$

Moreover, if $(\lambda - 1) 2^{j/2} \tau^{-1/2} > c(1 - K_2)$, *where* $c > 1$ *is a constant, there exists some constant* $c_1 > 0$ *such that*

$$\mathbb{E}(\hat{\theta}^G - \theta)^2 \geq c_1 \left[\left(\sum_{j=j_0}^{j_1-1} \sum_k \beta_{jk}^2 1_{\sum \beta_{jk}^2 < K_2 \epsilon^2 \sqrt{2^j \tau}} \right)^2 + \left(\sum_{j=j_1}^{+\infty} \sum_k \beta_{jk}^2 \right)^2 \right].$$

Proposition 4. *Consider a continuous function $u : [0,1] \longrightarrow \mathbb{R}_+$ such that (5.1) is satisfied. If $K_2 \geq 1/2$, $2^{j_0} = \epsilon^{-2\gamma}$, and if we suppose there exists a constant $m > 0$ such that, for any $j \geq j_0$, $\lambda - 1 \leq m\, 2^{-j/2}\sqrt{\tau}$,*

$$\sup_{\epsilon>0} u(\epsilon)^{-1} \sum_{j \geq j_0} 2^j (\lambda - 1)\epsilon^2 1_{\sum_k \beta_{jk}^2 > K_2 \epsilon^2 \sqrt{2^j \tau}} \leq \frac{2mM}{1 - 2^{-\delta/2}} \sup_{\epsilon>0} u(\epsilon)^{-1} \sum_{j \geq j_0} \sum_k \beta_{jk}^2 1_{\sum_k \beta_{jk}^2 \leq \epsilon^2 \sqrt{2^j \tau}}.$$

Remark 4. Using the first part of Proposition 1 with $\mu = 0$ and the first part of Proposition 3 with $\lambda = 1$, we easily see that the maxisets associated with the soft thresholding procedures are included in the maxisets associated with the hard thresholding procedures. Whether these inclusions are strict remains an open question.

8.2. Proofs of Theorems 4 and 5.

Using Proposition 1, Theorem 4 is obvious. We prove Theorem 5 for local thresholding estimates. Let $\hat{\theta}$ be $\hat{\theta}^L$ and

$$\mathcal{A} = W_\gamma^L(u) \quad \text{if} \quad j_1 = +\infty,$$
$$\mathcal{A} = W_\gamma^L(u) \cap \mathcal{B}_{2,\gamma',\infty}(u) \quad \text{if} \quad j_1 = -2\gamma' \log_2(\epsilon).$$

First, let us assume that $f \in MS(\hat{\theta}, u^2(\epsilon))(R)$. So, for any $\epsilon > 0$, $\mathbb{E}(\hat{\theta} - \theta)^2 \leq R^2 u^2(\epsilon)$. If $K = K_2\sqrt{\kappa} < 1$, using Proposition 1 with $\mu = \tau$,

$$\left(\sum_{j=-2\gamma \log_2(K\epsilon)}^{+\infty} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq K\epsilon\sqrt{j}} \right)^2 \leq \left(\sum_{j=-2\gamma \log_2(\epsilon)}^{+\infty} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq K_2\epsilon\sqrt{\tau}} \right)^2 \leq c_1^{-1} \mathbb{E}(\hat{\theta} - \theta)^2 \leq c_1^{-1} R^2 u^2(\epsilon).$$

If $K = K_2\sqrt{\kappa} \geq 1$,

$$\left(\sum_{j=-2\gamma \log_2(\epsilon)}^{+\infty} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq \epsilon\sqrt{j}} \right)^2 \leq \left(\sum_{j=-2\gamma \log_2(\epsilon)}^{+\infty} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq K_2\epsilon\sqrt{\tau}} \right)^2 \leq c_1^{-1} R^2 u^2(\epsilon).$$

Recalling that

$$\mathbb{E}(\hat{\theta} - \theta)^2 \geq c_1 \left(\sum_{j=j_1}^{+\infty} \sum_k \beta_{jk}^2 \right)^2,$$

we obtain $MS(\hat{\theta}, u^2(\epsilon)) \subset \mathcal{A}(R')$ with $(R')^2 = c_1^{-1} R^2 \max(M^2 K^{2\delta-4}, 1)$. Since u satisfies (5.1), if $f \in \mathcal{A}(R')$, with $K = \sqrt{\kappa} \geq 1$, and still using Proposition 1,

$$\left(\sum_{j=-2\gamma \log_2(\epsilon)}^{+\infty} \sum_k \beta_{jk}^2 1_{|\beta_{jk}| \leq K\epsilon\sqrt{j}} \right)^2 \leq (R')^2 M^2 K^{4-2\delta} u^2(\epsilon).$$

So, using Proposition 2, f belongs to $MS(\hat{\theta}, u^2(\epsilon))(R)$, with

$$R^2 = c_2 \left(\left(\frac{4M}{1-2^{-\delta}} \right)^2 + 1 \right) (R')^2 M^2 K^{4-2\delta}.$$

This completes the proof. Similarly, combining Proposition 3 and Proposition 4, one obtains Theorem 5 for the global thresholding procedure.

8.3. Notations.

In the sequel, to provide upper bounds for many terms. We use exponential inequalities. And, with an appropriate choice of constants, these terms will then be negligible. In the study of the local threshold estimate, the exponential inequality will deal with Gaussian variables. More precisely,

$$\mathbb{P}(|y_{jk} - \beta_{jk}| \geq x) = \mathbb{P}(|z_{jk}| \geq \epsilon^{-1} x) \leq 2\exp(-\frac{x^2}{2\epsilon^2}),$$

which is bounded by $2(2^{-j\kappa}) \wedge \epsilon^{\kappa}$ as soon as $x \geq \sqrt{2\log(2)\kappa}\,(\epsilon\sqrt{|\log_2 \epsilon|} \vee \epsilon\sqrt{j})$. This inequality is valid for every j, k. So, in the sequel, we use the notation $LD(j, \epsilon, \kappa)$ to denote a large deviation term that depends on j, ϵ and κ. The value of $LD(j, \epsilon, \kappa)$ may change from line to line, but the constant κ is chosen to ensure that

$$\lim_{\epsilon \to 0} \epsilon^{-2} \sum_{j=j_0}^{+\infty} |LD(j, \epsilon, \kappa)| = 0.$$

8.4. Proof of Proposition 1.

We note $\hat{\theta} = \hat{\theta}^L$. Since $\mathbb{E}\hat{\theta}_{jk} = \beta_{jk}^2$ if $j < j_0$, and $\hat{\theta}_{jk} = 0$ if $j \geq j_1$, the classical decomposition in variance and bias terms gives

$$\mathbb{E}(\hat{\theta} - \theta)^2$$
$$= \mathbb{E} \left(\sum_{j=-1}^{j_1-1} \sum_k \left(\hat{\theta}_{jk} - \mathbb{E}\hat{\theta}_{jk} \right) \right)^2 + \left(\sum_{j=j_0}^{j_1-1} \sum_k (\beta_{jk}^2 - \mathbb{E}\hat{\theta}_{jk}) + \sum_{j=j_1}^{+\infty} \sum_k \beta_{jk}^2 \right)^2.$$

Study of the upper bound

We have

$$\begin{aligned}\mathbb{E}(\hat\theta-\theta)^2 &\leq 2\Bigg[\mathbb{E}\left(\sum_{j=-1}^{j_0-1}\sum_k(\hat\theta_{jk}-\mathbb{E}\hat\theta_{jk})\right)^2+\mathbb{E}\left(\sum_{j=j_0}^{j_1-1}\sum_k\left(\hat\theta_{jk}-\mathbb{E}\hat\theta_{jk}\right)\right)^2\\ &\quad+\left(\sum_{j=j_0}^{j_1-1}\sum_k(\beta_{jk}^2-\mathbb{E}\hat\theta_{jk})\right)^2+\left(\sum_{j=j_1}^{+\infty}\sum_k\beta_{jk}^2\right)^2\Bigg].\end{aligned}$$

The variables $(\hat\theta_{jk}-\mathbb{E}\hat\theta_{jk})_{jk}$ are independent and for any $j\leq j_0-1$ and any k, $\mathbb{E}(\hat\theta_{jk}-\mathbb{E}\hat\theta_{jk})^2\leq c(\epsilon^2\beta_{jk}^2+\epsilon^4)$. So, the first term is bounded as follows.

$$\begin{aligned}\mathbb{E}\left(\sum_{j=-1}^{j_0-1}\sum_k(\hat\theta_{jk}-\mathbb{E}\hat\theta_{jk})\right)^2 &= \sum_{j=-1}^{j_0-1}\sum_k\mathbb{E}(\hat\theta_{jk}-\mathbb{E}\hat\theta_{jk})^2\leq c\sum_{j=-1}^{j_0-1}\sum_k(\epsilon^2\beta_{jk}^2+\epsilon^4)\\ &\leq c\left(\theta\,\epsilon^2+2^{j_0}\epsilon^4\right).\end{aligned}$$

For $j_0\leq j\leq j_1-1$, we use the following lemma of Cai and Low (2005) for $\mu=\tau$.

Lemma 1 (Cai and Low (2005)). *Let $\hat\theta_{jk}^{SL}$ be the local thresholding estimate with $\mu=\tau$. Then*

$$\begin{aligned}\left|\mathbb{E}\left(\hat\theta_{jk}^{SL}-\beta_{jk}^2\right)\right| &\leq 2\left(\epsilon^2\tau\,1_{|\beta_{jk}|>\epsilon\sqrt{\tau}}+\beta_{jk}^2 1_{|\beta_{jk}|\leq\epsilon\sqrt{\tau}}\right)\\ var(\hat\theta_{jk}^{SL}) &\leq c\left(\beta_{jk}^2\,\epsilon^2+\epsilon^4\tau^{1/2}\exp(-\tau/2)\right).\end{aligned}$$

We note that, for every $j\geq j_0$, and for every k,

$$\hat\theta_{jk}^L = \hat\theta_{jk}^{SL}-(\mu-\tau)\,\epsilon^2\left(1_{|y_{jk}|>\epsilon\sqrt{\tau}}-\mathbb{E}1_{|z_{jk}|>\sqrt{\tau}}\right).$$

Then, for $K_2<1$,

$$\begin{aligned}\mathrm{var}(\hat\theta_{jk}^L) &\leq 2\left[\mathrm{var}(\hat\theta_{jk}^{SL}+\epsilon^4(\mu-\tau)^2\mathrm{var}(1_{|y_{jk}|>\epsilon\sqrt{\tau}})\right]\\ &\leq c\Big[\mathrm{var}(\hat\theta_{jk}^{SL})+\epsilon^4(\mu-\tau)^2 1_{|\beta_{jk}|>K_2\epsilon\sqrt{\tau}}\\ &\qquad+\epsilon^4(\mu-\tau)^2\mathbb{P}(|y_{jk}-\beta_{jk}|>(1-K_2)\epsilon\sqrt{\tau})\Big]\\ &\leq c\left[\beta_{jk}^2\,\epsilon^2+\epsilon^4(\mu-\tau)^2 1_{|\beta_{jk}|>K_2\epsilon\sqrt{\tau}}+LD(j,\epsilon,\kappa)\right],\end{aligned}$$

$$\begin{aligned}
\left|\mathbb{E}(\hat{\theta}^{L}_{jk}-\beta^2_{jk})\right| &= c\left|\mathbb{E}(\hat{\theta}^{SL}_{jk}-\beta^2_{jk})-\epsilon^2(\mu-\tau)\mathbb{E}(1_{|y_{jk}|>\epsilon\sqrt{\tau}})+LD(j,\epsilon,\kappa)\right| \\
&\leq c\left|\mathbb{E}(\hat{\theta}^{SL}_{jk}-\beta^2_{jk})\right|+\epsilon^2|\mu-\tau|1_{|\beta_{jk}|>K_2\epsilon\sqrt{\tau}}]+ \\
&\qquad \epsilon^2|\mu-\tau|\mathbb{P}(|y_{jk}-\beta_{jk}|>(1-K_2)\epsilon\sqrt{\tau})+LD(j,\epsilon,\kappa) \\
&\leq c\left[\epsilon^2(|\mu-\tau|+\tau)1_{|\beta_{jk}|>K_2\epsilon\sqrt{\tau}}+\beta^2_{jk}1_{|\beta_{jk}|\leq\epsilon\sqrt{\tau}}+LD(j,\epsilon,\kappa)\right].
\end{aligned}$$

It follows that

$$\mathbb{E}\left(\sum_{j=j_0}^{j_1-1}\sum_k\left(\hat{\theta}_{jk}-\mathbb{E}\hat{\theta}_{jk}\right)\right)^2\leq c\left[\epsilon^2\theta+\left(\sum_{j=j_0}^{j_1-1}\sum_k|\mu-\tau|\epsilon^2 1_{|\beta_{jk}|>K_2\epsilon\sqrt{\tau}}\right)^2\right],$$

$$\begin{aligned}
&\left(\sum_{j=j_0}^{j_1-1}\sum_k(\beta^2_{jk}-\mathbb{E}\hat{\theta}_{jk})\right)^2 \\
&\quad\leq c\left[\left(\sum_{j=j_0}^{j_1-1}\sum_k(|\mu-\tau|+\tau)\epsilon^2 1_{|\beta_{jk}|>K_2\epsilon\sqrt{\tau}}\right)^2+\left(\sum_{j=j_0}^{j_1-1}\sum_k\beta^2_{jk}1_{|\beta_{jk}|\leq\epsilon\sqrt{\tau}}\right)^2\right].
\end{aligned}$$

The result follows.

Study of the lower bound

We focus on the bias terms. Obviously,

$$\mathbb{E}(\hat{\theta}-\theta)^2 \geq \left(\sum_{j=j_0}^{j_1-1}\sum_k(\beta^2_{jk}-\mathbb{E}\hat{\theta}_{jk})+\sum_{j=j_1}^{+\infty}\sum_k\beta^2_{jk}\right)^2.$$

Let $K_1>1$ and $K_2<1$ be positive constants. For any $j_0\leq j\leq j_1$ and any k, the following expansion holds:

$$\begin{aligned}
\mathbb{E}(\beta^2_{jk}-\hat{\theta}_{jk}) &= \mathbb{E}(\beta^2_{jk}-\hat{\theta}_{jk})\left[1_{|\beta_{jk}|\leq K_2\epsilon\sqrt{\tau}}+1_{K_2\epsilon\sqrt{\tau}<|\beta_{jk}|\leq K_1\epsilon\sqrt{\tau}}+1_{K_1\epsilon\sqrt{\tau}<|\beta_{jk}|}\right] \\
&= \beta^2_{jk}1_{|\beta_{jk}|\leq K_2\epsilon\sqrt{\tau}}+\mathbb{E}(\beta^2_{jk}-\hat{\theta}_{jk})1_{K_2\epsilon\sqrt{\tau}<|\beta_{jk}|\leq K_1\epsilon\sqrt{\tau}} \\
&\quad+\mathbb{E}(\beta^2_{jk}-\hat{\theta}_{jk})1_{|\beta_{jk}|>K_1\epsilon\sqrt{\tau}}+LD(j,\epsilon,\kappa).
\end{aligned}$$

First, we get

$$\begin{aligned}
&\mathbb{E}((\beta_{jk}^2 - \hat\theta_{jk}) 1_{|\beta_{jk}|>K_1\epsilon\sqrt{\tau}}) \\
&\quad = \mathbb{E}\left((\beta_{jk}^2 - \hat\theta_{jk}) \left[1_{|y_{jk}|>\epsilon\sqrt{\tau}} + 1_{|y_{jk}|\le\epsilon\sqrt{\tau}} \right] \right) 1_{|\beta_{jk}|>K_1\epsilon\sqrt{\tau}} \\
&\quad \ge LD(j,\epsilon,\kappa) + \mathbb{E}\left((\beta_{jk}^2 + \mu\,\epsilon^2 - y_{jk}^2) \left[1 - 1_{|y_{jk}|\le\epsilon\sqrt{\tau}} \right] \right) 1_{|\beta_{jk}|>K_1\epsilon\sqrt{\tau}} \\
&\quad = LD(j,\epsilon,\kappa) + (\mu-1)\epsilon^2\, 1_{|\beta_{jk}|>K_1\epsilon\sqrt{\tau}} \ge LD(j,\epsilon,\kappa),
\end{aligned}$$

since $\mu > 1$. Next, we have

$$\begin{aligned}
&\mathbb{E}(\beta_{jk}^2 - \hat\theta_{jk}) 1_{K_2\epsilon\sqrt{\tau}<|\beta_{jk}|\le K_1\epsilon\sqrt{\tau}} \\
&\quad \ge \mathbb{E}\left((\beta_{jk}^2 - \hat\theta_{jk}) 1_{|y_{jk}|>\epsilon\sqrt{\tau}} \right) 1_{K_2\epsilon\sqrt{\tau}<|\beta_{jk}|\le K_1\epsilon\sqrt{\tau}} \\
&\quad \ge \mathbb{E}\left((\beta_{jk}^2 - y_{jk}^2 + \mu\epsilon^2) 1_{|y_{jk}|>\epsilon\sqrt{\tau}} 1_{\beta_{jk}^2 - y_{jk}^2 + \mu\epsilon^2 < 0} \right) 1_{K_2\epsilon\sqrt{\tau}<|\beta_{jk}|\le K_1\epsilon\sqrt{\tau}} + LD(j,\epsilon,\kappa)
\end{aligned}$$

Note that

$$\left.\begin{array}{l} y_{jk}^2 - \mu\,\epsilon^2 - \beta_{jk}^2 > 0 \\ |\beta_{jk}| > K_2\epsilon\sqrt{\tau} \end{array}\right\} \Longrightarrow |y_{jk}| > \sqrt{K_2^2 + \frac{\mu}{\tau}}\,\epsilon\sqrt{\tau},$$

implying $\mathbb{E}(\beta_{jk}^2 - \hat\theta_{jk}) 1_{K_2\epsilon\sqrt{\tau}<|\beta_{jk}|\le K_1\epsilon\sqrt{\tau}} \ge LD(j,\epsilon,\kappa)$ as soon as $K_1 < \sqrt{K_2^2 + \frac{\mu}{\tau}}$. Therefore, we obtain $\mathbb{E}(\beta_{jk}^2 - \hat\theta_{jk}) \ge \beta_{jk}^2 1_{|\beta_{jk}|\le K_2\epsilon\sqrt{\tau}} + LD(j,\epsilon,\kappa)$, leading to

$$\mathbb{E}(\hat\theta - \theta)^2 \quad \ge \quad c\left(\sum_{j=j_0}^{j_1-1} \sum_k \beta_{jk}^2 1_{|\beta_{jk}|\le K_2\epsilon\sqrt{\tau}} + \sum_{j=j_1}^{+\infty} \sum_k \beta_{jk}^2 \right)^2,$$

where c is a positive constant. This completes the proof of the lower bound.

8.5. Proof of Proposition 2.

Let j_0 be defined by $2^{j_0} = \epsilon^{-2\gamma}$ for some $\gamma > 0$. Since $K_2 \ge 1/2$, for any $\epsilon > 0$,

$$\begin{aligned}
\sum_{j\ge j_0,k} \epsilon^2\tau 1_{|\beta_{jk}|>K_2\epsilon\sqrt{\tau}} &\le \sum_{j\ge\gamma\log_2(\epsilon^{-2})} \sum_k \epsilon^2\tau \sum_{x\ge -1} 1_{2^x\epsilon\sqrt{\tau}<|\beta_{jk}|\le 2^{x+1}\epsilon\sqrt{\tau}} \\
&\le \sum_{x\ge -1} 2^{-2x} \sum_{j\ge\gamma\log_2(\epsilon^{-2}2^{-2(x+1)})} \sum_k \beta_{jk}^2 1_{|\beta_{jk}|\le(\epsilon 2^{(x+1)})\sqrt{\tau}}.
\end{aligned}$$

If for any $\epsilon > 0$,

$$u(\epsilon)^{-1} \sum_{j\ge\gamma\log_2(\epsilon^{-2})} \sum_k \beta_{jk}^2 1_{|\beta_{jk}|\le\epsilon\sqrt{\tau}} \quad \le \quad S$$

where S lies in $\mathbb{R}_+ \cup \{+\infty\}$, then

$$\sup_{\epsilon>0} u(\epsilon)^{-1} \sum_{j\geq j_0} \sum_k \epsilon^2 \tau 1_{|\beta_{jk}|>K_2\epsilon\sqrt{\tau}} \leq \sup_{\epsilon>0} \sum_{x\geq -1} 2^{-2x} S \frac{u(\epsilon 2^{x+1})}{u(\epsilon)} \leq \frac{4SM}{1-2^{-\delta}},$$

which implies the result.

8.. Proof of Proposition 4

Using similar arguments to those of Proposition 2, and recalling that $\tau_j = 2^{j/2}\sqrt{\tau}$, we have for any $\epsilon > 0$,

$$\sum_{j\geq j_0} 2^j(\lambda-1)\epsilon^2 1_{\sum_k \beta_{jk}^2 > K_2\epsilon^2\tau_j}$$
$$\leq \sum_{x\geq -1} 2^{-x} \sum_{j\geq \gamma \log_2(\epsilon^{-2}2^{-(x+1)})} 2^{j/2}\frac{\lambda-1}{\sqrt{\tau}} \sum_k \beta_{jk}^2 1_{\sum_k \beta_{jk}^2 \leq (\epsilon 2^{(x+1)/2})^2 \tau_j}.$$

If there exists $m > 0$ such that for any $j \geq j_0$, $\lambda - 1 \leq m2^{-j/2}\sqrt{\tau}$, then as for the proof of Proposition 2 it is easy to see that

$$\sup_{\epsilon>0} u(\epsilon)^{-1} \sum_{j\geq j_0} 2^j(\lambda-1)\epsilon^2 1_{\sum_k \beta_{jk}^2 > K_2\epsilon^2\tau_j}$$
$$\leq \frac{2Mm}{1-2^{-\delta/2}} \sup_{\epsilon>0} u(\epsilon)^{-1} \sum_{j\geq \gamma\log_2(\epsilon^{-2})} \sum_k \beta_{jk}^2 1_{\sum_k \beta_{jk}^2 \leq \epsilon^2\tau_j}.$$

References

Autin, F., Picard, D. and Rivoirard, V. (2004). Maxiset comparisons of procedures, application to choosing priors in a Bayesian nonparametric setting. *Technical Report.*

Cai, T.T. and Low, M.G. (2005). Nonquadratic estimators of a quadratic functional. Ann. Statist. **33**, 2930–2956

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

Cohen, A., DeVore, R.A., Kerkyacharian, G. and Picard, D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms. *Applied and Computational Harmonic Analysis* **11**, 167–191.

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

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

Donoho, D.L. (1993). Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis* **1**, 100–115.

Johnstone, I.M. (1994). Minimax Bayes, asymptotic minimax and sparse wavelet priors. in *Statistical decision theory and related topics*, Springer, New York, 303-326.

Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases. *Test* **9**, 283–344.

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

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

Rivoirard, V. (2004). Maxisets for linear procedures *Statistics & Probability Letters* **67**, 3, 267–275.

Tribouley, K. (2000). Adaptive Estimation of Integrated Functionals. *Mathematical Methods of Statistics* **9**, 19–36.

Vincent Rivoirard, CNRS UMR 8628, Université Paris XI.

E-mail: (Vincent.Rivoirard@math.u-psud.fr)

Karine Tribouley, CNRS UMR 8628, Université Paris XI and Modal'X, Université Paris X.

E-mail: (Karine.Tribouley@u-paris10.fr)