Maxiset comparisons of procedures, application to choosing priors in a Bayesian nonparametric setting

Florent Autin, Dominique Picard, Vincent Rivoirard

Universités de Paris 6 & Paris 7 - CNRS (UMR 7599)

PRÉPUBLICATIONS DU LABORATOIRE

DE PROBABILITÉS & MODÈLES ALÉATOIRES

4, place Jussieu - Case 188 - 75 252 Paris cedex 05

http://www.proba.jussieu.fr

Maxiset comparisons of procedures, application to choosing priors in a Bayesian nonparametric setting

F. AUTIN, D. PICARD & V. RIVOIRARD

SEPTEMBRE 2004

Prépublication n^o 931

F. Autin : Département de Mathématiques, Université Paris X, 200 avenue de la République, F-92000 Nanterre & Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599, Université Paris VI & Université Paris VII, 4 place Jussieu, Case 188, F-75252 Paris Cedex 05.

D. Picard : Laboratoire de Probabilités et Modèles Aléatoires, CNRS-UMR 7599, Université Paris VI & Université Paris VII, 4 place Jussieu, Case 188, F-75252 Paris Cedex 05.

V. Rivoirard : Laboratoire de Mathématique, UMR 8628, Equipe Probabilités, Statistique et Modélisation, Université Paris-Sud 11, Bât. 425, F-91405 Orsay Cedex.

Maxiset comparisons of procedures, application to choosing priors in a Bayesian nonparametric setting.*

Florent Autin, Dominique Picard and Vincent Rivoirard
CNRS- Universités de Paris X-Nanterre, Paris VII and Paris XI

August 26, 2004

Abstract

In this paper our aim is to provide tools for easily calculating the maxisets of several procedures. Then we apply these results to perform a comparison between several Bayesian estimators in a non parametric setting. We obtain that many Bayesian rules can be described through a general behavior such as being shrinkage rules, limited, and/or elitist rules. This has consequences on their maxisets which happen to be automatically included in some Besov or weak Besov spaces, whereas other properties such as cautiousness imply that their maxiset conversely contains some of the spaces quoted above.

We compare Bayesian rules taking into account the sparsity of the signal with priors which are combination of a Dirac with a standard distribution. We consider the case of Gaussian and heavy tail priors. We prove that the heavy tail assumption is not necessary to attain maxisets equivalent to the thresholding methods. Finally we provide methods using the tree structure of the dyadic aspect of the multiscale analysis, and related to Lepki's procedure, achieving strictly larger maxisets than those of thresholding methods.

1 Introduction

Our aim in this paper is twofold. First, we provide tools for easily calculating the maxisets of several procedures, then we apply these results to perform a comparison between several Bayesian estimators in a non parametric setting.

Let us first briefly recall the definitions of maxisets. 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 with this sequence of models, a loss function $\rho(\hat{q}_n, q(\theta))$, and a rate of convergence α_n tending to 0. Then, we 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:

$$MS(\hat{q}_n, \rho, \alpha_n)(T) = \{\theta \in \Theta,\ \sup_n \mathbb{E}_\theta^n \rho(\hat{q}_n, q(\theta))(\alpha_n)^{-1} \le T\}$$

*Key Words and Phrases: minimax, maxiset, nonparametric estimation, bayesian methods.
AMS 2000 Subject Classification: 62G05, 62G07, 62G20.

In various parametric cases, we can easily prove in regular sequences of models and if ρ is a norm 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), the focus in this domain has mainly be 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 maxiset is associated with the procedure in a genuine way. The comparison of procedures using maxisets is not as famous as minimax comparison. However the results that have been obtained up to now are very promising since they generally show that the maxisets of well-known procedures are spaces which are well established and easily interpretable. For instance, it has been established in [Kerkyacharian and Picard, 1993] that the maxisets of most kernel linear methods are in fact Besov spaces, whereas the maxisets of thresholding estimates (see [Cohen et al., 2001b]) are Lorentz spaces reflecting extremely well the practical observation that wavelet thresholding performs well when the number of wavelet coefficients is small. It has also been observed (see [Kerkyacharian and Picard, 2002]) that there is a deep connection between oracle inequalities and maxisets.

Although looking for the maxiset of a procedure is something different from looking at minimax rates and proving that the procedure is asymptotically minimax, still there is a deep parallel between maxisets and minimax theory. For instance, facing a particular situation, the standard procedure to prove that a set B is the maxiset usually consists (exactly as in minimax theory) in two steps: first showing that $B \subset MS(\hat{q}_n, \rho, \alpha_n)(T)$, but this is generally obtained using similar arguments as for proving upper bound inequalities in minimax setting since it is simply needed to prove that if $\theta \in B$ then $\mathbb{E}_\theta^n \rho(\hat{q}_n, q(\theta)) \leq T\alpha_n$. The second inclusion $MS(\hat{q}_n, \rho, \alpha_n)(T) \subset B$ is of the same nature as lower bound inequalities in the minimax setting, but often much simpler.

In sections 3 and 4, we will precisely provide conditions ensuring that the maxiset of a procedure is necessarily larger than some fixed space, and conversely prove that other conditions restrict the procedure to have its maxiset smaller than a fixed space. This study will be performed on the class of shrinkage procedures in a white noise model. Among these procedures we will investigate the consequences for a procedure to be limited, elitist, cautious... (see the definitions in section 2.)

Moreover, it is important to notice that this study can obviously be generalized to different models (since the conditions on the model are in fact not very restrictive), and one can easily imagine conditions on kernel methods (for instance) translating the notions of shrinkage, limited, elitist, cautious or hereditary although it is certainly less natural. (see in section 6 the discussion establishing a link with Lepski procedure.)

The second part of the paper uses the results of the first one to perform a comparison among Bayesian estimates.

We chose to focus on Bayes rules precisely because Bayesian techniques have now become very popular to estimate signals decomposed on wavelet bases. From the practical point of view,

many authors have built Bayes estimates that outperform classical procedures and in particular thresholding procedures. See for instance, [Chipman et al., 1997], [Abramovich et al., 1998], [Clyde et al., 1998], [Johnstone and Silverman, 1998], [Vidakovic, 1998], [Clyde and George, 1998] or [Clyde and George, 2000] who discussed the choice of the Bayes model to capture the sparsity of the signal to be estimated and the choice of the Bayes rule (and among others, posterior mean or median). We also refer the reader to the very complete review paper of [Antoniadis et al., 2001] who provide descriptions and comparisons of various Bayesian wavelet shrinkage and wavelet thresholding estimators.
From the minimax point of view, recent works have proved that Bayes rules can achieve optimal rates of convergence. [Abramovich et al., 2004] investigated theoretical performance of the procedures introduced by [Abramovich et al., 1998]. More precisely, they considered a prior model based on a combination of a point mass at zero and a normal density. For the mean squared error, they proved that the non adaptive posterior mean and posterior median achieve optimal rates up to a logarithmic factor on the Besov space $\mathcal{B}^s_{p,q}$ when $p \geq 2$. When $p < 2$, these estimators can achieve only the best possible rates for linear estimates. As [Abramovich et al., 2004], [Johnstone and Silverman, 2004a] and [Johnstone and Silverman, 2004b] investigated minimax properties of Bayes rules, but the prior is based on heavy-tailed distributions and they consider an empirical Bayes setting. In this case, the posterior mean and median are optimal. Other more sophisticated results concerning minimax properties of Bayes rules have been established by [Zhang, 2002].
The goal of section 5 is to study some Bayesian procedures from the maxiset point of view in the light of the results of sections 3 and 4. To capture the sparsity of the signal, we introduce the following prior model on the wavelet coefficients:

$$\beta_{jk} \sim \pi_{j,\epsilon}\gamma_{j,\epsilon} + (1-\pi_{j,\epsilon})\delta(0), \tag{1}$$

where $0 \leq \pi_{j,\epsilon} \leq 1$, $\delta(0)$ is a point mass at zero and the β_{jk}'s are independent. The nonzero part of the prior $\gamma_{j,\epsilon}$ is assumed to be the dilation of a fixed symmetric, positive, unimodal and continuous density γ:

$$\gamma_{j,\epsilon}(\beta_{jk}) = \frac{1}{\tau_{j,\epsilon}}\gamma\left(\frac{\beta_{jk}}{\tau_{j,\epsilon}}\right),$$

where the dilation parameter $\tau_{j,\epsilon}$ is positive. The parameter $\pi_{j,\epsilon}$ can be interpreted as the proportion of non negligible coefficients. We also introduce the parameter

$$w_{j,\epsilon} = \frac{\pi_{j,\epsilon}}{1-\pi_{j,\epsilon}}.$$

When the signal is sparse, most of the $w_{j,\epsilon}$ are small. These priors or very close forms have extensively been used by the authors cited above and especially [Abramovich et al., 2004], [Johnstone and Silverman, 2004a] and [Johnstone and Silverman, 2004b]. To complete the definition of the prior model, we have to fix the hyperparameters $\tau_{j,\epsilon}$ and $w_{j,\epsilon}$ and the density γ. The most popular choice for γ is the normal density. However priors with heavy tails have proved also to work extremely well. One of our results will be to show that if some Bayesian procedures using Gaussian priors behave quite unwell (in terms of maxisets) compared to those with heavy tails, it is nevertheless possible to attain a maxiset as good as thresholding estimates, among procedures based on Gaussian priors, under the condition that the hyperparameter $\tau_{j,\epsilon}$ is "large". Under this assumption, the density $\gamma_{j,\epsilon}$ is then more spread around 0, which enables us to avoid considering heavy-tailed densities.

Now, a natural question arises. Does there exist a non linear procedure that outperforms the thresholding procedures in terms of maxiset comparisons? We will see in section 6 that the answer is yes. By making use of the dyadic structure of the wavelet bases (which has not been used before in fact) and performing algorithm with tree properties, we can prove that this provides a way of enlarging the maxisets.

2 Model, Shrinkage rules.

2.1 Model

We will consider a white noise setting: $X_\epsilon(.)$ is a random measure satisfying on $[0,1]$ the following equation:

$$X_\epsilon(dt) = f(t)dt + \epsilon W(dt)$$

where $0 < \epsilon < 1/e$ is the noise level and f is a function defined on $[0,1]$, $W(.)$ is a Brownian motion on $[0,1]$. As usual, to connect with the standard framework of sequences of experiments we put $\epsilon = n^{-1/2}$.
Let $\{\psi_{jk}(\cdot),\, j \geq -1, k \in \mathbb{Z}\}$ be a compactly supported wavelet basis of $\mathbb{L}_2([0,1])$, such that any $f \in \mathbb{L}_2([0,1])$ can be represented as:

$$f = \sum_{j\geq -1} \sum_{k} \beta_{jk}\psi_{jk}$$

where $\beta_{jk} = (f, \psi_{jk})_{\mathbb{L}_2}$. (As usual, ψ_{-1k} denotes the translations of the scaling function.) The model is reduced to a sequence space model if we put: $y_{jk} = X_\epsilon(\psi_{jk}) = \int f\psi_{jk} + \epsilon Z_{jk}$ where Z_{jk} are i.i.d $\mathcal{N}(0,1)$. Let us note that at each level $j \geq 0$, the number of non-zero wavelet coefficients is smaller than or equal to $2^j + l_\psi - 1$, where l_ψ is the maximal size of the supports of the scaling function and the wavelet. So, there exists a constant S_ψ such that at each level $j \geq -1$, there are less than or equal to $S_\psi \times 2^j$ coefficients to be estimated. In the sequel, we shall not distinguish between f and $\beta = (\beta_{jk})_{jk}$ its sequence of wavelet coefficients.

2.2 Classes of Estimators

Let us first consider the following very general class of shrinkage estimators:

$$\mathcal{F}_\epsilon = \left\{ \hat{f}_\epsilon(.) = \sum_{j\geq -1} \sum_{k} \gamma_{jk} y_{jk} \psi_{jk}(.);\quad \gamma_{jk}(\varepsilon) \in [0,1],\ \text{measurable} \right\}.$$

Let us observe here that the γ_{jk} may be constant (linear estimators) or data dependent. Among this class, we'll particularly focus on the following classes of estimators:

Definition 1. *We say that* $\hat{f}_\epsilon \in \mathcal{F}_\epsilon$ *is a* **limited rule** *if there exist a determinist function of* ϵ, λ_ϵ, *and a constant* $a \in [0, 1[$ *such that, for any* j, k,

$$\gamma_{jk} > a \Longrightarrow 2^{-j} > \lambda_\epsilon.$$

We note $\hat{f}_\epsilon \in \mathcal{L}(\lambda_\epsilon, a)$.

The simplest example to illustrate limited rules is provided by the projection estimator:

$$\gamma_{jk}(\epsilon) = \gamma_j^{(1)}(\lambda_\epsilon) = I\{2^{-j} > \lambda_\epsilon\},$$

which obviously belongs to $\mathcal{L}(\lambda_\epsilon, 0)$. But, more generally, the class of linear shrinkage estimates provides natural limited procedures. For instance, linear estimates associated with Tikhonov-Phillips weights:

$$\gamma_{jk}(\epsilon) = \gamma_j^{(2)}(\lambda_\epsilon) = \frac{1}{1 + (2^j \lambda_\epsilon)^\alpha}, \quad \alpha > 0,$$

or with Pinsker weights:

$$\gamma_{jk}(\epsilon) = \gamma_j^{(3)}(\lambda_\epsilon) = (1 - (2^j \lambda_\epsilon)^\alpha)_+, \quad \alpha > 0,$$

are limited rules respectively belonging to $\mathcal{L}(\lambda_\epsilon, 1/2)$ and $\mathcal{L}(\lambda_\epsilon, 0)$.
To detail other examples, let us introduce

$$\begin{aligned} t_\epsilon &= \epsilon\sqrt{\log(\epsilon^{-1})} \\ j_\epsilon \in \mathbb{N},\ 2^{-j_\epsilon} &\leq t_\epsilon^2 < 2^{1-j_\epsilon}. \end{aligned}$$

This will be denoted in the sequel by $2^{j_\epsilon} \sim t_\epsilon^{-2}$. We recall the hard thresholding $\hat{f}^T$ and the soft thresholding $\hat{f}^S$ rules respectively defined by

$$\hat{f}^T = \sum_{-1 \leq j < j_\epsilon} \sum_k y_{jk} I\{|y_{jk}| > m t_\epsilon\} \psi_{jk}, \tag{2}$$

$$\hat{f}^S = \sum_{-1 \leq j < j_\epsilon} \sum_k \left(1 - \frac{m t_\epsilon}{|y_{jk}|}\right) I\{|y_{jk}| > m t_\epsilon\} y_{jk} \psi_{jk}, \tag{3}$$

where m is a positive constant. It is obvious that these procedures belong to $\mathcal{L}(t_\epsilon^2, 0)$. In sections 5 and 6, we shall provide many more examples of limited rules.

Definition 2. *We say that* $\hat{f}_\epsilon \in \mathcal{F}_\epsilon$ *is an* **elitist rule** *if there exist a determinist function of* ϵ, λ_ϵ, *and a constant* $a \in [0, 1[$ *such that, for any* j, k

$$\gamma_{jk} > a \Longrightarrow |y_{jk}| > \lambda_\epsilon.$$

In the sequel, we note $\hat{f}_\epsilon \in \mathcal{E}(\lambda_\epsilon, a)$.

Remark: This definition generalizes the notion of *elitist rules* introduced by [Autin, 2003] in density estimation. $\diamond$
To give some examples of elitist rules, consider $\hat{f}^T$ and $\hat{f}^S$ defined in (2) and (3) that belong to $\mathcal{E}(m t_\epsilon, 0)$. Other examples of elitist rules will be given in section 5 by considering Bayesian procedures.

Definition 3. *We say that* $\hat{f}_\epsilon \in \mathcal{F}_\epsilon$ *is a* **cautious rule** *if there exist a determinist function of* ϵ*,* λ_ϵ *and a constant* $a \in]0,1]$ *such that, for any* $j < j_\epsilon$ *and any* k

$$\gamma_{jk} \leq a \Longrightarrow |y_{jk}| \leq \lambda_\epsilon,$$

where $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$*. In the sequel, we note* $\hat{f}_\epsilon \in \mathcal{C}(\lambda_\epsilon, a)$*.*

Remark: For instance, $\hat{f}^T$ and $\hat{f}^S$ defined in (2) and (3) belong respectively to $\mathcal{C}(mt_\epsilon, \frac{1}{2})$ and $\mathcal{C}(2mt_\epsilon, \frac{1}{2})$. $\diamond$

Until now, the dyadic aspect of the procedure has not been used. We'll see in section 6 that this can be taken into account with profit by, for instance, introducing the following notion of heredity. In the sequel we say that, for $j \in \mathbb{N}$, I is a j-dyadic interval if and only if $I = I_{jk} = [\frac{k}{2^j}, \frac{k+1}{2^j}]$. In this case, we shall note y_I (rep. β_I) instead of y_{jk} (rep. β_{jk}) and shall set $|I| = 2^{-j}$, its length. Note that for any j-dyadic interval I_{jk} and any $j' \geq j$, there exist at most $2^{(j'-j)}$ j'-dyadic intervals included in I_{jk}.

Definition 4. *Let* $\hat{f}_\epsilon \in \mathcal{F}_\epsilon$*. We say that* $\hat{f}_\epsilon$ *is a* **hereditary rule** *if there exist a determinist function of* ϵ*,* λ_ϵ*, and a constant* $a \in [0,1[$ *such that for any* $j < j_\epsilon$ *and any* k

$$\gamma_{jk} > a \Longrightarrow \exists I' \subset I_{jk} \ / \ |I'| > \lambda_\epsilon^2 \text{ and } |y_{I'}| > \lambda_\epsilon,$$

where $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$*. In the sequel, we note* $\hat{f}_\epsilon \in \mathcal{H}(\lambda_\epsilon, a)$*.*

Some examples of hereditary rules are given in subsection 6.1.

Remark: The limited rules as well as the elitist rules and the hereditary rules are forming a non decreasing class with respect to a. The cautious rules are forming a non increasing class with respect to a. We also have that any of the classes introduced above are convex. So they are obviously stable if we consider aggregation of procedures or as in learning algorithms, if we build a procedure averaging the opinions of different experts all belonging to one of the previous class. $\diamond$

3 Ideal maxisets for classes of estimators.

Proving lower bound inequalities in minimax theory consists in showing that if we consider the class of all estimators on a functional spaces, there exists a best achievable rate α_n. In this section our tactic will be of the same spirit, but somewhat different since we will fix the rate α_n, consider classes of procedures and prove that they have a best achievable maxiset. More precisely, we will prove that when a procedure belongs to one of the classes considered above, its maxiset is necessarily smaller than a simple functional class. Here, for simplicity, we shall restrict to the case where ρ is the square of the $\mathbb{L}_2$ norm, even though if a large majority of the following results can be extended to more general norms.

3.1 Functional spaces

We recall the definitions of the following functional spaces. They will play an important role in the sequel. Note that, here, they appear with definitions depending on the wavelet basis. However, as has been remarked in [Meyer, 1990] and [Cohen et al., 2001b], most of them also have different definitions proving that this dependence in the basis is not crucial at all. Here and later we set for all $\lambda > 0$, $2^{j_\lambda} \sim \lambda^{-2}$.

Definition 5. *Let $s > 0$. We say that a function $f \in \mathbb{L}_2([0,1])$ belongs to the Besov space $\mathcal{B}^s_{2,\infty}$, if and only if:*

$$\sup_{J \geq -1} 2^{2Js} \sum_{j \geq J} \sum_k \beta^2_{jk} < \infty.$$

We denote by $\mathcal{B}^s_{2,\infty}(R)$ the ball of radius R in this space.

Definition 6. *Let $0 < r < 2$. We say that a function f belongs to the weak Besov space W_r if and only if:*

$$\|f\|_{W_r} := [\sup_{\lambda > 0} \lambda^{r-2} \sum_{j \geq -1} \sum_k \beta^2_{jk} I\{|\beta_{jk}| \leq \lambda\}]^{1/2} < \infty.$$

We denote by $W_r(R)$, the ball of radius R in this space.

Definition 7. *Let $0 < r < 2$. We say that a function f belongs to the space W^*_r if and only if:*

$$\|f\|_{W^*_r} := [\sup_{0<\lambda<1} \lambda^r [\log(\frac{1}{\lambda})]^{-1} \sum_{-1 \leq j < j_\lambda} \sum_k I\{|\beta_{jk}| > \lambda\}]^{1/2} < \infty.$$

Remark: If $\subsetneq$ denotes the strict inclusion between two functional spaces, for all $0 < r < 2$, it is easy to see using Markov inequality that $\mathcal{B}^s_{2,\infty} \subsetneq W_r$ as soon as $s \geq \frac{1}{r} - \frac{1}{2}$ and $W_r \subsetneq W^*_r$. $\diamond$

Definition 8. *Let $0 < r < 2$. We say that a function f belongs to the tree-Besov space W^T_r if and only if:*

$$\|f\|_{W^T_r} := [\sup_{\lambda > 0} \lambda^{r-2} \sum_{0 \leq j < j_\lambda} \sum_k \beta^2_{jk} I\{\forall I' \subset I_{jk} \ / \ |I'| > \lambda^2, |\beta_{I'}| \leq \frac{\lambda}{2}\}]^{1/2} < \infty.$$

Remark: Obviously, $W_r \subset W^T_r$. These spaces taking account of the dyadic structure of the wavelet bases are very close to the oscillation spaces introduced in [Jaffard, 1998] and [Jaffard, 2004]. $\diamond$

For sake of simplicity, the result presented in the following section emphasizes the cases where the rate of convergence is linked in a direct way to either the limitation or to the threshold bound for elitist or cautious rules. This constraint can be relaxed. For instance, there are many cases where either the threshold bound or the rate contain logarithmic factors. In these cases the link is not so direct. Results can also be obtained in these cases, which may be less

aesthetic, but still useful. These results are given in Appendix.
Notation: For $\mathcal{A}$, a given normed space, the following notations:

$$\begin{array}{rcl} MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s}) & \subset & \mathcal{A} \\ (resp.) \quad \mathcal{A} & \subset & MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s}) \end{array}$$

will mean in the sequel

$$\begin{array}{rcl} \forall\, M \;\exists\, M',\; MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s})(M) & \subset & \mathcal{A}(M') \\ (resp.) \quad \forall\, M' \;\exists\, M,\; \mathcal{A}(M') & \subset & MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s})(M), \end{array}$$

where M and M' respectively denote the radii of balls of $MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s})$ and $\mathcal{A}$. $\diamond$

3.2 Ideal Maxisets for limited rules

In this section, we study the ideal maxisets for limited procedures. For this purpose, let us give a sequence $(\lambda_\epsilon)_\epsilon$ going to 0 as ϵ tending to 0.

Theorem 1 (Ideal maxiset for limited rules). *Let $\sigma > 0$ and $\hat{f}_\epsilon$ be a limited rule in $\mathcal{L}(\lambda_\epsilon, a)$, with $a \in [0,1[$. Then, if λ_ε is a non decreasing, continuous function such that $\lambda_0 = 0$,*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2\sigma}) \subset \mathcal{B}^\sigma_{2,\infty}$$

(with $M' = \frac{\sqrt{2M}}{(1-a)}$.)

Proof: Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2\sigma})(M)$. If we observe that if $2^{-j} \le \lambda_\epsilon$ then $\gamma_{jk} \le a$, we have:

$$\begin{array}{rl} & (1-a)^2 \sum_{j,k} \beta_{jk}^2 I\{2^{-j} \le \lambda_\epsilon\} \\ = & 2(1-a)^2 \sum_{j,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0) I\{\beta_{jk} \ge 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0) I\{\beta_{jk} < 0\}\right] I\{2^{-j} \le \lambda_\epsilon\} \\ \le & 2\mathbb{E} \sum_{j,k} \left[(\gamma_{jk} y_{jk} - \beta_{jk})^2 I\{\beta_{jk} \ge 0\} + (\gamma_{jk} y_{jk} - \beta_{jk})^2 I\{\beta_{jk} < 0\}\right] I\{2^{-j} \le \lambda_\epsilon\} \\ \le & 2\mathbb{E} \sum_{j,k} (\gamma_{jk} y_{jk} - \beta_{jk})^2 \\ \le & 2M\ \lambda_\epsilon^{2\sigma}. \end{array}$$

So, using the continuity of λ_ϵ in 0, we deduce

$$\sup_{J \ge -1} 2^{2J\sigma} \sum_{j \ge J} \sum_k \beta_{jk}^2 \;\le\; \frac{2M}{(1-a)^2},$$

and f belongs to $\mathcal{B}^\sigma_{2,\infty}$. $\square$

We have proved here that $\mathcal{B}^\sigma_{2,\infty}$ is a good candidate for an ideal maxiset among limited rules. We will prove in section 4 that it is reached by standard and well known limited procedures. So, as a consequence, $\mathcal{B}^\sigma_{2,\infty}$ is the ideal maxiset among limited rules with the relation between the limiting parameter and the rate of convergence above prescribed.
In the next subsection, we focus on elitist procedures.

3.3 Ideal maxisets for elitist rules

Theorem 2 (Ideal maxiset for elitist rules). *Let $\hat{f}_\epsilon$ be an elitist rule in $\mathcal{E}(\lambda_\epsilon, a)$ with $a \in [0,1[$. Then, if λ_ε is a non decreasing, continuous function such that $\lambda_0 = 0$, and $0 < r < 2$ is a real number,*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r}) \subset W_r$$

(with $M' = \frac{\sqrt{2M}}{(1-a)}$.)

Remark: It is important to notice that this inclusion will be mostly used for $\lambda_\epsilon = t_\epsilon$, $r = \frac{2}{1+2s}$, $2 - r = \frac{4s}{1+2s}$, where we find back the usual rates of convergence. $\diamond$

Proof: Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r})(M)$. If we observe that if $|y_{jk}| \leq \lambda_\epsilon$ then $\gamma_{jk} \leq a$, we have:

$$\begin{aligned}
& (1-a)^2 \sum_{j,k} \beta_{jk}^2 I\{|\beta_{jk}| \leq \lambda_\epsilon\} \\
= \;& 2(1-a)^2 \sum_{j,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0) I\{\beta_{jk} \geq 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0) I\{\beta_{jk} < 0\}\right] I\{|\beta_{jk}| \leq \lambda_\epsilon\} \\
\leq \;& 2\mathbb{E} \sum_{j,k} \left[(\beta_{jk} - \gamma_{jk} y_{jk})^2 I\{\beta_{jk} \geq 0\} + (\beta_{jk} - \gamma_{jk} y_{jk})^2 I\{\beta_{jk} < 0\}\right] I\{|\beta_{jk}| \leq \lambda_\epsilon\} \\
\leq \;& 2\mathbb{E} \sum_{j,k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 \\
\leq \;& 2M \; \lambda_\epsilon^{2-r}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that

$$\sup_{\lambda > 0} \lambda^{r-2} \sum_{j \geq -1} \sum_k \beta_{jk}^2 I\{|\beta_{jk}| \leq \lambda\} \; \leq \frac{2M}{(1-a)^2},$$

and f belongs to W_r. $\square$

In the next subsection, we focus on cautious procedures.

3.4 Ideal maxisets for cautious rules

Theorem 3 (Ideal maxiset for cautious rules). *Let $\hat{f}_\epsilon$ be a cautious rule in $\mathcal{C}(\lambda_\epsilon, a)$ with $a \in]0,1]$. Let us suppose that $0 < r < 2$ is a real number and λ_ε is a non decreasing, continuous function such that $\lambda_0 = 0$. Suppose that*

$$\exists\, c > 0, \quad \forall\, \epsilon > 0, \; \frac{\lambda_\epsilon}{\sqrt{\log(\frac{1}{\lambda_\epsilon})}} \leq c\epsilon. \tag{4}$$

Then

$$MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r}) \subset W_r^*$$

(with $M' = \frac{2c\sqrt{2M}}{a}$.)

Remark: Note that the case $\lambda_\epsilon = t_\epsilon$ (resp. $\lambda_\epsilon = \epsilon$) satisfies (4) with $c = \sqrt{2}$ (resp. $c = 1$)$\diamond$
Proof: It is a consequence of the following lemma:

Lemma 1. *Let $\epsilon > 0$ and suppose that $|\beta_{jk}| > \lambda_\epsilon$ and $sign(\beta_{jk})y_{jk} < |\beta_{jk}|$. Then,*

$$a|\beta_{jk} - y_{jk}| \leq 2|\beta_{jk} - \gamma_{jk}y_{jk}|.$$

Proof of the lemma: We only prove the case $\beta_{jk} > \lambda_\epsilon$ and $y_{jk} < \beta_{jk}$ since the case $\beta_{jk} < -\lambda_\epsilon$ and $y_{jk} > \beta_{jk}$ can be proved with the same arguments.
It is clear that,

a) if $y_{jk} \geq 0$, then, $a(\beta_{jk} - y_{jk}) \leq a(\beta_{jk} - \gamma_{jk}y_{jk})$

b) if $y_{jk} < -\lambda_\epsilon$, then, because the rule is cautious, $\gamma_{jk} > a$ and $a(\beta_{jk} - y_{jk}) \leq \gamma_{jk}(\beta_{jk} - y_{jk}) \leq (\beta_{jk} - \gamma_{jk}y_{jk})$

c) if $-\lambda_\epsilon \leq y_{jk} < 0$, then $a(\beta_{jk} - y_{jk}) \leq 2a\beta_{jk} \leq 2a(\beta_{jk} - \gamma_{jk}y_{jk})$.

Since $0 < a < 1$ we deduce from a) b) and c) that $a(\beta_{jk} - y_{jk}) \leq 2(\beta_{jk} - \gamma_{jk}y_{jk})$.

□

Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r})(M)$. Using (4),

$$a^2\lambda_\epsilon^2 \left[\log(\frac{1}{\lambda_\epsilon})\right]^{-1} \sum_{j<j_\epsilon,k} I\{|\beta_{jk}| > \lambda_\epsilon\} \leq a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} I\{|\beta_{jk}| > \lambda_\epsilon\}.$$

Now, let us recall that if X is a zero-mean Gaussian variable with variance ϵ^2, then

$$\mathbb{E}(X^2 I_{\{X<0\}}) = \mathbb{E}(X^2 I_{\{X>0\}}) = \frac{\epsilon^2}{2}.$$

So, from Lemma 1

$$\begin{aligned}
& a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} I\{|\beta_{jk}| > \lambda_\epsilon\} \\
= \; & a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} [I\{\beta_{jk} > \lambda_\epsilon\} + I\{\beta_{jk} < -\lambda_\epsilon\}] \\
= \; & 2a^2c^2 \; \mathbb{E} \sum_{j<j_\epsilon,k} (\beta_{jk} - y_{jk})^2 \left[I\{y_{jk} - \beta_{jk} < 0\}I\{\beta_{jk} > \lambda_\epsilon\} + I\{y_{jk} - \beta_{jk} > 0\}I\{\beta_{jk} < -\lambda_\epsilon\}\right] \\
\leq \; & 8c^2 \; \mathbb{E} \sum_{j<j_\epsilon,k} (\beta_{jk} - \gamma_{jk}y_{jk})^2 \\
\leq \; & 8c^2M \; \lambda_\epsilon^{2-r}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that

$$\sup_{\lambda>0} \lambda^r \left[\log(\frac{1}{\lambda})\right]^{-1} \sum_{j<j_\lambda,k} I\{|\beta_{jk}| > \lambda\} \;\leq\; \frac{8c^2M}{a^2}$$

and f belongs to W_r^*. □
In the next subsection, we focus on hereditary procedures.

3.5 Ideal maxisets for hereditary rules

Theorem 4. *Let $\hat{f}_\epsilon$ be a hereditary rule that belongs to $\mathcal{H}(\lambda_\epsilon, a)$ with $a \in [0,1[$. Let $0 < r < 2$ be a real number and λ_ε be a non decreasing, continuous function with $\lambda_0 = 0$ such that there exists a constant $C > 0$ which satisfies for any $\epsilon > 0$,*

$$\mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}) \leq C\lambda_\epsilon^4 \tag{5}$$

with $Z \sim \mathcal{N}(0,1)$. Then

$$MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r)}) \subset W_r^T$$

(with $M' = \frac{\sqrt{2(M+C)}}{1-a}$.)

Remark: For instance, for $\lambda_\epsilon = mt_\epsilon$, condition (5) is satisfied for any $m \geq 4\sqrt{2}$. $\diamond$
Proof: Let $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$ and $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2-r})(M)$. Denote

- $|\bar{y}_{jk}(\lambda_\epsilon)| := \max\{|y_{_I}|;\ I \subset I_{jk} \text{ and } |I| > \lambda_\epsilon^2\}$,
- $|\bar{\beta}_{jk}(\lambda_\epsilon)| := \max\{|\beta_{_I}|;\ I \subset I_{jk} \text{ and } |I| > \lambda_\epsilon^2\}$.
- $|\bar{\delta}_{jk}(\lambda_\epsilon)| := \max\{|y_I - \beta_{_I}|;\ I \subset I_{jk} \text{ and } |I| > \lambda_\epsilon^2\}$.

We have the following lemma:

Lemma 2. *If λ_ϵ satisfies (5) then, for any $0 \leq j < j_\epsilon$ and any k:*

$$\mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon) I\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} \leq \mathbb{P}(|\bar{\delta}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon/2) \leq 2C\ \lambda_\epsilon^2.$$

Proof of the lemma: Let $Z \sim \mathcal{N}(0,1)$. We have for any $0 \leq j < j_\epsilon$ and any k:

$$\begin{aligned}
\mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon) I\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} &\leq \mathbb{P}(|\bar{\delta}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon/2) \\
&\leq \sum_{I \subset I_{jk} \text{ and } |I| > \lambda_\epsilon^2} \mathbb{P}(|y_I - \beta_I| > \frac{\lambda_\epsilon}{2}) \\
&\leq 2^{j_\epsilon} \mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}) \\
&\leq 2\lambda_\epsilon^{-2}\ \mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}) \\
&\leq 2C\ \lambda_\epsilon^2
\end{aligned}$$

$\square$

Now, using the fact that the rule is hereditary and Lemma 2:

$$
\begin{aligned}
& (1-a)^2 \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 I\{\forall I' \subset I_{jk},\ /\ |I'| > \lambda_\epsilon^2, |\beta_{I'}| \le \frac{\lambda_\epsilon}{2}\} \\
= \;& (1-a)^2 \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 I\{|\bar\beta_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2}\} \\
= \;& 2(1-a)^2 \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk}-\beta_{jk}<0) I\{\beta_{jk}>0\} + \mathbb{P}(y_{jk}-\beta_{jk}>0) I\{\beta_{jk}<0\}\right] I\{|\bar\beta_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2} \\
\le \;& 2(1-a)^2 \mathbb{E} \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 \left[I\{y_{jk}-\beta_{jk}<0\} I\{\beta_{jk}>0\} + I\{y_{jk}-\beta_{jk}>0\} I\{\beta_{jk}<0\}\right] I\{|\bar y_{jk}(\lambda_\epsilon)| \le \lambda \\
& +2\mathbb{E} \sum_{0\le j<j_\epsilon,k} \beta_{jk}^2 \mathbb{P}(|\bar y_{jk}(\lambda_\epsilon)| > \lambda_\epsilon) I\{|\bar\beta_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2}\} \\
\le \;& 2\,\mathbb{E} \sum_{0\le j<j_\epsilon,k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 I\{|\bar y_{jk}(\lambda_\epsilon)| \le \lambda_\epsilon\} + \frac{\lambda_\epsilon^2}{2} \sum_{0\le j<j_\epsilon,k} \mathbb{P}(|\bar y_{jk}(\lambda_\epsilon)| > \lambda_\epsilon\} I\{|\bar\beta_{jk}(\lambda_\epsilon)| \le \frac{\lambda_\epsilon}{2}\} \\
\le \;& 2\,\mathbb{E} \sum_{j,k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 + 2C\lambda_\epsilon^2 \\
\le \;& 2(M+C)\ \lambda_\epsilon^{2-r}.
\end{aligned}
$$

So, using the continuity of λ_ϵ in 0, we deduce that

$$
\sup_{\lambda>0} \lambda^{r-2} \sum_{0\le j<j_\lambda,k} \beta_{jk}^2 I\{\forall I \subset I_{jk},\ /\ |I| > l_\psi \lambda^2, |\beta_I| \le \frac{\lambda}{2}\} \ \le \frac{2(M+C)}{(1-a)^2}.
$$

It comes that $f \in W_r^T$. $\square$

4 Rules ensuring that their maxiset contains a prescribed subset

In this section we prove three types of conditions ensuring that the maxiset of a given shrinkage rule contains either a Besov space, a weak Besov space or a tree-Besov space. This part is obviously strongly linked with upper bounds inequalities in minimax theory. Indeed, our technique of proof here will be to show that some classes of estimators satisfy an upper bound inequality associated with the considered subset.

4.1 When does the maxiset contain a Besov space?

We have the following result, which is a converse result to Theorem 1 with respect to the ideal maxiset result for limited rules:

Theorem 5. *Let* $s > 0$ *and* $(\gamma_j(\epsilon))_{jk}$ *a non increasing sequence of weights lying in* $[0, 1]$ *such that* $\hat{\beta}^L_\epsilon = (\gamma_j(\epsilon)y_{jk})_{jk}$ *belongs to* $\mathcal{L}(\lambda_\epsilon, a)$, *with* $a \in [0, 1[$, λ_ε *is continuous and* $\lambda_0 = 0$. *If there exist* C_1 *and* C_2 *in* $\mathbb{R}$ *such that, with* $\gamma_{-2} = 1$, $\forall\, \epsilon > 0$,

$$\sum_{j\geq -1} (\gamma_{j-1} - \gamma_j)(1 - \gamma_j)2^{-2js} I\{2^j < \lambda_\epsilon^{-1}\} \leq C_1 \lambda_\epsilon^{2s}$$

$$\sum_{j\geq -1} 2^j \gamma_j(\epsilon)^2 \leq C_2\, \epsilon^{-2}\, \lambda_\epsilon^{2s}$$

then,

$$\mathcal{B}^s_{2,\infty} \subset MS(\hat{\beta}^L_\epsilon, \|.\|_2^2, \lambda_\epsilon^{2s}).$$

Proof: This result is a simple consequence of Theorem 2 of [Rivoirard, 2004]. A more general result is established in Appendix. □

Combining Theorems 1 and 5, by straightforward computations, we obtain:

Corollary 1. *If we consider linear estimates associated with the weights* $\gamma_j^{(1)}(\lambda_\epsilon)$, $\gamma_j^{(2)}(\lambda_\epsilon)$ *with* $\alpha > (s \vee 1/2)$ *or* $\gamma_j^{(3)}(\lambda_\epsilon)$ *with* $\alpha > s$ *(see section 2.2), then for* $i \in \{1, 2, 3\}$

$$MS((\gamma_j^{(i)}(\lambda_\epsilon)y_{jk})_{jk}, \|.\|_2^2, \lambda_\epsilon^{2s}) = \mathcal{B}^s_{2,\infty},$$

as soon as $(\epsilon^2 \lambda_\epsilon^{-(1+2s)})_\epsilon$ *is bounded. In particular, for the polynomial rate* $\epsilon^{4s/(1+2s)}$, *corresponding to* $\lambda_\epsilon = \epsilon^{2/(1+2s)}$, $\mathcal{B}^s_{2,\infty}$ *is exactly the maxiset of these estimates.*

Remark: [Rivoirard, 2004] extended these results for a more general statistical model: the heteroscedastic white noise model that naturally appears in the literature of inverse problems. This last result illustrates the strong link between linear procedures (and more generally limited procedures) and Besov spaces. This has already been pointed out by [Kerkyacharian and Picard, 1993] who studied maxisets for linear procedures for the model of density estimation. ◇

4.2 When does the maxiset contain a weak Besov space?

We have the following result, which is a converse result to Theorems 1 and 2 with respect to the ideal maxiset results for limited and elitist rules:

Theorem 6. *Let* $s > 0$ *and* $\gamma_{jk}(\epsilon)$ *a sequence of random weights lying in* $[0, 1]$. *We assume that there exist positive constants* c, m *and* $K(\gamma)$ *such that for any* $\epsilon > 0$

$$\hat{\beta}(\epsilon) = (\gamma_{jk}(\epsilon)y_{jk})_{jk} \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(mt_\epsilon, ct_\epsilon), \tag{6}$$

$$(1 - \gamma_{jk}(\epsilon)) \leq K(\gamma) \left(\frac{t_\varepsilon}{|y_{jk}|} + t_\epsilon \right), \quad a.e. \quad \forall\, j < j_\epsilon,\ \forall\, k. \tag{7}$$

Then, as soon as $m \geq 8$,

$$\mathcal{B}^{\frac{s}{1+2s}}_{2,\infty} \cap W_{\frac{2}{1+2s}} \subset MS(\hat{f}_\epsilon, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}).$$

Remark: It is worthwhile to note that (7) is a condition implying that the procedure belongs to $\mathcal{C}(t_\epsilon, Dt_\epsilon)$, and can be considered as a refinement of the cautiousness condition.
It is enough to verify condition (7) for ϵ small enough without modifying the conclusion of the theorem. This remark will be useful in sections 5.2 and 5.3, where we apply Theorem 6 to Bayesian procedures. $\diamond$
This theorem, is an obvious consequence of the following two propositions concerning functional spaces inclusions and general upper bound results for shrinkage procedures.

Proposition 1. *Let $0 < r < 2$, $C > 0$ and $f \in W_r$. Then,*

$$\sup_{\lambda > 0} \lambda^r \sum_{j,k} I\{|\beta_{jk}| > \lambda\} \leq \frac{2^{2-r}\|f\|^2_{W_r}}{1 - 2^{-r}}.$$

The proof of this proposition is standard, see for instance in [Kerkyacharian and Picard, 2000], where it is proved that the condition above is in fact equivalent to the fact that $f \in W_r$.

Proposition 2. *Under the conditions of Theorem 6, we have the following inequality:*

$$\begin{aligned}\mathbb{E}\|\hat{f}_\epsilon - f\|_2^2 \quad \leq \quad & \Bigg[4c^2 S_\psi + 4(1 + K(\gamma)^2)\|f\|_2^2 + 4\sqrt{3}S_\psi + 2(2^{\frac{4s}{1+2s}} + 2^{\frac{-4s}{1+2s}})m^{\frac{4s}{1+2s}}\|f\|^2_{W_{\frac{2}{1+2s}}} + \\ & + \tfrac{8m^{-2/1+2s}}{(1-2^{-2/1+2s})}(1 + 8K(\gamma)^2)\|f\|^2_{W_{\frac{2}{1+2s}}} + \|f\|^2_{B^{\frac{s}{1+2s}}_{2,\infty}}\Bigg] t_\epsilon^{\frac{4s}{1+2s}}.\end{aligned}$$

Proof: Let $f \in \mathcal{B}^{\frac{s}{1+2s}}_{2,\infty} \cap W_{\frac{2}{1+2s}}$. Obviously, using the limitation assumption, we have for j_ϵ such that $2^{j_\epsilon} \sim t_\epsilon^{-2}$

$$\mathbb{E}\|\hat{f}_\epsilon - f\|_2^2 = \mathbb{E}\|\sum_{j<j_\epsilon,k}(\gamma_{jk}(\epsilon)y_{jk} - \beta_{jk})\psi_{j,k}\|_2^2 + \sum_{j \geq j_\epsilon,k}\beta_{jk}^2.$$

The second term is a bias term bounded by $t_\epsilon^{\frac{4s}{1+2s}}\|f\|^2_{B^{\frac{s}{1+2s}}_{2,\infty}}$, by definition of the Besov norm.
We split $\mathbb{E}\sum_{j<j_\epsilon,k}(\gamma_{jk}(\epsilon)y_{jk} - \beta_{jk})^2$ into $2(A + B)$ with

$$\begin{aligned} A &= \mathbb{E}\sum_{j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk} - \beta_{jk})^2 + (1 - \gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ I\{|y_{jk}| \leq mt_\epsilon\}, \\ B &= \mathbb{E}\sum_{j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk} - \beta_{jk})^2 + (1 - \gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ I\{|y_{jk}| > mt_\epsilon\}.\end{aligned}$$

Again, we split A into $A_1 + A_2$, and using $\hat{\beta}(\epsilon) \in \mathcal{E}(mt_\epsilon, ct_\epsilon)$, we have on $\{|y_{jk}| \leq mt_\epsilon\}$, $\gamma_{jk} \leq ct_\epsilon$. So,

$$\begin{aligned} A_1 &= \mathbb{E}\sum_{j<j_\epsilon,k}\gamma_{jk}(\epsilon)^2(y_{jk} - \beta_{jk})^2\ I\{|y_{jk}| \leq mt_\epsilon\} \\ &\leq c^2 S_\psi 2^{j_\epsilon}t_\epsilon^2\epsilon^2 \\ &\leq 2c^2 S_\psi t_\epsilon^2.\end{aligned}$$

$$
\begin{aligned}
A_2 &= \mathbb{E}\sum_{j<j_\epsilon,k}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\ I\{|y_{jk}|\le mt_\epsilon\}\\
&\le \mathbb{E}\sum_{j<j_\epsilon,k}\beta_{jk}^2\ I\{|y_{jk}|\le mt_\epsilon\}[I\{|\beta_{jk}|\le 2mt_\epsilon\}+I\{|\beta_{jk}|>2mt_\epsilon\}]\\
&\le (2mt_\epsilon)^{4s/1+2s}\|f\|^2_{W_{\frac{2}{1+2s}}}+\sum_{j<j_\epsilon,k}\beta_{jk}^2\mathbb{P}(|\beta_{jk}-y_{jk}|\ge mt_\epsilon)\\
&\le (2mt_\epsilon)^{4s/1+2s}\|f\|^2_{W_{\frac{2}{1+2s}}}+\|f\|_2^2\epsilon^{m^2/2}\\
&\le (2mt_\epsilon)^{4s/1+2s}\|f\|^2_{W_{\frac{2}{1+2s}}}+\|f\|_2^2t_\epsilon^2.
\end{aligned}
$$

We have used here the concentration property of the Gaussian distribution and the fact that $m^2\ge 4$.

$$
\begin{aligned}
B &:= B_1+B_2\\
&= \mathbb{E}\sum_{j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2+(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ I\{|y_{jk}|>mt_\epsilon\}[I\{|\beta_{jk}|\le mt_\epsilon/2\}\\
&\quad +I\{|\beta_{jk}|>mt_\epsilon/2\}].
\end{aligned}
$$

For B_1 we use the Schwartz inequality:

$$
\mathbb{E}(y_{jk}-\beta_{jk})^2I\{|y_{jk}-\beta_{jk}|>mt_\epsilon/2\}\le(\mathbb{P}(|y_{jk}-\beta_{jk}|>mt_\epsilon/2))^{1/2}(\mathbb{E}(y_{jk}-\beta_{jk})^4)^{1/2}.
$$

Now, observing that $\mathbb{E}(y_{jk}-\beta_{jk})^4=3\epsilon^4$ and that $\mathbb{P}(|y_{jk}-\beta_{jk}|>mt_\epsilon/2)\le\epsilon^{\frac{m^2}{8}}$, we have for $m^2\ge 32$:

$$
\begin{aligned}
B_1 &\le \sqrt{3}\sum_{j<j_\epsilon,k}\epsilon^2I\{|\beta_{jk}|\le mt_\epsilon/2\}\epsilon^{\frac{m^2}{16}}+\sum_{j<j_\epsilon,k}\beta_{jk}^2I\{|\beta_{jk}|\le mt_\epsilon/2\}\\
&\le 2\sqrt{3}S_\psi t_\epsilon^2+\left(\frac{m}{2}t_\epsilon\right)^{4s/1+2s}\|f\|^2_{W_{\frac{s}{1+2s}}}.
\end{aligned}
$$

For B_2, we use Proposition 1,

$$
\begin{aligned}
B_2 &= \mathbb{E}\sum_{j<j_\epsilon,k}[\gamma_{jk}(\epsilon)^2(y_{jk}-\beta_{jk})^2+(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ I\{|y_{jk}|>mt_\epsilon\}I\{|\beta_{jk}|>mt_\epsilon/2\}\\
&\le \sum_{j<j_\epsilon,k}[\epsilon^2I\{|\beta_{jk}|>mt_\epsilon/2\}+B_3\\
&\le \frac{4m^{-2/1+2s}}{(1-2^{-2/1+2s})}\|f\|^2_{W_{\frac{2}{1+2s}}}t_\epsilon^{4s/1+2s}+B_3.
\end{aligned}
$$

$$
\begin{aligned}
B_3 &:= \sum_{j<j_\epsilon,k}\mathbb{E}(1-\gamma_{jk}(\epsilon))^2\beta_{jk}^2\ I\{|y_{jk}|>mt_\epsilon\}I\{|\beta_{jk}|>mt_\epsilon/2\}[I\{|y_{jk}|\ge|\beta_{jk}|/2\}+I\{|y_{jk}|<|\beta_{jk}|/2
\\
&:= B_3'+B"_3.\\
B"_3 &\le \sum_{j<j_\epsilon,k}\beta_{jk}^2\mathbb{P}(|y_{jk}-\beta_{jk}|\ge mt_\varepsilon/4)\\
&\le \|f\|_2^2t_\epsilon^2.
\end{aligned}
$$

since $m^2 \geq 64$. We have used in the line above the concentration property of the Gaussian distribution. Now using (7) and Proposition 1, we get,

$$\begin{aligned} B_3' &\leq \sum_{j<j_\epsilon,k} \mathbb{E}\beta_{jk}^2(1-\gamma_{jk}(\epsilon))^2 I\{|y_{jk}| \geq |\beta_{jk}|/2\} I\{|\beta_{jk}| > mt_\epsilon/2\} I\{|y_{jk}| \geq mt_\epsilon\}] \\ &\leq \sum_{j<j_\epsilon,k} \mathbb{E}\beta_{jk}^2 K(\gamma)^2 \left(\frac{t_\varepsilon}{|y_{jk}|} + t_\epsilon\right)^2 I\{|y_{jk}| \geq |\beta_{jk}|/2\} I\{|\beta_{jk}| > mt_\epsilon/2\}) \\ &\leq K(\gamma)^2 \frac{32 m^{-2/1+2s}}{1-2^{-2/1+2s}} \|f\|^2_{W_{\frac{2}{1+2s}}} t_\epsilon^{4s/1+2s} + 2K(\gamma)^2 \|f\|_2^2 t_\epsilon^2. \end{aligned}$$

□

We deduce as a corollary the following results.

Corollary 2. *The Hard thresholding $\hat{f}_T$ and the Soft thresholding $\hat{f}_S$ rules as defined in (2) and (3) with $m \geq 8$ are satisfying:*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}) = \mathcal{B}_{2,\infty}^{s/1+2s} \cap W_{\frac{2}{1+2s}}.$$

The proof of this corollary is an elementary consequence of Theorems 1, 2 and 6. It proves that these procedures are optimal in the maxiset sense among elitist rules which are limited.

4.3 When does the maxiset contain a tree-Besov space?

In this paragraph, we give a converse result to Theorems 1 and 4 with respect to the ideal maxiset results for limited and hereditary rules. This result suppose that the chosen wavelet basis is the Haar one.

Theorem 7. *Let $s > 0$, $m > 0$, $c > 0$ and $\gamma_{jk}(\epsilon)$ a sequence of weights lying in $[0,1]$ such that $\hat{\beta}(\epsilon) = (\gamma_{jk}(\epsilon) y_{jk})_{jk}$ belongs to $\mathcal{L}((mt_\epsilon)^2, 0) \cap \mathcal{H}(mt_\epsilon, ct_\epsilon)$. Suppose in addition that there exists a constant $K(\gamma)$ such that for any $\epsilon > 0$, any $0 \leq j < j_\epsilon$ and any k*

$$\max\{|y_{_I}|;\ I \subset I_{jk} \text{ and } |I| > m^2 t_\epsilon^2\} > mt_\epsilon \Longrightarrow (1-\gamma_{jk}(\epsilon)) \leq K(\gamma)[\frac{\varepsilon}{|y_{jk}| \vee mt_\epsilon} + t_\epsilon], \quad a.e. \tag{8}$$

where $2^{j_\epsilon} \sim (mt_\epsilon)^{-2}$. Then, as soon as $m \geq 4\sqrt{3}$,

$$MS(\hat{f}_\epsilon, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}) \supseteq \mathcal{B}_{2,\infty}^{s/1+2s} \cap W^T_{\frac{2}{1+2s}}.$$

The proof of this theorem is parallel to the previous one. It follows from the following two propositions.

Proposition 3. *For any $0 < r < 2$ and any $f \in \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W_r^T$, then*

$$\sup_{0<\lambda<1/e} \lambda^r \left[\log(\frac{1}{\lambda})\right]^{-1} \sum_{0\leq j<j_\lambda,k} I\{\exists I' \subset I_{jk}\ /\ |I'| > \lambda^2 \text{ and } |\beta_{I'}| > \frac{\lambda}{2}\} \leq \frac{2^{6-r}\left(\|f\|^2_{W_r^T} + \|f\|^2_{\mathcal{B}_{2,\infty}^{(2-r)/4}}\right)}{(1-2^{-r})\log(2)}. \quad ($$

Moreover, we have the following inclusion spaces

$$W_r \subset W_r^T \quad \textit{and} \quad \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W_r^T \subset \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W_r^*.$$

Proof: The inclusion $W_r \subset W_r^T$ is easy to prove using the definitions of W_r and W_r^T. The second inclusion $\mathcal{B}_{2,\infty}^{(2-r)/4} \cap W_r^T \subset \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W_r^*$ is just a consequence of (9) which will be proved now. Let $f \in \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W_r^T$ and $0 < \lambda < 1/e$. We recall $2^{j_\lambda} \sim \lambda^{-2}$ and set for any $u \in \mathbb{N}$, $2^{j_{\lambda,u}} \sim (2^{1+u}\lambda)^{-2}$. Observing that, for any $j \geq 0$, any k there exist exactly $j+1$ dyadic interval I containing I_{jk}, we have

$$\begin{aligned}
& \sum_{0 \leq j < j_\lambda, k} I\{\exists I' \subset I_{jk} \;/\; |I'| > \lambda^2 \text{ and } |\beta_{I'}| > \frac{\lambda}{2}\} \\
\leq \; & \sum_{0 \leq j < j_\lambda, k} (j+1) I\{|\beta_{jk}| > \frac{\lambda}{2}, \;\; \forall I' \subsetneq I_{jk}, \;/\; |I'| > \lambda^2, |\beta_{I'}| \leq \frac{\lambda}{2}\} \\
\leq \; & \sum_{0 \leq j < j_\lambda, k} (j+1) I\{|\beta_{jk}| > \frac{\lambda}{2}, \;\; \forall I' \subsetneq I_{jk}, \;/\; |I'| > \lambda^2, |\beta_{I'}| \leq |\beta_{jk}|\} \\
\leq \; & \sum_{u \geq 0} \sum_{0 \leq j < j_\lambda, k} (j+1) I\{2^{u-1}\lambda < |\beta_{jk}| \leq 2^u \lambda, \;\; \forall I' \subsetneq I_{jk}, \;/\; |I'| > \lambda^2, |\beta_{I'}| \leq 2^u \lambda\} \\
\leq \; & j_\lambda \sum_{u \geq 0} \sum_{0 \leq j < j_\lambda, k} I\{2^{u-1}\lambda < |\beta_{jk}| \leq 2^u \lambda, \;\; \forall I' \subset I_{jk}, \;/\; |I'| > \lambda^2, |\beta_{I'}| \leq 2^u \lambda\} \\
\leq \; & \frac{2^4}{\log(2)} \log(\frac{1}{\lambda}) \sum_{u \geq 0} (2^u \lambda)^{-2} \sum_{0 \leq j < j_\lambda, k} \beta_{jk}^2 I\{\forall I' \subset I_{jk}, \;/\; |I'| > 4^{1+u}\lambda^2, |\beta_{I'}| \leq 2^u \lambda\} \\
\leq \; & \frac{2^4}{\log(2)} \log(\frac{1}{\lambda}) \sum_{u \geq 0} (\lambda 2^u)^{-2} \sum_{0 \leq j < j_{\lambda,u}, k} \beta_{jk}^2 I\{\forall I' \subset I_{jk}, \;/\; |I'| > 4^{1+u}\lambda^2, |\beta_{I'}| \leq 2^u \lambda\} \\
& + \frac{2^4}{\log(2)} \log(\frac{1}{\lambda}) \sum_{u \geq 0} (\lambda 2^u)^{-2} \sum_{j \geq j_{\lambda,u}, k} \beta_{jk}^2 \\
\leq \; & \frac{2^{6-r}}{(1-2^{-r})\log(2)} \left(\|f\|_{W_r^T}^2 + \|f\|_{\mathcal{B}_{2,\infty}^{(2-r)/4}}^2 \right) \log(\frac{1}{\lambda}) \lambda^{-r}.
\end{aligned}$$

The last inequalities use the fact that $f \in \mathcal{B}_{2,\infty}^{(2-r)/4} \cap W_r^T$. This ends the proof of the proposition.
□

Proposition 4. *Under the conditions of Theorem 7, we have the following inequality:*

$$\mathbb{E}\|\hat{f}_\epsilon - f\|_2^2 \leq t_\epsilon^{\frac{4s}{1+2s}} \quad \Big[\tfrac{4c^2}{m^2} + 2(\tfrac{2}{m^2} + 1 + 2K(\gamma)^2)\|f\|_2^2 + \tfrac{4\sqrt{6}}{m^3} \; + 2(4^{\frac{4s}{1+2s}} + 1) m^{\frac{4s}{1+2s}} \|f\|_{W_{\frac{2}{1+2s}}}^2$$
$$+ \tfrac{2^{7-2/1+2s} m^{-2/1+2s}}{(1-2^{-2/1+2s})\log(2)} (1 + 8K(\gamma)^2)(\|f\|_{W_{\frac{2}{1+2s}}}^2 + \|f\|_{B_{2,\infty}^{\frac{s}{1+2s}}}^2) + m^{\frac{4s}{1+2s}} (1 + 2 \times 4^{\frac{4s}{1+2s}}) \|f\|_{B_{2,\infty}^{\frac{s}{1+2}}}^2$$

Proof of the proposition. Obviously because of the limitation assumption, we have for $2^{j_\epsilon} \sim (mt_\epsilon)^{-2}$,

$$\mathbb{E}\|\hat{f}_\epsilon - f\|_2^2 = \mathbb{E}\| \sum_{j<j_\epsilon,k} (\gamma_{jk}(\epsilon)y_{jk} - \beta_{jk})\psi_{j,k}\|_2^2 + \sum_{j\geq j_\epsilon,k} \beta_{jk}^2.$$

The second term can be bounded by $(mt_\epsilon)^{4s/1+2s}\|f\|^2_{B^{\frac{s}{1+2s}}_{2,\infty}}$, by using the definition of the Besov norm.

Let us recall, for any $\lambda > 0$

- $|\bar{y}_{jk}(\lambda)| := \max\{|y_{_I}|;\ I \subset I_{jk} \text{ and } |I| > \lambda^2\}$,
- $|\bar{\beta}_{jk}(\lambda)| := \max\{|\beta_{_I}|;\ I \subset I_{jk} \text{ and } |I| > \lambda^2\}$,
- $|\bar{\delta}_{jk}(\lambda)| := \max\{|y_I - \beta_{_I}|;\ I \subset I_{jk} \text{ and } |I| > \lambda^2\}$.

The term $\mathbb{E} \sum_{0\leq j<j_\epsilon,k} (\gamma_{jk}(\epsilon)y_{jk} - \beta_{jk})^2$ can be bounded by $2(A + B)$, where

$$\begin{aligned} A + B &= \mathbb{E} \sum_{0\leq j<j_\epsilon,k} [\gamma_{jk}(\epsilon)^2(y_{jk} - \beta_{jk})^2 + (1 - \gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ I\{|\bar{y}_{jk}(mt_\epsilon)| \leq mt_\epsilon\} \\ &+ \mathbb{E} \sum_{0\leq j<j_\epsilon,k} [\gamma_{jk}(\epsilon)^2(y_{jk} - \beta_{jk})^2 + (1 - \gamma_{jk}(\epsilon))^2\beta_{jk}^2]\ I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} \end{aligned}$$

Again we split A into $A_1 + A_2$, and because of the condition $\mathcal{H}(mt_\epsilon, ct_\epsilon)$, we have that, on $\{|\bar{y}_{jk}(mt_\epsilon)| \leq mt_\epsilon\}$, $\gamma_{jk} \leq ct_\epsilon$. So,

$$\begin{aligned} A_1 &= \mathbb{E} \sum_{0\leq j<j_\epsilon,k} \gamma_{jk}(\epsilon)^2(y_{jk} - \beta_{jk})^2\ I\{|\bar{y}_{jk}(mt_\epsilon)| \leq mt_\epsilon\} \\ &\leq c^2 2^{j_\epsilon} t_\epsilon^2 \epsilon^2 \\ &\leq \frac{2c^2}{m^2} t_\epsilon^2. \end{aligned}$$

As for the proof of Proposition 3, and using lemma 2, we obtain

$$\begin{aligned} A_2 &\leq \mathbb{E} \sum_{0\leq j<j_\epsilon,k} \beta_{jk}^2\ I\{|\bar{y}_{jk}(mt_\epsilon)| \leq mt_\epsilon\}[I\{|\bar{\beta}_{jk}(mt_\epsilon)| \leq 2mt_\epsilon\} + I\{|\bar{\beta}_{jk}(mt_\epsilon)| > 2mt_\epsilon\}] \\ &\leq (4mt_\epsilon)^{4s/1+2s}(\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{B^{s/1+2s}_{2,\infty}}) + \sum_{0\leq j<j_\epsilon,k} \beta_{jk}^2 \mathbb{P}(|\bar{\delta}_{jk}(mt_\epsilon)| > mt_\epsilon) \\ &\leq (4mt_\epsilon)^{4s/1+2s}(\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{B^{s/1+2s}_{2,\infty}}) + 2^{j_\epsilon}\|f\|_2^2 \epsilon^{m^2/2} \\ &\leq (4mt_\epsilon)^{4s/1+2s}(\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{B^{s/1+2s}_{2,\infty}}) + \frac{2\|f\|_2^2}{m^2} t_\epsilon^2 \end{aligned}$$

We have used the fact that $m^2 \geq 8$.

$$\begin{aligned} B &= \mathbb{E} \sum_{0 \leq j < j_\epsilon, k} [\gamma_{jk}(\epsilon)^2 (y_{jk} - \beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2] \; I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\}[I\{|\bar{\beta}_{jk}(mt_\epsilon)| \leq mt_\epsilon/2\} \\ &\quad + I\{|\bar{\beta}_{jk}(mt_\epsilon) > mt_\epsilon/2\}] \\ &:= B_1 + B_2 \end{aligned}$$

For B_1 we use the Schwartz inequality:

$$\mathbb{E}(y_{jk} - \beta_{jk})^2 I\{|\bar{\delta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} \leq (\mathbb{P}(|\bar{\delta}_{jk}(mt_\epsilon)| > mt_\epsilon/2)^{1/2} (\mathbb{E}(y_{jk} - \beta_{jk})^4)^{1/2}$$

where $\mathbb{E}(y_{jk} - \beta_{jk})^4 = 3\epsilon^4$ and $\mathbb{P}(|\bar{\delta}_{jk}(mt_\epsilon)| > mt_\epsilon/2) \leq \epsilon^{m^2/8}$ (using lemma 2). So, choosing m such that $m^2 \geq 48$,

$$\begin{aligned} B_1 &\leq \sqrt{3}\; 2^{\frac{j_\epsilon}{2}} \textstyle\sum_{0 \leq j < j_\epsilon, k} \epsilon^2 I\{|\bar{\beta}_{jk}(mt_\epsilon)| \leq mt_\epsilon/2\} \epsilon^{m^2/16} + \textstyle\sum_{0 \leq j < j_\epsilon, k} \beta_{jk}^2 I\{|\bar{\beta}_{jk}(mt_\epsilon)| \leq mt_\epsilon/2\} \\ &\leq \sqrt{3}\; 2^{\frac{3j_\epsilon}{2}} t_\epsilon^{2+m^2/16} + \|f\|^2_{W^T_{\frac{2}{1+2s}}} (mt_\epsilon)^{4s/1+2s} \\ &\leq \frac{2\sqrt{6}}{m^3} t_\epsilon^2 + \|f\|^2_{W^T_{\frac{2}{1+2s}}} (mt_\epsilon)^{4s/1+2s} \end{aligned}$$

For B_2, we use, Proposition 3:

$$\begin{aligned} B_2 &= \mathbb{E} \textstyle\sum_{0 \leq j < j_\epsilon, k} [\gamma_{jk}(\epsilon)^2 (y_{jk} - \beta_{jk})^2 + (1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2] \; I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} \\ &\leq \textstyle\sum_{0 \leq j < j_\epsilon, k} [\epsilon^2 I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} + B_3 \\ &\leq \frac{2^{6-2/1+2s}}{(1-2^{-2/1+2s})\log(2)} \left(\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{\mathcal{B}^{s/1+2s}_{2,\infty}} \right) \epsilon^2 \log(\tfrac{1}{mt_\epsilon}) (mt_\epsilon)^{-\frac{2}{1+2s}} + B_3 \\ &\leq \frac{2^{6-2/1+2s} m^{-2/1+2s}}{(1-2^{-2/1+2s})\log(2)} \left(\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{\mathcal{B}^{s/1+2s}_{2,\infty}} \right) t_\epsilon^{4s/1+2s} + B_3 \end{aligned}$$

$$\begin{aligned} B_3 &:= \sum_{0 \leq j < j_\epsilon, k} \mathbb{E}(1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2 \; I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} \\ &\leq \sum_{0 \leq j < j_\epsilon, k} \mathbb{E}(1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2 \; I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} I\{|\beta_{jk}| < |y_{jk}| + mt_\epsilon\} \\ &\quad + \sum_{0 \leq j < j_\epsilon, k} \mathbb{E}(1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2 \; I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} I\{|y_{jk} - \beta_{jk}| \geq mt_\epsilon\} \\ &\leq \sum_{0 \leq j < j_\epsilon, k} \mathbb{E}(1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2 \; I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} I\{|\beta_{jk}| < 2(|y_{jk}| \vee mt_\epsilon)\} \\ &\quad + \sum_{0 \leq j < j_\epsilon, k} \mathbb{E}(1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2 \; I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} I\{|y_{jk} - \beta_{jk}| \geq mt_\epsilon\} \\ &:= B_3' + B"_3 \end{aligned}$$

$$B"_3 \leq \sum_{0 \leq j < j_\epsilon, k} \beta_{jk}^2 \mathbb{P}(|y_{jk} - \beta_{jk}| \geq mt_\epsilon) \leq \|f\|_2^2 \epsilon^{\frac{m^2}{2}} \leq \|f\|_2^2 \, t_\epsilon{}^2$$

since $m^2 \geq 4$. Now, using (8) and Proposition 3 we get,

$$\begin{aligned}
B_3' &\leq \sum_{0 \leq j < j_\epsilon, k} \mathbb{E}(1-\gamma_{jk}(\epsilon))^2 \beta_{jk}^2 I\{|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon\} I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} I\{|\beta_{jk}| < 2(|y_{jk}| \vee mt_\epsilon)\} \\
&\leq K(\gamma)^2 \sum_{0 \leq j < j_\epsilon, k} \mathbb{E}[t_\epsilon + \frac{\varepsilon}{|y_{jk}| \vee mt_\epsilon}]^2 \beta_{jk}^2 I\{|\bar{\beta}_{jk}(mt_\epsilon)| > mt_\epsilon/2\} I\{|\beta_{jk}| < 2(|y_{jk}| \vee mt_\epsilon)\} \\
&\leq 2K(\gamma)^2 \left[t_\epsilon^2 \|f\|_2^2 + \frac{2^{8-2/1+2s} m^{-2/1+2s}}{(1-2^{-2/1+2s}) \log(2)} (\|f\|^2_{W^T_{\frac{2}{1+2s}}} + \|f\|^2_{\mathcal{B}^{s/1+2s}_{2,\infty}}) t_\epsilon^{4s/1+2s} \right].
\end{aligned}$$

□

In section 6, we will provide two examples of procedures satisfying (8) that are also optimal among hereditary rules which are limited.

5 Maxisets for Bayesian procedures

In this section, we focus on the study of Bayes rules. We recall that we consider the prior model defined in Introduction.

5.1 Gaussian priors: a first approach

Let us consider the Bayes model (1) where γ is the Gaussian density, which is the most classical choice. In this case, we easily derive the Bayes rules of β_{jk} associated with the l^1-loss and the l^2-loss:

$$\breve{\beta}_{jk} = \mathrm{Med}(\beta_{jk}|y_{jk}) = \mathrm{sign}(y_{jk}) \max(0, \xi_{jk}),$$

$$\tilde{\beta}_{jk} = \mathbb{E}(\beta_{jk}|y_{jk}) = \frac{b_j}{1+\eta_{jk}} y_{jk},$$

where

$$\xi_{jk} = b_j |y_{jk}| - \epsilon \sqrt{b_j} \Phi^{-1} \left(\frac{1 + \min(\eta_{jk}, 1)}{2} \right),$$

$$b_j = \frac{\tau_{j,\epsilon}^2}{\epsilon^2 + \tau_{j,\epsilon}^2},$$

$$\eta_{jk} = \frac{1}{w_{j,\epsilon}} \frac{\sqrt{\epsilon^2 + \tau_{j,\epsilon}^2}}{\epsilon} \exp\left(-\frac{\tau_{j,\epsilon}^2 y_{jk}^2}{2\epsilon^2(\epsilon^2 + \tau_{j,\epsilon}^2)} \right),$$

and Φ is the normal cumulative distributive function. Both rules are then shrinkage rules. We also note that $\breve{\beta}_{jk}$ is zero whenever y_{jk} falls in an implicitly defined interval $[-\lambda_{j,\epsilon}, \lambda_{j,\epsilon}]$. So it is a thresholding rule. In the following, we study the maxisets of the previous estimates associated with the following very classical form for the hyperparameters:

$$\tau_{j,\epsilon}^2 = c_1 2^{-\alpha j}, \quad \pi_{j,\epsilon} = \min(1, c_2 2^{-bj}),$$

where c_1, c_2, α and b are positive constants. This particular form for the hyperparameters was suggested by [Abramovich et al., 1998] and then used by [Abramovich et al., 2004]. A nice interpretation was provided by these authors who explained how α, b, c_1 and c_2 can be derived for applications.

Remark: An alternative for eliciting these hyperparameters consists in using empirical Bayes methods and EM algorithm (see [Clyde and George, 1998], [Clyde and George, 2000] or [Johnstone and Silverman, 1998]). $\diamond$

In a minimax setting, [Abramovich et al., 2004] obtained the following result:

Theorem 8. *Let β^0 be $\breve{\beta}$ or $\tilde{\beta}$. With $\alpha = 2s+1$ and any $0 \leq b < 1$, there exist two positive constants C_1 and C_2 such that $\forall\ \epsilon > 0$,*

$$C_1(\epsilon\sqrt{\log(1/\epsilon)})^{4s/(2s+1)} \leq \sup_{\beta \in \mathcal{B}^s_{2,\infty}(M)} \mathbb{E}\|\beta^0 - \beta\|_2^2 \leq C_2 \log(1/\epsilon)\epsilon^{4s/(2s+1)}.$$

Now, let us consider the maxiset setting. Both previous Bayesian procedures are limited. Indeed, as soon as $\tau^2_{j,\epsilon} \leq \epsilon^2$ we have $b_j \leq 1/2$. So, each of these procedures belongs to $\mathcal{L}((c_1^{-1}\epsilon^2)^{1/\alpha}, 1/2)$. So, if $\alpha > 1$, by using Theorem 1, for $\beta^0 \in \{\breve{\beta}, \tilde{\beta}\}$,

$$MS(\beta^0, \|.\|_2^2, \epsilon^{2(\alpha-1)/\alpha}) \subset \mathcal{B}^{(\alpha-1)/2}_{2,\infty}.$$

With $s > 0$ and $\alpha = 1+2s$,

$$MS(\beta^0, \|.\|_2^2, \epsilon^{4s/(1+2s)}) \subset \mathcal{B}^s_{2,\infty}. \tag{10}$$

Actually, we have the following theorem:

Theorem 9. *For $s > 0$, $\alpha = 2s+1$, any $0 \leq b < 1$, and if β^0 is $\breve{\beta}$ or $\tilde{\beta}$,*

1. *for the rate $\epsilon^{4s/(1+2s)}$,*
$$MS(\beta^0, \|.\|_2^2, \epsilon^{4s/(1+2s)}) \subsetneq \mathcal{B}^s_{2,\infty},$$
2. *for the rate $(\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}$,*
$$MS(\beta^0, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) \subset \mathcal{B}^{*s}_{2,\infty},$$
3. *for the rate $\epsilon^{4s/(1+2s)}\log(1/\epsilon)$,*
$$\mathcal{B}^s_{2,\infty} \subset MS(\beta^0, \|.\|_2^2, \epsilon^{4s/(1+2s)}\log(1/\epsilon)).$$

with

$$\mathcal{B}^{*s}_{2,\infty} = \left\{ f \in \mathbb{L}^2 : \quad \sup_{J>0} 2^{2Js} J^{-2s/(1+2s)} \sum_{j \geq J} \sum_k \beta^2_{jk} < \infty \right\}.$$

Proof: The first point is a simple consequence of equation (10) and Theorem 8. The second one is easily obtained by using similar arguments as for the proof of Theorem 1. Finally, the proof of the last one is provided by Theorem 8. □

If we consider limited procedures, this theorem shows that the maxiset of these Bayesian procedures is not the ideal one. The first point of Theorem 9 and Corollary 1 show that they are also outperformed by linear estimates for polynomial rates of convergence. Furthermore, these procedures do not achieve the same performance as classical non linear procedures, since, obviously, $\mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W_{\frac{2}{2s+1}}$ is not included in $\mathcal{B}^{*s}_{2,\infty}$. The following theorem even reinforces this bad sentence by proving that these procedures are highly non robust with respect to the choice of α, which is a serious drawback in practise since s is generally unknown.

Theorem 10. *With the previous choice for the hyperparameters, for $s > 0$ and $\beta^0 \in \{\breve{\beta}, \tilde{\beta}\}$,*

- $\alpha > 2s + 1$ *implies $\mathcal{B}^s_{p,\infty}$ is not included in $MS(\beta^0, \|.\|_2^2, t_\epsilon^{4s/(1+2s)})$ for any $1 \le p \le \infty$.*
- $\alpha = 2s + 1$ *implies $\mathcal{B}^s_{p,\infty}$ is not included in $MS(\beta^0, \|.\|_2^2, t_\epsilon^{4s/(1+2s)})$ if $p < 2$,*

where

$$\mathcal{B}^s_{p,\infty} = \left\{ \beta : \quad \sup_{j \ge -1} 2^{jp(s+\frac{1}{2}-\frac{1}{p})} \sum_k |\beta_{jk}|^p < \infty \right\}.$$

Remark: Theorem 10 is established for the rate $t_\epsilon^{4s/(1+2s)}$ but it can be generalized for any rate of convergence of the form $\epsilon^{4s/(1+2s)}(\log(1/\epsilon))^m$, with $m \ge 0$. ◇
The proof of Theorem 10 is based on the following result:

Proposition 5. *If $\beta \in MS(\beta^0, \|.\|_2^2, t_\epsilon^{4s/(1+2s)})$ then there exists a constant C such that, for ϵ small enough:*

$$\sum_{j,k} \beta_{jk}^2 I\{\tau_{j,\epsilon}^2 \le \epsilon^2\} I\{|\beta_{jk}| > t_\epsilon\} \le C t_\epsilon^{\frac{4s}{1+2s}} \qquad (11)$$

Proof of the proposition:

Here we shall distinguish the cases of the posterior mean and median.

The posterior median can be written as follows:

$$\breve{\beta}_{jk} = \text{sign}(y_{jk})(b_j|y_{jk}| - g(\epsilon, \tau_{j,\epsilon}, y_{jk})),$$

with $0 \le g(\epsilon, \tau_{j,\epsilon}, y_{jk}) \le b_j|y_{jk}|$.
Let us assume that $b_j|y_{jk} - \beta_{jk}| \le (1-b_j)|\beta_{jk}|/2$ and $\tau_{j,\epsilon}^2 \le \epsilon^2$, so $b_j \le 1/2$.
First, let us suppose that $y_{jk} \ge 0$ so $\breve{\beta}_{jk} \ge 0$. If $\beta_{jk} \ge 0$, then

$$\begin{aligned} |\breve{\beta}_{jk} - \beta_{jk}| &= |b_j(y_{jk} - \beta_{jk}) - (1-b_j)\beta_{jk} - g(\epsilon, \tau_{j,\epsilon}, y_{jk})| \\ &= (1-b_j)\beta_{jk} - b_j(y_{jk} - \beta_{jk}) + g(\epsilon, \tau_{j,\epsilon}, y_{jk}) \\ &\ge \frac{1}{2}(1-b_j)\beta_{jk} \\ &\ge \frac{1}{4}\beta_{jk}. \end{aligned}$$

If $\beta_{jk} \leq 0$, then

$$|\breve{\beta_{jk}} - \beta_{jk}| \geq \frac{1}{4}|\beta_{jk}|.$$

The case $y_{jk} \leq 0$ is handled by using similar arguments and the particular form of the posterior median. So, we obtain:

$$\begin{aligned}
\mathbb{E}(\breve{\beta_{jk}} - \beta_{jk})^2 I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} &\geq \frac{1}{16}\beta_{jk}^2\, \mathbb{P}(b_j|y_{jk} - \beta_{jk}| \leq (1-b_j)|\beta_{jk}|/2) I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} \\
&\geq \frac{1}{16}\beta_{jk}^2\, \mathbb{P}(|y_{jk} - \beta_{jk}| \leq |\beta_{jk}|/2) I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\}.
\end{aligned}$$

So, we obtain:

$$\begin{aligned}
\mathbb{E}(\breve{\beta_{jk}} - \beta_{jk})^2 I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} &\geq \frac{1}{16}\beta_{jk}^2 \mathbb{P}(|y_{jk} - \beta_{jk}| \leq |\beta_{jk}|/2) I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} \\
&\geq \frac{1}{16}\beta_{jk}^2 (1 - \mathbb{P}(|y_{jk} - \beta_{jk}| > |\beta_{jk}|/2)) I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\}
\end{aligned}$$

Using the large deviations inequalities for the Gaussian variables, we obtain for ϵ small enough:

$$\begin{aligned}
\mathbb{E}(\breve{\beta_{jk}} - \beta_{jk})^2 I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} I\{|\beta_{jk}| > t_\epsilon\} &\geq \frac{1}{16}\beta_{jk}^2 (1 - \mathbb{P}(|y_{jk} - \beta_{jk}| > t_\epsilon/2)) I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} I\{|\beta_{jk}| > t_\epsilon\} \\
&\geq \frac{1}{32}\beta_{jk}^2 I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} I\{|\beta_{jk}| > t_\epsilon\}
\end{aligned}$$

This implies (11).

For the posterior mean, we have:

$$\begin{aligned}
\mathbb{E}(\tilde{\beta_{jk}} - \beta_{jk})^2 &= \mathbb{E}\left(\frac{b_j}{1+\eta_{jk}}(y_{jk} - \beta_{jk}) - (1 - \frac{b_j}{1+\eta_{jk}})\beta_{jk}\right)^2 \\
&\geq \frac{1}{4}\mathbb{E}\left((1 - \frac{b_j}{1+\eta_{jk}})\beta_{jk}\right)^2 I\left\{\frac{b_j}{1+\eta_{jk}}|y_{jk} - \beta_{jk}| \leq (1 - \frac{b_j}{1+\eta_{jk}})|\beta_{jk}|/2\right\}
\end{aligned}$$

So, we obtain:

$$\begin{aligned}
\mathbb{E}(\tilde{\beta_{jk}} - \beta_{jk})^2 I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} &\geq \frac{1}{16}\beta_{jk}^2 \mathbb{P}(|y_{jk} - \beta_{jk}| \leq |\beta_{jk}|/2) I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\} \\
&\geq \frac{1}{16}\beta_{jk}^2 (1 - \mathbb{P}(|y_{jk} - \beta_{jk}| > |\beta_{jk}|/2)) I\{\tau_{j,\epsilon}^2 \leq \epsilon^2\}
\end{aligned}$$

Finally, using similar arguments as those used for the posterior median, we obtain (11). Proposition 5 is proved. □

Now, let us prove Theorem 10. Let us first investigate the case $\alpha > 2s + 1$.
Let us take β such that all the β_{jk}'s are zero, except 2^j coefficients at each level j that are equal

to $2^{-j(s+\frac{1}{2})}$. Then, $\beta \in \mathcal{B}^s_{p,\infty}$. Since $\tau^2_{j,\epsilon} = c_1 2^{-j\alpha}$, if we put $2^{J_\alpha} \sim c_1^{\frac{1}{\alpha}} \epsilon^{-\frac{2}{\alpha}}$ and $2^{J_s} \sim t_\epsilon^{\frac{-2}{2s+1}}$, we observe that asymptotically $J_\alpha < J_s$. So, for ϵ small enough:

$$\begin{aligned} \sum_{j,k} \beta^2_{jk} I\{\tau^2_{j,\epsilon} \leq \epsilon^2\} I\{|\beta_{jk}| > t_\epsilon\} &= \sum_{J_\alpha \leq j < J_s} 2^{-2js} \\ &\geq c\epsilon^{\frac{4s}{\alpha}}, \end{aligned}$$

with c a positive constant. Using Proposition 5, β does not belong to $MS(\beta^0, \|.\|^2_2, t_\epsilon^{4s/(1+2s)})$.
Let us then investigate the case $\alpha = 2s+1$.
Let us take β such that all the β_{jk}'s are zero, except 1 coefficient at each level j that is equal to $2^{-j(s+\frac{1}{2}-\frac{1}{p})}$. Then, $\beta \in \mathcal{B}^s_{p,\infty}$. Similarly, we put $2^{J_\alpha} \sim c_1^{\frac{1}{\alpha}} \epsilon^{-\frac{2}{\alpha}}$ and $2^{\tilde{J}_s} \sim t_\epsilon^{-1/(s+\frac{1}{2}-\frac{1}{p})}$, we observe that asymptotically $J_\alpha < \tilde{J}_s$. So, for ϵ small enough:

$$\begin{aligned} \sum_{j,k} \beta^2_{jk} I\{\tau^2_{j,\epsilon} \leq \epsilon^2\} I\{|\beta_{jk}| > t_\epsilon\} &= \sum_{J_\alpha \leq j < \tilde{J}_s} 2^{-2j(s+\frac{1}{2}-\frac{1}{p})} \\ &\geq \tilde{c}\epsilon^{4(s+\frac{1}{2}-\frac{1}{p})/\alpha}, \end{aligned}$$

with $\tilde{c}$ a positive constant. Using Proposition 5, β does not belong to $MS(\beta^0, \|.\|^2_2, t_\epsilon^{4s/(1+2s)})$, since $p < 2$. □

The goal of the following subsections is to investigate a different choice for the hyperparameters $\tau_{j,\epsilon}$ and $w_{j,\epsilon}$ and for the density γ. Indeed, as in [Johnstone and Silverman, 2004a] and [Johnstone and Silverman, 2004b] in the minimax setting, we would like to point out posterior Bayes estimates stemmed from the prior model (1) that achieve the same performance as non linear ones in the maxiset approach. It is all the more natural since Bayesian procedures can achieve better performances than classical non linear ones from a practical point of view. More precisely, we investigate a choice for the hyperparameters and for the density γ that enables us to obtain maxisets at least as large as $\mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W_{\frac{2}{2s+1}}$. Two different ways will be investigated. In section 5.2, we give up Gaussian densities and we consider heavy-tailed densities γ, as in [Johnstone and Silverman, 2004a] and [Johnstone and Silverman, 2004b]. Not surprisingly, the modified Bayesian procedures achieve very good performances. We show this result by proving that the Bayesian procedures are both limited and elitist. Then, in section 5.3, we wonder whether heavy-tailed priors are unavoidable and we consider, once more, Gaussian priors but with a different choice for the hyperparameters.

5.2 Heavy-tailed priors

In this section, we still consider the prior model (1), but the density γ is no longer Gaussian. We assume that there exist two positive constants M and M_1 such that

$$\sup_{\beta \geq M_1} \left| \frac{d}{d\beta} \log \gamma(\beta) \right| = M < \infty. \tag{12}$$

The hypothesis (12) means that the tails of γ have to be exponential or heavier. Indeed, under (12), we have:

$$\forall\, u \geq M_1, \quad \gamma(u) \geq \gamma(M_1)\exp(-M(u-M_1)).$$

In the minimax approach of [Johnstone and Silverman, 2004a] and [Johnstone and Silverman, 2004b], the priors also verified (12). To complete the prior model, we assume that $\tau_{j,\epsilon} = \epsilon$ and $w_{j,\epsilon}$ depends only on ϵ with

$$w_{j,\epsilon} = w(\epsilon) \longrightarrow 0, \text{ as } \epsilon \longrightarrow 0$$

and w a positive continuous function. Using these assumptions, the following proposition describes the properties of the posterior median and mean:

Proposition 6. *We have:*

1. *The estimates $\breve{\beta}_{jk} = Med(\beta_{jk}|y_{jk})$ and $\tilde{\beta}_{jk} = \mathbb{E}(\beta_{jk}|y_{jk})$ are shrinkage rules: for $\beta^0_{jk} \in \{\breve{\beta}_{jk}, \tilde{\beta}_{jk}\}$, $y_{jk} \longrightarrow \beta^0_{jk}$ is antisymmetric, increasing on $(-\infty, +\infty)$ and*

$$0 \leq \beta^0_{jk} \leq y_{jk}, \quad \forall\, y_{jk} \geq 0.$$

2. *$\breve{\beta}_{jk}$ is a thresholding rule: there exists $\breve{t}_\epsilon$ such that*

$$\breve{\beta}_{jk} = 0 \iff |y_{jk}| \leq \breve{t}_\epsilon,$$

where the threshold $\breve{t}_\epsilon$ verifies for ϵ small enough, $\breve{t}_\epsilon \geq \epsilon\sqrt{2\log(1/w(\epsilon))}$ and

$$\lim_{\epsilon \to 0} \frac{\breve{t}_\epsilon}{\epsilon\sqrt{2\log(1/w(\epsilon))}} = 1.$$

3. *There exists a positive constant C such that*

$$\tilde{\beta}_{jk} = \tilde{\gamma}_{jk} y_{jk},$$

with

$$0 \leq \tilde{\gamma}_{jk} \leq C w(\epsilon)\exp(\frac{y_{jk}^2}{2\epsilon^2}).$$

4. *Let us consider the threshold $\breve{t}_\epsilon$ introduced previously. There exists a positive constant K such that for $\beta^0_{jk} \in \{\breve{\beta}_{jk}, \tilde{\beta}_{jk}\}$*

$$\limsup_{\epsilon \to 0} |\epsilon^{-1} y_{jk} - \epsilon^{-1}\beta^0_{jk}| I_{|y_{jk}| > 2\breve{t}_\epsilon} \leq K. \quad a.s.$$

Proof: The first point has been established by [Johnstone and Silverman, 2004a] and [Johnstone and Silverman, 2004b]. The second point is an immediate consequence of Proposition 3 of [Rivoirard, 2003]. To prove the third point, we use Proposition 4 and Remark 1 of

[Rivoirard, 2003] yielding that there exist two positive constants C_1 and C_2 and two positive functions $\tilde{e}_1$ and $\tilde{e}_2$ such that

$$\tilde{\beta}_{jk} = y_{jk} \times \frac{\tilde{e}_1(\epsilon^{-1}y_{jk})}{1 + w(\epsilon)^{-1}\exp(-\frac{y_{jk}^2}{2\epsilon^2})\gamma(\epsilon^{-1}y_{jk})^{-1}\tilde{e}_2(\epsilon^{-1}y_{jk})},$$

where

$$\forall\, x \geq 0, \quad C_1 \leq \tilde{e}_1(x), \tilde{e}_2(x) \leq C_2$$

So,

$$\tilde{\gamma}_{jk} \leq \frac{C_2\Gamma}{C_1} w(\epsilon)\exp(\frac{y_{jk}^2}{2\epsilon^2}),$$

where Γ is an upper bound for γ. The fourth point is easily derived by using Propositions 3 and 4 of [Rivoirard, 2003]. □

Now, let us introduce the following procedures. Given the previous prior model, we set

$$\breve{f}_\epsilon = \sum_{j<j_\epsilon}\sum_k \breve{\beta}_{jk}\psi_{jk}, \quad \breve{\beta}_{jk} = \mathrm{Med}(\beta_{jk}|y_{jk}), \tag{13}$$

and

$$\tilde{f}_\epsilon = \sum_{j<j_\epsilon}\sum_k \tilde{\beta}_{jk}\psi_{jk}, \quad \tilde{\beta}_{jk} = \mathbb{E}(\beta_{jk}|y_{jk}), \tag{14}$$

where j_ϵ is such that $2^{j_\epsilon} \sim t_\epsilon^{-2}$. Using the first three points of Proposition 6, we immediately obtain:

Corollary 3. *With C and $\breve{t}_\epsilon$ that have been introduced in Proposition 6, and $a \in]0,1[$, we have:*

$$\breve{f}_\epsilon \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(\breve{t}_\epsilon, 0),$$

$$\tilde{f}_\epsilon \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(\tilde{t}_\epsilon, a),$$

as soon as $\tilde{t}_\epsilon \leq \epsilon\sqrt{2\log(\frac{a}{Cw(\epsilon)})}$.

Remark: Proposition 6 also shows that the posterior median is a cautious procedure. Using a proper choice of the hyperparameters, we can easily prove that the procedure associated with the posterior mean is also cautious. ◇

We have the following consequences on the maxisets of the procedures:

Theorem 11. *Let $s > 0$. We suppose that there exist two positive constants ρ_1 and ρ_2 such that for $\epsilon > 0$ small enough,*

$$\epsilon^{\rho_1} \leq w(\epsilon) \leq \epsilon^{\rho_2}.$$

Then, we have:

$$MS(f_\epsilon^0, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) = \mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}},$$

where $f_\epsilon^0 \in \{\tilde{f}_\epsilon, \breve{f}_\epsilon\}$, as soon as $\rho_2 \geq 32$ for the posterior median and $\rho_2 \geq 33$ for the posterior mean.

Proof: The inclusions

$$MS(\breve{f}_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) \subset \mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}}$$

and

$$MS(\tilde{f}_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) \subset \mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}}$$

are provided by Theorems 1 and 2 and Corollary 3.
The inclusions

$$\mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}} \subset MS(\breve{f}_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)})$$

and

$$\mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}} \subset MS(\tilde{f}_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)})$$

are provided by the fourth point of Proposition 6, Corollary 3 and Theorem 6. □
So, the adaptive Bayesian procedures based on heavy-tailed prior densities are optimal among the class of limited and elitist procedures. We can also note that they outperform the Bayesian procedures of section 5.1 from the maxiset point of view.

5.3 Gaussian priors with large variance

The previous subsection has shown the power of the Bayes procedures built from heavy-tailed prior models in the maxiset setting. The goal of this section is then to answer the following questions. Are heavy-tailed priors unavoidable? Can we simultaneously consider Gaussian densities and ignore the empirical Bayes setting to build optimal Bayesian procedures? In other words, if γ is the Gaussian density, does there exist a fixed and adaptive choice of the hyperparameters $\pi_{j,\epsilon}$ and $w_{j,\epsilon}$ such that

$$MS(f^0_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) = \mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}},$$

where $f^0_\epsilon \in \{\breve{f}_\epsilon, \tilde{f}_\epsilon\}$ (see (13) and (14))?

This is a very important issue since calculation using Gaussian priors are mostly direct and obviously much easier than heavy tails priors.

The answers are provided by the following theorem:

Theorem 12. *We consider the prior model (1), where γ is the Gaussian density. We assume that $\tau_{j,\epsilon} = \tau(\epsilon)$ and $w_{j,\epsilon} = w(\epsilon)$ are independent of j with w a continuous positive function. We consider $\breve{f}_\epsilon$ and $\tilde{f}_\epsilon$ introduced in (13) and (14). If*

$$1 + \epsilon^{-2}\tau(\epsilon)^2 = t_\epsilon^{-1}$$

and there exist q_1 and q_2 such that for ϵ small enough

$$\epsilon^{q_1} \le w(\epsilon) \le \epsilon^{q_2}, \qquad (15)$$

we have:

$$MS(f^0_\epsilon, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) = \mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}},$$

where $f^0_\epsilon \in \{\tilde{f}_\epsilon, \breve{f}_\epsilon\}$ as soon as $q_2 > 63/2$ for the posterior median and $q_2 \ge 65/2$ for the posterior mean.

Whereas we usually consider $\tau_{j,\epsilon}^2 = \epsilon^2$ or $\tau_{j,\epsilon}^2 = 2^{-j\alpha}$, here we impose a "larger" variance. It is the key point of the proof of Theorem 12. In a sense, we re-create the heavy tails by increasing the variance.

Before giving it, let us prove that both Bayesian procedures belong to the class of limited and elitist procedures:

Proposition 7. *Under the assumptions of Theorem 12, we have for any $m > 0$ and for ϵ small enough,*

- *if $q_2 > \frac{m^2-1}{2}$, $\breve{f}_\epsilon \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(mt_\epsilon, 0)$,*
- *if $q_2 \geq \frac{m^2+1}{2}$, $\tilde{f}_\epsilon \in \mathcal{L}(t_\epsilon^2, 0) \cap \mathcal{E}(mt_\epsilon, t_\epsilon)$.*

Proof: Using the definition of j_ϵ, each Bayesian procedure belongs to $\mathcal{L}(t_\epsilon^2, 0)$. Now, let us assume that $|y_{jk}| \leq mt_\epsilon$. Then,

$$\begin{aligned} \eta_{jk} &= \frac{1}{w(\epsilon)}\frac{\sqrt{\epsilon^2+\tau(\epsilon)^2}}{\epsilon}\exp\left(-\frac{\tau(\epsilon)^2 y_{jk}^2}{2\epsilon^2(\epsilon^2+\tau(\epsilon)^2)}\right) \\ &\geq \frac{1}{w(\epsilon)} t_\epsilon^{-1/2}\exp\left(-\frac{m^2 t_\epsilon^2}{2\epsilon^2}\right) \\ &\geq \epsilon^{\frac{m^2}{2}-\frac{1}{2}}\frac{1}{w(\epsilon)}(\log(1/\epsilon))^{-1/4}. \end{aligned}$$

If $q_2 > \frac{m^2-1}{2}$, for ϵ small enough, $\eta_{jk} \geq 1$ and $\breve{\beta}_{jk} = 0$. So, $\breve{f}_\epsilon \in \mathcal{E}(mt_\epsilon, 0)$.
If $q_2 \geq \frac{m^2+1}{2}$, for ϵ small enough, $\eta_{jk} \geq t_\epsilon^{-1}$ and $\frac{b_j}{1+\eta_{jk}} \leq t_\epsilon$. So, $\tilde{f}_\epsilon \in \mathcal{E}(mt_\epsilon, t_\epsilon) \subset \mathcal{E}(mt_\epsilon, 1/2)$ for $\epsilon < 1$. □

Now let us prove the theorem:
Proof of Theorem 12: The inclusion

$$MS(f_\epsilon^0, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}) \subset \mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}}$$

is a direct consequence of Proposition 7 and Theorems 1 and 2.
Now, let us prove that

$$\mathcal{B}_{2,\infty}^{s/(2s+1)} \cap W_{\frac{2}{2s+1}} \subset MS(f_\epsilon^0, \|.\|_2^2, (\epsilon\sqrt{\log(1/\epsilon)})^{4s/(1+2s)}).$$

For this purpose, let us prove (7). Let us fix a constant $M \geq \sqrt{6+4q_1}$. We assume $|y_{jk}| > Mt_\epsilon$. Then, for ϵ small enough,

$$\begin{aligned} \eta_{jk} &= \frac{1}{w(\epsilon)}\frac{\sqrt{\epsilon^2+\tau(\epsilon)^2}}{\epsilon}\exp\left(-\frac{\tau(\epsilon)^2 y_{jk}^2}{2\epsilon^2(\epsilon^2+\tau(\epsilon)^2)}\right) \\ &\leq \frac{1}{w(\epsilon)}\frac{\sqrt{\epsilon^2+\tau(\epsilon)^2}}{\epsilon}\epsilon^{\frac{M^2}{4}} \\ &\leq \frac{1}{w(\epsilon)} t_\epsilon^{-1/2}\epsilon^{\frac{M^2}{4}} \\ &\leq t_\epsilon. \end{aligned}$$

Let us prove (7) for $\breve{\beta_{jk}}$. Using the previous inequality, we have for ϵ small enough, and for any $j < j_\epsilon$ and any k,

$$\epsilon\sqrt{b_j}\Phi^{-1}\left(\frac{1+\min(\eta_{jk},1)}{2}\right) \leq t_\epsilon.$$

So,

$$\begin{aligned}
|y_{jk}-\breve{\beta_{jk}}| &= |y_{jk}-\breve{\beta_{jk}}|I\{|y_{jk}|>Mt_\epsilon\}+|y_{jk}-\breve{\beta_{jk}}|I\{|y_{jk}|\leq Mt_\epsilon\} \\
&\leq ((1-b_j)|y_{jk}|+t_\epsilon)I\{|y_{jk}|>Mt_\epsilon\}+2|y_{jk}|I\{|y_{jk}|\leq Mt_\epsilon\} \\
&\leq t_\epsilon|y_{jk}|+(1+2M)t_\epsilon,
\end{aligned}$$

which implies the required inequlity. Now, let us deal with the posterior mean. For ϵ small enough, and for any $j < j_\epsilon$ and any k,

$$\begin{aligned}
|y_{jk}-\tilde{\beta_{jk}}| &= |y_{jk}-\tilde{\beta_{jk}}|I\{|y_{jk}|>Mt_\epsilon\}+|y_{jk}-\tilde{\beta_{jk}}|I\{|y_{jk}|\leq Mt_\epsilon\} \\
&\leq \left(1-\frac{b_j}{1+\eta_{jk}}\right)|y_{jk}|I\{|y_{jk}|>Mt_\epsilon\}+2|y_{jk}|I\{|y_{jk}|\leq Mt_\epsilon\} \\
&\leq (1-b_j+\eta_{jk})|y_{jk}|I\{|y_{jk}|>Mt_\epsilon\}+2|y_{jk}|I\{|y_{jk}|\leq Mt_\epsilon\} \\
&\leq 2t_\epsilon|y_{jk}|+2Mt_\epsilon,
\end{aligned}$$

which implies (7) for the posterior mean.
Now, using Proposition 7 and Theorem 6, we obtain the required inclusion. □
So, Theorem 12 provides optimal Bayesian procedures among limited and elitist procedures, based on Gaussian priors, under the condition that the hyperparameter $\tau_{j,\epsilon}$ is "large". Under this assumption, the density $\gamma_{j,\epsilon}$ is then more spread around 0, which enables us to avoid considering heavy-tailed densities. Since the maxiset of these estimates is the intersection of the Besov space $\mathcal{B}^{s/(2s+1)}_{2,\infty}$ and the Lorentz space $W_{\frac{2}{2s+1}}$, they achieve the same performance as thresholding ones. Now, a natural question arises: Does there exist a non linear procedure that outperforms these procedures?

Another way of asking this question is the following : Up to now the largest maxiset that we encountered is of the form $\mathcal{B}^{s/(2s+1)}_{2,\infty} \cap W_{\frac{2}{2s+1}}$. Is it the 'maxi' maxiset?

The purpose of the following section is to prove that the answer to this question is no and provide examples of procedures yielding larger maxisets.

6 Hereditary procedures and Lepski method

Still choosing for wavelet basis, the Haar basis, we focus now on hereditary rules. First of all, we give two examples of such rules which satisfy conditions of Theorem 7. The first one (resp. the second one) is based on hereditary constraints associated with hard (resp. soft) thresholds. Combining Theorems 1, 4 and 7, we prove that these rules are optimal (in the maxiset sense with the rate $t_\epsilon^{4s/1+2s}$) among hereditary and limited rules, since their maxiset is exactly the ideal one described in section 3.
Then we point out certain likenesses between hereditary rules and Lepski's procedure based on local bandwith selection.

6.1 Two examples of optimal hereditary rules

To give a first example of Hereditary rule, let us consider the following procedure (**hard tree rule**) defined by:

$$\tilde{f}_{HT}(t) = y_{-10}\psi_{-10}(t) + \sum_{0\leq j<j_\epsilon} \sum_k \gamma^{H}_{jk} y_{jk}\psi_{jk}(t) \tag{16}$$

where $2^{j_\epsilon} \sim (mt_\epsilon)^{-2}$, $\gamma^{H}_{jk} = 1$ if $|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon$ and $\gamma^{H}_{jk} = 0$ otherwise.
It is obvious that

$$\tilde{f}_{HT} \in \mathcal{L}((mt_\epsilon)^2, 0) \cap \mathcal{H}(mt_\epsilon, t_\epsilon).$$

Remark: The definition of this hereditary rule is directly inspired from tree-methods in Approximation Theory (see [Cohen et al., 2001a]). Precisely, this procedure is a tree rule (see [Engel, 1994]), since it satisfies the following hereditary constraints:

$$\gamma^{H}_{jk} = 1 \Longrightarrow \forall I \supset I_{jk}, \quad \gamma^{H}_{I} = 1,$$

$$\gamma^{H}_{jk} = 0 \Longrightarrow \forall I \subset I_{jk}, \quad \gamma^{H}_{I} = 0.$$

Notice that this rule is the *smallest*[1] tree rule containing:

$$\mathcal{T}(\lambda_\epsilon) = \{(j,k);\ 0 \leq j < j_\epsilon, 0 \leq k < 2^j \text{ and } |y_{jk}| > \lambda_\epsilon\}.$$

◇

To point out a second example of hereditary rule, let us consider the following procedure (**soft tree rule**) defined by:

$$\tilde{f}_{ST}(t) = y_{-10}\psi_{-10}(t) + \sum_{0\leq j<j_\epsilon} \sum_k \gamma^{S}_{jk} y_{jk}\psi_{jk}(t) \tag{17}$$

where $2^{j_\epsilon} \sim (mt_\epsilon)^{-2}$, $\gamma^{S}_{jk} = 1 - \frac{\epsilon}{|\bar{y}_{jk}(mt_\epsilon)|}$ if $|\bar{y}_{jk}(mt_\epsilon)| > mt_\epsilon$ and $\gamma^{S}_{jk} = 0$ otherwise.
It is obvious that

$$\tilde{f}_{ST} \in \mathcal{L}((mt_\epsilon)^2, 0) \cap \mathcal{H}(mt_\epsilon, t_\epsilon).$$

We have:

Theorem 13. *If m is large enough, hard tree and soft tree rules $\hat{f}_\epsilon$ as defined in (16) and (17) are satisfying*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, t_\epsilon^{4s/(1+2s)}) = \mathcal{B}^{s/1+2s}_{2,\infty} \cap W^{T}_{\frac{2}{1+2s}}.$$

The proof is an elementary consequence of Theorems 1, 4 and 7. It proves that these procedures are optimal in the maxiset sense among limited and hereditary rules.

[1] according to the number of empirical coefficients used in the reconstruction

6.2 A connection with Lepski's procedure

In this paragraph, we show that the hard tree rule can be viewed as the *wavelet-version* of the local bandwith procedure of Lepski using kernel methods. Indeed let $(\hat{f}_j)_{j\in\mathbb{N}}$ be the family of estimators defined as follows:

- $\hat{f}_0(t) = y_{-10}\psi_{-10}(t)$
- $\hat{f}_{j+1}(t) = \hat{f}_j(t) + \sum_k y_{jk}\psi_{jk}(t).$

For all $0 < \epsilon < 1/e$, it is clear that for any j, $\hat{f}_j \in \mathcal{F}_\epsilon$. Let us denote $\mathbf{I^t_j}$ the j-dyadic interval containing t and let us say that an integer j is t-admissible if:
either $j = j_\epsilon$ or, for all $j \le j' < j_\epsilon$, for all $\mathbf{t}' \in \mathbf{I^t_j}$: $|\hat{f}_{j'+1}(t') - \hat{f}_{j'}(t')| \le 2^{j'/2}mt_\epsilon$.
Denote $\hat{j}(t) = \inf\{j;\ j \text{ is t-admissible}\}$. There are certain likenesses between the hard tree rule and the local bandwidth selection procedure defined by Lepski who introduced admissibility for kernel estimators. Indeed, we can easily observe that:

$$\hat{f}_{\hat{j}(t)}(t) = \tilde{f}_{HT}(t)$$

when the wavelet is the Haar basis. In this sense, our procedure can be considered in this particular case as a hybrid version of Lepski's procedures, using wavelets.
Remark: Recall that [Kerkyacharian and Picard, 2002] pointed out another hybrid version $\hat{f}^H$ of local bandwidth selection using the following definition of admissibility:
j is t-admissible if, either $j = j_\epsilon$ or for all $j \le j' < j_\epsilon$, for all $\mathbf{t}' \in \mathbf{I^t_{j'}}$: $|\hat{f}_{j'+1}(t') - \hat{f}_{j'}(t')| \le 2^{j'/2}mt_\epsilon$. The maxiset of this procedure for the rate $(\epsilon\sqrt{\log(\epsilon^{-1})})^{4s/1+2s}$ has been shown to be at least as large as $\mathcal{B}^{s/1+2s}_{2,\infty} \cap W_{\frac{2}{1+2s}}$, too. But [Autin, 2004] has proved that the maxiset of the hard tree rule for the same rate is larger than the last one.◇

7 Appendix

In the previous sections, for sake of simplicity, the choice of the rates of convergence was often restricted. Indeed, the rate was linked in a direct way to either the limitation or to the threshold bound for elitist or cautious rules. But generally, it is not necessary and we show in this section how this constraint can be relaxed.

7.1 Maxisets for limited procedures

Definition 9. *Let $s > 0$ and u be an increasing continuous map of $\mathbb{R}_+$ such that $u(0) = 0$. We shall say that a function $f \in \mathbb{L}_2([0,1])$ belongs to the space $\mathcal{B}^s_{2,\infty}(u)$, if and only if:*

$$\sup_{\lambda>0}(u(\lambda))^{-2s}\sum_j\sum_k \beta^2_{jk} I\{2^{-j} \le \lambda\} < \infty.$$

Of course, when $u(x) = x$, $\mathcal{B}^s_{2,\infty}(u)$ is the classical Besov space $\mathcal{B}^s_{2,\infty}$. In this section, we study the ideal maxisets for limited procedures. We also provide estimates that are optimal among the class of limited ones. For this purpose, let λ_ε be a increasing continuous function with $\lambda_0 = 0$,

Theorem 14 (Ideal maxiset for limited rules). *Let $s > 0$ and $\hat{f}_\epsilon$ be a limited rule belonging to $\mathcal{L}(\lambda_\epsilon, a)$, with $a \in [0, 1[$. Then*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{2s}) \subset \mathcal{B}^s_{2,\infty}(u).$$

Proof: Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{2s})$. We have:

$$\begin{aligned}
& (1-a)^2 \sum_{j,k} \beta_{jk}^2 I\{2^{-j} \leq \lambda_\epsilon\} \\
= \; & 2(1-a)^2 \sum_{j,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0) I\{\beta_{jk} \geq 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0) I\{\beta_{jk} < 0\}\right] I\{2^{-j} \leq \lambda_\epsilon\} \\
\leq \; & 2\mathbb{E} \sum_{j,k} \left[(\gamma_{jk} y_{jk} - \beta_{jk})^2 I\{\beta_{jk} \geq 0\} + (\gamma_{jk} y_{jk} - \beta_{jk})^2 I\{\beta_{jk} < 0\}\right] I\{2^{-j} \leq \lambda_\epsilon\} \\
\leq \; & 2\mathbb{E} \sum_{j,k} (\gamma_{jk} y_{jk} - \beta_{jk})^2 \\
\leq \; & C \; (u(\lambda_\epsilon))^{2s},
\end{aligned}$$

where C is a positive constant. So, f belongs to $\mathcal{B}^s_{2,\infty}(u)$. □

Conversely, we have the following result:

Theorem 15. *Let $s > 0$ and $(\gamma_j(\epsilon))_{jk}$ be a non increasing sequence of weights lying in $[0, 1]$ such that $\hat{\beta}^L_\epsilon = (\gamma_j(\epsilon) y_{jk})_{jk}$ belongs to $\mathcal{L}(\lambda_\epsilon, a)$, with $a \in [0, 1[$. If there exist C_1 and C_2 in $\mathbb{R}$ such that, with $\gamma_{-2} = 1$, $\forall \; \epsilon > 0$,*

$$\sum_{j \geq -1} (\gamma_{j-1} - \gamma_j)(1 - \gamma_j)(u(2^{-j}))^{2s} I\{2^j < \lambda_\epsilon^{-1}\} \leq C_1 (u(\lambda_\epsilon))^{2s} \tag{18}$$

$$\sum_{j \geq -1} 2^j \gamma_j(\epsilon)^2 \leq C_2 \; \epsilon^{-2} \; (u(\lambda_\epsilon))^{2s} \tag{19}$$

then,

$$\mathcal{B}^s_{2,\infty}(u) \subset MS(\hat{\beta}^L_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{2s}).$$

Proof: With $s_l = \sum_j \sum_k \beta_{jk}^2 I\{2^{-j} \leq 2^{-l}\}$, we have, using (18) and (19):

$$\begin{aligned}
\sum_{j,k} \mathbb{E}(\gamma_j y_{jk} - \beta_{jk})^2 &= \sum_{j,k} \mathbb{E}(\gamma_j (y_{jk} - \beta_{jk}) - (1-\gamma_j)\beta_{jk})^2 \\
&= \sum_{j,k} \gamma_j^2 \epsilon^2 + \sum_{j,k} (1-\gamma_j)^2 \beta_{jk}^2 \\
&\leq S_\psi \epsilon^2 \sum_j 2^j \gamma_j^2 + \sum_{j,k} \beta_{jk}^2 I\{2^{-j} \leq \lambda_\epsilon\} + \sum_{j,k} (1-\gamma_j)^2 \beta_{jk}^2 I\{2^{-j} > \lambda_\epsilon\} \\
&\leq (S_\psi C_2 + M'^2)\,(u(\lambda_\epsilon))^{2s} + \sum_{j \geq -1} (1-\gamma_j)^2 (s_j - s_{j+1}) I\{2^{-j} > \lambda_\epsilon\} \\
&\leq (S_\psi C_2 + M'^2)\,(u(\lambda_\epsilon))^{2s} + 2 \sum_{j \geq -1} (\gamma_{j-1} - \gamma_j)(1-\gamma_j) s_j I\{2^{-j} > \lambda_\epsilon\} \\
&\leq (S_\psi C_2 + M'^2)\,(u(\lambda_\epsilon))^{2s} + 2M'^2 \sum_{j \geq -1} (\gamma_{j-1} - \gamma_j)(1-\gamma_j)(u(2^{-j}))^{2s} I\{2^{-j} > \lambda_\epsilon\} \\
&\leq (S_\psi C_2 + M'^2 + 2M'^2 C_1)(u(\lambda_\epsilon))^{2s}.
\end{aligned}$$

□

Combining Theorems 14 and 15, by straightforward computations, we obtain:

Corollary 4. *If we assume that $u(x) = x\tilde{u}(x)$ where*

$$\tilde{u}(x)^{-1} = O(1) \text{ as } x \text{ goes to } 0$$

and if we consider linear estimates associated with the weights $\gamma_j^{(1)}(\lambda_\epsilon)$, $\gamma_j^{(2)}(\lambda_\epsilon)$ with $\alpha > (s \vee 1/2)$ or $\gamma_j^{(3)}(\lambda_\epsilon)$ with $\alpha > s$ (see section 2.2), then for $i \in \{1,2,3\}$

$$MS((\gamma_j^{(i)}(\lambda_\epsilon) y_{jk})_{jk}, \|.\|_2^2, (u(\lambda_\epsilon))^{2s}) = \mathcal{B}_{2,\infty}^s(u),$$

as soon as $(\epsilon^2 \lambda_\epsilon^{-1} u(\lambda_\epsilon)^{-2s})_\epsilon$ is bounded.

To shed light on this result, let us take $\lambda_\epsilon = \epsilon^{2/(1+2s)}$. So, $(\epsilon^2 \lambda_\epsilon^{-1} u(\lambda_\epsilon)^{-2s})_\epsilon$ is bounded as soon as $(\epsilon^{4s/(1+2s)}\ u(\lambda_\epsilon)^{-2s})_\epsilon$ is bounded. So, for the rate $\epsilon^{4s/(1+2s)}(\log(1/\epsilon))^{2sm}$, $m \geq 0$, the maxisets of the linear estimates mentioned in Corollary 4 are the spaces $\mathcal{B}_{2,\infty}^s(u)$, where $u(x) = x(\log(1/x))^m$.

7.2 Ideal maxisets for elitist rules

Definition 10. *Let $0 < r < 2$ and u be an increasing continuous map of $\mathbb{R}_+$ such that $u(0) = 0$. We shall say that a function $f \in \mathbb{L}_2([0,1])$ belongs to the space $W_{r,u}$ if and only if:*

$$\sup_{\lambda > 0} (u(\lambda))^{r-2} \sum_j \sum_k |\beta_{jk}|^2 I_{\{|\beta_{jk}| \leqslant \lambda\}} < \infty.$$

Theorem 16 (Ideal maxiset for elitist rules). *Let $s > 0$ and $\hat{f}_\epsilon$ be an elitist rule that belongs to $\mathcal{E}(\lambda_\epsilon, a)$ with $a \in [0, 1[$, where λ_ϵ is an increasing continuous function of ϵ, such that $\lambda_0 = 0$. Then*

$$MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{4s/(1+2s)}) \subset W_{\frac{2}{1+2s},u}.$$

Proof: Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{4s/(1+2s)})(M)$. We have:

$$\begin{aligned}
& (1-a)^2 \sum_{j,k} \beta_{jk}^2 I\{|\beta_{jk}| \le \lambda_\epsilon\} \\
= \; & 2(1-a)^2 \sum_{j,k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0) I\{\beta_{jk} \ge 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0) I\{\beta_{jk} < 0\}\right] I\{|\beta_{jk}| \le \lambda_\epsilon\} \\
\le \; & 2\mathbb{E} \sum_{j,k} \left[(\beta_{jk} - \gamma_{jk} y_{jk})^2 I\{\beta_{jk} \ge 0\} + (\beta_{jk} - \gamma_{jk} y_{jk})^2 I\{\beta_{jk} < 0\}\right] I\{|\beta_{jk}| \le \lambda_\epsilon\} \\
\le \; & 2\mathbb{E} \sum_{j,k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 \\
\le \; & 2M \; (u(\lambda_\epsilon))^{4s/1+2s}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that $f \in W_{\frac{2}{1+2s},u}$.

7.3 Ideal maxisets for cautious procedures

Definition 11. *Let $0 < r < 2$ and u be a increasing continuous map of $\mathbb{R}_+$ such that $u(0) = 0$. We shall say that a function $f \in \mathbb{L}_2([0, 1])$ belongs to the space $W^*_{r,u}$ if and only if:*

$$\sup_{\lambda > 0} (u(\lambda))^{r-2} \lambda^2 \left[\log(\frac{1}{\lambda})\right]^{-1} \sum_{j < j_\lambda, k} I_{\{|\beta_{jk}| > \lambda\}} < \infty.$$

Theorem 17 (Ideal maxiset for cautious rules). *Let $s > 0$ and $\hat{f}_\epsilon$ be a cautious rule that belongs to $\mathcal{C}(\lambda_\epsilon, a)$ with $a \in]0, 1]$. Let λ_ε be an increasing continuous function with $\lambda_0 = 0$ such that:*

$$\exists\, c > 0, \quad \forall\, \epsilon > 0, \; \frac{\lambda_\epsilon}{\sqrt{\log(\frac{1}{\lambda_\epsilon})}} \le c\epsilon. \tag{20}$$

Then

$$MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{4s/1+2s}) \subset W^*_{\frac{2}{1+2s},u}.$$

Remark: Note that the case $\lambda_\epsilon = t_\epsilon$ (resp. $\lambda_\epsilon = \epsilon$) satisfies (20) with $c = \sqrt{2}$ (resp. $c = 1$)$\Diamond$

Proof: Let $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{4s/1+2s})(M)$. Using (20),

$$a^2 \lambda_\epsilon^2 \left[\log(\frac{1}{\lambda_\epsilon})\right]^{-1} \sum_{j < j_\epsilon, k} I\{|\beta_{jk}| > \lambda_\epsilon\} \le a^2 c^2 \epsilon^2 \sum_{j < j_\epsilon, k} I\{|\beta_{jk}| > \lambda_\epsilon\}.$$

Now, let us recall that if X is a zero-mean Gaussian variable with variance ϵ^2, then

$$\mathbb{E}(X^2 I_{\{X<0\}}) = \mathbb{E}(X^2 I_{\{X>0\}}) = \frac{\epsilon^2}{2}.$$

From Lemma 1,

$$\begin{aligned}
& a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} I\{|\beta_{jk}| > \lambda_\epsilon\} \\
= \; & a^2c^2\epsilon^2 \sum_{j<j_\epsilon,k} [I\{\beta_{jk} > \lambda_\epsilon\} + I\{\beta_{jk} < -\lambda_\epsilon\}] \\
= \; & 2a^2c^2 \,\mathbb{E} \sum_{j<j_\epsilon,k} (\beta_{jk} - y_{jk})^2 \left[I\{y_{jk} - \beta_{jk} < 0\} I\{\beta_{jk} > \lambda_\epsilon\} + I\{y_{jk} - \beta_{jk} > 0\} I\{\beta_{jk} < -\lambda_\epsilon\}\right] \\
\leq \; & 8c^2 \,\mathbb{E} \sum_{j<j_\epsilon,k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 \\
\leq \; & 8c^2 M \; (u(\lambda_\epsilon))^{4s/1+2s}.
\end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that f belongs to ${W^*}_{\frac{2}{1+2s},u}$. □

7.4 Ideal maxisets for hereditary procedures

Definition 12. *Let $0 < r < 2$ and u be an increasing continuous transformation of $\mathbb{R}^+$ such that $u(0) = 0$. We say that a function $f \in \mathbb{L}_2([0,1])$ belongs to the space $W^{T}_{r,u}$ if and only if:*

$$\sup_{\lambda>0} (u(\lambda))^{r-2} \sum_{0\leq j<j_\lambda} \sum_k \beta_{jk}^2 I\{\forall I \subset I_{jk}, \; / \; |I| > \lambda^2, |\beta_I| \leq \frac{\lambda}{2}\} \; < \infty.$$

Theorem 18 (Ideal maxiset for hereditary rules). *Let $s > 0$ and $\hat{f}_\epsilon$ be a hereditary rule that belongs to $\mathcal{H}(\lambda_\epsilon, a)$ with $a \in [0,1[$. Let λ_ε be an increasing continuous function with $\lambda_0 = 0$ such that there exists a constant $C > 0$ satisfying*

$$\mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}) \leq Cu(\lambda_\epsilon)^2 \lambda_\epsilon^2. \tag{21}$$

Then

$$MS(\hat{f}_\epsilon, \|.\|_2^2, (u(\lambda_\epsilon))^{4s/(1+2s)}) \subset W^{T}_{r,u}$$

with $Z \sim \mathcal{N}(0,1)$.

Remark: For instance, $\lambda_\epsilon = mt_\epsilon$ and $u = I_d$ satisfy (21) for any $m \geq 4\sqrt{2}$. ◇

Proof: Let $2^{j_\epsilon} \sim \lambda_\epsilon^{-2}$ and $f \in MS(\hat{f}_\epsilon, \|.\|_2^2, \lambda_\epsilon^{4s/(1+2s)})(M)$. Denote

- $|\bar{y}_{jk}(\lambda_\epsilon)| := \max\{|y_I|; \; I \subset I_{jk}, |I| > \lambda_\epsilon^2\}$,
- $|\bar{\beta}_{jk}(\lambda_\epsilon)| := \max\{|\beta_I|; \; I \subset I_{jk}, |I| > \lambda_\epsilon^2\}$.

We have the following lemma:

Lemma 3. *Let λ_ϵ that satisfies (21). Then, for any $j \geq 0$ and any k:*

$$\mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon\} I\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} \leq 2C\ u(\lambda_\epsilon)^2.$$

Proof of the lemma: Let $Z \sim \mathcal{N}(0,1)$. We have

$$\begin{aligned} \mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon\} I\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} &\leq 2^{j_\epsilon} \mathbb{P}(|y_{jk} - \beta_{jk}| > \frac{\lambda_\epsilon}{2}\} \\ &\leq 2\lambda_\epsilon^{-2}\ \mathbb{P}(|Z| > \frac{\lambda_\epsilon}{2\epsilon}\} \\ &\leq 2C\ u(\lambda_\epsilon)^2. \end{aligned}$$

□

Now, using Lemma 3:

$$\begin{aligned} & (1-a)^2 \sum_{0 \leq j < j_\epsilon, k} \beta_{jk}^2 I\{\forall I' \subset I_{jk},\ /\ |I'| > \lambda_\epsilon^2, |\beta_{I'}| \leq \frac{\lambda_\epsilon}{2}\} \\ = & (1-a)^2 \sum_{0 \leq j < j_\epsilon, k} \beta_{jk}^2 I\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} \\ = & 2(1-a)^2 \sum_{0 \leq j < j_\epsilon, k} \beta_{jk}^2 \left[\mathbb{P}(y_{jk} - \beta_{jk} < 0) I\{\beta_{jk} > 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0) I\{\beta_{jk} < 0\}\right] I\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2} \\ \leq & 2(1-a)^2 \mathbb{E} \sum_{0 \leq j < j_\epsilon, k} \beta_{jk}^2 \left[I\{y_{jk} - \beta_{jk} < 0\} I\{\beta_{jk} > 0\} + \mathbb{P}(y_{jk} - \beta_{jk} > 0) I\{\beta_{jk} < 0\}\right] I\{|\bar{y}_{jk}(\lambda_\epsilon)| \leq \lambda \\ & +2\mathbb{E} \sum_{0 \leq j < j_\epsilon, k} \beta_{jk}^2 \mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon\} I\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} \\ \leq & 2\ \mathbb{E} \sum_{0 \leq j, k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 + \frac{\lambda^2}{2} \sum_{0 \leq j < j_\epsilon, k} \mathbb{P}(|\bar{y}_{jk}(\lambda_\epsilon)| > \lambda_\epsilon\} I\{|\bar{\beta}_{jk}(\lambda_\epsilon)| \leq \frac{\lambda_\epsilon}{2}\} \\ \leq & 2\ \mathbb{E} \sum_{0 \leq j, k} (\beta_{jk} - \gamma_{jk} y_{jk})^2 + 2Cu(\lambda_\epsilon)^2 \\ \leq & 2(M + C)\ (u(\lambda_\epsilon))^{4s/1+2s}. \end{aligned}$$

So, using the continuity of λ_ϵ in 0, we deduce that $f \in W^T_{\frac{2}{1+2s}, u}$. □

References

[Abramovich et al., 2004] Abramovich, F., Amato, U., and Angelini, C. (2004). On optimality of Bayesian wavelet estimators. *Scand. J. Statist.*, 31(2):217–234.

[Abramovich et al., 1998] Abramovich, F., Sapatinas, T., and Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(4):725–749.

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

[Autin, 2003] Autin, F. (2003). Maxiset for density estimation. *Technical Report. Submitted to Bernoulli.*

[Autin, 2004] Autin, F. (2004). Lepski's procedure with the maxiset approach. *Technical Report.*

[Chipman et al., 1997] Chipman, H. A., Kolaczyk, E. D., and McCulloch, R. E. (1997). Adaptive bayesian wavelet shrinkage. *Journal of the American Statistical Association*, 92:1413–1421.

[Clyde and George, 1998] Clyde, M. and George, E. I. (1998). Robust empirical bayes estimation in wavelets. *Technical Report.*

[Clyde and George, 2000] Clyde, M. and George, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 62(4):681–698.

[Clyde et al., 1998] Clyde, M., Parmigiani, G., and Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika*, 85(2):391–401.

[Cohen et al., 2001a] Cohen, A., Dahmen, W., Daubechies, I., and DeVore, R. (2001a). Tree approximation and optimal encoding. *Appl. Comput. Harmon. Anal.*, 11(2):192–226.

[Cohen et al., 2001b] Cohen, A., DeVore, R., Kerkyacharian, G., and Picard, D. (2001b). Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11(2):167–191.

[Engel, 1994] Engel, J. (1994). A simple wavelet approach to nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.*, 49(2):242–254.

[Jaffard, 1998] Jaffard, S. (1998). Oscillation spaces: properties and applications to fractal and multifractral functions. *J. Math. Phys.*, 39(8):4129–4141.

[Jaffard, 2004] Jaffard, S. (2004). Beyond besov spaces: Oscilllation spaces. *Technical Report. To appear in Constructive Approximation.*

[Johnstone and Silverman, 1998] Johnstone, I. M. and Silverman, B. W. (1998). Empirical bayes approaches to mixture problems and wavelet regression. *Technical Report.*

[Johnstone and Silverman, 2004a] Johnstone, I. M. and Silverman, B. W. (2004a). Empirical bayes selection of wavelet thresholds. *Technical Report.*

[Johnstone and Silverman, 2004b] Johnstone, I. M. and Silverman, B. W. (2004b). Needles and hay in haystacks: Empirical bayes estimates of possibly sparse sequences. *Ann. Statist*, 32:1594–1649.

[Kerkyacharian and Picard, 1993] Kerkyacharian, G. and Picard, D. (1993). Density estimation by kernel and wavelets methods: optimality of Besov spaces. *Statist. Probab. Lett.*, 18(4):327–336.

[Kerkyacharian and Picard, 2000] Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2):283–344.

[Kerkyacharian and Picard, 2002] Kerkyacharian, G. and Picard, D. (2002). Minimax or maxisets? *Bernoulli*, 8(2):219–253.

[Meyer, 1990] Meyer, Y. (1990). *Ondelettes et opérateurs. I.* Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris.

[Rivoirard, 2003] Rivoirard, V. (2003). Bayesian modelization of sparse sequences and maxisets for bayes rules. *Technical Report. Submitted to Math. Methods Statist.*

[Rivoirard, 2004] Rivoirard, V. (2004). Maxisets for linear procedures. *Statist. Probab. Lett.*, 67(3):267–275.

[Vidakovic, 1998] Vidakovic, B. (1998). Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *J. Amer. Statist. Assoc.*, 93(441):173–179.

[Zhang, 2002] Zhang, C.-H. (2002). General empirical bayes wavelet methods. *Technical Report.*