Non linear estimation over weak Besov spaces and minimax Bayes method

VINCENT RIVOIRARD

Laboratoire de Mathématique, UMR 8628, Equipe Probabilités Statistique et Modélisation, Université Paris-Sud 11, Batiment 425, 15 rue Georges Clémenceau, 91405 Orsay Cedex, France. E-mail: Vincent.Rivoirard@math.u-psud.fr

Summary

Weak Besov spaces play important roles in statistics as maxisets of classical procedures or for measuring the sparsity of signals. The goal of this paper is to study weak Besov balls $\mathcal{WB}_{s,p,q}(C)$ from the statistical point of view by using the minimax Bayes method. In particular, we compare weak and strong Besov balls statistically. By building an optimal Bayes wavelet thresholding rule, we first establish that, under suitable conditions, the rate of convergence of the minimax risk for $\mathcal{WB}_{s,p,q}(C)$ is the same as for the strong Besov ball $\mathcal{B}_{s,p,q}(C)$ that is contained by $\mathcal{WB}_{s,p,q}(C)$. However, we show that the asymptotically least favorable priors of $\mathcal{WB}_{s,p,q}(C)$ that are based on Pareto distributions cannot be asymptotically least favorable priors for $\mathcal{B}_{s,p,q}(C)$. Finally, we present sample paths of such priors that provide representations of the worst functions to be estimated for classical procedures and we give an interpretation of the roles of the parameters s, p and q of $\mathcal{WB}_{s,p,q}(C)$.

Keywords: asymptotically least favorable priors, Bayes method, minimax risk, rate of convergence, thresholding rules, weak Besov spaces.

1 Introduction

In this paper, we study weak Besov spaces, denoted $\mathcal{WB}_{s,p,q}$ in the sequel, from the statistical point of view. First, let us point out the importance of these spaces in approximation theory and statistics and explain the interest in considering these spaces that are defined in section 2.2.

DeVore (1989), Donoho (1996), Donoho and Johnstone (1996) and Cohen, DeVore and Hochmuth (2000a) noticed that weak l_p spaces, denoted wl_p, can be viewed as collections of functions on $[0,1]$ that can be approximated in $L_2([0,1])$ at rate $N^{-\sigma}$, $\sigma = 1/p - 1/2$. Then, Cohen *et al.* (2000a) have linked the approach of non linear approximation not only to weak l_p spaces but also to weak Besov spaces that can be viewed as weighted weak l_p spaces. Indeed, Cohen *et al.* (2000a) showed that weak Besov spaces appear in the characterization of the approximation performance of wavelet thresholding. In statistics, weak Besov spaces naturally appear in the framework of the maxiset theory. This approach was proposed by Cohen, DeVore, Kerkyacharian and Picard (2000b) and is an alternative to the classical minimax theory. Indeed, the minimax criterion asks of a procedure 'what is the worst performance over a given class of functions?' The maxiset criterion asks instead 'what is the class of functions for which the procedure attains a given rate of convergence?' This class of functions is called maxiset. We can note that for the minimax approach, we have to choose the function class and this choice is quite subjective, whereas the maxiset approach provides function spaces directly connected to the estimation procedure. The maxiset approach revealed the following main fact. Roughly speaking, weak Besov spaces are the maxisets of many classical estimation procedures. See for instance, the maxiset results proved by Cohen *et al.* (2000b) for wavelet thresholding and by Rivoirard (2005) for general Bayesian procedures. These authors underlined that the performance of these procedures depends on the smoothness and on the sparsity of the underlying signal to be estimated. So, smoothness and sparsity are strongly linked to weak Besov spaces. This point will be extensively developed in section 2.2 and we shall see how to use weak Besov spaces for measuring the smoothness and, in particular, the sparsity of signals. Let us end this presentation of the spaces $\mathcal{WB}_{s,p,q}$ by justifying the terminology of 'weak Besov space'. Section 2.2 shows that $\mathcal{WB}_{s,p,q}$ is very close to the classical Besov space $\mathcal{B}_{s,p,q}$ that will be denoted 'strong Besov space' in the sequel to avoid any ambiguity. The definition of $\mathcal{B}_{s,p,q}$ and its characterization by using wavelet coefficients are recalled in section 2.1. Actually, the space $\mathcal{WB}_{s,p,q}$ is defined by slightly relaxing Besov constraints on the wavelet coefficients and we have $\mathcal{B}_{s,p,q} \subset \mathcal{WB}_{s,p,q}$. So, the space $\mathcal{WB}_{s,p,q}$ appears as a weak version of $\mathcal{B}_{s,p,q}$.

Our first issue is to point out the minimax rate of convergence for each weak Besov ball denoted $\mathcal{WB}_{s,p,q}(C)$, in the framework of the classical white noise model and with Besov norms as loss functions. More precisely, we focus on the $\mathcal{B}_{s',p',p'}$-loss, where $0 \leq s' < \infty$, $1 \leq p' < \infty$. So, our first goal is to generalize the results proved by Johnstone (1994) who obtained the asymptotic values of the minimax risk for weak l_p balls with the l_2-loss. Naturally, the next goal is to compare the rates of convergence of the minimax risk associated respectively with $\mathcal{WB}_{s,p,q}(C)$ and $\mathcal{B}_{s,p,q}(C)$. The results concerning minimax rates are given by Theorem 1 in section 4.2 under assumptions on the parameters s, p, q, s' and p'. We show that the rates for $\mathcal{B}_{s,p,q}(C)$ and $\mathcal{WB}_{s,p,q}(C)$ are the same up to constants. This result may seem surprising since, actually, the inclusion $\mathcal{B}_{s,p,q}(C) \subset \mathcal{WB}_{s,p,q}(C)$ is strict. But, we noted that strong Besov spaces and weak Besov spaces are close. So, this result generalizes Theorem 1 of Johnstone (1994) who proved that minimax rates for weak l_p balls are the same as for l_p balls.

The proofs of these results exploit the well known Bayesian approach proposed by Pinsker (1980). This approach consists in proving that the minimax risk is asymptotically equal to the Bayes risk associated with a prior model. See Pinsker (1980) for more details. Pinsker's paper inspired a considerable literature. Let us cite Casella and Strawderman (1981) and Bickel (1981) who used minimax Bayes methods for the estimation of a bounded normal mean and Donoho and Johnstone (1994b), Johnstone (1994) and Donoho and Johnstone (1998) respectively for the estimation over l_p balls, weak l_p balls and strong Besov balls. To get the upper bound of the minimax risk on $\mathcal{WB}_{s,p,q}(C)$, we exploit the approach developed by Johnstone (1994) by building a minimax Bayes wavelet thresholding estimator. More precisely, each wavelet coefficient is estimated, at large resolution levels j, by the soft thresholding rule and the threshold depends on the parameters of the prior model (see section 4.2). Section 4.4 briefly describes the method to apply this estimator for denoising discrete data. So, this paper provides a new contribution to the crucial problem of choosing thresholds for wavelet thresholding. This issue has often been investigated, in particular, in a Bayesian framework. See for instance Abramovich, Sapatinas and Silverman (1998) who used the posterior median of a Gaussian prior model and the Bayes Factor procedure of Vidakovic (1998) that mimics hard thresholding.

Our second issue deals with asymptotically least favorable priors of $\mathcal{WB}_{s,p,q}(C)$. Such priors, that maximize the Bayes risk on a given class of probability measures, have a Bayes risk that is asymptotically equal to the minimax risk and their support belongs asymptotically to $\mathcal{WB}_{s,p,q}(C)$ (see section 3). Since minimax risks for $\mathcal{B}_{s,p,q}(C)$ and $\mathcal{WB}_{s,p,q}(C)$ are the same, a natural question arises: are asymptotically least favorable priors for $\mathcal{B}_{s,p,q}(C)$ and $\mathcal{WB}_{s,p,q}(C)$ also the same? We shall prove that the answer is no. Indeed, for each weak Besov ball $\mathcal{WB}_{s,p,q}(C)$, we present asymptotically least favorable priors derived from Pareto(p) distributions, but these priors are not asymptotically least favorable priors for $\mathcal{B}_{s,p,q}(C)$ (see Theorem 1). Johnstone (1994) pointed out asymptotically least favorable priors for strong Besov spaces based on Gaussian distributions or on two or three points distributions. So, this paper shows that Pareto distributions are typical of weak Besov spaces.

Our last issue is to get a representation of the 'typical enemies' for classical procedures. Since maxisets for these procedures seem to be often characterized by weak Besov spaces, it is natural to look for these signals in weak Besov balls. For this purpose, we shall naturally use simulations of asymptotically least favorable priors of a given weak Besov ball $\mathcal{WB}_{s,p,q}(C)$ that provide a good representation of the worst functions of $\mathcal{WB}_{s,p,q}(C)$ to be estimated. These simulations will show relationships between the parameters s, p and q and smoothness or sparsity of signals of $\mathcal{WB}_{s,p,q}(C)$.

The paper is organized as follows. Section 2 is devoted to weak Besov spaces and sparsity after an overview of strong Besov spaces. Section 3 introduces asymptotically least favorable priors. Section 4 presents the results we get. Finally, in section 5, we recall the minimax Bayes approach and we give some elements of the proofs. Section 6 is devoted to the proof of Theorem 3.

2 Weak Besov spaces

2.1 Overview of strong Besov spaces

Let us recall the definition of strong Besov spaces $\mathcal{B}_{s,p,q}$ when $0 < s < \infty$, $1 \leq p \leq \infty$, $1 \leq q < \infty$ (we do not consider the case $q = \infty$ in this paper). Let $f \in L_p(\mathbb{R})$. For any $r \in \mathbb{N}^*$ and any $h \in \mathbb{R}$, we set

$$\Delta_h^r(f,x) = \sum_{k=0}^{r} \frac{r!}{k!(r-k)!}(-1)^{r-k} f(x+kh), \quad x \in \mathbb{R},$$

and introduce the r-th modulus of smoothness of f:

$$w_r(f,t)_p = \sup_{0<h\leq t} \|\Delta_h^r(f,.)\|_p, \quad t \geq 0.$$

Now, for $0 < s < \infty$, $1 \leq p \leq \infty$, $1 \leq q < \infty$,

$$\mathcal{B}_{s,p,q} = \{f: \quad \|f\|_p + |f|_{\mathcal{B}_{s,p,q}} < \infty\},$$

where

$$|f|_{\mathcal{B}_{s,p,q}} = \left(\int_0^\infty [t^{-s} w_r(f,t)_p]^q \frac{dt}{t}\right)^{1/q}, \quad r \in \mathbb{N}^* \text{ such that } s < r \leq s+1.$$

See DeVore and Lorentz (1993) for more details. Using a multiresolution analysis, we can connect Besov norms to sequence space norms. Let us suppose that we are given a pair of scaling function and wavelet ϕ and ψ and a function f having the following decomposition:

$$f(t) = \sum_{j\geq -1} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(t), \tag{1}$$

where $\psi_{jk}(t) = 2^{\frac{j}{2}}\psi(2^j t - k)$ if $j \in \mathbb{N}$, $\psi_{-1k}(t) = \phi(t-k)$, and $\beta_{jk} = \int f(t)\psi_{jk}(t)dt$. The sequences $(\beta_{-1k})_k$ and for $j \geq 0$, $(\beta_{jk})_k$ are respectively the approximation and the detail wavelet coefficients at level j. The following facts are true under standard properties of smoothness and moment vanishing of ϕ and ψ (see Meyer (1992)). If we are given $1 \leq p \leq \infty$, $1 \leq q < \infty$ and $0 < s < \infty$, the function f in (1) belongs to $\mathcal{B}_{s,p,q}$, if and only if $\beta = (\beta_{jk})_{jk}$ verify

$$\|\beta\|_{\mathcal{B}_{s,p,q}} = \left(\sum_{j=-1}^{+\infty} 2^{jq(s+\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\beta_{jk}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}} < \infty$$

(with the obvious modifications for $p = \infty$). In the following, we use this sequential characterization of strong Besov spaces, in particular for the evaluation of the minimax risk for $\mathcal{B}_{s',p',p'}$-norms as loss functions and we note $\|f\|_{\mathcal{B}_{s,p,q}} = \|\beta\|_{\mathcal{B}_{s,p,q}}$. This allows one to consider the case $s' = 0$ (we can note that when $s' = 0$ and $p' = 2$, $\|f\|_{\mathcal{B}_{s',p',p'}} = \|f\|_2$). Furthermore, we exploit Daubechies' construction that enables us to suppose in addition and without loss of generality that ϕ and ψ are both supported by the interval $[-A^{\phi,\psi}, B^{\phi,\psi}]$ (see Daubechies (1992)).

2.2 Sparsity and weak Besov spaces

Abramovich *et al.* (2000) introduced the notion of sparsity of an infinite vector $\theta \in \mathbb{R}^{\mathbb{N}}$ through the following approach. The vector θ is said to be sparse if there is a small proportion of relatively large entries. Therefore, they order the components of θ according to their size:

$$|\theta|_{(1)} \geq |\theta|_{(2)} \geq \dots \geq |\theta|_{(n)} \geq \dots$$

and they control the number of large entries by using the power-law bound: $\sup_n n^{\frac{1}{p}}|\theta|_{(n)} < \infty$, where $p > 0$. This condition is equivalent to say that θ belongs to the weak l_p space wl_p defined by:

$$wl_p = \left\{\theta \in \mathbb{R}^{\mathbb{N}}: \quad \sup_{\lambda>0} \lambda^p \sum_n \mathbb{1}_{|\theta_n|>\lambda} < \infty\right\}.$$

Now, let us introduce weak Besov spaces.

Definition 1. *Let us fix $0 < s,p,q < \infty$. We say that f in (1) belongs to the weak Besov space of parameters s,p,q, noted $\mathcal{WB}_{s,p,q}$ if*

$$\sup_{\lambda>0} \lambda^q \sum_j 2^{jq(s+\frac{1}{2}-\frac{1}{p})} N(j,\lambda)^{\frac{q}{p}} < \infty,$$

where $N(j,\lambda)$ is the number of wavelet coefficients at level j greater than λ:

$$N(j,\lambda)=\sum_k \mathbb{1}_{|\beta_{jk}|>\lambda}.$$

With each weak Besov space $\mathcal{WB}_{s,p,q}$, we associate the balls:

$$\mathcal{WB}_{s,p,q}(C)=\left\{f:\quad \sup_{\lambda>0}\lambda^q\sum_j 2^{jq(s+\frac{1}{2}-\frac{1}{p})}\,N(j,\lambda)^{\frac{q}{p}}\leq C^q\right\}.$$

So, we note that the weak Besov space $\mathcal{WB}_{s,p,p}$ can be viewed as a weighted weak l_p space. The weights penalize the counting of the β_{jk}'s greater than λ for the large scales according to the sign of $p(s+1/2)-1$. Obviously, $\mathcal{WB}_{s,p,p}$ with $s=1/p-1/2$ can be identified with wl_p. Therefore, the use of weak Besov spaces may appear as a good device to measure the smoothness, but in particular the sparsity of a wavelet expanded signal. Finally, using the Markov inequality and the sequential characterization of strong Besov spaces, it is easy to note that $\mathcal{B}_{s,p,q}(C)\subset\mathcal{WB}_{s,p,q}(C)$.

3 Minimax risk and asymptotically least favorable priors

We consider the white noise model:

$$dY_t=f(t)\,dt+\varepsilon dW_t,\quad t\in[0,1].$$

Restricting our attention to functions supported by $[0,1]$, β_{jk} is non-zero as soon as k is in $\mathcal{I}_j$, with

$$\mathcal{I}_j=\left\{-B^{\phi,\psi}+1,\ldots,\max(2^j,1)+A^{\phi,\psi}-1\right\}.$$

By setting for any $j\geq -1$ and $k\in\mathcal{I}_j$, $y_{jk}=\int\psi_{jk}(t)dY_t$, this model is reduced to a sequence space model, and we obtain the following sequence of independent variables:

$$y_{jk}=\beta_{jk}+\varepsilon z_{jk},\quad z_{jk}\sim\mathcal{N}(0,1),\quad j\geq -1,\ k\in\mathcal{I}_j.$$

When $0\leq s'<\infty$ and $1\leq p'<\infty$, the minimax risk is denoted:

$$\tilde{R}_\varepsilon\;=\;\inf_{\hat\beta}\sup_{\beta\in\mathcal{WB}_{s,p,q}(C)}\mathbb{E}_\beta\|\hat\beta-\beta\|^{p'}_{\mathcal{B}_{s',p',p'}}.$$

Let us note that evaluating the minimax risk for $\mathcal{B}_{s',p',p'}$-losses constitutes the first step to get the evaluation of the minimax risk for $L_{p'}$-losses. In particular, the value $s'=0$ provides a conjecture of the minimax rates for the $L_{p'}$-loss. The minimax risk will be evaluated by using 'asymptotically least favorable priors'. We say that π_ε, a prior distribution on $\beta=(\beta_{jk})_{jk}$, is an asymptotically least favorable prior associated with $\mathcal{WB}_{s,p,q}(C)$ and the $\mathcal{B}_{s',p',p'}$-loss, if following conditions are satisfied.

- π_ε maximizes the Bayes risk on an appropriate class of probability measures (see section 5.1 for the class of probability measures that is naturally introduced for the issues of this paper).
- The Bayes risk of π_ε, denoted $B(\pi_\varepsilon)$, must verify:

  $$C_1B(\pi_\varepsilon)\leq\tilde{R}_\varepsilon\leq C_2B(\pi_\varepsilon),$$

  where C_1 and C_2 are positive constants depending only on s,p,q,s',p'. We recall that

  $$B(\pi_\varepsilon)=\inf_{\hat\beta}\ \mathbb{E}_{\pi_\varepsilon}\ \mathbb{E}_\beta\|\hat\beta-\beta\|^{p'}_{\mathcal{B}_{s',p',p'}}.$$

- As ε tends to 0, $\mathbb{P}_{\pi_\varepsilon}(\beta\notin\mathcal{WB}_{s,p,q}(C))$ goes to 0 with an exponential rate of convergence.

Similarly, by replacing $\mathcal{WB}_{s,p,q}(C)$ with $\mathcal{B}_{s,p,q}(C)$, we obtain the definition of asymptotically least favorable priors for strong Besov balls. We easily see that the last two conditions introduced previously ensure that the 'typical enemies' of weak Besov balls $\mathcal{WB}_{s,p,q}(C)$ are well represented by simulations of asymptotically least favorable priors associated with theses spaces. In particular, the exponential rate for the support property is essential to provide a statistical sense for asymptotically least favorable priors. We can note that priors with similar properties have also been used by Johnstone (1994) who also used the terminology of 'asymptotically least favorable priors'.

4 Results, discussions and simulations

In this section, we give results concerning the minimax risk for $\mathcal{WB}_{s,p,q}(C)$. For this purpose, we need to introduce the following two distinct zones denoted respectively as the regular and the critical zones:

$$\mathcal{R} = \left\{(s,p,q) \in (0,+\infty)^3 : \quad p' > p,\ p\left(s+\frac{1}{2}\right) > p'\left(s'+\frac{1}{2}\right)\right\} \cup \left\{(s,p,q) \in (0,+\infty)^3 : \quad p' \leq p\right\},$$

$$\mathcal{C} = \left\{(s,p,q) \in (0,+\infty)^3 : \quad p' > p,\ p\left(s+\frac{1}{2}\right) = p'\left(s'+\frac{1}{2}\right)\right\}.$$

The logarithmic zone

$$\mathcal{L} = \left\{(s,p,q) \in (0,+\infty)^3 : \quad p' > p,\ p\left(s+\frac{1}{2}\right) < p'\left(s'+\frac{1}{2}\right)\right\}$$

is very different from the other ones and will not be considered in this paper. In the critical case, we need a minimal hypothesis on the smoothness of f to control the size of the β_{jk}'s at high levels. That is the reason why we suppose from now on until the end that in addition, in the critical case, f lies in $\mathcal{B}_{\eta,p',\infty}(C)$ (with $\eta > s'$ but $\eta - s'$ eventually very small). So, let us set

$$\Theta = \mathcal{WB}_{s,p,q}(C) \quad \text{ on } \mathcal{R},$$

$$\Theta = \mathcal{WB}_{s,p,q}(C) \cap \mathcal{B}_{\eta,p',\infty}(C) \quad \text{ on } \mathcal{C},$$

and the minimax risk we consider is from now on

$$R_\varepsilon = \inf_{\hat\beta} \sup_{\beta\in\Theta} \mathbb{E}_\beta \|\hat\beta - \beta\|^{p'}_{\mathcal{B}_{s',p',p'}}. \tag{2}$$

4.1 Notations and technical tools

In this section, we introduce some notations that will be useful in the following. For this purpose, we suppose that we are given $0 < s,p,q,C < \infty$. If h_1 and h_2 denote two positive functions of ε, $h_1(\varepsilon) \overset{\varepsilon\to 0}{\sim} h_2(\varepsilon)$ means that $\lim_{\varepsilon\to 0}\frac{h_1(\varepsilon)}{h_2(\varepsilon)} = 1$. The notation $h_1(\varepsilon) \approx h_2(\varepsilon)$ means that there exist positive constants A and B, depending only on s,p,q,s',p' such that

$$\forall\, \varepsilon > 0, \quad A\, h_2(\varepsilon) < h_1(\varepsilon) \leq B\, h_2(\varepsilon),$$

and $h_1(\varepsilon) \simeq h_2(\varepsilon)$, means that $A = 1$, $B = 2$. For any x, δ_x denotes Dirac measure at the point x and for any real number y, $\lfloor y \rfloor$ denotes the greatest integer smaller or equal to y.

Let us fix κ a real number belonging to $\left(1, \frac{2p'(s'+\frac{1}{2})}{2p'(s'+\frac{1}{2})-1}\right)$. We also consider a function of ε, γ_ε such that $\gamma_\varepsilon - 1$ is positive and goes to 0 as ε tends to 0 but not faster than a logarithmic rate. Finally, we set for any $\varepsilon > 0$, $(j_*, j^*) \in \mathbb{N}^2$ such that

$$2^{j_*} \simeq \left(\frac{C}{\varepsilon}\right)^{\frac{1}{s+\frac{1}{2}}}, \quad 2^{j^*} \simeq \left(\frac{C}{\varepsilon}\right)^{\frac{\kappa}{s+\frac{1}{2}}}.$$

Now, we introduce a sequence of non negative real numbers $(\alpha_j)_j$ and two integers j_1 and j_2 depending on the zone. When $j < j_*$, we set $\alpha_j = V_\varepsilon 2^{-j(s+\frac{1}{2})}$, where V_ε is defined subsequently.

- If (s,p,q) lies in $\mathcal{C}$ and if $q > p$, we set $j_1 = j_*$, $j_2 = j^*$ and for $j \geq j_*$, $\alpha_j = V_\varepsilon 2^{-j(s+\frac{1}{2})} j$ if $j \in \{j_1, \ldots, j_2\}$ and $\alpha_j = 0$ otherwise. Then, V_ε is the real number independent of j such that

$$\sum_j 2^{jq(s+\frac{1}{2})} \alpha_j^q = \left(\frac{C}{\gamma_\varepsilon \varepsilon}\right)^q. \tag{3}$$

- If (s,p,q) lies in $\mathcal{C}$ and if $q \leq p$, we set $j_1 = j_2 = j^*$, and for $j \geq j_*$, $\alpha_j = 0$ if $j \neq j_1$, and α_{j_1} is such that (3) is satisfied.

- If (s,p,q) lies in $\mathcal{R}$, we set $j_1 = j_2 = j_*$ and for $j \geq j_*$, $\alpha_j = 0$ if $j \neq j_1$, and α_{j_1} is such that (3) is satisfied.

With each sequence $(\alpha_j)_j$, we associate two sequences $(c_j)_j$ and $(\mu_j)_j$ such that $\mu_j^p \phi(\mu_j + c_j) = \alpha_j^p \, \phi(c_j)$, where ϕ denotes the standard Gaussian density function. Furthermore, if $(\alpha_j)_j$ tends to 0, we require in addition that $(c_j)_j$ and $(\mu_j)_j$ tend to $+\infty$ with $c_j = o(\mu_j)$, which yields that $\mu_j^2 \sim -2\log\alpha_j^p$.

4.2 Main results and comments

By using notations defined in section 4.1, we have the following main results:

Theorem 1. *Let us fix $0 < s, p, q, C < \infty$, such that $s > s'$.*

- *If $\vartheta = \frac{s-s'}{s+\frac{1}{2}}$, and*

$$\Psi(C,\varepsilon) = \begin{cases} C^{p'(1-\vartheta)} \varepsilon^{p'\vartheta} & \text{on } \mathcal{R} \\ C^{p'(1-\vartheta)} \varepsilon^{p'\vartheta} \left(\log\left(\frac{C}{\varepsilon}\right)\right)^{\frac{p'\vartheta}{2} + (1-\frac{p}{q})_+} & \text{on } \mathcal{C}, \end{cases}$$

we have

$$C_1 \leq \liminf_{\varepsilon \to 0} R_\varepsilon \Psi(C,\varepsilon)^{-1} \leq \limsup_{\varepsilon \to 0} R_\varepsilon \Psi(C,\varepsilon)^{-1} \leq C_2,$$

where C_1 and C_2 are positive constants depending on s, p, q, s', p'. On $\mathcal{C}$, C_2 also depends on η.

- *If we set π_ε as the distribution of a sequence of independent variables $(\beta_{jk})_{j,k}$ such that the distribution of β_{jk} is symmetric about 0 and*

$$|\beta_{jk}| = \begin{cases} \varepsilon \alpha_j & \text{if } j < j_* \\ \varepsilon \min(\alpha_j X_{jk}, \mu_j) & \text{otherwise}, \end{cases}$$

where X_{jk} is a Pareto(p)-variable, then π_ε is an asymptotically least favorable prior for $\mathcal{WB}_{s,p,q}(C)$. Furthermore, $\mathbb{P}_{\pi_\varepsilon}(\beta \in \mathcal{B}_{s,p,q}(C)) \overset{\varepsilon\to 0}{\longrightarrow} 0$.

Remark 1. *The upper bound of the minimax risk is obtained by pointing out a minimax thresholding rule defined in (9) of the form*

$$\hat{f}_\varepsilon^* = \sum_j \sum_k sign(y_{jk}) \left(|y_{jk}| - \varepsilon\lambda_j\right)_+ \psi_{jk},$$

where $\lambda_j = 0$ if $j < j_*$ and $\lambda_j = (-2\log\tilde{\alpha}_j^p)^{\frac{1}{2}}$ otherwise, where $\tilde{\alpha}_j$ is a dilation parameter of the prior model based on Pareto distributions used in section 5.2 and that strongly depends on s, p, q, p', s', η, C and ε. See section 5.2 for a precise definition of $\tilde{\alpha}_j$.

We can now compare strong and weak Besov spaces statistically. On the one hand, the previous theorem and Theorem 1 of Donoho *et al.* (1996) enable us to conclude that the rates of convergence for $\mathcal{B}_{s,p,q}(C)$ and $\mathcal{WB}_{s,p,q}(C)$ on regular and critical zones are the same. On the other hand, Theorem 1 shows that π_ε is an asymptotically least favorable prior for $\mathcal{WB}_{s,p,q}(C)$ but π_ε is not an asymptotically least favorable prior for $\mathcal{B}_{s,p,q}(C)$. Let us note that the construction of π_ε uses Pareto distributions that are dense. We note that this is not necessarily the case for the asymptotically least favorable priors of $\mathcal{B}_{s,p,q}(C)$ exploited by Johnstone (1994). If for Pinsker's case ($p = q = 2$), Johnstone uses Gaussian distributions (so, they are dense), when $p < 2$, the least favorable priors are based on three or two point distributions (for the coarsest levels, the prior distributions are dense, whereas they are sparse for high levels, with a few wavelet coefficients carrying all the energy). Let us note that we can certainly build a lot of asymptotically least favorable priors for $\mathcal{WB}_{s,p,q}(C)$ (in particular, asymptotically least favorable priors for $\mathcal{B}_{s,p,q}(C)$ are also asymptotically least favorable priors for $\mathcal{WB}_{s,p,q}(C)$) but priors based on Pareto distributions naturally appear when we investigate the upper bound of the risk.

4.3 Representations of the 'typical enemies' of weak Besov balls

The goal of this subsection is to provide a good representation of the worst functions to be estimated and belonging to weak Besov balls $\mathcal{WB}_{s,p,q}(C)$. We use realizations of asymptotically least favorable priors for the regular and the critical cases and for different values of s, p and q.

Before presenting our approach, let us briefly describe the discrete wavelet transform (DWT) that is the natural algorithm associated with the multiresolution framework introduced previously (see for instance Antoniadis (1994), Johnstone (1994), Mallat (1998) or Abramovich *et al.* (1998)). Indeed, given the approximation coefficients of a signal f at a resolution level j, the DWT provides the approximation coefficients and the detail coefficients of f at level $j-1$. This wavelet decomposition (as far as the reconstruction) is easily calculated by using discrete convolutions with the appropriate wavelet filters and can be performed as many times as desired. This cascade algorithm has to be initialized. For this purpose, let us assume we are given the following n-sample of a 1-periodic signal f:

$$\mathbf{f} = \left\{ f\left(\frac{i}{n}\right): \quad 1 \le i \le n = 2^N \right\}.$$

By supposing that the wavelet basis is based on a scaling function ϕ that has a high number of vanishing moments (for instance a coiflet), Lemma 3.1 of Antoniadis (1994) proves that if, in addition, f is regular enough, we have the following approximation:

$$f\left(\frac{i}{n}\right) \asymp \sqrt{n} \times \int_0^1 f(t) 2^{N/2} \phi(2^N t - i) dt, \quad 1 \le i \le n.$$

Starting with $\mathbf{f}$, the decomposition step described above is performed N times. Thus, we obtain an algorithm that is invertible and whose outputs consist of approximation coefficients (the remaining ones) and all the detail coefficients that were accumulated along the way. All these coefficients that can be denoted d_{jk}, $-1 \le j \le N-1$, $k \in \mathcal{I}_j$ satisfy $d_{jk} \asymp \sqrt{n} \times \beta_{jk}$. Let us note that both the DWT and the inverse DWT are available for instance by using the wavelet TOOLBOX of MATLAB.

So, our approach is logically the following. First, we set $C = 1$, $n = 2^{12}$, the classical calibration $\varepsilon = n^{-1/2}$ and $\gamma_\varepsilon = 1$. Using Theorem 1 (and its notations), to represent 'typical enemies' of weak Besov balls, we simulate coefficients d_{jk} with the following prior:

$$d_{jk} \sim \begin{cases} F_j & j \ge j_* \\ \frac{1}{2}(\delta_{\alpha_j} + \delta_{-\alpha_j}) & j < j_*, \end{cases} \tag{4}$$

where $F_j = \frac{1}{2}(F_j^+ + F_j^-)$, F_j^+ is the distribution of $\min(\alpha_j X_{jk}, \mu_j)$ and F_j^- is the reflection of F_j^+ about 0. To complete the definition of this prior model, and according to section 4.1, we set $j_* \in \mathbb{N}$ such that $2^{j_*} \simeq \left(\frac{C}{\varepsilon}\right)^{\frac{1}{s+\frac{1}{2}}}$, and for the critical case, we set $j^* \in \mathbb{N}$ such that $2^{j^*} \simeq \left(\frac{C}{\varepsilon}\right)^{\frac{\kappa}{s+\frac{1}{2}}}$, where

$$\kappa = \frac{2p'(s' + \frac{1}{2})}{2p'(s' + \frac{1}{2}) - 1} = \frac{2p(s + \frac{1}{2})}{2p(s + \frac{1}{2}) - 1},$$

since on $\mathcal{C}$, $p'(s'+\frac{1}{2}) = p(s+\frac{1}{2})$. This allows one to define the integers j_1 and j_2, the sequence $(\alpha_j)_j$ as in section 4.1 and we take $\forall\ j \in \{j_1, \dots, j_2\}$, $\mu_j = \left(-2\log(\alpha_j^p)\right)^{\frac{1}{2}}$. Naturally, for the reconstruction, we use the wavelet filters associated with coiflets of order 5. To deal with the boundary problems that naturally appear when we consider compactly supported signals, we use the periodized form of the inverse DWT. But of course, this choice does not alter the qualitative phenomena we wish to present here (see Mallat (1998) for instance). Thus, we obtain a n-sample of a periodic signal f that is a good representation of the worst functions to be estimated and belonging to $\mathcal{WB}_{s,p,q}(C)$.

Figure 1 shows sample paths we obtain. Of course, realizations for the regular case are more regular than realizations for the critical case. This fact is illustrated by the comparison between (c) and (a). As expected, we note that realizations are more regular when s is great (compare (a) and (f)) or when q is small (compare (a) and (b) or (d) and (e)), even if q is less influential. Finally, when p decreases, the realizations are less regular and in particular, we also note that the size of the peaks increases (compare (a) and (d) or (b) and (e)). Figures (a) and (b) show that for small values of p, we obtain very high peaks. Thus, we verify that the parameter p controls the proportion of large coefficients and consequently, measures the sparsity of the signal.

Figure 1 here

4.4 How to apply $\hat{f}_\varepsilon^*$ for denoising discrete data?

In this paper, we have pointed out a minimax thresholding rule $\hat{f}_\varepsilon^*$ (see Remark 1 and (9)). Many problems arise to study this estimator from a practical point of view, since $\hat{f}_\varepsilon^*$ is not adaptive. These problems have been tackled by Rivoirard (2004) whose main ideas are recalled now. In the standard non parametric regression model reduced to the following one by applying the DWT

$$Y_{jk} = d_{jk} + \sigma Z_{jk}, \quad -1 \le j \le N-1, \quad k \in \mathcal{I}_j, \quad Z_{jk} \overset{iid}{\sim} \mathcal{N}(0,1),$$

the signal to be estimated is assumed to belong to the class of the worst functions of $\mathcal{WB}_{s,p,q}$ to be estimated, where $\mathcal{WB}_{s,p,q}$ is an unknown weak Besov space. Then, a prior model is fixed on its discrete wavelet coefficients (the d_{jk}'s) that is very close to (4). To estimate each d_{jk}, Rivoirard (2004) proposes an estimator of the form $d_{jk}^*(Y_{jk}) = \mathrm{sign}(Y_{jk})(|Y_{jk}| - \sigma\lambda_j)_+$, where σ is supposed to be known and $\lambda_j = (\max(0, -2\log\alpha_j^p))^{\frac{1}{2}}$ (as in (4), α_j is the dilation parameter in the prior model). Of course, this procedure is not exactly equal to $\hat{f}_\varepsilon^*$. However, it takes into account the main characteristics of $\hat{f}_\varepsilon^*$ (so, it is close to $\hat{f}_\varepsilon^*$) and necessary adaptations (see Rivoirard (2004) for more justifications). Then, instead of estimating the parameters s, p and q, Rivoirard (2004) proposes data-driven methods to estimate p and each α_j. We do not describe these methods in detail, which would be too tedious, but we mention that the number of the Y_{jk}'s that are greater than a chosen threshold is used, as it could be expected in a framework based on weak Besov spaces. The resulting procedure is called ParetoThresh. The results obtained by ParetoThresh are quite satisfactory when compared with other efficient wavelet thresholding algorithms. Indeed, ParetoThresh outperforms VisuShrink and SureShrink of Donoho and Johnstone (1994a) and its performances are quite similar to the procedure 'BayesThresh' proposed by Abramovich *et al.* (1998) and 'BayesFactor' proposed by Vidakovic (1998). But unlike to them, ParetoThresh does not require a high computational time. See Rivoirard (2004) for a precise description of advantages and drawbacks of ParetoThresh.

5 Proof of Theorem 1

In this section, the notations K, K_1 and K_2, will keep designating all the positive constants depending only on s, p, q, s', p' we could need. From now on, we suppose for sake of simplicity and without loss of generality, that for any j, $|\mathcal{I}_j| = 2^j$.
For any j_η in $\{0, 1, \dots, +\infty\}$, we consider

$$\Theta^* = \{\beta : \quad \beta \in \Theta \text{ with } \beta_{jk} = 0,\ \forall\ (j \ge j_\eta, k \in \mathcal{I}_j)\},$$

and

$$R^*_\varepsilon = \inf_{\hat\beta} \sup_{\beta\in\Theta^*} \mathbb{E}_\beta \sum_{j<j_\eta} 2^{jp'(s'+\frac{1}{2}-\frac{1}{p'})} \sum_k |\hat\beta_{jk} - \beta_{jk}|^{p'}. \tag{5}$$

Using (2), (5) and the properties of $\mathcal{B}_{\eta,p',\infty}(C)$, we have

$$R^*_\varepsilon \le R_\varepsilon \le R^*_\varepsilon + \sup_{\beta\in\mathcal{B}_{\eta,p',\infty}(C)} \sum_{j\ge j_\eta} 2^{jp'(s'+\frac{1}{2}-\frac{1}{p'})} \sum_k |\beta_{jk}|^{p'} \le R^*_\varepsilon + K\, C^{p'} 2^{p'(s'-\eta)j_\eta}.$$

In the critical case, we choose j_η so that $j_\eta = \lfloor \frac{\vartheta}{\eta-s'} \log_2\left(\frac{C}{\varepsilon}\right)\rfloor$. In the regular case, we set $j_\eta = \infty$ and $R^*_\varepsilon = R_\varepsilon$. So, the first point of Theorem 1 will be proved as soon as we prove that

$$C_1 \le \liminf_{\varepsilon\to 0} R^*_\varepsilon \Psi(C,\varepsilon)^{-1} \le \limsup_{\varepsilon\to 0} R^*_\varepsilon \Psi(C,\varepsilon)^{-1} \le C_2.$$

For this purpose, we use the minimax Bayes method that is often exploited in the literature (see the references cited in Introduction). That is the reason why most of the details are omitted in the next section.

5.1 Minimax Bayes method

When $q \ge p$, let us consider M, the natural set of probability measures associated with $\mathcal{WB}_{s,p,q}(C)$:

$$M = \left\{ \pi : \quad \sum_{j<j_\eta} 2^{jq(s+\frac{1}{2}-\frac{1}{p})} [\mathbb{E}_\pi N(j,\lambda)]^{\frac{q}{p}} \le \left(\frac{C}{\lambda}\right)^q \quad \forall\, \lambda > 0 \right\},$$

that is convex and compact for the Prohorov metric, and

$$B(M,\varepsilon) = \inf_{\hat\beta} \sup_{\pi\in M} \mathbb{E}_\pi\ \mathbb{E}_\beta \sum_{j<j_\eta} 2^{jp'(s'+\frac{1}{2}-\frac{1}{p'})} \sum_k |\hat\beta_{jk} - \beta_{jk}|^{p'},$$

the minimax Bayes risk for M. Now, by applying the minimax theorem, we have

$$B(M,\varepsilon) = \sup_{\pi\in M} B(\pi),$$

where $B(\pi)$ is the Bayes risk of π. For each π in M, we construct $\tilde\pi$, the distribution of $\beta = (\beta_{jk})_{jk}$, such that under $\tilde\pi$, the β_{jk}'s are independent and for any $j \ge -1$ and any $k \in \mathcal{I}_j$, the distribution of β_{jk} is $\tilde\pi_j$, where

$$\tilde\pi_j = \frac{1}{2^j} \sum_{l\in\mathcal{I}_j} \pi_{jl},$$

and we set $\mathcal{M} = \{\tilde\pi : \pi \in M\}$. It is relevant to note that under a prior of $\mathcal{M}$, the Bayes estimator of β_{jk} depends only on y_{jk}.

For F a probability measure, we note $s_\varepsilon F$ the probability measure defined by $s_\varepsilon F(\mathcal{A}) = F(\varepsilon^{-1}\mathcal{A})$. Using standard arguments, we get:

$$B(M,\varepsilon) = \varepsilon^{p'} \sup_{s_\varepsilon F\in\mathcal{M}} \inf_{(d_j)} \sum_{j<j_\eta} 2^{jp'(s'+\frac{1}{2})}\, \mathbb{E}_{F_j}\ \mathbb{E}_{u_{j1}} |d_j(x_{j1}) - u_{j1}|^{p'}, \tag{6}$$

where $x_{j1} = \varepsilon^{-1} y_{j1}$ and $u_{j1} = \varepsilon^{-1}\beta_{j1}$ whose distribution is denoted F_j. Then, the upper bound of R^*_ε relies on $B(M,\varepsilon)$. Indeed, we obviously have $R^*_\varepsilon \le B(M,\varepsilon)$. In section 5.2, to get the upper bound of $B(M,\varepsilon)$, we exploit the method exhibited by Johnstone (1994) and Donoho and Johnstone (1994b) based on a Bayes risk restricted to soft thresholding rules that enables us to obtain minimax thresholding rules. Since arguments are given by Johnstone (1994) and Donoho and Johnstone (1994b), many details are omitted. Furthermore, we naturally rely on the worst prior pointed out in section 5.2 to build the asymptotically least favorable prior π_ε.

On $\mathcal{R}$, the lower bound of R^*_ε is provided by the lower bound of the minimax risk for $\mathcal{B}_{s,p,q}(C)$ included into $\mathcal{WB}_{s,p,q}(C)$ (see Donoho *et al.* (1996)). On $\mathcal{C}$, we cannot use this argument, since $\Theta = \mathcal{WB}_{s,p,q}(C) \cap \mathcal{B}_{\eta,p',\infty}(C)$ does not contain $\mathcal{B}_{s,p,q}(C)$. The lower bound of R^*_ε is provided by the lower bound of the Bayes risk of π_ε whose support is asymptotically included into Θ when η is close to s'. Since the proof is standard (see Johnstone (1994)), it is omitted.

Remark 2. *When $q < p$, we use the same arguments but with the closure of the generalized convex hull of M instead of M.*

5.2 Upper bound

Our first goal is to prove that

$$B(M, \varepsilon) \leq K\Psi(C, \varepsilon), \tag{7}$$

where K may depend on η on $\mathcal{C}$. We omit the case $q < p$ that will be handled by the case $q = p$ (let us note that $\mathcal{WB}_{s,p,q}(C) \subset \mathcal{WB}_{s,p,p}(C)$ if $q < p$). To reach our goal, we use the risk associated with the soft thresholding estimator

$$d_\lambda(x) = \mathrm{sign}(x)\ (|x| - \lambda)_+$$

denoted by

$$r(\lambda, \xi) = \mathbb{E}_\xi |d_\lambda(x) - \xi|^{p'},$$

where $x \sim \mathcal{N}(\xi, 1)$. We recall the properties of $\xi \longrightarrow r(\lambda, \xi)$ we use in the following:

Proposition 1. *For any $\lambda > 0$, if ϕ denotes the standard Gaussian density function and $\Phi(y) = \int_{-\infty}^{y} \phi(z)dz$ is the standard Gaussian cumulative distribution function,*

1. *$\xi \to r(\lambda, \xi)$ is symmetric about 0.*
2. *$\forall\ \xi > 0$, $\frac{\partial r}{\partial \xi}(\lambda, \xi) = p'\ \xi^{p'-1}\Phi([-\lambda - \xi, \lambda - \xi])$ and $\xi \longrightarrow r(\lambda, \xi)$ is a strictly increasing function.*
3. *$\lim_{\xi \to +\infty} r(\lambda, \xi) = \int_{-\infty}^{+\infty} |z|^{p'}\phi(z + \lambda)dz$.*
4. *$K_1 \exp(-\frac{\lambda^2}{2})\lambda^{-p'-1} \leq r(\lambda, 0) \leq K_2 \exp(-\frac{\lambda^2}{2})\lambda^{-p'-1}$.*

The proof of this proposition is omitted since it is just an extension of the classical case $p' = 2$ (see Donoho and Johnstone (1994a)).
By (6), we have

$$B(M, \varepsilon) \leq \varepsilon^{p'} \sup_{s_\varepsilon F \in \mathcal{M}} \inf_{(\lambda_j)} \sum_{j < j_\eta} 2^{jp'(s' + \frac{1}{2})} \mathbb{E}_{F_j} r(\lambda_j, \xi). \tag{8}$$

The condition $s_\varepsilon F \in \mathcal{M}$ means that

$$\sum_{j < j_\eta} 2^{jq(s+\frac{1}{2})} \left(\mathbb{E}_{F_j} (\mathbb{1}_{\ |\xi| > \lambda \varepsilon^{-1}})\right)^{\frac{q}{p}} \leq \left(\frac{C}{\lambda}\right)^q, \quad \forall\ \lambda > 0.$$

Since $\xi \longrightarrow r(\lambda_j, \xi)$ is symmetric about 0, we assume without loss of generality that F_j is supported by $\mathbb{R}_+$. And using the second point of Proposition 1, we prove that the supremum for (8) is reached with F_j such that it is absolutely continuous with respect to the Lebesgue measure and has the following density:

$$f_{\alpha_j}(\xi) = p\mathbb{1}_{\ \xi\ \geq\ \alpha_j}\ \alpha_j^p\ \xi^{-p-1}$$

such that

$$\sum_{j<j_\eta} \alpha_j^q \, 2^{jq(s+\frac{1}{2})} \le \left(\frac{C}{\varepsilon}\right)^q .$$

And we can conclude by using (8) that

$$B(M,\varepsilon) \le \varepsilon^{p'} \sup \left\{ \sum_{j<j_\eta} 2^{jp'(s'+\frac{1}{2})} \inf_{\lambda_j} \mathbb{E}_{f_{\alpha_j}} r(\lambda_j,\xi) : \quad \alpha_j \ge 0, \quad \sum_{j<j_\eta} \alpha_j^q \, 2^{jq(s+\frac{1}{2})} \le \left(\frac{C}{\varepsilon}\right)^q \right\}.$$

Let $\tilde\alpha = (\tilde\alpha_j)_{-1\le j<j_\eta}$ be the point where the supremum is reached. We have

$$\begin{aligned}
\sum_{j<j_\eta} 2^{jp'(s'+\frac{1}{2})} \inf_{\lambda_j} \mathbb{E}_{f_{\tilde\alpha_j}} r(\lambda_j,\xi) &= \sum_{j<j_*} 2^{jp'(s'+\frac{1}{2})} \inf_{\lambda_j} \mathbb{E}_{f_{\tilde\alpha_j}} r(\lambda_j,\xi) + \sum_{j_*\le j<j_\eta} 2^{jp'(s'+\frac{1}{2})} \inf_{\lambda_j} \mathbb{E}_{f_{\tilde\alpha_j}} r(\lambda_j,\xi) \\
&:= S_{\varepsilon,1} + S_{\varepsilon,2},
\end{aligned}$$

where j_* is defined in section 4.1. By using the second and the third points of Proposition 1 and the value of ϑ given in Theorem 1,

$$S_{\varepsilon,1} \le \sum_{j<j_*} 2^{jp'(s'+\frac{1}{2})} \inf_{\lambda_j} \int_{-\infty}^{+\infty} |z|^{p'} \phi(z+\lambda_j) dz \int_{\tilde\alpha_j}^{+\infty} \xi^{-p-1} \, p \, \tilde\alpha_j^p d\xi \le K \left(\frac{C}{\varepsilon}\right)^{p'(1-\vartheta)} .$$

To get an upper bound for $S_{\varepsilon,2}$, we need to evaluate the Bayes threshold risk $\mathbb{E}_{f_{\tilde\alpha_j}} r(\lambda_j,\xi)$ when $\tilde\alpha_j$ tends to 0. We have the following theorem, which generalizes the case $p'=2$ and $p<2$ investigated by Johnstone (1994):

Theorem 2. *Let us suppose that we are given $1 \le p' < \infty$ and $p>0$. When α tends to 0 and for any threshold $\lambda \ge (-2\,\log \alpha^{p\wedge p'})^{\frac{1}{2}}$, we have $\mathbb{E}_{f_\alpha} r(\lambda,\xi) \sim J(\alpha,\lambda)$, where*

$$J(\alpha,\lambda) = \begin{cases} \frac{p'}{p'-p}\, \alpha^p \, \lambda^{p'-p} & \text{if } p<p', \\ p\, \alpha^p \, \log\left(\frac{\lambda}{\alpha}\right) & \text{if } p=p', \\ \frac{p}{p-p'}\, \alpha^{p'} & \text{if } p>p'. \end{cases}$$

Proof: Using Proposition 1, we have

$$\begin{aligned}
\mathbb{E}_{f_\alpha} r(\lambda,\xi) &= r(\lambda,\alpha) + p'\alpha^p \int_\alpha^{+\infty} \xi^{p'-p-1} \Phi([-\lambda-\xi, \lambda-\xi]) d\xi \\
&= r(\lambda,0) + \int_0^\alpha p' \, \xi^{p'-1} \Phi([-\lambda-\xi,\lambda-\xi]) d\xi + p'\alpha^p \int_\alpha^{+\infty} \xi^{p'-p-1} \Phi([-\lambda-\xi,\lambda-\xi]) d\xi.
\end{aligned}$$

When α tends to 0 and λ to $+\infty$, we prove that

$$\int_\alpha^{+\infty} \xi^{p'-p-1} \Phi([-\lambda-\xi,\lambda-\xi]) d\xi \sim I(\alpha,\lambda),$$

where

$$I(\alpha,\lambda) = \begin{cases} \frac{1}{p'-p} \lambda^{p'-p} & \text{if } p<p', \\ \log\left(\frac{\lambda}{\alpha}\right) & \text{if } p=p', \\ \frac{1}{p-p'} \alpha^{p'-p} & \text{if } p>p'. \end{cases}$$

Finally, by using again Proposition 1, we easily obtain the required result. $\square$

Noting that $\forall\, j_* \le j < j_\eta, \quad \tilde\alpha_j < 1$, we have:

$$S_{\varepsilon,2} \le K \sup \left\{ \sum_{j=j_*}^{j_\eta-1} 2^{jp'(s'+\frac{1}{2})} J\left(\tilde\alpha_j, (-2\log(\tilde\alpha_j^p))^{\frac{1}{2}}\right) : \quad \tilde\alpha_j \ge 0, \quad \sum_{j=j_*}^{j_\eta-1} \tilde\alpha_j^q \, 2^{jq(s+\frac{1}{2})} = \left(\frac{C}{\varepsilon}\right)^q \right\},$$

with the notations of Theorem 2. Computations of this supremum enables us to obtain (7). On $\mathcal{C}\cap\{q>p\}$, the supremum is reached at the point whose coordinates $\tilde\alpha_j$ are equivalent to $T2^{-j(s+\frac12)}j^a$, where $a=\frac{(p'-p)}{2(q-p)}$ and T is a constant depending on ε and η. If $(s,p,q)\in\mathcal{R}\cup(\mathcal{C}\cap\{q=p\})$, all the coordinates equal zero except one: $\tilde\alpha_{j_*}$ on $\mathcal{R}$ and $\tilde\alpha_{j_*}=\left(\frac{C}{\varepsilon}\right)2^{-j_*(s+\frac12)}$, $\tilde\alpha_{j_\eta-1}$ on $\mathcal{C}$ and if $q=p$ and $\tilde\alpha_{j_\eta-1}=\left(\frac{C}{\varepsilon}\right)2^{-(j_\eta-1)(s+\frac12)}$. Finally, the thresholding rule

$$\hat f^*_\varepsilon=\sum_j\sum_k \mathrm{sign}(y_{jk})\left(|y_{jk}|-\varepsilon\lambda_j\right)_+\psi_{jk}, \tag{9}$$

where $\lambda_j=0$ if $j<j_*$ and $\lambda_j=(-2\log\tilde\alpha_j^p)^{\frac12}$ otherwise and $\tilde\alpha_j$ is defined as before, attains the minimax rate up to constants. □

5.3 The prior π_ε is an asymptotically least favorable prior for $\mathcal{WB}_{s,p,q}(C)$

We give asymptotic values of $B(\pi_\varepsilon)$ by proving the following proposition.

Proposition 2. *π_ε maximizes the Bayes risk on M and we have*

$$B(\pi_\varepsilon)\approx\Psi(C,\varepsilon).$$

Proof: Let us recall that $\eta>s'$, but $\eta-s'$ is small. So, we can assume that $j_2\le j_\eta$.
On the one hand, we easily show that $\pi_\varepsilon\in M$. So, $B(\pi_\varepsilon)\le K\Psi(C,\varepsilon)$.
On the other hand, we have:

$$B(\pi_\varepsilon)\;=\;\inf_{\hat\beta}\sum_{j=-1}^{j_2}2^{jp'(s'+\frac12-\frac1{p'})}\sum_k\mathbb{E}_{\pi_\varepsilon}\,\mathbb{E}_{\beta_{jk}}|\hat\beta_{jk}-\beta_{jk}|^{p'}\ge\varepsilon^{p'}\sum_{j=j_1}^{j_2}2^{jp'(s'+\frac12)}\,b(\alpha_j,p'),$$

where $\forall\, j\in\{j_1,\dots,j_2\}$, $b(\alpha_j,p')$ is the univariate Bayes risk for F_j, the distribution of $\varepsilon^{-1}\beta_{jk}$, and for the $L_{p'}$-loss:

$$b(\alpha_j,p')=\inf_d\,\mathbb{E}_{F_j}\int|d(x)-\xi|^{p'}\phi(x-\xi)dx.$$

When $p'=1$, the Bayesian estimator $d(x)$ is easily available since it is the median of the posterior distribution. We have the following result proved in Appendix:

Theorem 3. *If $p'=1$, when α tends to 0, $b(\alpha,1)\sim L(\alpha)$, where*

$$L(\alpha)=\begin{cases}\frac{1}{1-p}\,\alpha^p(-2\log\alpha^p)^{\frac{1-p}{2}} & \text{if } p<1,\\ \alpha\log\left(\frac1\alpha\right) & \text{if } p=1,\\ \frac{p}{p-1}\,\alpha & \text{if } p>1.\end{cases}$$

When p' is arbitrary and $p<p'$, when α tends to 0,

$$L(\alpha)=\frac{p'}{p'-p}\,\alpha^p(-2\log\alpha^p)^{\frac{p'-p}{2}}.$$

When $p<p'$, straightforward computations show that $B(\pi_\varepsilon)\ge K\Psi(C,\varepsilon)$. To get this lower bound when $p\ge p'$, we use in addition the Jensen inequality, which shows that:

$$b(\alpha_j,p')\;\;\ge\;\;b(\alpha_j,1)^{p'}.$$

□

We show now the support property for $\mathcal{WB}_{s,p,q}(C)$. We only deal with the case $p'>p$, $p(s+\frac12)=p'(s'+\frac12)$ and $q>p$. The other cases follow easily from the same arguments.

Proposition 3. $\mathbb{P}_{\pi_\varepsilon}(\beta \in \mathcal{WB}_{s,p,q}(C)) \overset{\varepsilon \longrightarrow 0}{\longrightarrow} 1$ *with an exponential rate.*

Proof: It will be a consequence of the asymptotic evaluation of $(\mathbb{P}_{\pi_\varepsilon}(A_j))_{j_1 \le j \le j_2}$, where

$$A_j = \bigcap_{\lambda > 0} \left\{ \frac{1}{2^j} \sum_k \mathbb{1}_{|\beta_{jk}| > \lambda} \le \left(\frac{\varepsilon}{\lambda}\right)^p \gamma_\varepsilon^p \, \alpha_j^p \right\},$$

given by the following lemma:

Lemma 1. *There exists $\tilde{\kappa} > 0$ such that, for ε small enough and for any j in $\{j_1, \dots, j_2\}$,*

$$\mathbb{P}_{\pi_\varepsilon}(A_j^c) \le 2 \exp\left(-2^{j+1} (\gamma_\varepsilon^p - 1)^2 \left(\frac{\alpha_j}{\mu_j}\right)^{2p} \right) \le 2 \exp\left(-(\gamma_\varepsilon^p - 1)^2 \left(\frac{C}{\varepsilon}\right)^{\tilde{\kappa}} \right).$$

Proof of Lemma 1: For the last inequality, we use the fact that κ belongs to $\left(1, \frac{2p'(s'+\frac{1}{2})}{2p'(s'+\frac{1}{2})-1}\right)$. We have:

$$A_j = \bigcap_{\left(\frac{\alpha_j}{\mu_j}\right)^p \le v \le 1} \left\{ \frac{1}{2^j} \sum_k \mathbb{1}_{\left(\frac{|\varepsilon^{-1}\beta_{jk}|}{\alpha_j}\right)^{-p} < v} \le \gamma_\varepsilon^p \, v \right\}.$$

As the distribution of $|\varepsilon^{-1}\beta_{jk}|$ is F_j^+, the distribution of $\left(\frac{|\varepsilon^{-1}\beta_{jk}|}{\alpha_j}\right)^{-p}$ is

$$\left(\frac{\alpha_j}{\mu_j}\right)^p \delta_{\left(\frac{\alpha_j}{\mu_j}\right)^p}(\xi) + \mathbb{1}_{\left(\frac{\alpha_j}{\mu_j}\right)^p \le \xi \le 1} d\xi.$$

Since we consider the values of v greater than $\left(\frac{\alpha_j}{\mu_j}\right)^p$, the distribution of $\left(\frac{|\varepsilon^{-1}\beta_{jk}|}{\alpha_j}\right)^{-p}$ matches that of U_{jk}, where the U_{jk}'s are independent uniform observations over $[0,1]$. Therefore,

$$\begin{aligned}
\mathbb{P}_{\pi_\varepsilon}(A_j^c) &\le \mathbb{P}\left(\sup_{\left(\frac{\alpha_j}{\mu_j}\right)^p \le v \le 1} |F_{2^j}^u(v) - v| > (\gamma_\varepsilon^p - 1) \left(\frac{\alpha_j}{\mu_j}\right)^p \right) \\
&\le \mathbb{P}\left(\sup_{v \in [0,1]} |F_{2^j}^u(v) - v| > (\gamma_\varepsilon^p - 1) \left(\frac{\alpha_j}{\mu_j}\right)^p \right),
\end{aligned}$$

where $F_{2^j}^u$ is the empirical distribution of the U_{jk}'s for k in $\mathcal{I}_j$. Hence, using the DKW inequality proved by Massart (1990),

$$\mathbb{P}_{\pi_\varepsilon}(A_j^c) \le 2 \exp\left(-2^{j+1} (\gamma_\varepsilon^p - 1)^2 \left(\frac{\alpha_j}{\mu_j}\right)^{2p} \right),$$

which ends the proof of the lemma. □

Since

$$\sum_{j \le j_2} 2^{jq(s+\frac{1}{2})} \, \alpha_j^q = \left(\frac{C}{\gamma_\varepsilon \varepsilon}\right)^q,$$

$$\mathbb{P}_{\pi_\varepsilon}\left[\sum_{j \le j_2} 2^{jq(s+\frac{1}{2}-\frac{1}{p})} \left(\sum_k \mathbb{1}_{|\beta_{jk}| > \lambda} \right)^{\frac{q}{p}} \le \left(\frac{C}{\lambda}\right)^q, \quad \forall\, \lambda > 0 \right] \ge \mathbb{P}_{\pi_\varepsilon}\left(A_{-1} \bigcap \left[\bigcap_{j=j_1}^{j_2} A_j \right] \right),$$

where

$$A_{-1} = \bigcap_{\lambda>0} \left\{ \sum_{j<j_1} 2^{jq(s+\frac{1}{2}-\frac{1}{p})} \left(\sum_k \mathbb{1}_{|\beta_{jk}|>\lambda} \right)^{\frac{q}{p}} \leq \gamma_\varepsilon^q \left(\frac{\varepsilon}{\lambda}\right)^q \sum_{j<j_1} 2^{jq(s+\frac{1}{2})} \alpha_j^q \right\}.$$

We have $\mathbb{P}_{\pi_\varepsilon}(A_{-1}) = 1$, and

$$\mathbb{P}_{\pi_\varepsilon} \left(\bigcap_{j=j_1}^{j_2} A_j \right) \quad \geq \quad \exp\left[-\exp\left(-\left(\frac{C}{\varepsilon}\right)^{\frac{\tilde{\kappa}}{2}} \right) \right],$$

for ε small enough. Therefore,

$$\mathbb{P}_{\pi_\varepsilon} \left[\sum_{j\leq j_2} 2^{jq(s+\frac{1}{2}-\frac{1}{p})} \left(\sum_k \mathbb{1}_{|\beta_{jk}|>\lambda} \right)^{\frac{q}{p}} \leq \left(\frac{C}{\lambda}\right)^q, \quad \forall\, \lambda > 0 \right] \geq \exp\left[-\exp\left(-\left(\frac{C}{\varepsilon}\right)^{\frac{\tilde{\kappa}}{2}} \right) \right].$$

Proposition 3 is proved. □

Then, it is easy to show that π_ε cannot be an asymptotically least favorable prior for $\mathcal{B}_{s,p,q}(C)$. Indeed, straightforward computations lead easily to the following proposition.

Proposition 4.

$$\mathbb{P}_{\pi_\varepsilon}\left(\beta \in \mathcal{B}_{s,p,q}(C)\right) \overset{\varepsilon\to 0}{\longrightarrow} 0.$$

6 Appendix: Proof of Theorem 3

As Johnstone (1994), we shall use the results of the following lemma that can be easily proved:

Lemma 2. *We can introduce functions of α, μ and c defined on $(0,+\infty)$, such that when α goes to 0, $\mu \longrightarrow +\infty$, $c \longrightarrow +\infty$, $c = o(\mu)$, and $\phi(\mu+c) = \left(\frac{\alpha}{\mu}\right)^p \phi(c)$, where ϕ denotes the standard Gaussian density function. These four conditions entail*

$$\mu(\alpha) \overset{\alpha\to 0}{\sim} (-2\log\alpha^p)^{\frac{1}{2}}.$$

We will need the following functions of α: $T = \mu + \frac{c}{2} \longrightarrow \infty$, and $\mu^- = \frac{1}{2T} \longrightarrow 0$. In the following, the notation $o_\alpha(1)$ will keep designating any function that is bounded by a function depending only on α and tending to 0 when α tends to 0. We evaluate

$$b(\alpha, p') = \inf_d \ \mathbb{E}_{f_\alpha} \int |d(x) - \xi|^{p'} \phi(x-\xi)dx,$$

with

$$f_\alpha^+(\xi)d\xi = p\alpha^p\, \xi^{-1-p} \mathbb{1}_{\alpha\leq\xi\leq\mu} d\xi + \left(\frac{\alpha}{\mu}\right)^p \delta_\mu(\xi),$$

and

$$f_\alpha(\xi) = \frac{f_\alpha^+(\xi) + f_\alpha^+(-\xi)}{2}.$$

To get the upper bounds of Theorem 3, we notice that, with $d(x) = 0$,

$$b(\alpha, p') \leq \mathbb{E}_{f_\alpha} \int |\xi|^{p'} \phi(x-\xi)dx,$$

which allows one to obtain the result. The following lemmas will be useful in the following:

Lemma 3. *When α tends to 0, for any $\mu^- \leq x \leq T$,*

$$\int_{2\mu^-}^{\mu} \phi(x-\xi)p\alpha^p\xi^{-p-1}d\xi \leq \phi(x)\exp(-\frac{1}{2}\mu c)\mu^p T^p = \phi(x)o_\alpha(1).$$

Proof:

$$\begin{aligned}\int_{2\mu^-}^{\mu} \phi(x-\xi)p\alpha^p\xi^{-p-1}d\xi &\leq \phi(x)\int_{2\mu^-}^{\mu}\exp(-\frac{\xi^2}{2}+T\xi)p\alpha^p\xi^{-1-p}d\xi \\ &\leq \phi(x)\exp(-\frac{\mu^2}{2}+T\mu)\alpha^pT^p \\ &\leq \phi(x)\exp(-\frac{1}{2}\mu c)\mu^pT^p.\end{aligned}$$

As $\exp(-\frac{1}{2}\mu c)\mu^pT^p$ tends to 0, the lemma is proved. □

Lemma 4. *When α tends to 0, for any $\mu^- \leq x \leq T$,*

$$\int_{\alpha}^{2\mu^-} \phi(x-\xi)p\alpha^p\xi^{-p-1}d\xi = \phi(x)(1+o_\alpha(1)).$$

Proof: To prove this lemma, we suppose that the random variable ξ has the density

$$g(z) = p(\alpha^{-p}-(2\mu^-)^{-p})^{-1}\mathbb{1}_{\alpha\leq z\leq 2\mu^-}z^{-p-1}.$$

We have

$$\int_{\alpha}^{2\mu^-} \phi(x-\xi)p\alpha^p\xi^{-p-1}d\xi \sim \phi(x)\,\mathbb{E}_g\exp\left(x\xi-\frac{\xi^2}{2}\right).$$

As $x\geq\mu^-$, for any ξ in $[\alpha,2\mu^-]$, $\quad x\xi-\frac{\xi^2}{2}\geq 0$, and

$$\left|\exp\left(x\xi-\frac{\xi^2}{2}\right)-1\right| \leq \exp\left(2T\mu^- - \frac{(2\mu^-)^2}{2}\right)\left(x\xi-\frac{\xi^2}{2}\right)\leq \exp(1)T\xi$$

For any $\varepsilon>0$, for $\alpha<\varepsilon\exp(-1)T^{-1}$,

$$\mathbb{P}_g\left(\left|\exp\left(x\xi-\frac{\xi^2}{2}\right)-1\right|>\varepsilon\right)\leq\mathbb{P}_g\left(\xi>\varepsilon\exp(-1)T^{-1}\right)\leq\alpha^p\varepsilon^{-p}\exp(p)T^p,$$

which tends to 0 when α tends to 0.
Therefore, as for any ξ in $[\alpha,2\mu^-]$, $\quad\left|\exp\left(x\xi-\frac{\xi^2}{2}\right)-1\right|\leq\exp(1)$,

$$\mathbb{E}_g\exp\left(x\xi-\frac{\xi^2}{2}\right)=1+o_\alpha(1).$$

The lemma is proved. □

In the following, we will consider the Bayesian estimators associated with f_α and f_α^+:

$$d(x)=\arg\inf_m\int_{\mathbb{R}}f_\alpha(\xi)\phi(x-\xi)|\xi-m|^{p'}d\xi \quad d^+(x)=\arg\inf_m\int_{\mathbb{R}}f_\alpha^+(\xi)\phi(x-\xi)|\xi-m|^{p'}d\xi.$$

Now, we prove that for any $\mu^-\leq x\leq T$, $d^+(x)\leq 3\mu^-$. For any m in $[3\mu^-,\mu]$, using Lemma 4,

$$\begin{aligned}\int_{\alpha}^{2\mu^-}\phi(x-\xi)p\alpha^p\xi^{-p-1}|\xi-m|^{p'}d\xi &\geq (\mu^-)^{p'}\int_{\alpha}^{2\mu^-}\phi(x-\xi)p\alpha^p\xi^{-p-1}d\xi \\ &= \phi(x)(\mu^-)^{p'}(1+o_\alpha(1)),\end{aligned}$$

using Lemma 3,

$$\begin{aligned}\int_{2\mu^-}^{\mu}\phi(x-\xi)p\alpha^p\xi^{-p-1}|\xi-m|^{p'}d\xi &\leq \mu^{p'}\int_{2\mu^-}^{\mu}\phi(x-\xi)p\alpha^p\xi^{-p-1}d\xi \\ &\leq \phi(x)\exp(-\frac{1}{2}\mu c)\mu^{p'+p}T^p,\end{aligned}$$

and

$$\left(\frac{\alpha}{\mu}\right)^p \int \phi(x-\xi)|\xi-m|^{p'}\delta_\mu(\xi) \;\le\; \phi(x)\alpha^p\mu^{p'-p}\exp(-\frac{\mu^2}{2}+T\mu) \;\le\; \phi(x)\mu^{p'}\exp(-\frac{1}{2}\mu c).$$

As $\exp(-\frac{1}{2}\mu c)\mu^{p'+p}T^p = o((\mu^-)^{p'})$, we have for any m in $[3\mu^-, \mu]$,

$$\begin{aligned}\int_{\mathbb{R}} f_\alpha^+(\xi)\phi(x-\xi)|\xi-m|^{p'}d\xi &\sim \int_\alpha^{2\mu^-} f_\alpha^+(\xi)\phi(x-\xi)|\xi-m|^{p'}d\xi \\ &\ge \int_\alpha^{2\mu^-} f_\alpha^+(\xi)\phi(x-\xi)|\xi-3\mu^-|^{p'}d\xi \\ &\sim \int_{\mathbb{R}} f_\alpha^+(\xi)\phi(x-\xi)|\xi-3\mu^-|^{p'}d\xi\end{aligned}$$

and for any x in $[\mu^-, T]$, $\alpha \le d^+(x) \le 3\mu^-$. With $f_\alpha^-(\xi) = f_\alpha^+(-\xi)$, for any ξ in $\mathbb{R}$,

$$\begin{aligned}b(\alpha,p') &= \mathbb{E}_{f_\alpha}\int |d(x)-\xi|^{p'}\phi(x-\xi)dx \\ &\ge \frac{1}{2}\inf_m \mathbb{E}_{f_\alpha^+}\int |m-\xi|^{p'}\phi(x-\xi)dx + \frac{1}{2}\inf_m \mathbb{E}_{f_\alpha^-}\int |m-\xi|^{p'}\phi(x-\xi)dx \\ &= \mathbb{E}_{f_\alpha^+}\int |d^+(x)-\xi|^{p'}\phi(x-\xi)dx \\ &\ge \int_{\frac{c}{2}}^{\mu} r(d^+,\xi)p\alpha^p\xi^{-1-p}d\xi + \left(\frac{\alpha}{\mu}\right)^p r(d^+,\mu),\end{aligned}$$

with

$$r(d^+,\xi) = \int |d^+(x)-\xi|^{p'}\phi(x-\xi)dx.$$

For any $\frac{c}{2} \le \xi \le \mu$,

$$r(d^+,\xi) \ge \int_{\mu^-}^{T} |d^+(x)-\xi|^{p'}\phi(x-\xi)dx \ge (\xi-3\mu^-)^{p'}s_\alpha,$$

with s_α that tends to 1 when α tends to 0. Finally, we have when $p < p'$,

$$\begin{aligned}b(\alpha,p') &\ge s_\alpha p\alpha^p \int_{\frac{c}{2}}^{\mu}(\xi-3\mu^-)^{p'}\xi^{-1-p}d\xi + s_\alpha\left(\frac{\alpha}{\mu}\right)^p(\mu-3\mu^-)^{p'} \\ &\ge \frac{p'}{p'-p}\,\alpha^p(-2\log\alpha^p)^{\frac{p'-p}{2}}(1+o_\alpha(1)).\end{aligned}$$

The second part of Theorem 3 is proved.
Now, we suppose that $p' = 1$ and we prove the following lemma:

Lemma 5. *For any x such that $|x|$ lies in $[\mu^-, T]$, $\alpha \le |d(x)| \le \alpha(1+o_\alpha(1))$.*

Proof: Without loss of generality, we suppose that $x > 0$. Since $p' = 1$, $d(x)$ is the median of the posterior distribution:

$$\int_{-\infty}^{d(x)} f_\alpha(\xi)\phi(x-\xi)d\xi = \frac{1}{2}\int_{-\infty}^{+\infty} f_\alpha(\xi)\phi(x-\xi)d\xi.$$

Since $x > 0$, we obviously have that $d(x) \ge \alpha$. From Lemma 3 and Lemma 4, it follows that

$$\int_\alpha^{\mu} f_\alpha^+(\xi)\phi(x-\xi)d\xi = \phi(x)(1+o_\alpha(1)).$$

By using similar arguments as previously, we have

$$\frac{1}{2}\left(\frac{\alpha}{\mu}\right)^p \phi(x+\mu) + \frac{1}{2}\left(\frac{\alpha}{\mu}\right)^p \phi(x-\mu) \leq \phi(x)\exp(-\frac{1}{2}\mu c) = \phi(x)o_\alpha(1).$$

As

$$\begin{aligned}\int_{-\mu}^{-\alpha} f_\alpha^+(-\xi)\phi(x-\xi)d\xi &= \phi(x)\int_\alpha^\mu p\alpha^p\xi^{-1-p}e^{-\frac{\xi^2}{2}-x\xi}d\xi \\ &= \phi(x)(1+o_\alpha(1)),\end{aligned}$$

we have

$$\int_{-\infty}^{+\infty} f_\alpha(\xi)\phi(x-\xi)d\xi = \phi(x)(1+o_\alpha(1)).$$

Since

$$\begin{aligned}\int_\alpha^{d(x)} f_\alpha(\xi)\phi(x-\xi)d\xi &\geq \int_\alpha^{d(x)} p\alpha^p\xi^{-1-p}d\xi \times \frac{1}{2}\phi(x-\alpha) \\ &\geq \left(1-\left(\frac{\alpha}{d(x)}\right)^p\right) \times \frac{1}{2}\int_{-\infty}^{+\infty} f_\alpha(\xi)\phi(x-\xi)d\xi \times (1+o_\alpha(1)),\end{aligned}$$

we have $d(x) \leq \alpha\ \tau(\alpha)$, where τ is a function that does not depend on x and that tends to 1 when α tends to 0. □

Now, to get the lower bound of $b(\alpha,1)$, we write

$$b(\alpha,1) \geq \int_{m_1(\alpha)}^{m_2(\alpha)} f_\alpha(\xi)\ r(d,\xi)d\xi + \int_{-m_2(\alpha)}^{-m_1(\alpha)} f_\alpha(\xi)\ r(d,\xi)d\xi,$$

with

$$r(d,\xi) = \int |d(x)-\xi|\phi(x-\xi)dx,$$

$$m_1(\alpha) = \alpha, \quad m_2(\alpha) = \log\left(\frac{1}{\alpha}\right)^{-1}, \text{ if } p \geq 1,$$

$$m_1(\alpha) = \frac{c}{2}, \quad m_2(\alpha) = \mu, \text{ if } p < 1.$$

From similar computations as previously, we obtain the required inequalities. □

Acknowledgements: I would like to thank professor Dominique Picard for a wealth of advice and encouragement, and the three referees for comments leading to improvements in the paper.

References

Abramovich, F., Benjamini, Y., Donoho, D.L. and Johnstone, I.M. (2000) Adapting to unknown sparsity by controlling the false discovery rate. Technical report.

Abramovich, F., Sapatinas, T. and Silverman, B.W. (1998) Wavelet thresholding via a bayesian approach. *J. Roy. Statist. Soc. Ser. B,* **60**, 725–749.

Antoniadis, A. (1994) Smoothing noisy data with coiflets. *Statist. Sinica,* **4**, 651–678.

Bickel, P.J. (1981) Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.,* **9**, 1301–1309.

Casella, G. and Strawderman, W.E. (1981) Estimating a bounded normal mean. *Ann. Statist.,* **9**, 870–878.

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

Cohen, A., DeVore, R.A., Kerkyacharian, G. and Picard, D. (2000b) Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.,* **11**, 167–191.

Daubechies, I. (1992) *Ten Lectures on Wavelets.* Philadelphia: SIAM.

DeVore, R.A. (1989) Degree of nonlinear approximation. In C.K. Chui, L.L. Schumaker and J.D. Wards (eds), *Approximation Theory VI,* Vol. 1, pp. 175–201. New York: Academic Press.

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

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

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

Donoho, D.L. and Johnstone, I.M. (1994b) Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields,* **99**, 277–303.

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

Donoho, D.L. and Johnstone, I.M. (1998) Minimax estimation via wavelet shrinkage. *Ann. Statist.,* **26**, 879–921.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996) Universal near minimaxity of wavelet shrinkage. In D. Pollard, E. Torgersen and G. Yangs (eds), *Festschrift for Lucien Le Cam*, pp. 183–218. New York: Springer-Verlag.

Johnstone, I.M. (1994) Minimax Bayes, asymptotic minimax and sparse wavelet priors. In S.S. Gupta and J.O. Berger (eds), *Statistical Decision Theory and Related Topics*, Vol. 5, pp. 303–326. New York: Springer-Verlag.

Mallat, S. (1998) *A wavelet tour of signal processing.* San Diego: Academic Press.

Massart, P. (1990) The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.*, **18**, 1269–1283.

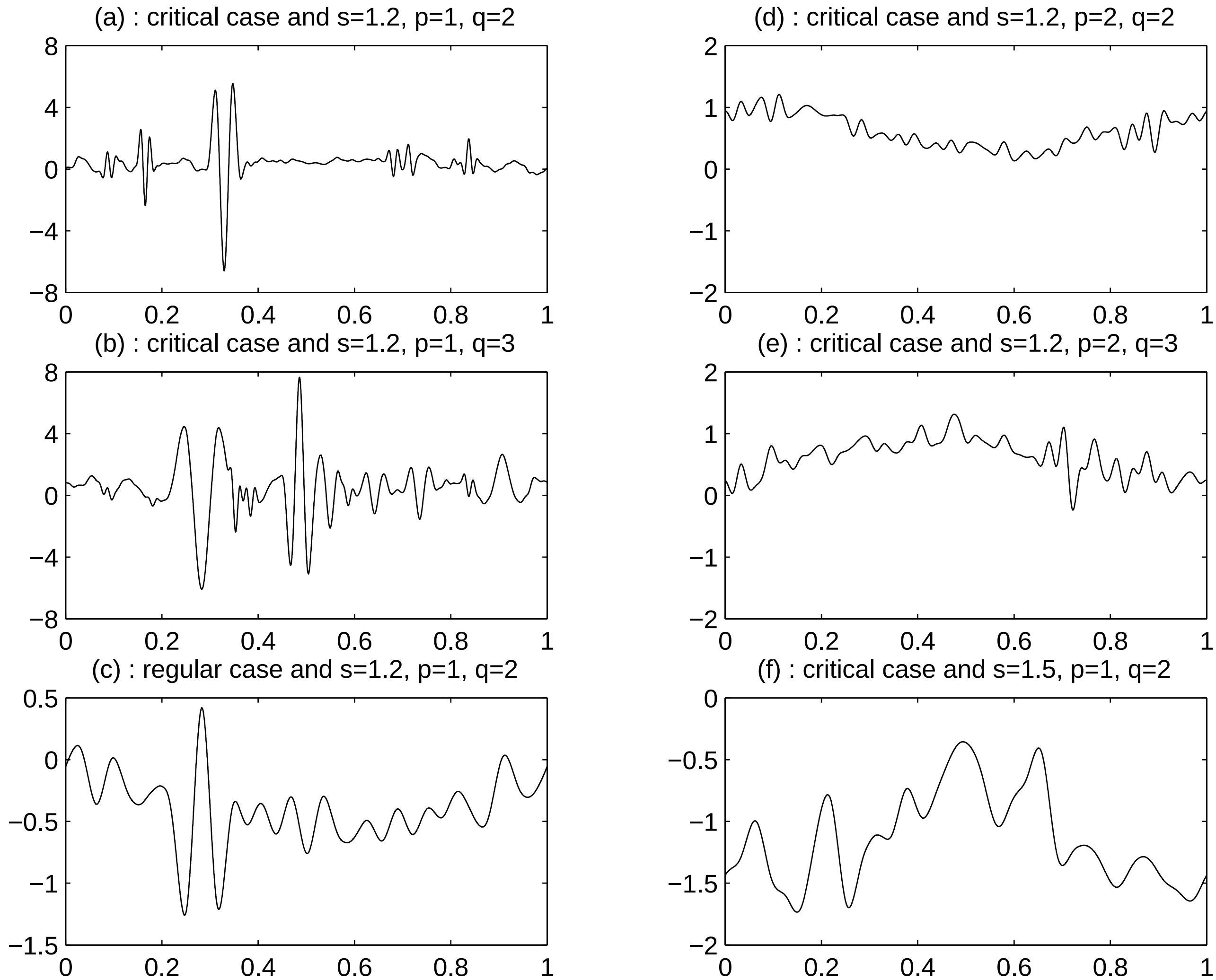
Meyer, Y. (1992) *Wavelets and Operators.* Cambridge: Cambridge University Press.

Pinsker, M.S. (1980) Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission,* **16**, 120–133.

Rivoirard, V. (2004) Thresholding procedure with priors based on Pareto distributions. *Test,* **13**, 213–246.

Rivoirard, V. (2005) Bayesian modelling of sparse sequences and maxisets for Bayes rules. To appear in *Mathematical Methods of Statistics.*

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



Caption for Figure 1:

Realizations with various values of parameters s, p and q, with $C = 1$ and $n = 2^{12} = 4096$ plotting points. The critical case is illustrated by (a), (b), (d), (e) and (f) and the regular case, by (c). (a): $s = 1.2$, $p = 1$, $q = 2$; (b): $s = 1.2$, $p = 1$, $q = 3$; (c): $s = 1.2$, $p = 1$, $q = 2$; (d): $s = 1.2$, $p = 2$, $q = 2$; (e): $s = 1.2$, $p = 2$, $q = 3$; (f): $s = 1.5$, $p = 1$, $q = 2$.