# Maxisets for linear procedures

Vincent Rivoirard [1]

*Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII, France.*

**Abstract**

We study maxisets for linear procedures in the framework of the heteroscedastic white noise model. This enables us to point out the nature of the spaces naturally connected to these procedures and to compare the performance of linear and non linear estimates by comparing their respective maxisets.

*Key words:* Besov spaces, heteroscedastic white noise model, linear estimators, maxisets, rate of convergence

## 1 Introduction

One of the most traditional methods to measure the performance of an estimation procedure consists in investigating its minimax rates of convergence over large function classes. However, this minimax approach has undoubtedly two drawbacks: the choice for the function class is quite subjective and providing an estimator well adapted to the worst functions of this class seems

[1] Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII, UFR de Mathématiques, case 7012, 2, place Jussieu, 75251 Paris Cedex 05, France.
  *Email address:* `rivoirar@math.jussieu.fr` (Vincent Rivoirard).

too pessimistic for practical purposes. Among others, these are the reasons why Kerkyacharian and Picard (2000, 2002) focused on an alternative to the minimax setting: the maxiset approach that consists in investigating the maximal space (called maxiset) where an estimation procedure achieves a given rate of convergence. For instance, these authors applied this theory for two well known efficient procedures: wavelet thresholding and local bandwidth selection. Under some conditions, since the maxiset for the first procedure is embedded in the maxiset for the second one, they could conclude that local bandwidth selection is at least as efficient as the thresholding procedure. For the statistical model considered in this paper (see below), Rivoirard (2002) proved that the Bayesian procedures for the mean and the median associated to a very general Bayesian model, achieve the same performance as the thresholding procedure for the maxiset approach. We can note that this approach is less pessimistic and, above all, it provides function spaces directly connected to the estimation procedure. For instance, Kerkyacharian and Picard (1993) who pointed out a maxiset-type result for linear estimates under the model of density estimation, underlined the strong link between Besov spaces $\mathcal{B}_{\tau,p,\infty}$ and linear estimates for this statistical model.

In this paper, and as Kerkyacharian and Picard (2000) for thresholding procedures or Rivoirard (2002) for Bayesian procedures, we consider the following heteroscedastic white noise model:

$$x_k = \theta_k + \varepsilon \sigma_k \xi_k, \quad \xi_k \overset{iid}{\sim} \mathcal{N}(0,1), \quad k \in \mathbb{N}^*, \tag{1}$$

where $\theta = (\theta_k)_{k \in \mathbb{N}^*}$ is a sequence to be estimated and $\varepsilon > 0$. This heteroscedastic white noise model has extensively been used, for instance, as a sequence space framework for statistical inverse problems where $(\sigma_k^{-2})_{k \in \mathbb{N}^*}$ is the eigenvalues sequence of a known operator (see the references cited in section 2).

2

So, along this paper, we assume that $\sigma = (\sigma_k)_{k \in \mathbb{N}^*}$ is a known sequence of positive real numbers.

We can add that for inverse problems considered in the framework of the heteroscedastic white noise model, many authors consider linear estimates of the form

$$\hat{\theta}^\lambda(x) = (\lambda_k x_k)_{k \in \mathbb{N}^*}, \tag{2}$$

where $(\lambda_k)_{k \in \mathbb{N}^*}$, that can be interpreted as a smoothing parameter, is a sequence of non random weights that may depend on the parameters of the problem. Consequently, under the heteroscedastic white noise model, it seems relevant to determine maxisets for this linear procedure. It will be the main goal of this paper. Another goal will be to answer the following questions:

- is the nature of the maxisets obtained in the framework of the heteroscedastic white noise model the same as the one associated to the density estimation model ?
- can we compare the performances of linear and non linear procedures by comparing their respective maxisets ?

The results we obtain are the following. We consider the model (1) and we evaluate the risk for linear estimates for a weighted $l_p$-norm ($1 \leq p < \infty$), that can be the $l_p$-norm, a Sobolev norm or a Besov norm. The maxisets are obtained under very general assumptions on the $\lambda_k$'s that are verified by commonly used weights. Furthermore, if we consider the $l_p$-risk and a rate of convergence of the form $(\varepsilon \sqrt{\log(1/\varepsilon)})^{2s/(2s+1)}$, and if $\sum_k \sigma_k^p < \infty$ or if $\sigma$ has a polynomial growth or decay, the maxisets for the linear procedures are strictly embedded in the maxisets for the thresholding procedure. It means that linear estimators are outperformed by non linear ones. Then, we study maxisets

3

associated to the problem of the $L_p$-estimation $(1 < p < \infty)$ with projection weights of functions $f$ decomposed on a basis $(\psi_k)_{k \in \mathbb{N}^*}$: $f = \sum_{k \in \mathbb{N}^*} \theta_k \psi_k$. Once more, we prove from the maxiset point of view that if we have $\sigma_k = 1$, $k \in \mathbb{N}^*$, linear estimates are outperformed by thresholding ones. Let us note that in a different context (see Donoho, Johnstone, Kerkyacharian and Picard (1996)), the minimax approach allowed us to draw the same conclusion.

Finally, this paper precisely describes the nature of the function spaces naturally linked to linear procedures under the model (1). Indeed, whatever the choice of the risk, the maxisets that we obtain are of the same type. And for the $L_p$-risk, under some conditions on the basis, the maxisets are exactly Besov spaces $\mathcal{B}_{s,p,\infty}$ as for the model of density estimation.

The paper is organized as follows. Section 2 is devoted to the description of the statistical model and the function spaces involved in this paper and in section 3, we describe the maxisets obtained for linear estimates.

## 2   Model and function spaces

In the following, we assume that we are given the model (1) where $\sigma = (\sigma_k)_{k \in \mathbb{N}^*}$ is a known sequence of positive real numbers. This heteroscedastic white noise model can be viewed as a generalization of the classical white noise model (for which we have $\sigma_k = 1$, $k \in \mathbb{N}^*$) often considered by statisticians. Such a statistical model naturally appears when we have to estimate the solution of a linear operator equation $g = Af$, with noisy observations of $g$. Indeed, most of the time, to deal with such a problem, the singular value decomposition of $A$ is used and the sequence $(\sigma_k^{-2})_{k \in \mathbb{N}^*}$ is then the eigenvalues sequence of the operator $A^*A$, where $A^*$ is the adjoint of $A$. Some well-posed

4

inverse problems with noise can be reduced to (1) with $\sigma_k \to 0$ (as $k \to +\infty$). The condition $\sigma_k \to +\infty$ characterizes ill-posed problems. For instance, the sequences $(\sigma_k)_{k \in \mathbb{N}^*}$ associated to operators such as the Radon transform or the convolution (see Cavalier and Tsybakov (2002)) or operators for some elliptic differential equations (see Mair and Ruymgaart (1996)) have a polynomial growth. But, the $\sigma_k$'s may grow exponentially. See for instance Pereverzev and Schock (1999) who considered the problem of satellite geodesy or the inverse problems associated to partial differential equations such as the heat equation (see Mair and Ruymgaart (1996)).

As explained in Introduction, the final outcome of the maxiset approach is a function space where an estimation procedure achieves a given rate of convergence. So, let us introduce the following function spaces that will be useful throughout this paper.

**Definition 1** *For all $1 \leq p < \infty$ and $0 < \eta < \infty$, and for any non negative measure $\mu$ on $\mathbb{N}^*$, we set:*

$$B_{p,\infty}^{\eta}(\mu) = \left\{ \theta = (\theta_k)_{k \in \mathbb{N}^*} : \quad \sup_{\lambda > 0} \lambda^{p\eta} \sum_{k \geq \lambda} \mu(k) |\theta_k|^p < \infty \right\}.$$

*When $D$ is an interval of $\mathbb{R}^d$, $d \geq 1$, and $\mathcal{E} = (\psi_k)_{k \in \mathbb{N}^*}$ is an unconditional basis of $L_p(D)$, $1 < p < \infty$, we set*

$$\mathcal{B}_{p,\infty}^{\eta}(D, \mathcal{E}) = \left\{ f = \sum_k \theta_k \psi_k : \quad \sup_{\lambda > 0} \lambda^{\eta} \| \sum_{k \geq \lambda} \theta_k \psi_k \|_{L_p} < \infty \right\}.$$

Under some conditions on $D$ and on the basis $\mathcal{E}$, the space $\mathcal{B}_{p,\infty}^{\eta}(D, \mathcal{E})$ is a Besov space (see section 3). We shall see that the spaces $B_{p,\infty}^{\eta}(\mu)$ and $\mathcal{B}_{p,\infty}^{\eta}(D, \mathcal{E})$, that are of the same nature, are strongly connected to linear procedures. Now, let us give here the maxisets associated to a classical thresholding procedure that are very different from the spaces introduced previously. We

5

still consider the model (1) and for any $\varepsilon > 0$, we assume that we are given a real number $\Lambda_\varepsilon > 0$ only depending on $\varepsilon$ with $\Lambda_\varepsilon \to +\infty$ as $\varepsilon \to 0$, and we set:

$$\hat{\theta}_k^t(x_k) = \begin{cases} x_k \mathbf{1}_{|x_k| \geq \kappa_t t_{k,\varepsilon}} & \text{if } k < \Lambda_\varepsilon, \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\kappa_t$ is a constant and $t_{k,\varepsilon} = \sigma_k \varepsilon \sqrt{\log(1/\varepsilon)}$ is the universal threshold. The maxiset for this procedure has been obtained by Kerkyacharian and Picard (2000):

**Theorem 1** *Let* $1 \leq p < \infty$ *and* $0 < r < \infty$ *be two fixed real numbers. We suppose that* $\Lambda_\varepsilon = \left(\varepsilon \sqrt{\log(1/\varepsilon)}\right)^{-r}$, *for* $0 < \varepsilon \leq \varepsilon_0$ *where* $\varepsilon_0$ *is such that* $\varepsilon_0 \sqrt{\log(1/\varepsilon_0)} = 1$, *and there exists a positive constant* $C_t$, *such that*

$$\varepsilon^{\frac{\kappa_t^2}{16}} \log(1/\varepsilon)^{-\frac{1}{4}-\frac{p}{2}} \sum_{k < \Lambda_\varepsilon} \sigma_k^p \leq C_t, \quad 0 < \varepsilon \leq \varepsilon_0.$$

*Let* $0 < \nu < \infty$ *be a fixed real number. Then, if* $\kappa_t \geq \sqrt{2p}$, *there exists a positive constant* $C$ *such that for any* $0 < \varepsilon \leq \varepsilon_0$

$$\left(\mathbb{E}\|\hat{\theta}^t - \theta\|_{l_p}^p\right)^{\frac{1}{p}} \leq C \left(\varepsilon \sqrt{\log(1/\varepsilon)}\right)^{2\nu/(2\nu+1)}$$

*if and only if*

$$\theta \in wl_{p,p/(2\nu+1)}(\sigma) \cap B_{p,\infty}^{\frac{2\nu}{r(2\nu+1)}}(\mu),$$

*where* $\mu$ *is the counting measure on* $\mathbb{N}^*$ *and for any* $0 < q < p$,

$$wl_{p,q}(\sigma) = \left\{\theta = (\theta_k)_{k \in \mathbb{N}^*} : \sup_{\lambda > 0} \lambda^{q-p} \sum_k \mathbf{1}_{|\theta_k| \leq \lambda \sigma_k} |\theta_k|^p < \infty\right\}.$$

**Notations:** In the sequel, we shall note $[m]$ the integer such that $m - 1 \leq [m] < m$. Given two sets $\mathcal{A}_1$ and $\mathcal{A}_2$, the notation $\mathcal{A}_1 \subset \mathcal{A}_2$ means that $\mathcal{A}_1$ is embedded in $\mathcal{A}_2$, the notation $\mathcal{A}_1 \subsetneq \mathcal{A}_2$ means that this embedding is strict.

## 3 Maxisets and linear estimates

Along this section, we study linear estimators of the form given in (2) which risk is evaluated under weighted $l_p$-norms ($1 \leq p < \infty$): for any non negative measure $\mu$ on $\mathbb{N}^*$, we set:

$$R_{p,\mu}(\hat{\theta}^\lambda) = \left( \mathbb{E} \| \hat{\theta}^\lambda - \theta \|_{l_p(\mu)}^p \right)^{\frac{1}{p}} = \left( \sum_{k \in \mathbb{N}^*} \mu(k) \mathbb{E} |\lambda_k x_k - \theta_k|^p \right)^{\frac{1}{p}}.$$

Of course, if $\mu$ is the counting measure on $\mathbb{N}^*$, then the risk is studied under the $l_p$-norm. Other choices for $\mu$ and $p$ allow us to consider Sobolev or Besov norms. The maxisets obtained in this framework are the following:

**Theorem 2** *Let $1 \leq p < \infty$ and $0 < s < \infty$ be two real numbers. For any $m > 0$, we suppose we are given $(\lambda_k(m))_{k \in \mathbb{N}^*}$, a non increasing sequence of weights lying in $[0,1]$ such that*

- *$(A_1)$: there exists $\tilde{C} < 1$ such that for all $m > 0$ and $k \geq m$, $\lambda_k(m) \leq \tilde{C}$,*

- *$(A_2)$: there exists $C_s \in \mathbb{R}$ such that for any $m > 1$, with $\lambda_0 = 1$,*

$$\sum_{1 \leq k < m} (\lambda_{k-1}(m) - \lambda_k(m)) (1 - \lambda_k(m))^{p-1} \left( \frac{k}{m} \right)^{-ps} \leq C_s.$$

*Then, for any $\varepsilon > 0$, we suppose we are given $m_\varepsilon > 0$ such that*

- *$(M_1)$: there exists $\varepsilon$ such that $m_\varepsilon \leq 1$,*

- *$(M_2)$: $\varepsilon \to m_\varepsilon$ is continuous,*

- *$(M_3)$: $m_\varepsilon \to +\infty$ as $\varepsilon \to 0$.*

*We consider the model (1) and the estimator $\hat{\theta}^{\lambda(m_\varepsilon)} = (\lambda_k(m_\varepsilon) \times x_k)_{k \in \mathbb{N}^*}$. We suppose that there exists a positive constant $T_1$ such that for any $\varepsilon > 0$*

$$(\varepsilon m_\varepsilon^s)^p \sum_{k \in \mathbb{N}^*} \mu(k) \sigma_k^p (\lambda_k(m_\varepsilon))^p \leq T_1. \tag{3}$$

*Then, there exists a positive constant $C$ such that for any $\varepsilon > 0$*

$$R_{p,\mu}(\hat{\theta}^{\lambda(m_\varepsilon)}) \le C \, m_\varepsilon^{-s} \tag{4}$$

*if and only if $\theta \in B_{p,\infty}^s(\mu)$.*

**Proof of Theorem 2:** We have:

$$\sum_k \mu(k)(1 - \lambda_k)^p|\theta_k|^p \le 2^{p-1} \sum_k \mu(k) \left( \mathbb{E}|\lambda_k \, x_k - \theta_k|^p + \mathbb{E}|\lambda_k(x_k - \theta_k)|^p \right)$$

$$\le 2^{p-1} \left[ \mathbb{E}\|\hat{\theta}^\lambda - \theta\|_{l_p(\mu)}^p + \varepsilon^p \, \mathbb{E}|\xi_1|^p \sum_k \mu(k)\lambda_k^p \sigma_k^p \right].$$

Therefore, under (3), if (4) is true, for $\varepsilon > 0$,

$$(1 - \tilde{C})^p \sum_{k \ge m_\varepsilon} \mu(k)|\theta_k|^p \le \sum_k \mu(k)(1 - \lambda_k)^p|\theta_k|^p \le Km_\varepsilon^{-ps},$$

where $K$ is a positive constant. This implies that $\theta \in B_{p,\infty}^s(\mu)$.

Conversely, let us assume that $\theta \in B_{p,\infty}^s(\mu)$, then, there exists a positive constant $C$, such that for $k \ge 1$, $\sum_{l \ge k} \mu(l)|\theta_l|^p \le C \, k^{-ps}$. We have, for $\varepsilon > 0$:

$$\sum_{k < m_\varepsilon} \mu(k)(1 - \lambda_k)^p|\theta_k|^p \le pC \sum_{k=1}^{[m_\varepsilon]} (\lambda_{k-1} - \lambda_k)(1 - \lambda_k)^{p-1}k^{-ps} \le pCC_s m_\varepsilon^{-ps}.$$

Then, for $\varepsilon > 0$,

$$\mathbb{E}\|\hat{\theta}^{\lambda(m_\varepsilon)} - \theta\|_{l_p(\mu)}^p \le 2^{p-1} \left[ \sum_k \mu(k)(1 - \lambda_k)^p|\theta_k|^p + \sum_k \mu(k)\mathbb{E}|\lambda_k(x_k - \theta_k)|^p \right]$$

$$\le 2^{p-1} \left[ \sum_{k < m_\varepsilon} \mu(k)(1 - \lambda_k)^p|\theta_k|^p + \sum_{k \ge m_\varepsilon} \mu(k)|\theta_k|^p \right]$$

$$+ 2^{p-1}\varepsilon^p\mathbb{E}|\xi_1|^p \sum_k \mu(k)\lambda_k^p\sigma_k^p$$

$$\le Km_\varepsilon^{-ps},$$

where $K$ is a positive constant. Theorem 2 is proved. □

To shed light on this result, we consider the $l_p$-norm (we note $\mu$ the count-

ing measure on $\mathbb{N}^*$) and $\sigma_k = k^b$, $b > -1/p$. To check (3), we take $m_\varepsilon = \varepsilon^{-p/(ps+pb+1)}$, and with $\lambda_k(m) = \mathbb{1}_{k<m}$, we obtain:

$$\text{for any } \varepsilon > 0, \quad R_{p,\mu}(\hat{\theta}^{\lambda(m_\varepsilon)}) \leq C\varepsilon^{\frac{ps}{ps+pb+1}} \iff \theta \in B^s_{p,\infty}(\mu).$$

In particular, when $p = 2$, $\varepsilon^{2s/(2s+2b+1)}$ is the minimax rate achieved on the balls of the Sobolev space $W_s$ (see Cavalier and Tsybakov (2002)) but also on the balls of $B^s_{2,\infty}(\mu)$ that contains $W_s$. Let us note that $B^s_{2,\infty}(\mu)$ can be identified with the Besov space $\mathcal{B}_{s,2,\infty}([0,1])$, and when $b = 0$ ($\sigma_k = 1$, $k \in \mathbb{N}^*$), we obtain the classical rate $\varepsilon^{2s/(2s+1)}$. Of course, this theorem can be applied with the most commonly used weights, such as

- the projection weights: $\lambda_k^{(1)}(m) = \mathbb{1}_{k<T(m)}$,
- the Tikhonov-Phillips weights: $\lambda_k^{(2)}(m) = \frac{1}{1+(k/T(m))^\alpha}$,
- the Pinsker weights: $\lambda_k^{(3)}(m) = (1 - (k/T(m))^\alpha)_+$,

where the function $T$ and the parameter $\alpha > 0$ can be chosen to ensure that the weights $\lambda_k^{(i)}(m)$ verify assumptions $(A_1)$ and $(A_2)$ of Theorem 2.

Now, let us compare linear estimates with adaptive thresholding estimates from the maxiset point of view for the $l_p$-norm and with the rate of convergence $(\varepsilon\sqrt{\log(1/\varepsilon)})^{2s/(2s+1)}$, $(0 < s < \infty)$. Most of the time, the thresholding procedure of Theorem 1 is applied with $r = 2$. Nevertheless, we shall use the results of Theorem 1 with $r \geq 2$ and $\kappa_t \geq \sqrt{2p}$ and those of Theorem 2 with the projection weights $\lambda_k(m) = \mathbb{1}_{k<m}$ and we take $m_\varepsilon = (\varepsilon\sqrt{\log(1/\varepsilon)})^{-2/(2s+1)}$. We note $\mu$ the counting measure on $\mathbb{N}^*$.

First, let us assume that $\sum_k \sigma_k^p < \infty$. This assumption yields that condition (3) of Theorem 2 is fulfilled and straightforward computations show that

$$B^s_{p,\infty}(\mu) \subsetneqq B^{\frac{2s}{r(2s+1)}}_{p,\infty}(\mu) \cap wl_{p,p/(2s+1)}(\sigma) = B^{\frac{2s}{r(2s+1)}}_{p,\infty}(\mu),$$

9

which means that linear estimators are outperformed by non linear ones.

Now, let us suppose that there exist $c_1 > 0$, $c_2 > 0$ and $b \in \mathbb{R}$ such that,

$$c_1 k^b \leq \sigma_k \leq c_2 k^b, \quad k \in \mathbb{N}^*.$$

This case includes the homoscedastic case ($\sigma_k = 1$, $k \in \mathbb{N}^*$). Condition (3) of Theorem 2 is fulfilled if and only if $2(bp + 1) \leq p$ and we take $\kappa_t \geq \sqrt{\max(2p; 16r(pb + 1))}$.

**Lemma 1** *Since $r \geq 2$ and $2(bp + 1) \leq p$,*

$$B_{p,\infty}^s(\mu) \subsetneq B_{p,\infty}^{\frac{2s}{r(2s+1)}}(\mu) \cap wl_{p,p/(2s+1)}(\sigma).$$

**Proof of Lemma 1:** When $bp < -1$, $\sum_k \sigma_k^p < \infty$, so we assume that $bp \geq -1$. Then, for $\theta = (\theta_k)_{k \in \mathbb{N}^*}$, $0 < \lambda \leq 1$, $N_\lambda \geq 1$,

$$\lambda^{-\frac{2ps}{2s+1}} \sum_k \mathbf{1}_{|\theta_k| \leq \sigma_k \lambda} |\theta_k|^p \leq \lambda^{-\frac{2ps}{2s+1}} \sum_{k \geq N_\lambda} |\theta_k|^p + c_2^p \lambda^{\frac{p}{2s+1}} \sum_{k < N_\lambda} k^{pb}.$$

With $N_\lambda = \lambda^{-2/(2s+1)}$, and if $\theta \in B_{p,\infty}^s(\mu)$,

$$\sup_{\lambda \leq 1} \lambda^{-\frac{2ps}{2s+1}} \sum_k \mathbf{1}_{|\theta_k| \leq \sigma_k \lambda} |\theta_k|^p < \infty.$$

Furthermore, for $\theta = (\theta_k)_{k \in \mathbb{N}^*}$,

$$\sup_{\lambda \geq 1} \lambda^{-\frac{2ps}{2s+1}} \sum_k \mathbf{1}_{|\theta_k| \leq \sigma_k \lambda} |\theta_k|^p \leq \sum_k |\theta_k|^p.$$

So, $B_{p,\infty}^s(\mu) \subset wl_{p,p/(2s+1)}(\sigma)$. Now, since $\frac{p}{2} - pb \geq 1$, there exists $u > 0$ such that

$$\frac{1}{2} \leq \frac{(pb + u^{-1})(2s + 1)}{p} < \frac{(2s + 1)}{2},$$

so we set

$$v = \frac{(upb + 1)(2s + 1)}{p} \in \left[\frac{u}{2}, \frac{u(2s + 1)}{2}\right),$$

10

and for $l \in \mathbb{N}^*$ $\theta_l = \sigma_l \times \frac{1}{k^v}$ if there exists $k \in \mathbb{N}^*$ such that $l = [k^u + 1]$, and $\theta_l = 0$ otherwise. It is easy to see that $\theta \in B_{p,\infty}^{\frac{2s}{r(2s+1)}}(\mu) \cap wl_{p,p/(2s+1)}(\sigma)$, but $\theta \notin B_{p,\infty}^s(\mu)$, which proves the lemma. □

The result of Lemma 1 means that linear estimators are outperformed by non linear ones. Let us note that the assumption $r \geq 2$ is essential to claim this statement. Indeed, if $r < 2$ and for small values of $s$, $B_{p,\infty}^{\frac{2s}{r(2s+1)}}(\mu) \subsetneq B_{p,\infty}^s(\mu)$. Finally, if $\sigma$ has an exponential growth, then we could apply neither the thresholding procedure, nor the linear one with $m_\varepsilon = (\varepsilon\sqrt{\log(1/\varepsilon)})^{-2/(2s+1)}$. But with a logarithmic rate of convergence and by adapting the results of Theorem 1, we expect from the maxiset point of view, thresholding procedures to be preferable to linear ones. Let us also note that these last results remain valid for Pinsker and Tikhonov-Phillips weights for an appropriate choice of $\alpha$.

Now, we estimate functions of $L_p(D)$, where $D = [0,1]^d$ or $D = \mathbb{R}^d$. For this purpose, we use a wavelet basis of $L_2(D)$ noted $\mathcal{E} = (\psi_k)_{k \in \mathbb{N}^*}$. More precisely, we assume that $\mathcal{E}$ is a wavelet-tensor product basis constructed on compactly supported wavelets. So, if $1 < p < \infty$, Meyer (1992) proved that $\mathcal{E}$ is an unconditional basis of $L_p(D)$. Furthermore, we assume that $(\sigma_k \psi_k)_{k \in \mathbb{N}^*}$ verifies the following superconcentration inequality: for any $0 < r < \infty$, there exists a constant $C(p,r)$ such that for any $F \subset \mathbb{N}^*$,

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

**Remark 1** *Kerkyacharian and Picard (2000) noted that if the $\sigma_k$'s depend only on the resolution levels and if at level $j$, the noise level is proportional to $2^{bj}$, with $b > -\frac{d}{2}$, then $(\sigma_k \psi_k)_{k \in \mathbb{N}^*}$ verifies this superconcentration inequality.*

Under these hypotheses, we can get the maxiset associated to the linear procedure with projection weights for the $L_p$-risk. We have:

**Theorem 3** *Let $0 < s < \infty$ be a fixed real number. For any $\varepsilon > 0$, we suppose we are given $m_\varepsilon > 0$ such that Assumptions $(M_1)$, $(M_2)$ and $(M_3)$ are checked (see Theorem 2). To estimate the function $f = \sum_{k \in \mathbb{N}^*} \theta_k \psi_k$ under the model (1), we suppose that there exists a positive constant $T_2$ such that*

$$(\varepsilon m_\varepsilon^s)^p \sum_{k < m_\varepsilon} \int |\psi_k|^p \sigma_k^p \leq T_2, \quad \text{for any } \varepsilon > 0.$$

*If $\lambda_k(m) = \mathbf{1}_{k<m}$, there exists a positive constant $C$ such that for any $\varepsilon > 0$*

$$\left[ \mathbb{E} \| \sum_k (\lambda_k(m_\varepsilon) x_k - \theta_k) \psi_k \|_{L_p}^p \right]^{1/p} \leq C \, m_\varepsilon^{-s}$$

*if and only if $f = \sum_k \theta_k \psi_k \in \mathcal{B}_{p,\infty}^s(D, \mathcal{E})$.*

**Remark 2** *The proof of Theorem 3 uses the particular form of the projection weights that can take only two values: 0 or 1. To obtain a similar result for other weights $\lambda_k(m_\varepsilon)$, it is enough to check that for any $\varepsilon > 0$, $(\lambda_k(m_\varepsilon) \sigma_k \psi_k)_{k \in \mathbb{N}^*}$ verifies the superconcentration inequality.*

The proof of this theorem only requires some of the arguments of the proof of Theorem 2 and some of the arguments used by Kerkyacharian and Picard (2000) for proving Theorems 5.1 and 5.2. So, it is omitted. To shed light on this result, we assume until the end of this section that $\sigma_k = 1$, $k \in \mathbb{N}^*$, and $\mathcal{E}$ is a wavelet-tensor basis on $[0,1]^d$, denoted $\mathcal{E} = (\psi_{j\mathbf{l}})_{j \in \mathbb{N}, \mathbf{l} \in A_j}$, where for any $j \in \mathbb{N}$, $A_j$ is a set with cardinality proportional to $2^{jd}$. Then, under standard properties of regularity and moment vanishing of the wavelet (see Meyer (1992)), $\mathcal{B}_{p,\infty}^{\eta/d}([0,1]^d, \mathcal{E})$ can be identified with the Besov space

$$\mathcal{B}_{\eta,p,\infty}([0,1]^d) = \left\{ f = \sum_{j \in \mathbb{N}} \sum_{\mathbf{l} \in A_j} \beta_{j\mathbf{l}} \psi_{j\mathbf{l}} : \quad \sup_{J \geq 0} 2^{J\eta} \| \sum_{j \geq J} \sum_{\mathbf{l} \in A_j} \beta_{j\mathbf{l}} \psi_{j\mathbf{l}} \|_{L_p} < \infty \right\},$$

and this shows that whatever the choice of the model (density estimation or heteroscedastic white noise model), Besov spaces $\mathcal{B}_{\eta,p,\infty}$ are strongly connected

12

to linear procedures. Furthermore, using $\int |\psi_{j\mathbf{l}}|^p \approx 2^{jd(\frac{p}{2}-1)}$, the upper bound obtained for $\mathcal{B}_{s,p,\infty}([0,1]^d)$ and with the $L_p$-risk is $\varepsilon^{2s/(d+2s)}$. What is more, and as previously, using Theorems 5.1, 5.2 and 6.2 of Kerkyacharian and Picard (2000), we show that linear estimators are outperformed by non linear ones for the rate $(\varepsilon\sqrt{\log(1/\varepsilon)})^{2s/(d+2s)}$. With the classical calibration $\varepsilon \approx n^{-1/2}$, this rate is, up to a logarithmic term, the minimax rate that appears for the model of density estimation (see Donoho, Johnstone, Kerkyacharian and Picard (1996)) or for the model of non parametric regression (see Donoho, Johnstone, Kerkyacharian and Picard (1997)). Let us note that for the model (1), Kerkyacharian and Picard (2000) provided an adaptive thresholding procedure that achieves the rate $(\varepsilon\sqrt{\log(1/\varepsilon)})^{2s/(d+2s)}$ on the balls of $\mathcal{B}_{s,p,\infty}([0,1]^d)$.

## References

Cavalier, L., Tsybakov, A.B., 2002. Sharp adaptation for inverse problems with random noise. Probab. Theory Related Fields 123, 323–354.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D., 1996. Density estimation by wavelet thresholding. Ann. Statist. 24, 508–539.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D., 1997. Universal near minimaxity of wavelet shrinkage, in: Festschrift for Lucien Le Cam, Springer, New York, pp. 183–218.

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

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

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

Mair, B.A., Ruymgaart, F.H., 1996. Statistical inverse estimation in Hilbert scales. SIAM J. Appl. Math. 56, 1424–1444.

Meyer, Y., 1992. Wavelets and Operators. Cambridge University Press, Cambridge.

Pereverzev, S.V., Schock, E., 1999. Error estimates for band-limited spherical regularization wavelets in an inverse problem of satellite geodesy. Inverse problems 15, 881–890.

Rivoirard, V., 2002. Bayesian modelization of sparse sequences and maxisets for Bayes rules. Preprint Laboratoire de Probabilités et Modèles aléatoires, Universités Paris VI et Paris VII. Submitted to Math. Methods Statist.