Volume 14, No. 3 (2005), pp. Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

BAYESIAN MODELLING OF SPARSE SEQUENCES AND MAXISETS FOR BAYES RULES

V. Rivoirard

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

In this paper, our aim is to estimate sparse sequences in the framework of the heteroscedastic white noise model. To model sparsity, we consider a Bayesian model composed of a mixture of a heavy-tailed density and a point mass at zero. To evaluate the performance of the Bayes rules (the median or the mean of the posterior distribution), we exploit an alternative to the minimax setting developed in particular by Kerkyacharian and Picard: we determine the maxisets for each of these estimators. Using this approach, we compare the performance of Bayesian procedures with thresholding ones. Furthermore, the maxisets obtained can be viewed as weighted versions of weak l_q spaces that naturally model sparsity. This remark leads us to investigate the following problem: how can we choose the prior parameters to build typical realizations of weighted weak l_q spaces?

Key words: Bayes rules, Bayesian model, heteroscedastic white noise model, maxisets, rate of convergence, sparsity, thresholding rules, weighted weak l_q spaces.

2000 Mathematics Subject Classification: 62C10, 62G05, 62G20.

1. Introduction

1.1. Model. In this paper, we consider the following heteroscedastic white noise model:

$$x_k = \theta_k + \varepsilon\sigma_k\xi_k, \qquad k = 1, 2, \dots, \tag{1}$$

where $\theta = (\theta_k)_{k\geq 1}$ is an unknown sequence to be estimated by using observations $(x_k)_{k\geq 1}$, $\varepsilon > 0$ is a small parameter, and $(\xi_k)_{k\geq 1}$ is an independent and identically distributed (i.i.d.) sequence of Gaussian variables with mean zero and unit variance. Along this paper, we assume that $\sigma = (\sigma_k)_{k\geq 1}$ is a known sequence of positive real numbers. This heteroscedastic white noise model, that appears as a generalization of the classical white noise model (for which, we have $\forall\ k \geq 1$, $\sigma_k = 1$), is extensively used by statisticians. Let us briefly recall the reasons for this large use and provide references. Given a known linear operator A, we use the heteroscedastic white noise model when we have to estimate the solution f of the linear equation $g = Af$, with noisy observations of g. Most of the time, to deal with such a problem, we exploit the singular value decomposition of A and the sequence $(\sigma_k^{-2})_{k\geq 1}$ is then the eigenvalues sequence of the operator A^*A, with A^* the adjoint of A. There is a considerable literature about statistical inverse problems. Let us cite Korostelev and Tsybakov [28], Donoho [14], Golubev and Khasminskii [19], Goldenshluger and Pereverzev [18], Cavalier, Golubev, Picard and Tsybakov [7] and the references cited below. Some well-posed inverse problems with noise can be reduced to (1) with $\sigma_k \overset{k\to\infty}{\longrightarrow} 0$. The condition $\sigma_k \overset{k\to\infty}{\longrightarrow} +\infty$ characterizes ill-posed problems. For instance, the sequences $(\sigma_k)_{k\geq 1}$ associated with operators such as integration considered by Ruymgaart [35], the Radon transform (see Cavalier and Tsybakov [8]), convolution for the case studied by Cavalier and Tsybakov [8] or operators for some elliptic differential equations (see Mair and Ruymgaart [29]) have a polynomial growth. But, the σ_k's may grow exponentially. See, for instance, Pereverzev and Schock [32] who considered the problem of satellite geodesy or the inverse problems associated with partial differential equations such as the heat equation (see Mair and Ruymgaart [29]).

In the wavelet context, Johnstone [21] and Johnstone and Silverman [22] explained that the heteroscedastic white noise model can also be used to represent direct observations with correlated structure. More precisely, let us assume that we are given the following nonparametric regression model:

$$Y_i = f(i/n) + e_i, \qquad i \in \{1, 2, \ldots, n\}, \tag{2}$$

where n is an integer, f is the signal to be estimated, and the e_i's are drawn from a stationary Gaussian process. By studying the autocorrelation function of the e_i's, Johnstone [21] and Johnstone and Silverman [22] showed that under a good choice of ε and $\sigma = (\sigma_k)_{k\geq 1}$, the model (1) appears as a good approximation of the model (2) when n is large.

1.2. Bayesian model and Bayes rules. In this paper, we suppose that the sequence θ to be estimated is sparse. It means that only a small proportion of the θ_k's are non-negligible. When the θ_k's are wavelet coefficients, this assumption is natural since the underlying property of wavelets is that a function can be well approximated by a function with a relatively small proportion of nonzero wavelet coefficients. In this paper, we model the sparsity by using a Bayesian approach. For this purpose, we assume that the θ_k's are random and the distribution of θ is such that the θ_k's are independent and for any $k \geq 1$, there exist a fixed parameter $w_{k,\varepsilon} \in (0, 1)$ depending on ε and k and a fixed density γ, such that, with probability $1 - w_{k,\varepsilon}$, θ_k is equal to 0 and with probability $w_{k,\varepsilon}$, the density of θ_k is $\gamma_{k,\varepsilon}$, where

$$\gamma_{k,\varepsilon}(\theta) = s_{k,\varepsilon}\gamma(s_{k,\varepsilon}\theta), \qquad \forall\ \theta \in \mathbb{R}$$

and

$$s_{k,\varepsilon} = (\varepsilon\sigma_k)^{-1}.$$

If δ_0 denotes the Dirac mass at 0, this model is written as follows:

$$(M_1) \qquad \theta_k \sim (1-w_{k,\varepsilon})\delta_0(\theta_k) + w_{k,\varepsilon}\gamma_{k,\varepsilon}(\theta_k), \qquad k \geq 1.$$

So, roughly speaking, the first term models the negligible components and the second one non-negligible ones.

Bayesian procedures have now become very popular in signal estimation, since they often outperform classical procedures and in particular thresholding procedures from the practical point of view. See the very complete review paper of Antoniadis, Bigot and Sapatinas [4] who provide descriptions and practical comparisons of various Bayesian wavelet shrinkage and wavelet thresholding estimators. It is relevant to note that most of works about Bayesian procedures take place in the practical framework. However, we can cite Johnstone and Silverman [24, 25] and Abramovich, Amato and Angelini [1] who studied Bayesian procedures from the minimax point of view.

Most of the authors consider quite similar Bayesian models and often, priors are based on normal distributions. For instance, in the wavelet framework, Johnstone and Silverman [23] following Abramovich, Sapatinas and Silverman [3] and Clyde, Parmigiani and Vidakovic [10] consider a mixture of a normal component and a point mass at zero for the wavelet coefficients. Chipman, Kolaczyk and McCulloch [9] impose a mixture of two Gaussian distributions with different variances for negligible and non-negligible wavelet coefficients. Let us add that, often, properties of conjugate families enable statisticians to point out easily Bayes rules when Gaussian priors are considered in the classical Gaussian white noise model. However, Johnstone and Silverman [24, 25] did not use Gaussian distributions and showed the advantages from the minimax point of view in considering heavy-tailed distributions. In the maxiset framework, we shall draw similar conclusions concerning γ.

The posterior distribution of θ_k given x_k is

$$(3) \qquad \gamma^{\phi}_{k,\varepsilon}(\theta_k \mid x_k) = \frac{\phi_k(x_k-\theta_k)[w_{k,\varepsilon}\gamma_{k,\varepsilon}(\theta_k) + (1-w_{k,\varepsilon})\delta_0(\theta_k)]}{w_{k,\varepsilon}\int_{-\infty}^{+\infty}\phi_k(x_k-\theta)\gamma_{k,\varepsilon}(\theta)\,d\theta + (1-w_{k,\varepsilon})\phi_k(x_k)},$$

where ϕ_k denotes the density of $\varepsilon\sigma_k\xi_k$, using notations of Section 1.1. For all $\varepsilon > 0$, we assume that we are given a real number $\Lambda_\varepsilon > 1$ depending only on ε and tending to $+\infty$ as $\varepsilon \to 0$. We estimate each θ_k by $\hat{\theta}^{b_1}_k(x_k)$ or by $\hat{\theta}^{b_2}_k(x_k)$ defined by the following procedure.

- If $k < \Lambda_\varepsilon$, $\hat{\theta}^{b_1}_k(x_k)$ (respectively $\hat{\theta}^{b_2}_k(x_k)$) is the median (respectively the mean) of the posterior distribution of θ_k given x_k. Therefore, these rules satisfy: for any $\hat{\theta}_k < \hat{\theta}^{b_1}_k(x_k)$,

$$F_{\gamma^{\phi}_{k,\varepsilon}}(\hat{\theta}_k) < 0.5 \leq F_{\gamma^{\phi}_{k,\varepsilon}}(\hat{\theta}^{b_1}_k(x_k)), \qquad \hat{\theta}^{b_2}_k(x_k) = \int \theta_k \gamma^{\phi}_{k,\varepsilon}(\theta_k \mid x_k)\,d\theta_k,$$

where $F_{\gamma^{\phi}_{k,\varepsilon}}$ denotes the cumulative distribution function of $\gamma^{\phi}_{k,\varepsilon}(\cdot \mid x_k)$.

- If $k \geq \Lambda_\varepsilon$, then $\hat{\theta}^{b_1}_k(x_k) = \hat{\theta}^{b_2}_k(x_k) = 0$.

The values of the hyperparameters $w_{k,\varepsilon}$, γ, and Λ_ε will be chosen later. Other properties of these Bayes rules are given in Section 2.1. So, the choice for Bayes rules is very classical. Indeed, most of the time, in practice, the Bayes rules used by statisticians are the median, and more frequently the mean of the posterior distribution, which are generally better estimates than the mode (see Berger [5], p. 101). Note that the posterior mean and the posterior median are built by using the posterior distribution on its whole support. For instance, let us cite Chipman, Kolaczyk and McCulloch [9] and Clyde, Parmigiani and Vidakovic [10] who used the posterior mean. But Abramovich, Sapatinas and Silverman [3] considered the posterior median under a Bayesian model that has the same form as (M_1). In their Bayesian framework, unlike the posterior mean, the posterior median is a true thresholding rule. In the wavelet context, they showed several simulated examples for which this approach improves most of the traditional methods. If from the practical point of view, the median seems preferable to the mean, what happens under a theoretical approach? From the minimax point of view, Theorem 1 of Johnstone and Silverman [24] showed that the posterior median of their Bayes model achieves optimal rates of convergence under Besov body constraints and for l_q-losses, with $0 < q \leq 2$. If the posterior mean is used, optimal rates are also achieved but only if $1 < q \leq 2$ (see Section 7.3 of [24]). This provides some theoretical justification for preferring the posterior median over the posterior mean. In this paper, to evaluate the performance of $\hat{\theta}^{b_1} = (\hat{\theta}^{b_1}(x_k))_{k\geq 1}$ and $\hat{\theta}^{b_2} = (\hat{\theta}^{b_2}(x_k))_{k\geq 1}$, we use neither a practical approach nor the minimax theory, which have been extensively considered, but the maxiset theory that we describe now.

1.3. The maxiset theory and functional spaces. Let us first motivate the introduction of the maxiset point of view. When nonparametric problems are explored, the minimax theory is the most popular point of view: it consists in ensuring that the used procedure $\hat{\theta} = (\hat{\theta}_k(x_k))_{k\geq 1}$ achieves the best rate on a given sequence space $\mathcal{S}$. But, at first, the choice of $\mathcal{S}$ is arbitrary (what kind of spaces has to be considered: Sobolev spaces? Besov spaces? why?), secondly, $\mathcal{S}$ could contain sequences very difficult to estimate. Since the unknown quantity $\theta = (\theta_k)_{k\geq 1}$ could be easier to estimate, the used procedure could be too pessimistic and not adapted to the data. More embarrassing in practice, several minimax procedures may be proposed and the practitioner has no way to decide but his experiment. To answer these issues, another point of view has recently appeared: the maxiset point of view introduced by Cohen, DeVore, Kerkyacharian and Picard [12] and Kerkyacharian and Picard [26]. Given an estimate $\hat{\theta}$, it consists in assessing the accuracy of $\hat{\theta}$ by fixing a prescribed rate ρ_ε and pointing out the set of all the sequences θ that can be estimated by the procedure $\hat{\theta}$ at the target rate ρ_ε. So, under the statistical model (1), we introduce the following definition.

Definition 1. Let $1 \leq p < \infty$ and let $\hat{\theta} = (\hat{\theta}_k(x_k))_{k\geq 1}$ be an estimator. The maxiset of $\hat{\theta}$ associated with the rate ρ_ε and the l_p-loss is

$$MS(\hat{\theta}, \rho_\varepsilon, p) = \left\{\theta = (\theta_k)_{k\geq 1} \colon \sup_{\varepsilon} \left[\left(\mathbb{E} \sum_{k\geq 1} |\hat{\theta}_k(x_k) - \theta_k|^p\right)^{1/p} \rho_\varepsilon^{-1}\right] < \infty\right\}.$$

The maxiset point of view brings answers to the previous issues. Indeed, there is no a priori assumption on θ and then, the practitioner does not need to restrict his study to an arbitrary sequence space. The practitioner states the desired accuracy and then, knows the quality of the used procedure. Obviously, he chooses the procedure with the largest maxiset. Let us give first examples of maxiset results in the statistical framework of this paper. For this purpose, we need to introduce the following sequence spaces.

Definition 2. For all $1 \leq p < \infty$ and $0 < \eta < \infty$, we set

$$B^{\eta}_{p,\infty} = \left\{ \theta = (\theta_k)_{k\geq 1} : \sup_{\lambda > 0} \lambda^{p\eta} \sum_{k \geq \lambda} |\theta_k|^p < \infty \right\},$$

and if q is a real number such that $0 < q < p$, we set

$$wl_{p,q}(\sigma) = \left\{ \theta = (\theta_k)_{k\geq 1} : \sup_{\lambda > 0} \lambda^{q} \sum_{k} \mathbf{1}_{|\theta_k| > \lambda\sigma_k} \sigma_k^p < \infty \right\}.$$

Now, let us focus on thresholding rules associated with the universal threshold $\lambda_{k,\varepsilon} = \sigma_k \varepsilon \sqrt{|\log \varepsilon|}$ (see Donoho and Johnstone [16]): for all $\varepsilon > 0$, we assume that we are given a real number $\Lambda^*_\varepsilon > 0$ only depending on ε and tending to $+\infty$ as $\varepsilon \to 0$, and we set

$$\hat{\theta}^t_k(x_k) = \begin{cases} x_k \mathbf{1}_{|x_k| \geq \varkappa_* \lambda_{k,\varepsilon}} & \text{if} \quad k < \Lambda^*_\varepsilon, \\ 0 & \text{otherwise,} \end{cases}$$

where $\varkappa_*$ is a constant. Kerkyacharian and Picard [26] have studied the maxisets for this procedure. They obtained the following result for $\hat{\theta}^t = (\hat{\theta}^t_k(x_k))_{k\geq 1}$ (Theorems 3.1 and 3.2 of Kerkyacharian and Picard [26]):

Theorem 1. *Let $1 \leq p < \infty$ be a fixed real number and $0 < r < \infty$. We suppose that*

$$\forall\, \varepsilon, \quad \Lambda^*_\varepsilon = \left(\varepsilon \sqrt{|\log \varepsilon|}\,\right)^{-r}, \tag{4}$$

and there exists a positive constant T such that $\forall\, \varepsilon$

$$\varepsilon^{\varkappa_*^2/16} |\log \varepsilon|^{-1/4 - p/2} \sum_{k < \Lambda^*_\varepsilon} \sigma_k^p \leq T. \tag{5}$$

Let q be a fixed positive real number such that $q < p$. Then, if $\varkappa_* \geq \sqrt{2p}$,

$$MS\Big(\hat{\theta}^t, \big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{(1-q/p)}, p\Big) = wl_{p,q}(\sigma) \cap B^{\frac{1}{r}(1-q/p)}_{p,\infty}.$$

Remark 1. In this last result, of course, to avoid problems of definitions in (4) and (5), it is implicitly assumed, without loss of generality, that ε remains smaller than a positive constant ε_0 strictly smaller than 1 (equal to 1/2 for instance).

For the model (1), we can prove that under some conditions, the maxisets associated with linear estimates of the form $(l_k x_k)_{k\geq 1}$, where $(l_k)_{k\geq 1}$ is a non-increasing sequence of weights lying in $[0,1]$, are the spaces $B^{\eta}_{p,\infty}$, called Besov bodies. These conditions are satisfied, for instance, by projection weights, Tikhonov–Phillips weights or Pinsker weights. For further details see Theorem 2 of Rivoirard [34]. We can add that Lemma 1 of Rivoirard [34] proves that for the rate $(\varepsilon\sqrt{|\log\varepsilon|})^{(1-q/p)}$, the maxisets of linear estimates are strictly contained in the maxisets of thresholding rules. It means that from the maxiset point of view, linear estimates are outperformed by thresholding ones. Kerkyacharian and Picard [27] also applied the maxiset theory for local bandwidth selection in the framework of the Gaussian white noise model. Under some conditions, they proved that local bandwidth selection is at least as good as the thresholding procedure (see Section 5 of [27]).

The comparison of procedures based on maxisets is not as widely used 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 that are well established and easily interpretable. Indeed, in the maxiset approach, the Besov bodies (the spaces $B^{\eta}_{p,\infty}$) control the θ_k's for the large values of k. As for the spaces $wl_{p,q}(\sigma)$, they can be viewed as weighted weak l_q spaces. The weak l_q space is the space $wl_{p,q}(\sigma)$ when $\sigma_k = 1$ for any $k \geq 1$, so we denote it $wl_{p,q}(1)$, and it was considered in statistics by Johnstone [20], Donoho and Johnstone [17] or Abramovich, Benjamini, Donoho and Johnstone [2]. This space was also studied in approximation theory and coding by DeVore [13], Donoho [15], or Cohen, DeVore and Hochmuth [11]. Abramovich, Benjamini, Donoho and Johnstone [2] proved that if we order the components of a sequence θ according to their size:

$$|\theta|_{(1)} \geq |\theta|_{(2)} \geq \cdots \geq |\theta|_{(n)} \geq \ldots,$$

then

$$\theta \in wl_{p,q}(1) \iff \sup_n n^{1/q}|\theta|_{(n)} < \infty$$

(see Section 1.2 of [2]). So, $wl_{p,q}(1)$ spaces naturally measure the sparsity of a signal. Of course, the weighted versions of these spaces, the $wl_{p,q}(\sigma)$ spaces, share the same property.

So, since Bayesian Procedures, commonly used in practice, have been barely studied from a theoretical point of view, it seems relevant to investigate the maxiset results for the Bayesian procedures $\hat{\theta}^{b_1}$ and $\hat{\theta}^{b_2}$ introduced in Section 1.2 and our first two goals in this paper will be the following:

1. to point out the maxisets of the Bayesian procedures $\hat{\theta}^{b_1}$ and $\hat{\theta}^{b_2}$,
2. to compare these estimators with traditional procedures in the maxiset approach by comparing their respective maxisets.

We shall draw interesting conclusions from the maxiset results of these classical Bayes rules that are extensively used in practice.

1.4. Maxiset results. Given $1 \leq p < \infty$ and ρ_ε the prescribed rate, our first issue is to determine for $i \in \{1,2\}$, $MS(\hat{\theta}^{b_i}, \rho_\varepsilon, p)$ once we have fixed assumptions on the hyperparameters $w_{k,\varepsilon}$, γ, and Λ_ε. First, throughout this paper, we take the density γ to be unimodal, symmetric about 0, positive, and absolutely continuous on $\mathbb{R}$. We also assume that there exist two positive constants M and M_1 such that

$$(H_1) \qquad \sup_{\theta \geq M_1} \left| \frac{d}{d\theta} \log \gamma(\theta) \right| = M < \infty.$$

It implies that the tails of γ have to be exponential or heavier. This enables us to establish asymptotic properties of $\hat{\theta}_k^{b_1}(x_k)$ and $\hat{\theta}_k^{b_2}(x_k)$. We prove that the posterior median $\hat{\theta}_k^{b_1}(x_k)$ is a thresholding rule. The posterior mean does not have this thresholding property, but is a shrinkage rule (see Propositions 1, 2, and 3 in Section 2). We also assume that $w_{k,\varepsilon} = w_\varepsilon$ depends only on ε and we set

$$\pi_\varepsilon = (1 - w_\varepsilon) w_\varepsilon^{-1}.$$

Then, the maxisets for these procedures can be pointed out if we take in addition ρ_ε and Λ_ε of the form

$$\rho_\varepsilon = (\varepsilon \sqrt{\log \pi_\varepsilon})^{1-q/p}, \qquad \Lambda_\varepsilon = (\varepsilon \sqrt{\log \pi_\varepsilon})^{-r},$$

with $0 < r < \infty$ and $0 < q < p$. Corollaries 1 and 2 show that under mild assumptions on π_ε and on the size of the σ_k's (see Assumptions (7), (8), (10), and (11), for $i \in \{1, 2\}$,

$$MS\big(\hat{\theta}^{b_i}, (\varepsilon \sqrt{\log \pi_\varepsilon})^{1-q/p}, p\big) = wl_{p,q}(\sigma) \cap B_{p,\infty}^{1/r(1-q/p)}.$$

In particular, it is possible to choose $w_\varepsilon = \varepsilon^\nu$, as soon as ν is great enough to satisfy Assumptions (7), (8), (10), and (11). So, as far as the maxiset point of view is concerned, and roughly under the same conditions, both Bayesian procedures achieve exactly the same performance as the thresholding one for the rate $(\varepsilon \sqrt{|\log \varepsilon|})^{(1-q/p)}$. And for this last rate, we can then claim that each of the Bayesian procedures outperforms the linear algorithm. We note in Section 2.4 that Assumption (H_1) on the tails of γ is essential to get maxisets as large as possible. Finally, in Section 2.5, the previous results are readily extended to the case where we want to estimate functions f decomposed in an appropriate unconditional fixed basis $\mathcal{B} = \{\psi_k, k \geq 1\}$ as $f = \sum_{k \geq 1} \theta_k \psi_k$. We assume that θ is still observed through the model (1) and still estimated by Bayes rules. In this case, the maxisets are no longer sequence spaces but real functional spaces. See Section 2.5 for more details.

1.5. Connections between the proposed Bayesian model and the spaces pointed out. Starting from a Bayesian model, the outcomes of the study of the associated natural Bayes rules under the maxiset approach are the spaces $wl_{p,q}(\sigma)$. Thus, the Bayesian model (M_1) and $wl_{p,q}(\sigma)$ spaces are connected throughout the maxiset approach. We can wonder whether or not this connection is "artificial". It is worthwhile to observe that the Bayesian model has been constructed to model the sparsity of the sequences to be estimated. And as recalled in Section 1.3, $wl_{p,q}(\sigma)$ spaces are natural spaces to measure the sparsity of a sequence by controlling the proportion of non-negligible θ_k's. The third goal of this paper is then the following.

3. Can we establish a "direct" connection between (M_1) and $wl_{p,q}(\sigma)$ spaces?

Actually, we would like to prove a result similar to the one obtained by Abramovich, Sapatinas and Silverman [3] and exploited by Abramovich, Amato and Angelini [1]. Abramovich, Sapatinas and Silverman considered in the wavelet framework a Bayesian model (denoted (M_1^*)) similar to (M_1), where γ is fixed in advance and is the density of a Gaussian variable with mean zero and unit variance. Then, they established a necessary and sufficient condition on the other hyperparameters of (M_1^*) to ensure that the signal built from the wavelet coefficients coming from (M_1^*) belongs, almost surely, to a prescribed Besov space (see Theorem 1 of Abramovich, Sapatinas and Silverman [3]). We would like to do the same job with (M_1) and $wl_{p,q}(\sigma)$ spaces, but without fixing γ in advance. Theorem 7 of Section 3 gives answers to this issue. In particular, we point out the condition $\sup_{\lambda>0}\lambda^q\int_\lambda^{+\infty}\gamma(x)\,dx<\infty$, which means that the tails of γ cannot be heavier than those of a Pareto(q)-variable. Consequently, similarly to the result presented in Section 2.2 of Rivoirard [33], this result illustrates the strong connections between Pareto(q)-distributions and $wl_{p,q}(\sigma)$ spaces. Theorem 7 is proved by using results on weighted empirical distribution process established by Marcus and Zinn [30].

1.6. Contents. In Section 2, we give some properties of the Bayes rules and evaluate the maxisets obtained for the Bayesian procedures $\hat\theta^{b_1}$ and $\hat\theta^{b_2}$. Section 3 investigates the relationships between the Bayesian model and $wl_{p,q}(\sigma)$ spaces. Finally, in Section 4, we prove the results concerning the asymptotic properties of the Bayes rules.

2. Maxisets for Bayesian Procedures

2.1. Properties of the Bayes rules. In the Introduction, we defined the Bayes procedures $\hat\theta^{b_1}$ and $\hat\theta^{b_2}$ used in this paper. Recall that for $k<\Lambda_\varepsilon$, $\hat\theta_k^{b_1}(x_k)$ (respectively $\hat\theta_k^{b_2}(x_k)$) is the median (respectively the mean) of the posterior distribution of θ_k given x_k given by (3). We also mention that if $\tilde\gamma_{k,\varepsilon}$ denotes the prior distribution of θ_k, then for any $1\le p<\infty$ and any estimator $\hat\theta_k(x_k)$ of θ_k, the Bayes risk $B(\hat\theta_k,\tilde\gamma_{k,\varepsilon},p)$ of $\hat\theta_k(x_k)$ with respect to $\tilde\gamma_{k,\varepsilon}$ associated with the l_p-loss defined by

$$B(\hat\theta_k,\tilde\gamma_{k,\varepsilon},p)=\iint\tilde\gamma_{k,\varepsilon}(\theta_k)|\hat\theta_k(x_k)-\theta_k|^p\phi_k(x_k-\theta_k)\,d\theta_k\,dx_k,$$

where ϕ_k still denotes the density of $\varepsilon\sigma_k\xi_k$, satisfies:

$$B(\hat\theta_k^{b_1},\tilde\gamma_{k,\varepsilon},1)\le B(\hat\theta_k,\tilde\gamma_{k,\varepsilon},1),$$

and

$$B(\hat\theta_k^{b_2},\tilde\gamma_{k,\varepsilon},2)\le B(\hat\theta_k,\tilde\gamma_{k,\varepsilon},2).$$

These Bayes rules are shrinkage rules. In particular, they satisfy the following property, which will be capital for description of their maxisets. For further details, see Lemma 2, inequality (62), and Section 5.5 of Johnstone and Silverman [25].

Proposition 1. *For all $k\ge1$, since γ is symmetric, absolutely continuous, positive, and unimodal, we have for $i\in\{1,2\}$,*

- $\hat\theta_k^{b_i}(x_k^1)\le\hat\theta_k^{b_i}(x_k^2)$ *for any (x_k^1,x_k^2) such that $x_k^1\le x_k^2$,*

- $\hat{\theta}_k^{b_i}(-x_k) = -\hat{\theta}_k^{b_i}(x_k)$ *for any* x_k,
- $0 \leq \hat{\theta}_k^{b_i}(x_k) \leq x_k$ *for any* $x_k \geq 0$,
- $|\hat{\theta}_k^{b_i}(x_k) - \theta_k| \leq \max(|\theta_k|; |x_k - \theta_k|)$ *for any* (x_k, θ_k).

Assumptions imposed on the Bayesian model enable us to deduce other properties. Assumption (H_1) implies that

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

It means that the tails of γ have to be exponential or heavier. We shall see below (see Section 2.4) that this assumption is essential to get maxisets as large as possible. Furthermore, $w_{k,\varepsilon} = w_\varepsilon$ depends only on ε and we shall assume throughout this paper that $\pi_\varepsilon = (1 - w_\varepsilon)w_\varepsilon^{-1}$ satisfies the following mild assumptions, globally denoted (H_2):

1. $\varepsilon \longrightarrow \pi_\varepsilon$ is continuous,
2. $\inf_{\varepsilon>0} \pi_\varepsilon > 1$,
3. $\pi_1 = \exp(1)$,
4. $\pi_\varepsilon \overset{\varepsilon\to 0}{\longrightarrow} +\infty$,
5. $\varepsilon\sqrt{\log \pi_\varepsilon} \overset{\varepsilon\to 0}{\longrightarrow} 0$.

Then, we deduce the asymptotic behavior of $\hat{\theta}_k^{b_1}(x_k)$.

Proposition 2. *Assume that* (H_1) *and* (H_2) *hold. For all* $k < \Lambda_\varepsilon$, $\hat{\theta}_k^{b_1}(x_k)$ *is a thresholding rule, i.e., there exists a uniquely defined* $t(\pi_\varepsilon)$ *such that*

$$\hat{\theta}_k^{b_1}(x_k) = 0 \iff s_{k,\varepsilon}|x_k| \leq t(\pi_\varepsilon),$$

where the threshold $t(\pi_\varepsilon)$ *satisfies* $t(\pi_\varepsilon) \geq \sqrt{2\log(\pi_\varepsilon)}$ *for* π_ε *large enough, and*

$$\lim_{\pi_\varepsilon\to+\infty} \frac{t(\pi_\varepsilon)}{\sqrt{2\log(\pi_\varepsilon)}} = 1.$$

Furthermore, there exists a positive constant C *such that*

$$\limsup_{\pi_\varepsilon\to+\infty} |s_{k,\varepsilon}x_k - s_{k,\varepsilon}\hat{\theta}_k^{b_1}(x_k)|\mathbf{1}_{|s_{k,\varepsilon}x_k|>2t(\pi_\varepsilon)} \leq C.$$

Remark 2. The threshold $t(\pi_\varepsilon)$ introduced in Proposition 2 will be used throughout this paper, even though it is only implicitly defined.

The proof of this proposition is given in the Appendix. Since $\forall\, k < \Lambda_\varepsilon$, $\hat{\theta}_k^{b_1}(x_k)$ is a thresholding rule, it will be easy to evaluate the maxiset for $\hat{\theta}^{b_1}$. As for the posterior mean, we have the following useful result.

Proposition 3. *Assume that* (H_1) *and* (H_2) *hold. Let* $k < \Lambda_\varepsilon$ *be fixed. There exist two functions* ε_1 *and* ε_2 *bounded on* $[1, +\infty)$ *such that* $\varepsilon_1(x) \overset{x\to\infty}{\longrightarrow} 0$, *and*

$$\hat{\theta}_k^{b_2}(x_k) = x_k \frac{1 + \varepsilon_1(s_{k,\varepsilon}x_k)}{1 + \pi_\varepsilon\phi(s_{k,\varepsilon}x_k)\gamma(s_{k,\varepsilon}x_k)^{-1}\varepsilon_2(s_{k,\varepsilon}x_k)}, \tag{6}$$

where ϕ denotes the density of a $(0,1)$ Gaussian variable. If $t(\pi_\varepsilon)$ is the threshold introduced in Proposition 2, there exists a positive constant C such that

$$\limsup_{\pi_\varepsilon\to+\infty} |s_{k,\varepsilon}x_k - s_{k,\varepsilon}\hat{\theta}_k^{b_2}(x_k)|\mathbf{1}_{|s_{k,\varepsilon}x_k|>2t(\pi_\varepsilon)} \le C.$$

Proposition 3 is proved in the Appendix.

Remark 3. By using the results of Proposition 3, we have for π_ε large enough,

$$\left|\hat{\theta}_k^{b_2}\left(s_{k,\varepsilon}^{-1}\frac{t(\pi_\varepsilon)}{2}\right)\right| \le \pi_\varepsilon^{-\frac{1}{2}} s_{k,\varepsilon}^{-1}\sqrt{\log\pi_\varepsilon}.$$

Unlike $\hat{\theta}_k^{b_1}(x_k)$, for all $k<\Lambda_\varepsilon$, $\hat{\theta}_k^{b_2}(x_k)$ is not a thresholding rule, since $\hat{\theta}_k^{b_2}(x_k)\neq 0$ if $x_k\neq 0$, and we can easily prove that under (H_1),

$$\lim_{s_{k,\varepsilon}x_k\to 0}\left(x_k\frac{\int_{-\infty}^{+\infty}u^2\exp(-\frac{1}{2}u^2)\gamma(u)\,du}{\int_{-\infty}^{+\infty}\exp(-\frac{1}{2}u^2)\gamma(u)\,du+\pi_\varepsilon}\right)^{-1}\hat{\theta}_k^{b_2}(x_k)=1.$$

Now, we are ready to evaluate the maxisets for both Bayesian rules. In Sections 2.2 and 2.3, for $i\in\{1,2\}$, the risk of $\hat{\theta}^{b_i}$ studied under the l_p-norm ($1\le p<\infty$), will be denoted $R_p(\hat{\theta}^{b_i})$. We have:

$$R_p(\hat{\theta}^{b_i}) = (\mathbb{E}\,\|\hat{\theta}^{b_i}-\theta\|_{l_p}^p)^{1/p} = \left(\sum_{k\ge 1}\mathbb{E}\,|\hat{\theta}_k^{b_i}(x_k)-\theta_k|^p\right)^{1/p}.$$

2.2. Maxisets for the posterior median. Since $\forall\, k<\Lambda_\varepsilon$, $\hat{\theta}_k^{b_1}(x_k)$ is a thresholding rule, it is easy to evaluate the maxiset for $\hat{\theta}^{b_1}$. On the one hand, we have:

Theorem 2. *Assume that (H_1) and (H_2) hold. Let $0<r<\infty$ and $1\le p<\infty$ be two fixed real numbers. Suppose that $\forall\,\varepsilon>0$,*

$$\Lambda_\varepsilon = (\varepsilon\sqrt{\log\pi_\varepsilon})^{-r}$$

and there exist two positive constants T_1 and T_2 such that $\forall\,\varepsilon>0$,

$$\varepsilon^{-p}\,\pi_\varepsilon^{-1}(\log\pi_\varepsilon)^{-\frac{1}{2}-\frac{p}{2}} \le T_1, \tag{7}$$

$$\pi_\varepsilon^{-\frac{1}{8}}(\log\pi_\varepsilon)^{-\frac{1}{4}-\frac{p}{2}}\sum_{k<\Lambda_\varepsilon}\sigma_k^p \le T_2. \tag{8}$$

Let q be a fixed positive real number such that $q<p$. If $\theta\in B_{p,\infty}^{\frac{1}{r}(1-q/p)}\cap wl_{p,q}(\sigma)$, then there exists a positive constant C such that

$$\forall\,\varepsilon>0,\qquad R_p(\hat{\theta}^{b_1}) \le C(\varepsilon\sqrt{\log\pi_\varepsilon})^{1-q/p}.$$

The proof of this theorem that uses Proposition 2 is omitted since it is inspired by the proof of Theorem 3.1 of Kerkyacharian and Picard [26] and is very similar to the proof of Theorem 4 of Section 2.3.

On the other hand, we have:

Theorem 3. *Assume that* (H_1) *and* (H_2) *hold. Let* $0 < r < \infty$ *and* $1 \le p < \infty$ *be two fixed real numbers. Suppose that* $\forall\, \varepsilon > 0$,

$$\Lambda_\varepsilon = (\varepsilon\sqrt{\log \pi_\varepsilon})^{-r}.$$

Let q *be a fixed positive real number such that* $q < p$. *If there exists a positive constant* C *such that*

$$\forall\, \varepsilon > 0, \qquad R_p(\hat{\theta}^{b_1}) \le C(\varepsilon\sqrt{\log \pi_\varepsilon})^{1-q/p},$$

then $\theta \in B_{p,\infty}^{\frac{1}{r}(1-q/p)} \cap wl_{p,q}(\sigma)$.

Before proving Theorem 3, let us recall the following result (Lemma 2.2 of Kerkyacharian and Picard [26]):

Proposition 4. *For any* $1 \le p < \infty$ *and* $0 < q < p$,

$$wl_{p,q}(\sigma) = \Big\{ \theta = (\theta_k)_{k\ge 1} \colon \sup_{\lambda > 0} \lambda^{q-p} \sum_k \mathbf{1}_{|\theta_k| \le \lambda\sigma_k} |\theta_k|^p < \infty \Big\}.$$

Proof of Theorem 3. Since $\pi_1 = \exp(1)$, we have $\Lambda_1 = 1$ and $\|\theta\|_{l_p}^p \le C^p$. For any $\varepsilon > 0$,

$$\sum_{k \ge \Lambda_\varepsilon} |\theta_k|^p \le \mathbb{E}\, \|\hat{\theta}^{b_1} - \theta\|_{l_p}^p \le C^p \Lambda_\varepsilon^{-(p-q)/r},$$

and since $\varepsilon \to \Lambda_\varepsilon$ is continuous with $\lim_{\varepsilon\to 0} \Lambda_\varepsilon = +\infty$, we have:

$$\sup_{\lambda > 0} \lambda^{\frac{p}{r}(1-\frac{q}{p})} \sum_{k \ge \lambda} |\theta_k|^p \le C^p,$$

and $\theta \in B_{p,\infty}^{\frac{1}{r}(1-\frac{q}{p})}$.

In the following, we shall use $t(\pi_\varepsilon)$ (denoted t if there is no risk of confusion) introduced in Proposition 2 and the following inequality: for any $m > 0$, with $Z \sim \mathcal{N}(0,1)$, for ε small enough,

$$\begin{aligned}
(9)\qquad \mathbb{P}\Big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k| \ge \frac{t}{m}\Big) &\le \mathbb{P}\Big(|Z| \ge \frac{1}{m}\sqrt{2\log \pi_\varepsilon}\Big) \\
&\le 2\int_{\frac{1}{m}\sqrt{2\log\pi_\varepsilon}}^{\infty} e^{-u^2/2} \frac{du}{\sqrt{2\pi}} \\
&\le 2\exp\Big(-\frac{1}{2}\Big(\frac{1}{m}\sqrt{2\log\pi_\varepsilon}\Big)^2 \Big) \int_0^\infty e^{-u\frac{1}{m}\sqrt{2\log\pi_\varepsilon}} \frac{du}{\sqrt{2\pi}} \\
&\le K(\log\pi_\varepsilon)^{-1/2}\, \pi_\varepsilon^{-1/m^2},
\end{aligned}$$

where K depends only on m. So, for $k < \Lambda_\varepsilon$, and for ε small enough,

$$\begin{aligned}|\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le\frac{t}{2}} &= |\theta_k|^p \mathbb{E}\left(\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le\frac{t}{2}} \mathbf{1}_{|s_{k,\varepsilon}x_k|\ge t}\right) + \mathbb{E}\left(|\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le\frac{t}{2}} \mathbf{1}_{|s_{k,\varepsilon}x_k|<t}\right)\\ &\le |\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le\frac{t}{2}} \mathbb{P}\Big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k| \ge \frac{t}{2}\Big) + \mathbb{E}\,|\hat{\theta}_k^{b_1}(x_k) - \theta_k|^p\\ &\le \frac{1}{2}|\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le\frac{t}{2}} + \mathbb{E}\,|\hat{\theta}_k^{b_1}(x_k) - \theta_k|^p.\end{aligned}$$

Therefore, for ε small enough,

$$\sum_k |\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le\frac{t}{2}} \le 2\mathbb{E}\,\|\hat{\theta}^{b_1} - \theta\|_{l_p}^p \le 2C^p(\varepsilon\sqrt{\log \pi_\varepsilon})^{p-q}.$$

Since $\varepsilon \to \pi_\varepsilon$ is continuous, it implies that there exists $\lambda_0 > 0$ such that

$$\sup_{\lambda<\lambda_0} \lambda^{q-p} \sum_k |\theta_k|^p \mathbf{1}_{|\theta_k|\le\sigma_k\lambda} < \infty.$$

Since $\theta \in B_{p,\infty}^{\frac{1}{r}(1-\frac{q}{p})}$,

$$\sup_{\lambda\ge\lambda_0} \lambda^{q-p} \sum_k |\theta_k|^p \mathbf{1}_{|\theta_k|\le\sigma_k\lambda} \le \lambda_0^{q-p} \sum_k |\theta_k|^p < \infty.$$

Using Proposition 4, we have proved that $\theta \in B_{p,\infty}^{\frac{1}{r}(1-\frac{q}{p})} \cap wl_{p,q}(\sigma)$. $\square$

Finally, from Theorems 2 and 3, we deduce easily:

Corollary 1. *Assume that* (H_1) *and* (H_2) *hold. Let* $0 < r < \infty$ *and* $1 \le p < \infty$ *be two fixed real numbers. Suppose that* $\forall\, \varepsilon > 0$,

$$\Lambda_\varepsilon = (\varepsilon\sqrt{\log \pi_\varepsilon})^{-r}$$

such that (7) *and* (8) *are satisfied. Let* q *be a fixed positive real number such that* $q < p$. *Then,*

$$MS\big(\hat{\theta}^{b_1}, (\varepsilon\sqrt{\log \pi_\varepsilon})^{1-q/p}, p\big) = wl_{p,q}(\sigma) \cap B_{p,\infty}^{\frac{1}{r}(1-q/p)}.$$

We can conclude that the spaces $B_{p,\infty}^{\frac{1}{r}(1-\frac{q}{p})} \cap wl_{p,q}(\sigma)$ appear as maximal spaces where $\hat{\theta}^{b_1}$ attains specific rates of convergence.

2.3. Maxisets for the posterior mean. As in Section 2.2, we prove the following results:

Theorem 4. *Assume that* (H_1) *and* (H_2) *hold. Let* $0 < r < \infty$ *and* $1 \le p < \infty$ *be two fixed real numbers. Suppose that* $\forall\, \varepsilon > 0$,

$$\Lambda_\varepsilon = (\varepsilon\sqrt{\log \pi_\varepsilon})^{-r}$$

and there exist two positive constants T_1 and T_2 such that $\forall\, \varepsilon > 0$,

$$\varepsilon^{-p}\ \pi_\varepsilon^{-\frac{1}{4}}(\log \pi_\varepsilon)^{-\frac{1}{2}-\frac{p}{2}} \leq T_1, \tag{10}$$

$$\pi_\varepsilon^{-\frac{1}{32}}(\log \pi_\varepsilon)^{-\frac{1}{4}-\frac{p}{2}} \sum_{k<\Lambda_\varepsilon} \sigma_k^p \leq T_2. \tag{11}$$

Let q be a fixed positive real number such that $q < p$. If $\theta \in B_{p,\infty}^{\frac{1}{r}(1-q/p)} \cap wl_{p,q}(\sigma)$, then there exists a positive constant C such that

$$\forall\, \varepsilon > 0, \quad R_p(\hat{\theta}^{b_2}) \leq C(\varepsilon\sqrt{\log \pi_\varepsilon})^{1-q/p}.$$

Proof. In the following, K will denote a constant independent of ε that may be different at each line. Let $t = t(\pi_\varepsilon)$ be the threshold introduced in Proposition 2. For all $k < \Lambda_\varepsilon$,

$$\mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p = A + B + C,$$

with

$$A = \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}x_k| \leq \frac{t}{2}}, \qquad B = \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p \mathbf{1}_{\frac{t}{2} < |s_{k,\varepsilon}x_k| \leq 2t},$$

and

$$C = \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}x_k| > 2t}.$$

For the first term, we have, using Remark 3 and (9),

$$\begin{aligned}
A &= \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}x_k| \leq \frac{t}{2}} \\
&\leq 2^{p-1}\mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k)|^p \mathbf{1}_{|s_{k,\varepsilon}x_k| \leq \frac{t}{2}} + 2^{p-1}|\theta_k|^p \mathbb{E}\,\mathbf{1}_{|s_{k,\varepsilon}x_k| \leq \frac{t}{2}} \\
&\leq K s_{k,\varepsilon}^{-p} \pi_\varepsilon^{-\frac{p}{2}}(\log \pi_\varepsilon)^{\frac{p}{2}} + 2^{p-1}|\theta_k|^p\ \mathbb{E}\,[\mathbf{1}_{|s_{k,\varepsilon}x_k| \leq \frac{t}{2}} \mathbf{1}_{|s_{k,\varepsilon}\theta_k| > t}] \\
&\qquad + 2^{p-1}|\theta_k|^p\ \mathbb{E}\,[\mathbf{1}_{|s_{k,\varepsilon}x_k| \leq \frac{t}{2}} \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq t}] \\
&\leq K\sigma_k^p \varepsilon^p \pi_\varepsilon^{-\frac{p}{2}}(\log \pi_\varepsilon)^{\frac{p}{2}} + 2^{p-1}|\theta_k|^p\ \mathbb{P}\Big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k| \geq \frac{t}{2}\Big) \\
&\qquad + 2^{p-1}|\theta_k|^p\ \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq t} \\
&\leq K\sigma_k^p \varepsilon^p \pi_\varepsilon^{-\frac{p}{2}}(\log \pi_\varepsilon)^{\frac{p}{2}} + K|\theta_k|^p \pi_\varepsilon^{-\frac{1}{4}}(\log \pi_\varepsilon)^{-\frac{1}{2}} + 2^{p-1}|\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq t}.
\end{aligned}$$

The second term can be split into $B = B_1 + B_2 + B_3$, with

$$\begin{aligned}
B_1 &= \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p \mathbf{1}_{\frac{t}{2} < |s_{k,\varepsilon}x_k| \leq 2t} \mathbf{1}_{|s_{k,\varepsilon}\theta_k| > 3t}, \\
B_2 &= \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p \mathbf{1}_{\frac{t}{2} < |s_{k,\varepsilon}x_k| \leq 2t} \mathbf{1}_{\frac{t}{4} < |s_{k,\varepsilon}\theta_k| \leq 3t}, \\
B_3 &= \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p \mathbf{1}_{\frac{t}{2} < |s_{k,\varepsilon}x_k| \leq 2t} \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq \frac{t}{4}}.
\end{aligned}$$

Using Proposition 1 and (9),

$$B_1 = \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p \mathbf{1}_{\frac{t}{2} < |s_{k,\varepsilon}x_k| \leq 2t} \mathbf{1}_{|s_{k,\varepsilon}\theta_k| > 3t}$$

$$
\begin{aligned}
&\le \mathbb{E}\,|x_k-\theta_k|^p \mathbf{1}_{|x_k-\theta_k|\ge|\theta_k|}\mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>3t}\\
&\qquad + |\theta_k|^p \mathbb{E}\,\mathbf{1}_{|x_k-\theta_k|<|\theta_k|}\mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>3t}\\
&\le (\mathbb{E}\,|x_k-\theta_k|^{2p})^{\frac12}\mathbb{P}(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|\ge t)^{\frac12} + |\theta_k|^p\,\mathbb{P}(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|\ge t)\\
&\le K\varepsilon^p\sigma_k^p\pi_\varepsilon^{-\frac12}(\log\pi_\varepsilon)^{-\frac14} + K|\theta_k|^p\pi_\varepsilon^{-1}(\log\pi_\varepsilon)^{-\frac12},
\end{aligned}
$$

$$
\begin{aligned}
B_2 &= \mathbb{E}\,|\hat\theta_k^{b_2}(x_k)-\theta_k|^p\mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{\frac{t}{4}<|s_{k,\varepsilon}\theta_k|\le 3t}\\
&\le \mathbb{E}\,|x_k-\theta_k|^p\mathbf{1}_{|x_k-\theta_k|\ge|\theta_k|}\mathbf{1}_{\frac{t}{4}<|s_{k,\varepsilon}\theta_k|} + |\theta_k|^p\ \mathbb{E}\,\mathbf{1}_{|x_k-\theta_k|<|\theta_k|}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le 3t}\\
&\le K\varepsilon^p\sigma_k^p\ \mathbb{P}\Big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|\ge\frac{t}{4}\Big)^{\frac12} + |\theta_k|^p\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le 3t}\\
&\le K\varepsilon^p\sigma_k^p\pi_\varepsilon^{-\frac{1}{32}}(\log\pi_\varepsilon)^{-\frac14} + |\theta_k|^p\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le 3t},
\end{aligned}
$$

$$
\begin{aligned}
B_3 &= \mathbb{E}\,|\hat\theta_k^{b_2}(x_k)-\theta_k|^p\mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le\frac{t}{4}}\\
&\le \mathbb{E}\,|x_k-\theta_k|^p\mathbf{1}_{|s_{k,\varepsilon}x_k-s_{k,\varepsilon}\theta_k|\ge\frac{t}{4}}\\
&\le (\mathbb{E}\,|x_k-\theta_k|^{2p})^{\frac12}\mathbb{P}\Big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|\ge\frac{t}{4}\Big)^{\frac12}\\
&\le K\varepsilon^p\sigma_k^p\pi_\varepsilon^{-\frac{1}{32}}(\log\pi_\varepsilon)^{-\frac14}.
\end{aligned}
$$

For the last term, we denote $\tau(k,\varepsilon)=s_{k,\varepsilon}x_k - s_{k,\varepsilon}\hat\theta_k^{b_2}(x_k)$ and recall that ξ_k is the noise. Using Proposition 3 and (9), we have

$$
\begin{aligned}
C &= \mathbb{E}\,|\hat\theta_k^{b_2}(x_k)-\theta_k|^p\mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\\
&= \mathbb{E}\,|\hat\theta_k^{b_2}(x_k)-\theta_k|^p\mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le t} + \mathbb{E}\,|\hat\theta_k^{b_2}(x_k)-\theta_k|^p\mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}\\
&\le s_{k,\varepsilon}^{-p}\mathbb{E}\,(|\xi_k-\tau(k,\varepsilon)|^{2p}\mathbf{1}_{|s_{k,\varepsilon}x_k|>2t})^{\frac12}\mathbb{P}(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|\ge t)^{\frac12}\\
&\qquad + s_{k,\varepsilon}^{-p}\mathbb{E}\,|\xi_k-\tau(k,\varepsilon)|^p\mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}\\
&\le Ks_{k,\varepsilon}^{-p}\mathbb{P}(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|\ge t)^{\frac12} + Ks_{k,\varepsilon}^{-p}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}\\
&\le K\varepsilon^p\sigma_k^p(\pi_\varepsilon^{-\frac12}(\log\pi_\varepsilon)^{-\frac14} + \mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}).
\end{aligned}
$$

Finally, for ε small enough, and $\forall\ k<\Lambda_\varepsilon$,

$$
\begin{aligned}
\mathbb{E}\,|\hat\theta_k^{b_2}(x_k)-\theta_k|^p \le K\Big[&\varepsilon^p\sigma_k^p\pi_\varepsilon^{-\frac{1}{32}}(\log\pi_\varepsilon)^{-\frac14} + \varepsilon^p\sigma_k^p\mathbf{1}_{|\theta_k|>\varepsilon\sigma_k\sqrt{2\log\pi_\varepsilon}}\\
&+ |\theta_k|^p\mathbf{1}_{|\theta_k|\le 4\varepsilon\sigma_k\sqrt{2\log\pi_\varepsilon}} + |\theta_k|^p\pi_\varepsilon^{-\frac14}(\log\pi_\varepsilon)^{-\frac12}\Big].
\end{aligned}
$$

We conclude by using Proposition 4 and by observing that

$$
\mathbb{E}\,\|\hat\theta^{b_2}-\theta\|_{l_p}^p = \sum_{k<\Lambda_\varepsilon}\mathbb{E}\,|\hat\theta_k^{b_2}(x_k)-\theta_k|^p + \sum_{k\ge\Lambda_\varepsilon}|\theta_k|^p \le K(\varepsilon\sqrt{\log\pi_\varepsilon})^{p-q},
$$

since $\theta\in B_{p,\infty}^{\frac1r(1-\frac{q}{p})}\cap wl_{p,q}(\sigma)$ and $\Lambda_\varepsilon=(\varepsilon\sqrt{\log\pi_\varepsilon})^{-r}$. $\square$

As in Section 2.2, we have a converse result, but unlike Theorem 3, we need to control the size of the σ_k's.

Theorem 5. *Assume that* (H_1) *and* (H_2) *hold. Let* $0 < r < \infty$ *and* $1 \le p < \infty$ *be two fixed real numbers. Suppose that* $\forall\, \varepsilon > 0$,

$$\Lambda_\varepsilon = (\varepsilon\sqrt{\log \pi_\varepsilon})^{-r}.$$

Let q be a fixed positive real number such that $q < p$. If there exists a positive constant C such that

$$\forall\, \varepsilon > 0, \qquad R_p(\hat{\theta}^{b_2}) \le C(\varepsilon\sqrt{\log \pi_\varepsilon})^{1-q/p},$$

then $\theta \in B_{p,\infty}^{\frac{1}{r}(1-\frac{q}{p})} \cap wl_{p,q}(\sigma)$ *as soon as there exists a positive constant T such that*

$$\forall\, \varepsilon > 0, \qquad \pi_\varepsilon^{-\frac{p}{2}} \sum_{k<\Lambda_\varepsilon} \sigma_k^p \le T. \tag{12}$$

Proof. To prove that $\theta \in B_{p,\infty}^{\frac{1}{r}(1-\frac{q}{p})}$, we refer the reader to the proof of Theorem 3. Then, we want to show that

$$\sup_{\lambda>0} \lambda^{q-p} \sum_k |\theta_k|^p \mathbf{1}_{|\theta_k| \le \lambda\sigma_k} < \infty.$$

For this, we still use the threshold $t = t(\pi_\varepsilon)$ of Proposition 2. Using (9), for any $k < \Lambda_\varepsilon$ and for ε small enough,

$$\begin{aligned}
|\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}} &= |\theta_k|^p \mathbb{E}\,(\mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}} \mathbf{1}_{|s_{k,\varepsilon}x_k| \ge \frac{t}{2}}) + |\theta_k|^p \mathbb{E}\,(\mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}} \mathbf{1}_{|s_{k,\varepsilon}x_k| < \frac{t}{2}}) \\
&\le |\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}}\, \mathbb{P}\Big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k| \ge \frac{t}{4}\Big) \\
&\qquad + |\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}} \mathbb{P}\Big(|s_{k,\varepsilon}x_k| < \frac{t}{2}\Big) \\
&\le \frac{1}{2} |\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}} + |\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}} \mathbb{P}\Big(|s_{k,\varepsilon}x_k| < \frac{t}{2}\Big).
\end{aligned}$$

By using Remark 3, we have for ε small enough,

$$\Big|\hat{\theta}_k^{b_2}\Big(s_{k,\varepsilon}^{-1}\frac{t}{2}\Big)\Big| \le \pi_\varepsilon^{-\frac{1}{2}} s_{k,\varepsilon}^{-1} \sqrt{\log \pi_\varepsilon}.$$

Therefore,

$$\begin{aligned}
\sum_k |\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}} &\le 2 \sum_k |\theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \le \frac{t}{4}} \mathbb{P}\Big(|s_{k,\varepsilon}x_k| < \frac{t}{2}\Big) \\
&\le 2^p \sum_k \mathbb{E}\left[|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p + |\hat{\theta}_k^{b_2}(x_k)|^p\right] \mathbf{1}_{|s_{k,\varepsilon}x_k| < \frac{t}{2}} \\
&\le 2^p \sum_k \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p + 2^p \sum_{k<\Lambda_\varepsilon} \mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k)|^p \mathbf{1}_{|s_{k,\varepsilon}x_k| < \frac{t}{2}} \\
&\le 2^p\, \mathbb{E}\,\|\hat{\theta}^{b_2} - \theta\|_{l_p}^p + 2^p \sum_{k<\Lambda_\varepsilon} s_{k,\varepsilon}^{-p} \pi_\varepsilon^{-\frac{p}{2}} (\log \pi_\varepsilon)^{\frac{p}{2}} \\
&\le 2^p\, C^p (\varepsilon\sqrt{\log \pi_\varepsilon})^{p-q} + 2^p \varepsilon^p \pi_\varepsilon^{-\frac{p}{2}} (\log \pi_\varepsilon)^{\frac{p}{2}} \sum_{k<\Lambda_\varepsilon} \sigma_k^p.
\end{aligned}$$

Consequently, under condition (12), for ε small enough,

$$\sum_k |\theta_k|^p \mathbf{1}_{|\theta_k|\leq \frac{1}{4}\sigma_k \varepsilon\sqrt{\log \pi_\varepsilon}} \leq 2^p(C^p+T)(\varepsilon\sqrt{\log \pi_\varepsilon})^{p-q}.$$

Using the same arguments as for the proof of Theorem 3, the last inequality implies that $\theta \in wl_{p,q}(\sigma)$. $\square$

Finally, from Theorems 4 and 5, observing that condition (12) is less restrictive than condition (11) since $p/2 \geq 1/2 > 1/32$, we deduce easily:

Corollary 2. *Assume that* (H_1) *and* (H_2) *hold. Let* $0 < r < \infty$ *and* $1 \leq p < \infty$ *be two fixed real numbers. Suppose that* $\forall\, \varepsilon > 0$,

$$\Lambda_\varepsilon = (\varepsilon\sqrt{\log \pi_\varepsilon})^{-r}$$

such that (10) *and* (11) *are satisfied. Let* q *be a fixed positive real number such that* $q < p$. *Then,*

$$MS\big(\hat{\theta}^{b_2}, (\varepsilon\sqrt{\log \pi_\varepsilon})^{1-q/p}, p\big) = wl_{p,q}(\sigma) \cap B_{p,\infty}^{\frac{1}{r}(1-q/p)}.$$

Once more, we can conclude that the spaces $B_{p,\infty}^{\frac{1}{r}(1-\frac{q}{p})} \cap wl_{p,q}(\sigma)$ appear as maximal spaces where $\hat{\theta}^{b_2}$ attains specific rates of convergence.

2.4. First conclusions. Before going further, let us compare the various procedures involved in this paper. We recall that unlike $\hat{\theta}_k^{b_1}(x_k)$, $\hat{\theta}_k^{b_2}(x_k)$ does not possess the advantage of being a thresholding rule ($k < \Lambda_\varepsilon$) and this explains the differences between the assumptions that are needed to determine the respective maxisets associated with the Bayesian procedures $\hat{\theta}^{b_1}$ and $\hat{\theta}^{b_2}$. For instance, this explains why, unlike $\hat{\theta}^{b_1}$, if $\hat{\theta}^{b_2}$ achieves the given rate of convergence, we need a condition on the σ_k's to prove that $\theta \in B_{p,\infty}^{\frac{1}{r}(1-\frac{q}{p})} \cap wl_{p,q}(\sigma)$. Furthermore, to obtain the upper bound for $R_p(\hat{\theta}^{b_2})$, we use a decomposition into eleven terms for $\mathbb{E}\,|\hat{\theta}_k^{b_2}(x_k) - \theta_k|^p$. Since it is a thresholding rule, the corresponding decomposition for $\hat{\theta}_k^{b_1}$ is simpler. This explains why the assumptions of Theorem 4 are a bit more restrictive than those of Theorem 2. Actually, to obtain the assumptions of Theorem 4, we just have to replace π_ε with $\pi_\varepsilon^{1/4}$ in the assumptions of Theorem 2. But, since we consider a rate of the form $\varepsilon\sqrt{\log \pi_\varepsilon}$ without focusing on the optimal constant, the Bayesian procedures achieve exactly the same performance from the maxiset point of view. When π_ε is a power of ε, then, by using Theorem 1, we can compare the Bayesian procedures $\hat{\theta}^{b_1}$ and $\hat{\theta}^{b_2}$ with the thresholding one. We can conclude that each of them achieves the same performance as the thresholding one. Finally, since linear estimates are outperformed by thresholding ones, they are also outperformed by $\hat{\theta}^{b_1}$ and $\hat{\theta}^{b_2}$.

Let us show now the importance of Assumption (H_1). Section 7.2 of Johnstone and Silverman [24] proves that if γ is a normal density, whatever the value of $w_{k,\varepsilon}$, the posterior median satisfies

$$|\hat{\theta}^{b_1}(x_k)| \leq (1-\alpha)|x_k|,$$

for some $\alpha > 0$ and the same inequality holds for the posterior mean. It yields for $\theta_k > 0$,

$$\mathbb{E}\,|\hat{\theta}_k^{b_1}(x_k) - \theta_k|^p \geq \frac{1}{2}\alpha^p \theta_k^p.$$

Moreover, when γ has tails equivalent to $\exp(-C|t|^\lambda)$ for some $\lambda \in (1,2)$, Johnstone and Silverman showed that for large θ_k,

$$|\hat{\theta}_k^{b_1}(x_k) - \theta_k| \geq C|\theta_k|^{\lambda-1}.$$

Thus (H_1) cannot be essentially relaxed without obtaining smaller maxisets.

2.5. Maxisets of Bayesian procedures for estimating functions of L_p spaces. In this section, we estimate functions of $L_p(D) = \big\{f\colon \|f\|_{L_p} = (\int_D |f(x)|^p dx)^{1/p} < \infty\big\}$, where $D = [0,1]^d$ or $D = \mathbb{R}^d$. For this purpose, we exploit a wavelet basis of $L_2(D)$ denoted $\mathcal{B} = \{\psi_k, k \geq 1\}$. More precisely, we assume that $(\psi_k)_{k\geq 1}$ is the wavelet-tensor product constructed on compactly supported wavelets (see Meyer [31]). So, if $1 < p < \infty$, Meyer [31] proved that $\mathcal{B}$ is an unconditional basis of $L_p(D)$, which means that:

- for any $f \in L_p(D)$, there exists a unique sequence θ such that $f = \sum_k \theta_k \psi_k$,
- there exists an absolute constant K such that if $\forall\, k \geq 1$, $|\theta_k| \leq |\theta'_k|$, then

$$\Big\|\sum_k \theta_k \psi_k\Big\|_{L_p} \leq K \Big\|\sum_k \theta'_k \psi_k\Big\|_{L_p}.$$

Remark 4. The restriction $1 < p < \infty$ is due to the fact that there is no unconditional basis if $p \notin (1,\infty)$.

Furthermore, we assume that $\{\sigma_k \psi_k,\ k \geq 1\}$ satisfies the following inequality, called a superconcentration inequality: for any $0 < r_1 < \infty$, there exists a constant $C(p, r_1)$ such that for all $F \subset \{1, 2, \dots\}$,

$$\Big\|\Big[\sum_{k\in F} |\sigma_k \psi_k|^{r_1}\Big]^{\frac{1}{r_1}}\Big\|_{L_p} \leq C(p, r_1)\big\|\sup_{k\in F} |\sigma_k \psi_k|\big\|_{L_p}.$$

Remark 5. Theorem 4.2 of Kerkyacharian and Picard [26] gives conditions on the σ_k's and on $\mathcal{B}$ to satisfy the above superconcentration inequality.

We consider the model (1), and the function $f = \sum_k \theta_k \psi_k$ is estimated by $\hat{f}^{b_1} = \sum_k \hat{\theta}_k^{b_1}(x_k)\psi_k$ or $\hat{f}^{b_2} = \sum_k \hat{\theta}_k^{b_2}(x_k)\psi_k$. In this framework, we set:

Definition 3. For $1 < p < \infty$ and any $i \in \{1, 2\}$, the maxiset of $\hat{f}^{b_i}$ associated with the rate ρ_ε and the L_p-loss is

$$MS(\hat{f}^{b_i}, \rho_\varepsilon, p) = \Big\{f = \sum_k \theta_k \psi_k\colon\ \sup_\varepsilon \Big[(\mathbb{E}\,\|\hat{f}^{b_i} - f\|_{L_p}^p)^{\frac{1}{p}} \rho_\varepsilon^{-1}\Big] < \infty\Big\}.$$

To study maxisets of $\hat{f}^{b_1}$ and $\hat{f}^{b_2}$, we introduce for any $\eta > 0$ and any $0 < q < p$:

$$B^{\eta}_{p,\infty}(\mathcal{B}) = \Big\{ f = \sum_k \theta_k \psi_k \colon \sup_{\lambda>0} \lambda^{\eta} \Big\| \sum_{k \geq \lambda} \theta_k \psi_k \Big\|_{L_p} < \infty \Big\},$$
$$wl_{p,q}(\sigma)(\mathcal{B}) = \Big\{ f = \sum_k \theta_k \psi_k \colon \sup_{\lambda>0} \lambda^{q} \sum_k \mathbf{1}_{|\theta_k| > \lambda\sigma_k} \sigma_k^p \, \|\psi_k\|^p_{L_p} < \infty \Big\}.$$

If $\mathcal{B}$ is a standard wavelet basis regular enough, the space $B^{\eta}_{p,\infty}(\mathcal{B})$ can be identified with a real Besov space. See Meyer [31] for further details.

Theorem 6. *Assume that (H_1) and (H_2) hold. Let $0 < r < \infty$ be a fixed real number. Suppose that*

$$\forall\, \varepsilon > 0, \qquad \Lambda_\varepsilon = \big(\varepsilon \sqrt{\log \pi_\varepsilon}\big)^{-r},$$

and there exist two positive constants T_1 and T_2 such that for any $\varepsilon > 0$,

$$\varepsilon^{-p} \left[\pi_\varepsilon^{-1} (\log \pi_\varepsilon)^{-\frac{1}{2}} \right]^{\frac{1}{2}\min(p;2)} (\log \pi_\varepsilon)^{-\frac{p}{2}} \leq T_1$$

and

$$\pi_\varepsilon^{-\frac{1}{8}} (\log \pi_\varepsilon)^{-\frac{1}{4}-\frac{p}{2}} \sum_{k<\Lambda_\varepsilon} \sigma_k^p \|\psi_k\|^p_{L_p} \leq T_2.$$

Let q be a fixed positive real number such that $q < p$. Then, under the model (1),

$$MS\big(\hat{f}^{b_1}, (\varepsilon\sqrt{\log \pi_\varepsilon})^{1-q/p}, p\big) = wl_{p,q}(\sigma)(\mathcal{B}) \cap B^{\frac{1}{r}(1-q/p)}_{p,\infty}(\mathcal{B}).$$

The analogous result for $\hat{f}^{b_2}$ is obtained if we assume that for any $\varepsilon > 0$,

$$\varepsilon^{-p} \left[\pi_\varepsilon^{-\frac{1}{4}} (\log \pi_\varepsilon)^{-\frac{1}{2}} \right]^{\frac{1}{2}\min(p;2)} (\log \pi_\varepsilon)^{-\frac{p}{2}} \leq T_3,$$

and

$$\pi_\varepsilon^{-\frac{1}{32}} (\log \pi_\varepsilon)^{-\frac{1}{4}-\frac{p}{2}} \sum_{k<\Lambda_\varepsilon} \sigma_k^p \|\psi_k\|^p_{L_p} \leq T_4,$$

where T_3 and T_4 are two positive constants.

Once more, when π_ε is a power of ε (i.e., for the rate $\varepsilon\sqrt{\log(1/\varepsilon)}$), by using Theorems 5.1 and 5.2 of Kerkyacharian and Picard [26] we can conclude that each of the Bayesian procedures achieves the same performance as the thresholding one.

Proof. Again, we only give the proof for the procedure associated with the mean. Recall that $\mathcal{B}$ is an unconditional basis of $L_p(D)$ if and only if there exists $M > 0$ such that for any set $F \subset \{1, 2, \dots\}$ and any choice of the coefficients c_k's

$$M^{-1} \Big\| \sum_{k\in F} c_k \psi_k \Big\|_{L_p} \leq \Big\| \Big(\sum_{k \in F} |c_k \psi_k|^2 \Big)^{\frac{1}{2}} \Big\|_{L_p} \leq M \Big\| \sum_{k\in F} c_k \psi_k \Big\|_{L_p}. \tag{13}$$

Furthermore, since $\{\sigma_k\psi_k,\, k \geq 1\}$ satisfies a superconcentration inequality, there exist two positive constants c_p and C_p such that for any $F \subset \{1, 2, \dots\}$, we have:

$$c_p \int \sum_{k\in F} |\sigma_k\psi_k|^p \leq \int \Big(\sum_{k\in F} |\sigma_k\psi_k|^2\Big)^{\frac{p}{2}} \leq C_p \int \sum_{k\in F} |\sigma_k\psi_k|^p. \tag{14}$$

In the following, K will denote a constant independent of ε that may be different at each line. We use the threshold $t = t(\pi_\varepsilon)$ introduced in Proposition 2 and the results of Proposition 1. Let us assume that $f \in wl_{p,q}(\sigma)(\mathcal{B}) \cap B_{p,\infty}^{\frac{1}{r}(1-q/p)}(\mathcal{B})$. Then, $\mathbb{E}\,\|\sum_k(\hat{\theta}_k^{b_2} - \theta_k)\psi_k\|_{L_p}^p$ is bounded by $K\sum_{i=1}^{11} A_i$, with the A_i's defined as follows:

$$A_1 = \Big\|\sum_{k\geq\Lambda_\varepsilon} \theta_k\psi_k\Big\|_{L_p}^p \leq K\big(\varepsilon\sqrt{\log\pi_\varepsilon}\big)^{(p-q)},$$

since $f \in B_{p,\infty}^{\frac{1}{r}(1-q/p)}(\mathcal{B})$. Next,

$$\begin{aligned} A_2 &= \mathbb{E}\Big\|\sum_{k<\Lambda_\varepsilon} \theta_k\psi_k \mathbf{1}_{|s_{k,\varepsilon}x_k|\leq\frac{t}{2}}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}\Big\|_{L_p}^p \\ &\leq K\mathbb{E}\int\Big(\sum_{k<\Lambda_\varepsilon} \theta_k^2\psi_k^2\mathbf{1}_{|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|\geq\frac{t}{2}}\Big)^{\frac{p}{2}}, \end{aligned}$$

by using (13).

If $p \leq 2$, by using (9), the Jensen inequality, and (13),

$$\begin{aligned} A_2 &\leq K\int\Big(\sum_{k<\Lambda_\varepsilon} \theta_k^2\psi_k^2\mathbb{P}\big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k| \geq t/2\big)\Big)^{\frac{p}{2}} \\ &\leq K\big(\pi_\varepsilon^{-1/4}(\log\pi_\varepsilon)^{-\frac{1}{2}}\big)^{\frac{p}{2}}\Big\|\sum_{k<\Lambda_\varepsilon}\theta_k\psi_k\Big\|_{L_p}^p. \end{aligned}$$

If $p \geq 2$, by using (9), the generalized Minkowski inequality, and (13)

$$\begin{aligned} A_2 &\leq K\int\Big(\sum_{k<\Lambda_\varepsilon} \theta_k^2\psi_k^2\mathbb{P}\big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k| \geq t/2\big)^{\frac{2}{p}}\Big)^{\frac{p}{2}} \\ &\leq K\pi_\varepsilon^{-1/4}(\log\pi_\varepsilon)^{-\frac{1}{2}}\Big\|\sum_{k<\Lambda_\varepsilon}\theta_k\psi_k\Big\|_{L_p}^p \end{aligned}$$

and

$$A_3 = \mathbb{E}\Big\|\sum_{k<\Lambda_\varepsilon}\theta_k\psi_k\mathbf{1}_{|s_{k,\varepsilon}x_k|\leq\frac{t}{2}}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\leq t}\Big\|_{L_p}^p \leq K\int\Big(\sum_{k<\Lambda_\varepsilon}\theta_k^2\psi_k^2\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\leq t}\Big)^{\frac{p}{2}},$$

by using (13). Further,

$$\begin{aligned} A_4 &= \mathbb{E}\Big\|\sum_{k<\Lambda_\varepsilon}\hat{\theta}_k^{b_2}\psi_k\mathbf{1}_{|s_{k,\varepsilon}x_k|\leq\frac{t}{2}}\Big\|_{L_p}^p \leq K\mathbb{E}\int\Big(\sum_{k<\Lambda_\varepsilon}(\hat{\theta}_k^{b_2})^2\mathbf{1}_{|s_{k,\varepsilon}x_k|\leq\frac{t}{2}}\psi_k^2\Big)^{\frac{p}{2}} \\ &\leq K\varepsilon^p\pi_\varepsilon^{-p/2}(\log\pi_\varepsilon)^{\frac{p}{2}}\int\Big(\sum_{k<\Lambda_\varepsilon}\sigma_k^2\psi_k^2\Big)^{\frac{p}{2}} \leq K\varepsilon^p\pi_\varepsilon^{-p/2}(\log\pi_\varepsilon)^{\frac{p}{2}}\sum_{k<\Lambda_\varepsilon}\sigma_k^p\|\psi_k\|_{L_p}^p, \end{aligned}$$

by using (13), Remark 3, and (14). Using (13), we have:

$$A_5 = \mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{\frac{t}{4}<|s_{k,\varepsilon}\theta_k|\le 3t}\mathbf{1}_{|x_k-\theta_k|\ge|\theta_k|}\Big\|_{L_p}^p$$
$$\le K\mathbb{E}\int\Big(\sum_{k<\Lambda_\varepsilon}(x_k-\theta_k)^2\psi_k^2\mathbf{1}_{|s_{k,\varepsilon}x_k-s_{k,\varepsilon}\theta_k|\ge\frac{t}{4}}\Big)^{\frac{p}{2}}.$$

If $p \le 2$,

$$A_5 \le K\int\sum_{k<\Lambda_\varepsilon}|\psi_k|^p\mathbb{E}\left(|x_k-\theta_k|^p\mathbf{1}_{|s_{k,\varepsilon}x_k-s_{k,\varepsilon}\theta_k|\ge\frac{t}{4}}\right)$$
$$\le K\varepsilon^p\pi_\varepsilon^{-\frac{1}{32}}(\log\pi_\varepsilon)^{-\frac{1}{4}}\int\sum_{k<\Lambda_\varepsilon}|\psi_k|^p\sigma_k^p,$$

by using (9) and the Cauchy–Schwarz inequality.

If $p \ge 2$, by using (9), the generalized Minkowski inequality, and (14),

$$A_5 \le K\int\Big(\sum_{k<\Lambda_\varepsilon}\psi_k^2\big(\mathbb{E}\,|x_k-\theta_k|^p\mathbf{1}_{|s_{k,\varepsilon}x_k-s_{k,\varepsilon}\theta_k|\ge\frac{t}{4}}\big)^{\frac{2}{p}}\Big)^{\frac{p}{2}}$$
$$\le K\varepsilon^p\pi_\varepsilon^{-132}(\log\pi_\varepsilon)^{-\frac{1}{4}}\sum_{k<\Lambda_\varepsilon}\int|\psi_k|^p\sigma_k^p.$$

Now we have

$$A_6 = \mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{\frac{t}{4}<|s_{k,\varepsilon}\theta_k|\le 3t}\mathbf{1}_{|x_k-\theta_k|<|\theta_k|}\Big\|_{L_p}^p$$
$$\le K\int\Big(\sum_{k<\Lambda_\varepsilon}\theta_k^2\psi_k^2\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le 3t}\Big)^{\frac{p}{2}},$$

by using (13). By using (13), the following terms:

$$A_7 = \mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>3t}\mathbf{1}_{|x_k-\theta_k|\ge|\theta_k|}\Big\|_{L_p}^p$$
$$\le K\mathbb{E}\int\Big(\sum_{k<\Lambda_\varepsilon}(x_k-\theta_k)^2\psi_k^2\mathbf{1}_{|s_{k,\varepsilon}x_k-s_{k,\varepsilon}\theta_k|\ge t}\Big)^{\frac{p}{2}},$$
$$A_8 = \mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>3t}\mathbf{1}_{|x_k-\theta_k|<|\theta_k|}\Big\|_{L_p}^p$$
$$\le K\mathbb{E}\int\Big(\sum_{k<\Lambda_\varepsilon}\theta_k^2\psi_k^2\mathbf{1}_{|s_{k,\varepsilon}x_k-s_{k,\varepsilon}\theta_k|\ge t}\Big)^{\frac{p}{2}},$$
$$A_9 = \mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \mathbf{1}_{\frac{t}{2}<|s_{k,\varepsilon}x_k|\le 2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le\frac{t}{4}}\Big\|_{L_p}^p$$
$$\le K\mathbb{E}\int\Big(\sum_{k<\Lambda_\varepsilon}(x_k-\theta_k)^2\psi_k^2\mathbf{1}_{|s_{k,\varepsilon}x_k-s_{k,\varepsilon}\theta_k|\ge\frac{t}{4}}\Big)^{\frac{p}{2}}$$

are bounded, up to a constant, by the upper bounds of A_2 or A_5. Next,

$$\begin{aligned}
A_{10} &= \mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|\le t}\Big\|_{L_p}^p \\
&\le K\mathbb{E} \int \Big(\sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)^2\psi_k^2 \mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|>t}\Big)^{\frac{p}{2}},
\end{aligned}$$

by using (13).

If $p \le 2$, by using (9), the Cauchy–Schwarz inequality, and Proposition 3,

$$\begin{aligned}
A_{10} &\le K \int \sum_{k<\Lambda_\varepsilon} |\psi_k|^p \mathbb{E}\,|\hat{\theta}_k^{b_2} - \theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|>t} \\
&\le K \int \sum_{k<\Lambda_\varepsilon} |\psi_k|^p \varepsilon^p \sigma_k^p \mathbb{P}\big(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k| > t\big)^{\frac{1}{2}} \\
&\le K\varepsilon^p \pi_\varepsilon^{-1/2} (\log \pi_\varepsilon)^{-\frac{1}{4}} \sum_{k<\Lambda_\varepsilon} \sigma_k^p \int |\psi_k|^p.
\end{aligned}$$

If $p \ge 2$, by using the generalized Minkowski inequality, the Cauchy–Schwarz inequality, (9), (14), and Proposition 3,

$$\begin{aligned}
A_{10} &\le K \int \Big(\sum_{k<\Lambda_\varepsilon} \psi_k^2 \big(\mathbb{E}\,|\hat{\theta}_k^{b_2} - \theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k|>t}\big)^{\frac{2}{p}}\Big)^{\frac{p}{2}} \\
&\le K \int \Big(\sum_{k<\Lambda_\varepsilon} \psi_k^2 \sigma_k^2 \varepsilon^2 \mathbb{P}(|s_{k,\varepsilon}x_k - s_{k,\varepsilon}\theta_k| > t)^{\frac{1}{p}}\Big)^{\frac{p}{2}} \\
&\le K\varepsilon^p \pi_\varepsilon^{-1/2} (\log \pi_\varepsilon)^{-1/4} \sum_{k<\Lambda_\varepsilon} \int |\psi_k|^p \sigma_k^p.
\end{aligned}$$

Finally,

$$\begin{aligned}
A_{11} &= \mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}\Big\|_{L_p}^p \\
&\le K\mathbb{E} \int \Big(\sum_{k<\Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)^2\psi_k^2 \mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}\Big)^{\frac{p}{2}}.
\end{aligned}$$

If $p \le 2$, by using Proposition 3,

$$\begin{aligned}
A_{11} &\le K \int \sum_{k<\Lambda_\varepsilon} |\psi_k|^p \mathbb{E}\,|\hat{\theta}_k^{b_2} - \theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}x_k|>2t}\mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t} \\
&\le K\varepsilon^p \sum_{k<\Lambda_\varepsilon} \int |\psi_k|^p \sigma_k^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}.
\end{aligned}$$

If $p \geq 2$, by using (14) and Proposition 3,

$$A_{11} \leq K \int \Big(\sum_{k<\Lambda_\varepsilon} \psi_k^2 \mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t} \big(\mathbb{E}\, |\hat{\theta}_k^{b_2} - \theta_k|^p \mathbf{1}_{|s_{k,\varepsilon}x_k|>2t} \big)^{\frac{2}{p}} \Big)^{\frac{p}{2}}$$
$$\leq K\varepsilon^p \int \Big(\sum_{k<\Lambda_\varepsilon} \psi_k^2 \sigma_k^2 \mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t} \Big)^{\frac{p}{2}} \leq K\varepsilon^p \sum_{k<\Lambda_\varepsilon} \int |\psi_k|^p \sigma_k^p \mathbf{1}_{|s_{k,\varepsilon}\theta_k|>t}.$$

Thus we conclude that there exists a positive constant C such that

$$\forall\, \varepsilon > 0, \quad \mathbb{E} \Big\| \sum_k (\hat{\theta}_k^{b_2}(x_k) - \theta_k)\psi_k \Big\|_{L_p}^p \leq C^p \big(\varepsilon \sqrt{\log \pi_\varepsilon}\big)^{(p-q)},$$

by using the following result proved by Kerkyacharian and Picard [26] (upper bound of the term B_2, p. 311): if $f = \sum_k \theta_k \psi_k \in wl_{p,q}(\sigma)(\mathcal{B})$, then $\forall\, \lambda > 0$,

$$\int \Big(\sum_k \theta_k^2 \psi_k^2 \mathbf{1}_{|\theta_k| \leq \sigma_k \lambda} \Big)^{\frac{p}{2}} \leq K\lambda^{p-q}.$$

Now, let us assume that there exists a positive constant C such that

$$\forall\, \varepsilon > 0, \quad \mathbb{E} \Big\| \sum_k (\hat{\theta}_k^{b_2}(x_k) - \theta_k)\psi_k \Big\|_{L_p}^p \leq C^p \big(\varepsilon \sqrt{\log \pi_\varepsilon}\big)^{(p-q)}.$$

To bound the following term, we just use (13):

$$A_{12} = \Big\| \sum_{k \geq \Lambda_\varepsilon} \theta_k \psi_k \Big\|_{L_p}^p = \mathbb{E} \Big\| \sum_{k \geq \Lambda_\varepsilon} (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \Big\|_{L_p}^p \leq K\mathbb{E} \int \Big(\sum_k (\hat{\theta}_k^{b_2} - \theta_k)^2 \psi_k^2 \Big)^{\frac{p}{2}}$$
$$\leq K\mathbb{E} \Big\| \sum_k (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \Big\|_{L_p}^p \leq K \big(\varepsilon \sqrt{\log \pi_\varepsilon}\big)^{p-q},$$

which proves that $f = \sum_k \theta_k \psi_k \in B_{p,\infty}^{\frac{1}{r}(1-q/p)}(\mathcal{B})$. Now, using (13),

$$\Big\| \sum_k \theta_k \psi_k \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq \frac{t}{4}} \Big\|_{L_p}^p \leq K(A_{12} + A_{13}),$$

with

$$A_{13} = \Big\| \sum_{k<\Lambda_\varepsilon} \theta_k \psi_k \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq \frac{t}{4}} \Big\|_{L_p}^p$$
$$\leq 2^{p-1} \mathbb{E} \Big\| \sum_{k<\Lambda_\varepsilon} \theta_k \psi_k \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq \frac{t}{4}} \mathbf{1}_{|s_{k,\varepsilon}x_k| > \frac{t}{2}} \Big\|_{L_p}^p + 2^{p-1} A_{14}$$
$$\leq \frac{1}{2} \Big\| \sum_{k<\Lambda_\varepsilon} \theta_k \psi_k \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq \frac{t}{4}} \Big\|_{L_p}^p + 2^{p-1} A_{14} \leq \frac{1}{2} A_{13} + 2^{p-1} A_{14} \leq 2^p A_{14},$$

for ε small enough (see the upper bound of A_2), where

$$A_{14} = \mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} \theta_k \psi_k \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq \frac{t}{4}} \mathbf{1}_{|s_{k,\varepsilon}x_k| \leq \frac{t}{2}} \Big\|_{L_p}^p .$$

But, using (13), we obtain

$$\begin{aligned} A_{14} &\leq K\mathbb{E}\Big\| \sum_k (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \Big\|_{L_p}^p + K\mathbb{E}\Big\| \sum_{k<\Lambda_\varepsilon} \hat{\theta}_k^{b_2} \psi_k \mathbf{1}_{|s_{k,\varepsilon}x_k| \leq \frac{t}{2}} \Big\|_{L_p}^p \\ &\leq K\mathbb{E}\Big\| \sum_k (\hat{\theta}_k^{b_2} - \theta_k)\psi_k \Big\|_{L_p}^p + KA_4 \leq K\big(\varepsilon\sqrt{\log \pi_\varepsilon}\big)^{(p-q)}, \end{aligned}$$

which implies that

$$\Big\| \sum_k \theta_k \psi_k \mathbf{1}_{|s_{k,\varepsilon}\theta_k| \leq \frac{t}{4}} \Big\|_{L_p}^p \leq K\big(\varepsilon\sqrt{\log \pi_\varepsilon}\big)^{(p-q)}.$$

Now, we use Lemma 5.1 of Kerkyacharian and Picard [26], which ends the proof of the theorem. $\square$

3. Relationships between (M_1) and $wl_{p,q}(\sigma)$ Spaces

In this paper, our aim is to estimate sparse sequences, and we model sparsity within a Bayes approach, and more precisely, by using the model (M_1). We noticed that under this model, the maxisets for the previous Bayes rules are defined by using $wl_{p,q}(\sigma)$ spaces. To some extent, this result is not surprising, since we recalled in the Introduction that these spaces are weighted versions of weak l_q spaces that naturally measure sparsity. Then, the model (M_1) and $wl_{p,q}(\sigma)$ spaces are connected via a maxiset approach. So, it is natural to wonder whether we can establish other more natural connections between our Bayesian approach to model sparsity and $wl_{p,q}(\sigma)$ spaces. The following result gives a positive answer.

Theorem 7. *Suppose that we are given $1 \leq p < \infty$ and $0 < q < p$. We again consider the model (M_1) with $\varepsilon = 1$. Denote $w_k = w_{k,1}$ and $\forall\ \lambda \geq 0$, $\tilde{F}(\lambda) = 2\int_\lambda^{+\infty} \gamma(x)\,dx$. If there exists a constant C such that*

$$\sup_{\lambda>0} \lambda^q \sum_k \sigma_k^p \mathbf{1}_{|\theta_k| > \sigma_k\lambda} \leq C^q \quad a.s.,$$

then

$$\sup_{\lambda>0} \lambda^q \tilde{F}(\lambda) \sum_k w_k \sigma_k^p \leq C^q.$$

Conversely, if there exists a constant C such that

$$\sup_{\lambda>0} \lambda^q \tilde{F}(\lambda) \sum_k w_k \sigma_k^p \leq C^q, \tag{15}$$

then

$$\sup_{\lambda>0} \lambda^q \sum_k \sigma_k^p \mathbf{1}_{|\theta_k| > \sigma_k\lambda} < \infty \quad a.s.$$

Proof. If
$$\sup_{\lambda>0}\lambda^q\sum_k\sigma_k^p\mathbf{1}_{|\theta_k|>\sigma_k\lambda}\le C^q\quad\text{a.s.},$$
we have:
$$\begin{aligned}\sup_{\lambda>0}\lambda^q\tilde F(\lambda)\sum_k w_k\sigma_k^p&=\sup_{\lambda>0}\mathbb{E}\left(\lambda^q\sum_k\sigma_k^p\mathbf{1}_{|\theta_k|>\sigma_k\lambda}\right)\\&\le\mathbb{E}\left(\sup_{\lambda>0}\lambda^q\sum_k\sigma_k^p\mathbf{1}_{|\theta_k|>\sigma_k\lambda}\right)\le C^q.\end{aligned}$$

Conversely, suppose that (15) is true. To establish the required inequality, we exploit Theorem 0.3 of Marcus and Zinn [30]. Let $(r_k)_{k\ge1}$ be a Rademacher sequence (i.e., a sequence of i.i.d. random variables taking values $+1$ and -1 with probability $1/2$ each) independent of $(\theta_k)_{k\ge1}$ and let S_n be the partial sum of the symmetrized random variables $(\sigma_k^{p-q}|\theta_k|^q)_{k\ge1}$:
$$S_n=\sum_{k=1}^n Y_k,$$
where
$$Y_k=r_k\,\sigma_k^{p-q}|\theta_k|^q.$$
We have
$$\begin{aligned}\mathbb{E}\left(Y_k\mathbf{1}_{|Y_k|\le1}\right)&=0,\\\mathbb{P}(|Y_k|>1)&=\mathbb{P}\left(\left|\frac{\theta_k}{\sigma_k}\right|^q>\sigma_k^{-p}\right)=w_k\tilde F(\sigma_k^{-p/q})\le C^q\Big(\sum_k w_k\sigma_k^p\Big)^{-1}w_k\sigma_k^p,\\\operatorname{var}(Y_k\mathbf{1}_{|Y_k|\le1})&=\mathbb{E}\left(Y_k^2\mathbf{1}_{|Y_k|\le1}\right)=\sigma_k^{2p}\mathbb{E}\left[(\sigma_k^{-1}|\theta_k|)^{2q}\mathbf{1}_{\sigma_k^{-1}|\theta_k|\le\sigma_k^{-p/q}}\right]\\&=2w_k\sigma_k^{2p}\int_0^{\sigma_k^{-p/q}}x^{2q}\gamma(x)\,dx\le2w_k\sigma_k^{2p}\int_0^{\sigma_k^{-p/q}}qx^{2q-1}\tilde F(x)\,dx\\&\le2C^q\Big(\sum_k w_k\sigma_k^p\Big)^{-1}w_k\sigma_k^{2p}\int_0^{\sigma_k^{-p/q}}qx^{q-1}\,dx\\&\le2C^q\Big(\sum_k w_k\sigma_k^p\Big)^{-1}w_k\sigma_k^p.\end{aligned}$$

Using the three series theorem (see Theorem 22.8 of Billingsley [6]), S_n converges with probability 1 as $n\to+\infty$. Therefore, if μ is a fixed positive real number, we have for any increasing sequence of positive real numbers $(b_n)_n$ with $\lim_{n\to+\infty}b_n=+\infty$,
$$\limsup_{n\to+\infty}\frac{1}{b_n}|S_n|\le\mu\quad\text{a.s.}$$

Obviously,

$$\limsup_{n\to+\infty}\sup_{1\le j\le n}\sup_{\lambda>0}\frac{1}{b_n}\lambda^q\Big|\sum_{k=1}^{j}\sigma_k^p\mathbb{E}\left(\mathbf{1}_{|\theta_k|>\sigma_k\lambda}\right)\Big|\le\mu\quad\text{a.s.}$$

Then, we can apply Theorem 0.3 of Marcus and Zinn [30], which shows that

$$\limsup_{n\to+\infty}\frac{1}{b_n}\sup_{\lambda>0}\lambda^q\Big|\sum_{k=1}^{n}\sigma_k^p\big(\mathbf{1}_{|\theta_k|>\sigma_k\lambda}-\mathbb{P}(|\theta_k|>\sigma_k\lambda)\big)\Big|\le 1160\,\mu.$$

Since

$$\limsup_{n\to+\infty}\frac{1}{b_n}\sup_{\lambda>0}\lambda^q\sum_{k=1}^{n}\sigma_k^p\mathbb{P}(|\theta_k|>\sigma_k\lambda)=0,$$

it yields

$$\limsup_{n\to+\infty}\frac{1}{b_n}\sup_{\lambda>0}\lambda^q\sum_{k=1}^{n}\sigma_k^p\mathbf{1}_{|\theta_k|>\sigma_k\lambda}\le 1160\,\mu.$$

With $\mu\to 0$, we have proved that for any increasing sequence of positive real numbers $(b_n)_n$ with $\lim_{n\to+\infty}b_n=+\infty$,

$$\lim_{n\to+\infty}\frac{1}{b_n}\sup_{\lambda>0}\lambda^q\sum_{k=1}^{n}\sigma_k^p\mathbf{1}_{|\theta_k|>\sigma_k\lambda}=0\quad\text{a.s.}$$

If the random sequence $A_n=\sup_{\lambda>0}\lambda^q\sum_{k=1}^n\sigma_k^p\mathbf{1}_{|\theta_k|>\sigma_k\lambda}$ were not bounded, we could construct an increasing function Φ, with $\lim_{n\to+\infty}\Phi(n)=+\infty$, such that $A_{\Phi(n)}>n$. By considering an increasing sequence $(b_n)_n$, with $b_{\Phi(n)}=n$, we obtain a contradiction. So, there exists a finite random variable Y such that

$$\forall\, n\ge 1,\qquad A_n\le Y,\quad\text{a.s.}$$

It implies that

$$\sup_{\lambda>0}\lambda^q\sum_k\sigma_k^p\mathbf{1}_{|\theta_k|>\sigma_k\lambda}<\infty\quad\text{a.s.}$$

The result is proved. □

So, to ensure that a sequence coming from the Bayesian model (M_1) belongs to $wl_{p,q}(\sigma)$ almost surely, we should not consider densities γ having tails heavier than those of Pareto(q)-distributions. In the wavelet framework, with special values for the σ_k's, Rivoirard [33] has already noted the strong connections between Pareto(q)-distributions and $wl_{p,q}(\sigma)$ spaces, since in Section 2.2 of that paper, least favorable priors for these spaces are presented and it is explained how these priors are built from Pareto(q)-variables.

Concluding remarks. In this paper, we discussed the modelling of sparsity. The form of our Bayesian model was the following:

$$\theta_k\sim w_{k,\varepsilon}\gamma_{k,\varepsilon}(\theta_k)+(1-w_{k,\varepsilon})\delta_0(\theta_k),\qquad k\ge 1.$$

Provided the tails of $\gamma_{k,\varepsilon}$ are exponential or heavier, the maxisets of the Bayes rules are $wl_{p,q}(\sigma)$ spaces that naturally measure the sparsity of a signal. So, our choice for the Bayesian modelling seems appropriate. It is all the more appropriate since this model enables us to build typical realizations of $wl_{p,q}(\sigma)$ spaces.

The main goal of this paper was to compare in the maxiset approach the performances of classical Bayes estimators: the posterior median and mean of our Bayesian model. We proved that for a large range of loss functions, the maxisets of these estimators coincide with the maxisets associated with thresholding estimators. These results have been established for the heteroscedastic white noise model (1), where the ξ_k's are assumed to be Gaussian and independent. It would be interesting to study the maxisets of the Bayes rules without these assumptions. It would also be interesting to try to find Bayes estimators that outperform $\hat{\theta}^{b_1}$ and $\hat{\theta}^{b_2}$ under the maxiset approach, if possible. Note that if the maxiset theory seems to provide advantages, the problem of optimality in this approach remains an entirely open issue. Can we introduce a meaningful notion of optimality? If yes, what are optimal estimators? Other natural questions arise: what do the maxisets become when the σ_k's are unknown and they are, for instance, estimated by using a Bayes approach? Do we obtain larger maxisets when the θ_k's are gathered in non-overlapping blocks, each of which is provided with a prior?

Since the outcome of the maxiset approach is a functional space (or a sequence space), we have not focused on the Bayes risks of the estimators. But, inspired by the maxiset point of view used in this paper, we could investigate the maximal set of prior distributions such that the associated Bayes risk of a given estimator achieves a prescribed rate. This provides an interesting topic for further research.

4. Appendix

In this section, $O_x(1)$ will denote any function of x that is bounded as $x \to +\infty$. We write $o_{\pi_\varepsilon}(1)$ for any function that is bounded by a function depending only on π_ε and that tends to 0 as π_ε tends to $+\infty$. Furthermore, ϕ denotes the density of a $(0,1)$ Gaussian variable and γ is the density introduced in (M_1). We shall exploit the following lemma:

Lemma 1. *For any $x > 0$ and $0 < \tau < x$, define*:

$$K_\tau(x) = \int_\tau^x \exp\Big(-\frac{1}{2}v^2\Big)\gamma(x-v)\,dv$$

and

$$I(x) = \int_0^{+\infty} \exp\Big(-\frac{1}{2}v^2 + xv\Big)\gamma(v)\,dv.$$

Under (H_1), there exist four positive constants M_2, M_3, C_1, and C_2 such that

$$M_2 \int_\tau^x \exp\Big(-\frac{1}{2}v^2 - Mv\Big)\,dv \le K_\tau(x)\gamma(x)^{-1} \le M_3 \int_\tau^{+\infty} \exp\Big(-\frac{1}{2}v^2 + Mv\Big)\,dv,$$

and

$$C_1 \le \liminf_{x\to+\infty} I(x)\gamma(x)^{-1}\phi(x) \le \limsup_{x\to+\infty} I(x)\gamma(x)^{-1}\phi(x) \le C_2 < \infty.$$

Proof. Under (H_1), since γ is positive, absolutely continuous, symmetric, and unimodal, it is easy to show that there exist two constants M_2 and M_3 such that,

$$\forall\,(a,b) \in \mathbb{R}^2, \quad M_2 \exp(-M|a-b|) \le \gamma(a)\gamma(b)^{-1} \le M_3 \exp(M|a-b|).$$

We immediately get the first inequality. Now, let us define $\forall\, x > 0$,

$$J(x) = \int_0^{+\infty} \exp\Big(-\frac{1}{2}v^2\Big)\gamma(x+v)\,dv.$$

As before,

$$M_2 \int_0^{+\infty} \exp\Big(-\frac{1}{2}v^2 - Mv\Big)\,dv \le J(x)\gamma(x)^{-1} \le M_3 \int_0^{+\infty} \exp\Big(-\frac{1}{2}v^2 + Mv\Big)\,dv.$$

By simple computations, we have:

$$I(x) = \exp\Big(\frac{1}{2}x^2\Big)\big(J(x) + K_0(x)\big),$$

which implies the result. $\square$

Now, let us give the proof of Propositions 2 and 3.

Proof of Proposition 2. Without loss of generality, we can assume that $x_k > 0$. Then, using Proposition 1,

$$\begin{aligned}
\hat\theta_k^{b_1}(x_k) = 0 &\iff \mathbb{P}(\theta_k > 0 \mid x_k) < \frac{1}{2} \\
&\iff 2w_\varepsilon \int_0^{+\infty} s_{k,\varepsilon}\phi\big(s_{k,\varepsilon}(x_k - \theta)\big)s_{k,\varepsilon}\gamma(s_{k,\varepsilon}\theta)\,d\theta \\
&\qquad < w_\varepsilon \int_{-\infty}^{+\infty} s_{k,\varepsilon}\phi\big(s_{k,\varepsilon}(x_k - \theta)\big)s_{k,\varepsilon}\gamma(s_{k,\varepsilon}\theta)\,d\theta + (1 - w_\varepsilon)s_{k,\varepsilon}\phi(s_{k,\varepsilon}x_k) \\
&\iff \int_0^{+\infty} s_{k,\varepsilon}\phi\big(s_{k,\varepsilon}(x_k - \theta)\big)s_{k,\varepsilon}\gamma(s_{k,\varepsilon}\theta)\,d\theta \\
&\qquad - \int_{-\infty}^{0} s_{k,\varepsilon}\phi\big(s_{k,\varepsilon}(x_k - \theta)\big)s_{k,\varepsilon}\gamma(s_{k,\varepsilon}\theta)\,d\theta < \pi_\varepsilon s_{k,\varepsilon}\phi(s_{k,\varepsilon}x_k).
\end{aligned}$$

Then $\hat\theta_k^{b_1}(x_k) = 0 \iff s_{k,\varepsilon}x_k \le t$, with t such that

$$\int_0^{+\infty} \exp\Big(-\frac{1}{2}u^2\Big)\gamma(u)\big(\exp(tu) - \exp(-tu)\big)\,du = \pi_\varepsilon.$$

We have that t is a function depending only on π_ε and as $\pi_\varepsilon \longrightarrow +\infty$,

$$\int_0^{+\infty} \exp\Big(-\frac{1}{2}u^2\Big)\gamma(u)\exp(tu)\,du = \pi_\varepsilon(1 + o_{\pi_\varepsilon}(1)).$$

Using Lemma 1, we have, for π_ε large enough,

$$\sqrt{2\log(\pi_\varepsilon)} \le t(\pi_\varepsilon) \le \sqrt{2\log(\pi_\varepsilon)}(1+o_{\pi_\varepsilon}(1)),$$

and the first statement of Proposition 2 is proved.

For the second statement, we assume that $s_{k,\varepsilon}x_k > 2t$, which implies that $\hat\theta_k^{b_1}(x_k) > 0$. Using (13), we have:

$$\begin{aligned}
\mathbb{P}\big(\theta_k \le \hat\theta_k^{b_1}(x_k) \mid x_k\big) = \frac{1}{2} &\iff 2w_\varepsilon \int_{-\infty}^{\hat\theta_k^{b_1}} \phi_k(x_k-\theta)\gamma_{k,\varepsilon}(\theta)\,d\theta + (1-w_\varepsilon)\phi_k(x_k)\\
&\qquad = w_\varepsilon \int_{-\infty}^{+\infty} \phi_k(x_k-\theta)\gamma_{k,\varepsilon}(\theta)\,d\theta\\
&\iff 2\int_{-\infty}^{s_{k,\varepsilon}\hat\theta_k^{b_1}} \exp\Big(s_{k,\varepsilon}x_k u - \frac{1}{2}u^2\Big)\gamma(u)\,du + \pi_\varepsilon\\
&\qquad = \int_{-\infty}^{+\infty} \exp\Big(s_{k,\varepsilon}x_k u - \frac{1}{2}u^2\Big)\gamma(u)\,du.
\end{aligned}$$

Using Lemma 1, since $s_{k,\varepsilon}x_k \ge 2t$, we prove easily that

$$\pi_\varepsilon I(s_{k,\varepsilon}x_k)^{-1} = o_{\pi_\varepsilon}(1).$$

Therefore,

$$2\int_{-\infty}^{s_{k,\varepsilon}\hat\theta_k^{b_1}} \exp\Big(s_{k,\varepsilon}x_k u - \frac{1}{2}u^2\Big)\gamma(u)\,du = I(s_{k,\varepsilon}x_k)(1+o_{\pi_\varepsilon}(1)).$$

By using again Lemma 1, it implies that there exists a positive constant V such that for π_ε large enough,

$$\begin{aligned}
&\int_{-\infty}^{s_{k,\varepsilon}\hat\theta_k^{b_1} - s_{k,\varepsilon}x_k} \exp\Big(-\frac{1}{2}v^2\Big)\gamma(v+s_{k,\varepsilon}x_k)\,dv \cdot \gamma(s_{k,\varepsilon}x_k)^{-1} \ge V\\
&\qquad \iff K_{s_{k,\varepsilon}x_k - s_{k,\varepsilon}\hat\theta_k^{b_1}}(s_{k,\varepsilon}x_k)\gamma(s_{k,\varepsilon}x_k)^{-1} \ge V,
\end{aligned}$$

in notations of Lemma 1. Finally, there exists a positive constant C such that

$$\limsup_{\pi_\varepsilon\to+\infty} |s_{k,\varepsilon}x_k - s_{k,\varepsilon}\hat\theta_k^{b_1}(x_k)|\mathbf{1}_{|s_{k,\varepsilon}x_k|>2t(\pi_\varepsilon)} \le C. \qquad \square$$

Proof of Proposition 3. Without loss of generality, we can assume that $x_k > 0$. We have:

$$\begin{aligned}
\hat\theta_k^{b_2}(x_k) &= \int_{-\infty}^{+\infty} \theta\gamma_{k,\varepsilon}^{\phi}(\theta \mid x_k)\,d\theta\\
&= \frac{\int_{-\infty}^{+\infty} s_{k,\varepsilon}\theta\phi(s_{k,\varepsilon}(x_k-\theta))s_{k,\varepsilon}\gamma(s_{k,\varepsilon}\theta)\,d\theta}{\int_{-\infty}^{+\infty} s_{k,\varepsilon}\phi(s_{k,\varepsilon}(x_k-\theta))s_{k,\varepsilon}\gamma(s_{k,\varepsilon}\theta)\,d\theta + \pi_\varepsilon s_{k,\varepsilon}\phi(s_{k,\varepsilon}x_k)}\\
&= \frac{1}{s_{k,\varepsilon}} \frac{\int_{-\infty}^{+\infty} u\exp\big(-\frac{1}{2}u^2 + s_{k,\varepsilon}x_k u\big)\gamma(u)\,du}{\int_{-\infty}^{+\infty} \exp\big(-\frac{1}{2}u^2 + s_{k,\varepsilon}x_k u\big)\gamma(u)\,du + \pi_\varepsilon}.
\end{aligned}$$

Set

$$I_1(x) = \int_{-\infty}^{+\infty} u \exp\Big(-\frac{1}{2}u^2 + xu\Big)\gamma(u)\,du$$

and

$$I_2(x) = \int_{-\infty}^{+\infty} \exp\Big(-\frac{1}{2}u^2 + xu\Big)\gamma(u)\,du,$$

which implies that

$$\hat{\theta}_k^{b_2}(x_k) = \frac{1}{s_{k,\varepsilon}} \frac{I_1(s_{k,\varepsilon}x_k)}{I_2(s_{k,\varepsilon}x_k) + \pi_\varepsilon}. \tag{16}$$

On the one hand, as $x \to +\infty$, using Lemma 1,

$$C_1 \le \liminf_{x\to+\infty} I_2(x)\gamma(x)^{-1}\phi(x) \le \limsup_{x\to+\infty} I_2(x)\gamma(x)^{-1}\phi(x) \le C_2. \tag{17}$$

On the other hand, Lemma 1 yields,

$$\begin{aligned}
\exp\Big(-\frac{1}{2}x^2\Big)I_1(x) &= \int_{-\infty}^{+\infty} u \exp\Big(-\frac{1}{2}(x-u)^2\Big)\gamma(u)\,du \\
&= \int_{-\infty}^{+\infty} (v+x)\exp\Big(-\frac{1}{2}v^2\Big)\gamma(v+x)\,dv \\
&= x\exp\Big(-\frac{1}{2}x^2\Big)I_2(x) + \int_{-\infty}^{+\infty} v\exp\Big(-\frac{1}{2}v^2\Big)\gamma(v+x)\,dv.
\end{aligned}$$

But it is easy to prove that

$$\begin{aligned}
\int_0^{+\infty} v\exp\Big(-\frac{1}{2}v^2\Big)\gamma(v+x)\,dv &= \gamma(x)O_x(1), \\
\int_{-x}^{0} v\exp\Big(-\frac{1}{2}v^2\Big)\gamma(v+x)\,dv &= \gamma(x)O_x(1),
\end{aligned}$$

and obviously,

$$\lim_{x\to\infty} \gamma(x)^{-1} \int_{-\infty}^{-x} v\exp\Big(-\frac{1}{2}v^2\Big)\gamma(v+x)\,dv = 0.$$

Therefore,

$$\exp\Big(-\frac{1}{2}x^2\Big)I_1(x) = x\exp\Big(-\frac{1}{2}x^2\Big)I_2(x) + \gamma(x)O_x(1). \tag{18}$$

Using (17), we obtain equation (6). Now, let us prove the second statement of Proposition 3. Suppose that $s_{k,\varepsilon}x_k \ge 2t(\pi_\varepsilon)$. Using (16) and (18),

$$\begin{aligned}
0 \le s_{k,\varepsilon}x_k - s_{k,\varepsilon}\hat{\theta}_k^{b_2}(x_k) &= s_{k,\varepsilon}x_k - \frac{I_1(s_{k,\varepsilon}x_k)}{I_2(s_{k,\varepsilon}x_k) + \pi_\varepsilon} \\
&= \frac{\pi_\varepsilon s_{k,\varepsilon}x_k + I_2(s_{k,\varepsilon}x_k)T(s_{k,\varepsilon}x_k)}{I_2(s_{k,\varepsilon}x_k) + \pi_\varepsilon},
\end{aligned}$$

where T is a bounded function. But we suppose that $s_{k,\varepsilon} x_k \geq 2t(\pi_\varepsilon)$. So, (17) implies that for π_ε large enough

$$I_2(s_{k,\varepsilon} x_k) \geq C_1 \gamma(s_{k,\varepsilon} x_k) \phi(s_{k,\varepsilon} x_k)^{-1}.$$

Therefore, by using again $s_{k,\varepsilon} x_k \geq 2t(\pi_\varepsilon)$,

$$\limsup_{\pi_\varepsilon \to \infty} \frac{\pi_\varepsilon s_{k,\varepsilon} x_k}{I_2(s_{k,\varepsilon} x_k)} < \infty,$$

which ends the proof of the proposition. $\square$

Acknowledgments. I would like to thank Professor Dominique Picard for a wealth of advice and encouragement and an anonymous referee for helpful comments and valuable suggestions.

References

[1] F. Abramovich, U. Amato, and C. Angelini, *On optimality of Bayesian wavelet estimators*, Scandinavian J. Statist. Theory and Appl., 31 (2004), 217–234.

[2] F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone, *Adapting to Unknown Sparsity by Controlling the False Discovery Rate*, Technical report, 2000.

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

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

[5] J. O. Berger, *Statistical Decision Theory: Foundations, Concepts, and Methods*, Springer, New York, 1980.

[6] P. Billingsley, *Probability and Measure*, Wiley, New York, 1995.

[7] L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov, *Oracle inequalities for inverse problems*, Ann. Statist., 30 (2002), 843–874.

[8] L. Cavalier and A. B. Tsybakov, *Sharp adaptation for inverse problems with random noise*, Probab. Theory Rel. Fields, 123 (2002), 323–354.

[9] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, *Adaptive Bayesian wavelet shrinkage*, J. Amer. Statist. Assoc., 92 (1997), 1413–1421.

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

[11] A. Cohen, R. A. DeVore, and R. Hochmuth, *Restricted nonlinear approximation*, Constructive Approximation, 16 (2000), 85–113.

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

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

[14] D. L. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Appl. Comput. Harmonic Anal., 2 (1995), 101–126.

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

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

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

[18] A. Goldenshluger and S. V. Pereverzev, *Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations*, Probab. Theory Rel. Fields, 118 (2000), 169–186.

[19] G. K. Golubev and R. Z. Khasminskii, *Statistical approach to some inverse boundary problems for partial differential equations*, Problems Inform. Transmission, 35 (1999), 51–66.

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

[21] I. M. Johnstone, *Wavelet shrinkage for correlated data and inverse problems: adaptivity results*, Statistica Sinica, 9 (1999), 51–83.

[22] I. M. Johnstone and B. W. Silverman, *Wavelet threshold estimators for data with correlated noise*, J. Roy. Statist. Soc. B, 59 (1997), 319–351.

[23] I. M. Johnstone and B. W. Silverman, *Empirical Bayes Approaches to Mixture Problems and Wavelet Regression*, Technical report, 1998.

[24] I. M. Johnstone and B. W. Silverman, *Empirical Bayes selection of wavelet thresholds*, 2002, available on http://www.stats.ox.ac.uk/ silverma/papers.html. To appear.

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

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

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

[28] A. P. Korostelev and A. B. Tsybakov, *Minimax Theory of Image Reconstruction*, Lecture Notes in Statist., Vol. 82, Springer, New York, 1993.

[29] B. A. Mair and F. H. Ruymgaart, *Statistical inverse estimation in Hilbert scales*, SIAM J. Appl. Math., 56 (1996), 1424–1444.

[30] M. B. Marcus and J. Zinn, *The bounded law of the iterated logarithm for the weighted empirical distribution process in the non-i.i.d. case*, Ann. Probab., 12 (1984), 335–360.

[31] Y. Meyer, *Wavelets and Operators*, Cambridge Univ. Press, Cambridge. 1992.

[32] S. Pereverzev and E. Schock, *Error estimates for band-limited spherical regularization wavelets in an inverse problem of satellite geodesy*, Inverse Problems, 15 (1999), 881–890.

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

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

[35] F. H. Ruymgaart, *A unified approach to inversion problems in statistics*, Math. Methods Statist., 2 (1993), 130–146.

[*Received February* 2003; *revised July* 2005]