



**HAL**  
open science

# Non asymptotic minimax rates of testing in signal detection with heterogeneous variances

Béatrice Laurent, Jean-Michel Loubes, Clément Marteau

► **To cite this version:**

Béatrice Laurent, Jean-Michel Loubes, Clément Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. 2009. hal-00440825v1

**HAL Id: hal-00440825**

**<https://hal.science/hal-00440825v1>**

Preprint submitted on 11 Dec 2009 (v1), last revised 9 Feb 2010 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Non asymptotic minimax rates of testing in signal detection with heterogeneous variances

B. Laurent, J.M. Loubes, C. Marteau \*

## Abstract

The aim of this paper is to establish non-asymptotic minimax rates of testing for goodness-of-fit hypotheses in a heteroscedastic setting. More precisely, we deal with sequences  $(Y_j)_{j \in J}$  of independent Gaussian random variables, having mean  $(\theta_j)_{j \in J}$  and variance  $(\sigma_j)_{j \in J}$ . The set  $J$  will be either finite or countable. In particular, such a model covers the inverse problem setting where few results in test theory have been obtained. The rates of testing are obtained with respect to  $l_2$  and  $l_\infty$  norms, without assumption on  $(\sigma_j)_{j \in J}$  and on several functions spaces. Our point of view is completely non-asymptotic.

**AMS subject classifications:** 62G05, 62G20

**Key words and phrases:** Goodness-of-fit tests, heterogeneous variances, inverse problems.

## 1 Introduction

We consider the following statistical model :

$$Y_j = \theta_j + \sigma_j \epsilon_j, \quad j \in J, \quad (1.1)$$

where  $\theta = (\theta_j)_{j \in J}$  is unknown,  $\sigma = (\sigma_j)_{j \in J}$  is assumed to be known, and the variables  $(\epsilon_j)_{j \in J}$  are i.i.d. standard normal variables. The set  $J$  is either  $\{1, \dots, N\}$  for some  $N \in \mathbb{N}^*$  (which corresponds to a Gaussian regression model) or  $\mathbb{N}^*$  (which corresponds to the Gaussian sequence model). The sequence  $\theta$  has to be tested from the observations  $(Y_j)_{j \in J}$ . The particular case  $\sigma_j = \sigma$  for all  $j \in J$  corresponds to a classical statistical model: the variance of the observations is always the same. It has been widely considered in the literature, both for test and estimation approaches. In this paper, we consider a slightly different setting in the sense that the variance of the sequence is allowed to depend on  $j$ . For instance, the model (1.1) can describe inverse problems. Indeed, for a linear operator  $T$  on an Hilbert space  $H$  with inner product  $(\cdot, \cdot)$ , consider an unknown function  $f$  indirectly observed in a Gaussian white noise model

$$Y(g) = (Tf, g) + \sigma \epsilon(g), \quad g \in H \quad (1.2)$$

where  $\epsilon(g)$  is a centered Gaussian variable with variance  $\|g\|^2 := (g, g)$ . If  $T$  is assumed to be compact, it admits a singular value decomposition (SVD)  $(b_j, \psi_j, \phi_j)_{j \geq 1}$  in the sense that

$$T\phi_j = b_j\psi_j, \quad T^*\psi_j = b_j\phi_j,$$

---

\*Institut de Mathématiques de Toulouse, INSA de Toulouse, Université de Toulouse.

with  $T^*$  the adjoint operator of  $T$ . Hence considering the observations  $Y(\psi_j)$ , model (1.2) becomes

$$Z_j = b_j \theta_j + \sigma \epsilon_j, \quad j \in \mathbb{N}^*, \quad (1.3)$$

with  $\epsilon_j = \epsilon(\psi_j)$  and  $(Tf, \psi_j) = b_j \theta_j$  and  $\theta_j = (f, \phi_j)$ . This model is often considered in the inverse problem literature, see eg [6]. Setting  $Y_j = b_j^{-1} Z_j$  and  $\sigma_j = \sigma b_j^{-1}$  for all  $j \in \mathbb{N}^*$ , we obtain (1.1). Hence inference on observations from model (1.1) provides the same results for inverse problems.

For all  $\theta \in l_2(J)$ , we set  $\|\theta\|_2^2 = \sum_{j \in J} \theta_j^2$  and for all  $\theta \in l_\infty(J)$ ,  $\|\theta\|_\infty = \sup_{j \in J} |\theta_j|$ . The purpose of this paper is to determine the rates of testing to test the hypothesis " $\theta = 0$ " against the alternative " $\|\theta\|_q \geq \rho$ " where  $q = 2$  or  $q = \infty$ . More precisely, given  $\alpha, \beta \in ]0, 1[$ , a level  $\alpha$  test  $\Phi_\alpha$  of the null hypothesis " $\theta = 0$ ", and a class of vectors  $\mathcal{F} \subset l_q(J)$ , we define the uniform separation rate  $\rho_q(\Phi_\alpha, \mathcal{F}, \beta)$  of the test  $\Phi_\alpha$  over the class  $\mathcal{F}$  with respect to the  $l_q$  norm as the smallest radius  $\rho$  such that the test guarantees a power greater than  $1 - \beta$  for all alternatives  $\theta \in \mathcal{F}$  such that  $\|\theta\|_q \geq \rho$ . More formally

$$\rho_q(\Phi_\alpha, \mathcal{F}, \beta) = \inf \left\{ \rho > 0, \inf_{\theta \in \mathcal{F}, \|\theta\|_q \geq \rho} \mathbb{P}_\theta(\Phi_\alpha \text{ rejects}) > 1 - \beta \right\}.$$

Let us now define  $\rho_q(\mathcal{F}, \alpha, \beta)$  as the infimum over all level  $\alpha$  test  $\Phi_\alpha$  of the quantity  $\rho_q(\Phi_\alpha, \mathcal{F}, \beta)$ . This quantity will be called the  $(\alpha, \beta)$  minimax rate of testing over the class  $\mathcal{F}$ . The aim of the paper is to determine this minimax rate of testing over various classes of alternatives  $\mathcal{F}$ , for the test of null hypothesis " $\theta = 0$ " in Model (1.1) with respect to the  $l_2$  and  $l_\infty$  norms.

The main reference for computing minimax rates of testing over non parametric alternatives is the series of paper due to Ingster[10], where various statistical models and a wide range of sets of alternatives are considered. Lepski and Spokoiny [13] obtained minimax rates of testing over Besov bodies  $\mathcal{B}_{s,p,q}(R)$  in the irregular case (when  $0 < p < 2$ ), see also Ingster and Suslina [11]. Ermakov [9] determines a family of asymptotic minimax tests for testing that the signal belongs to a parametric set against nonparametric sets of alternatives in the heteroscedastic Gaussian white noise. In all these references, asymptotic minimax rates of testing are established. In Model (1.1), with  $\sigma_j = \sigma$  for all  $j \in J$ , Baraud [2] take a non asymptotic point of view, which means that the noise level  $\sigma$  is not assumed to converge towards 0. This is the point of view that we adopt in this paper. We give a precise expression of the dependency of the minimax rates of testing with respect to the sequence  $(\sigma_j)_{j \in J}$ . The particular cases of interest correspond to polynomial and exponentially increasing sequences, which in the case of (1.3) leads to the mildly and severely ill-posed inverse problems.

The paper is organised as follows. In Section 2, we provide lower bounds for the minimax separation rate over classes of vectors  $\theta$  with a finite number of non-zero coefficients, which yet covers sparse signals. In Section 3, we determine upper bounds for those minimax rates. In Section 4, we compute minimax rates of testing over ellipsoids and  $l_p$  balls. The proofs are gathered in Section 5.

To end this introduction, let us define some notations. Let  $Y = (Y_j)_{j \in J}$  obey to Model (1.1). We denote by  $\theta$  the vector (or sequence)  $(\theta_j)_{j \in J}$  and by  $\mathbb{P}_\theta$  the distribution of  $Y$ . All along the

paper, we consider the test of null hypothesis " $\theta = 0$ ". Let  $\alpha \in ]0, 1[$  be some prescribed level. A test function  $\Phi_\alpha$  is a measurable function of the observation  $Y$ , with values in  $\{0, 1\}$ . The null hypothesis is accepted if  $\Phi_\alpha = 0$  and rejected if  $\Phi_\alpha = 1$ .

## 2 Lower bounds

### 2.1 Lower bounds in $l_2$ norm

We consider Model (1.1). In this section, we generalize the results obtained by Baraud [2] in an homoscedastic model to the heteroscedastic Model (1.1).

We first give a lower bound for the minimax rate of testing over the set  $S_D$ , defined for all  $D \geq 1$  by

$$S_D = \{\theta \in l_2(J), \forall j > D, \theta_j = 0\}.$$

When  $J = \{1, \dots, N\}$ , we assume that  $D \leq N$ .

**Proposition 1** *Assume that  $Y = (Y_j)_{j \in J}$  obeys to Model (1.1). Let  $\beta \in ]0, 1 - \alpha[$ ,  $c(\alpha, \beta) = (2 \ln(1 + 4(1 - \alpha - \beta)^2))^{1/2}$  and*

$$\rho_D^2 = c(\alpha, \beta) \left( \sum_{j=1}^D \sigma_j^4 \right)^{1/2}.$$

*The following result holds:*

$$\forall \rho \leq \rho_D, \inf_{\Phi_\alpha} \sup_{\theta \in S_D, \|\theta\|_2 = \rho} \mathbb{P}_\theta(\Phi_\alpha = 0) \geq \beta.$$

*This implies that the minimax rate of signal detection over  $S_D$  with respect to the  $l_2$  norm satisfies*

$$\rho_2(S_D, \alpha, \beta) \geq \rho_D.$$

**Comment :** The results obtained in Proposition 1 coincide with the lower bound established by Baraud [2] in the homoscedastic model ( $\sigma_j = \sigma \quad \forall j \in J$ ).

Let us now consider the problem of sparse signal detection. Let  $k, n \in \mathbb{N}^*$  with  $k \leq n$ . When  $J = \{1, \dots, N\}$ , we assume that  $n \leq N$ . We want to obtain lower bounds for the minimax separation rate of signal detection over the set  $\mathcal{S}_{k,n}$  defined by

$$\mathcal{S}_{k,n} = \{\theta \in l_2(J), \forall j > n, \theta_j = 0, \text{Card } \{j \leq n, \theta_j \neq 0\} \leq k\}. \quad (2.4)$$

**Theorem 1** *Assume that  $Y = (Y_j)_{j \in J}$  obeys to Model (1.1). Let  $\sigma_{(1)} \leq \sigma_{(2)} \leq \dots \leq \sigma_{(n)}$ , we define for all  $l \in \{0, \dots, n - k\}$ ,*

$$\Sigma_{l,k}^2 = \sum_{j=l+1}^{l+k} \sigma_{(j)}^2. \quad (2.5)$$

Let  $\beta \in ]0, 1 - \alpha[$ , such that  $\alpha + \beta \leq 59\%$ . Let

$$\rho_{k,n}^2 = \left[ \max_{0 \leq l \leq n-k} \Sigma_{l,k}^2 \ln \left( 1 + \frac{n-l}{k^2} \vee \sqrt{\frac{n-l}{k^2}} \right) \vee \left( \sum_{j=n-k+1}^n \sigma_{(j)}^4 \right)^{1/2} \right]. \quad (2.6)$$

For all level  $\alpha$  test  $\Phi_\alpha$ , there exists  $\theta \in \mathcal{S}_{k,n}$  such that  $\|\theta\|_2 \geq \rho_{k,n}$  and  $\mathbb{P}_\theta(\Phi_\alpha = 1) \leq 1 - \beta$ . This implies that the minimax rate of signal detection over  $\mathcal{S}_{k,n}$  with respect to the  $l_2$  norm satisfies

$$\rho_2(\mathcal{S}_{k,n}, \alpha, \beta) \geq \rho_{k,n}.$$

**Comments :** Let us consider three cases concerning the behaviour of the sequence  $(\sigma_j)_{j \in J}$ .

1. In the homoscedastic case,  $\sigma_j = \sigma$  for all  $j \in J$ . In this case,  $\Sigma_{l,k}^2 = \sigma^2 k$  for all  $l$  and, taking  $l = 0$ , we obtain that

$$\rho_{k,n}^2 \geq \sigma^2 k \ln \left( 1 + \frac{n}{k^2} \vee \sqrt{\frac{n}{k^2}} \right),$$

which corresponds to the lower bound established by Baraud [2].

2. When  $\Sigma_{n/2,k}^2 \geq C \Sigma_{n-k,k}^2$  for some absolute constant  $C$  (independent of  $k$  and  $n$ ), we obtain that

$$\rho_{k,n}^2 \geq \left[ \Sigma_{n-k,k}^2 \ln \left( 1 + \frac{n}{2k^2} \vee \sqrt{\frac{n}{2k^2}} \right) \vee \left( \sum_{j=n-k+1}^n \sigma_{(j)}^4 \right)^{1/2} \right]. \quad (2.7)$$

At the price of a factor 2 in the logarithm ( $n$  is replaced by  $n/2$ ), the variance term appearing in the lower bound for  $\rho_{k,n}^2$  is  $\Sigma_{n-k,k}^2$  which corresponds to the largest possible variance for a set of cardinality  $k$  in  $\{1, \dots, n\}$ , indeed

$$\Sigma_{n-k,k}^2 = \max_{m \in \mathcal{M}_{k,n}} \sum_{j \in m} \sigma_j^2,$$

where  $\mathcal{M}_{k,n}$  denotes the set of all subsets of  $\{1, \dots, n\}$  with cardinality  $k$ .

This situation occurs for example when  $(\sigma_j)_{j \in J}$  grows at a polynomial rate :  $\sigma_j = j^{2\gamma}$  for some  $\gamma > 0$ . In this case,

$$\Sigma_{n-k,k}^2 \leq k n^{2\gamma}, \quad \Sigma_{n/2,k}^2 \geq k \left( \frac{n}{2} \right)^{2\gamma} \geq \frac{1}{2^{2\gamma}} \Sigma_{n-k,k}^2.$$

3. When  $(\sigma_j)_{j \in J}$  grows at an exponential rate :  $\sigma_j = \exp(\gamma j)$  for some  $\gamma > 0$ , we obtain that  $\rho_{k,n}^2 \geq \sigma_{(n)}^2$

## 2.2 Lower bounds in $l_\infty$ norm

In this section, we give lower bounds for the minimax rate of signal detection in Model 1.1 with respect to the  $l_\infty$  norm. These bounds can be derived from the Theorem 1. Indeed, let  $\mathcal{S}_{k,n}$  be defined by (2.4). For  $\theta \in \mathcal{S}_{1,n}$ ,  $\|\theta\|_2 = \|\theta\|_\infty$ . Hence, denoting by  $\rho_\infty(\mathcal{F}, \alpha, \beta)$  the  $(\alpha, \beta)$  minimax rate of testing over the class  $\mathcal{F}$  with respect to the  $l_\infty$  norm, we obtain that

$$\rho_\infty(\mathcal{S}_{1,n}, \alpha, \beta) = \rho_2(\mathcal{S}_{1,n}, \alpha, \beta).$$

By definition of the  $(\alpha, \beta)$  minimax rate of testing it is obvious that if  $\mathcal{S} \subset \mathcal{S}'$ , then  $\rho_\infty(\mathcal{S}) \leq \rho_\infty(\mathcal{S}')$ .  $\mathcal{S}_{1,n} \subset \mathcal{S}_n$  and  $\mathcal{S}_{1,n} \subset \mathcal{S}_{k,n}$  for all  $k \geq 1$ . This implies that

$$\rho_\infty(\mathcal{S}_n, \alpha, \beta) \geq \rho_2(\mathcal{S}_{1,n}, \alpha, \beta)$$

and that for all  $k \geq 1$ ,

$$\rho_\infty(\mathcal{S}_{k,n}, \alpha, \beta) \geq \rho_2(\mathcal{S}_{1,n}, \alpha, \beta).$$

This leads to the following corollary of Theorem 1 :

**Corollary 1** *Assume that  $Y = (Y_j)_{j \in J}$  obeys to Model (1.1). Let  $\beta \in ]0, 1 - \alpha[$ , such that  $\alpha + \beta \leq 59\%$ . Let*

$$\mathcal{S}_n = \{\theta \in l_2(J), \forall j > n, \theta_j = 0\}$$

and let

$$\mathcal{S}_{k,n} = \{\theta \in l_2(J), \forall j > n, \theta_j = 0, \text{Card } \{j \leq n, \theta_j \neq 0\} \leq k\}.$$

Let  $\sigma_{(1)} \leq \sigma_{(2)} \leq \dots \leq \sigma_{(n)}$ . We define

$$\rho_{n,\infty} = \max_{0 \leq l \leq n-1} \sigma_{(l+1)} \sqrt{\ln(1+n-l)}. \quad (2.8)$$

The following results hold :

$$\begin{aligned} \forall n \geq 1, \rho_\infty(\mathcal{S}_n, \alpha, \beta) &\geq \rho_{n,\infty}, \\ \forall 1 \leq k \leq n, \rho_\infty(\mathcal{S}_{k,n}, \alpha, \beta) &\geq \rho_{n,\infty}. \end{aligned}$$

The proof of the corollary follows directly from the arguments given above and will therefore be omitted.

**Comments :**

1. When  $\sigma_j = \sigma$  for all  $1 \leq j \leq n$ , we obtain  $\rho_{n,\infty} = \sigma \sqrt{\ln(n+1)}$ .
2. When  $(\sigma_j)_{j \in J}$  grows at a polynomial rate :  $\sigma_j = j^\gamma$  for some  $\gamma > 0$ , we obtain that  $\rho_{n,\infty} \geq C(\gamma)n^\gamma \sqrt{\ln(n)}$  by taking  $l = n/2$  in (2.8).
3. When  $(\sigma_j)_{j \in J}$  grows at an exponential rate :  $\sigma_j = \exp(j\gamma)$  for some  $\gamma > 0$ , we obtain that  $\rho_{n,\infty} \geq C \exp(n\gamma)$  for some constant  $C$ .

### 3 Upper bounds

In this section, we give upper bounds for the  $(\alpha, \beta)$  minimax rates of testing over the sets  $S_D$  and  $\mathcal{S}_{k,n}$  that we compare with the lower bounds obtained in the previous section.

In order to show that the  $(\alpha, \beta)$  minimax rate of testing with respect to the  $l_q$  norm over a set  $\mathcal{F}$  is bounded from above by  $\rho$ , it suffices to define a test statistic  $\Phi_\alpha$  such that the power of the test at each point  $\theta$  in  $\mathcal{F}$  satisfying  $\|\theta\|_q \geq \rho$  is greater than  $1 - \beta$ .

**Proposition 2** *Assume that  $Y = (Y_j)_{j \in J}$  obeys to Model (1.1). Let  $\alpha, \beta \in ]0, 1[$ , and let  $t_{D,1-\alpha}(\sigma)$  denote the  $1 - \alpha$  quantile of  $\sum_{j=1}^D \sigma_j^2 \epsilon_j^2$  :*

$$\mathbb{P} \left( \sum_{j=1}^D \sigma_j^2 \epsilon_j^2 \geq t_{D,1-\alpha}(\sigma) \right) = \alpha.$$

Let  $\Phi_\alpha$  be the test defined by

$$\Phi_\alpha = \mathbf{1}_{\sum_{j=1}^D Y_j^2 > t_{D,1-\alpha}(\sigma)}. \quad (3.9)$$

Then,  $\Phi_\alpha$  is a level- $\alpha$  test :

$$\mathbb{P}_{\theta=0}(\Phi_\alpha = 1) = \alpha.$$

Moreover, there exists a constant  $C(\alpha, \beta)$  such that for all  $\theta \in S_D$ ,

$$\|\theta\|_2^2 \geq C(\alpha, \beta) \left( \sum_{j=1}^D \sigma_j^4 \right)^{1/2} \Rightarrow \mathbb{P}_\theta(\Phi_\alpha = 1) > 1 - \beta.$$

Hence, we obtain that

$$\rho_2^2(S_D, \alpha, \beta) \leq C(\alpha, \beta) \left( \sum_{j=1}^D \sigma_j^4 \right)^{1/2}.$$

This result shows that the lower bound established in Proposition 1 is sharp.

Let us now propose a testing procedure for sparse signal detection. This procedure will be defined by a combination of two test. The first one is based on a thresholding method, which was already used for detection of irregular alternatives in Baraud et al [3] and in Fromont et al [8]. The second one is the test considered in Proposition 2 with  $D = n$ , which will be powerful when  $k$  is larger than  $\sqrt{n}$ .

**Theorem 2** *Assume that  $Y = (Y_j)_{j \in J}$  obeys to Model (1.1). Let  $\alpha, \beta \in ]0, 1[$ , and let  $t_{n,1-\alpha}(\sigma)$  denote the  $1 - \alpha$  quantile of  $\sum_{j=1}^n \sigma_j^2 \epsilon_j^2$ . Let  $\Phi_\alpha^{(1)}$  be the test defined by*

$$\Phi_\alpha^{(1)} = \mathbf{1}_{\sum_{j=1}^n Y_j^2 > t_{n,1-\alpha}(\sigma)}.$$

Let  $q_{n,1-\alpha}$  denote the  $1 - \alpha$  quantile of  $\max_{1 \leq j \leq n} \epsilon_j^2$ . Let  $\Phi_\alpha^{(2)}$  be the test defined by

$$\begin{aligned} \Phi_\alpha^{(2)} &= 1 \text{ if } \max_{1 \leq j \leq n} \left( \frac{Y_j^2}{\sigma_j^2} \right) > q_{n,1-\alpha} \\ &= 0 \text{ otherwise.} \end{aligned}$$

We define  $\Phi_\alpha = \max\left(\Phi_{\alpha/2}^{(1)}, \Phi_{\alpha/2}^{(2)}\right)$ . Then,  $\Phi_\alpha$  is a level- $\alpha$  test :

$$\mathbb{P}_{\theta=0}(\Phi_\alpha = 1) = \alpha.$$

Let  $\Sigma_{l,k}^2$  be defined by (2.5). There exists a constant  $C(\alpha, \beta)$  such that for all  $\theta \in \mathcal{S}_{k,n}$  satisfying

$$\|\theta\|_2^2 \geq C(\alpha, \beta) \left[ \left( \sum_{j=1}^n \sigma_j^4 \right)^{1/2} \wedge \left( \sum_{j, \theta_j \neq 0} \sigma_j^2 \right) \ln(n) \right], \quad (3.10)$$

we have

$$\mathbb{P}_\theta(\Phi_\alpha = 1) > 1 - \beta.$$

Hence, we obtain that

$$\rho_2^2(\mathcal{S}_{k,n}, \alpha, \beta) \leq C(\alpha, \beta) \left[ \left( \sum_{j=1}^n \sigma_j^4 \right)^{1/2} \wedge \Sigma_{n-k,k}^2 \ln(n) \right]. \quad (3.11)$$

**Comments :** Let us compare these results with the lower bounds obtained in Theorem 1.

1. We first assume that  $(\sigma_j)_{j \in J}$  grows at a polynomial rate :  $\forall j \in J, \sigma_j = \sigma j^\gamma$  for some  $\gamma \geq 0$  (this includes the homoscedastic case). In this case, there exists a constant  $C > 0$  such that  $\Sigma_{n/2,k}^2 \geq C \Sigma_{n-k,k}^2$ . A lower bound for the  $(\alpha, \beta)$  minimax separation rate of signal detection over  $\mathcal{S}_{k,n}$  is given by (2.7). This lower bound has to be compared with the upper bound (3.11).

- When  $k = n^l$  with  $l < 1/2$ , the upper and lower bounds coincide and are of order  $\Sigma_{n-k,k}^2 \ln(n)$ .
- When  $k = n^l$  with  $l \geq 1/2$ , the lower bound is of order  $\Sigma_{n-k,k}^2 \sqrt{n}/k$  and  $\Sigma_{n-k,k}^2 \geq Ck\sigma^2 n^{2\gamma}$ , which leads to a lower bound of order  $C\sigma^2 n^{2\gamma+1/2}$ . The upper bound is smaller than  $\left(\sum_{j=1}^n \sigma_j^4\right)^{1/2}$ , which is smaller than  $\sigma^2 n^{2\gamma+1/2}$ . Hence, the two bounds coincide.
- When  $k = \sqrt{n}/\phi(n)$  where  $\phi(n) \rightarrow +\infty$  and  $\phi(n)/n \rightarrow 0$  as  $n \rightarrow +\infty$  (typically  $\phi(n) = \ln(n)$ ), the lower bound is of order  $\Sigma_{n-k,k}^2 \ln(\phi(n))$  and the upper bound is of order  $\Sigma_{n-k,k}^2 \ln(n)$ . In this case, the upper and lower bound do not coincide, up to a logarithmic term.

2. Let us now assume that  $(\sigma_j)_{j \in J}$  grows at an exponential rate :  $\forall j \in J, \sigma_j = \sigma \exp(\gamma j)$  for some  $\gamma > 0$ . The lower bound is greater than  $\sigma_n^2 = \sigma^2 \exp(2\gamma n)$  and the upper bound is smaller than  $C(\alpha, \beta) \left(\sum_{j=1}^n \sigma_j^4\right)^{1/2}$ , which is bounded from above by  $C(\alpha, \beta, \gamma) \sigma^2 \exp(2\gamma n)$ . Hence the two bounds coincide. Note that in this case, the test  $\Phi_\alpha^{(2)}$  based on thresholding is useless and one can simply consider that test

$$\Phi_\alpha = \Phi_\alpha^{(1)},$$

which achieves the lower bound for the separation rate.



3. The result stated in (3.10) is more precise than the minimax upper bound given in (3.11). If the set  $J_1 = \{j, \theta_j \neq 0\}$  corresponds to small values for the variances  $(\sigma_j)_{j \in J_1}$ , it is not required that  $\|\theta\|_2^2$  is greater than the right hand term in (3.11) for the test to be powerful for this value of  $\theta$ . The minimax bound given in (3.11) corresponds to the worst situation, that is the case where the set  $J_1$  corresponds to the largest values for the variances.

Let us now present an upper bound for the minimax separation rate with respect to the  $l_\infty$  norm. This upper bound is obtained via the thresholding procedure presented in Theorem 2.

**Corollary 2** *Assume that  $Y = (Y_j)_{j \in J}$  obeys to Model (1.1). Let  $\sigma_{(1)} \leq \dots \leq \sigma_{(n)}$ . Let  $\alpha, \beta \in ]0, 1[$ . Let  $q_{n,1-\alpha}$  denote the  $1 - \alpha$  quantile of  $\max_{1 \leq j \leq n} \epsilon_j^2$ . Let  $\Phi_\alpha^{(2)}$  be the test defined by*

$$\begin{aligned} \Phi_\alpha^{(2)} &= 1 \text{ if } \max_{1 \leq j \leq n} \left( \frac{Y_j^2}{\sigma_j^2} \right) > q_{n,1-\alpha} \\ &= 0 \text{ otherwise.} \end{aligned}$$

*There exists a constant  $C(\alpha, \beta)$  such that for all  $k \in \{1, \dots, n\}$ , for all  $\theta \in \mathcal{S}_{k,n}$  such that*

$$\|\theta\|_\infty \geq C(\alpha, \beta) \sigma_{(n)} \sqrt{\ln(n)}$$

*we have*

$$\mathbb{P}_\theta(\Phi_\alpha^{(2)} = 1) > 1 - \beta.$$

*This implies that for all  $k \in \{1, \dots, n\}$ ,*

$$\rho_\infty(\mathcal{S}_{k,n}, \alpha, \beta) \leq C(\alpha, \beta) \sigma_{(n)} \sqrt{\ln(n)}$$

**Comments :** In particular, when  $k = n$ , we obtain

$$\rho_\infty(\mathcal{S}_n, \alpha, \beta) \leq C(\alpha, \beta) \sigma_{(n)} \sqrt{\ln(n)}.$$

This upper bound coincide with the lower bound obtained in Corollary 1 when the sequence  $(\sigma_j)_{j \in J}$  is constant or grows at a polynomial rate. When the sequence  $(\sigma_j)_{j \in J}$  grows at an exponential rate, we obtain a logarithmic loss with respect to the lower bound.

## 4 Minimax rates over ellipsoids and $l_p$ balls

In the previous sections, the only constraint on the signal was expressed through the number of non-zero coefficients. In several situations, one deal instead with infinite sequences having a finite number of significant coefficients, the reminder being considered as negligible (in a sense which will be precised later on). To this end, we consider in this section a slightly different framework. Our aim is to study the link between the decay of the  $\theta_k$  and the associated rate of testing. We consider in the following two different kinds of function spaces: ellipsoids and  $l_p$ -bodies.

## 4.1 Minimax rates of testing over ellipsoids

In the following, we assume that the sequence  $\theta = (\theta_j)_{j \in J}$  belongs to the ellipsoid  $\mathcal{E}_{a,2}(R)$  defined as

$$\mathcal{E}_{a,2}(R) = \left\{ \nu \in l_2(J), \sum_{j \in J} a_j^2 \nu_j^2 \leq R^2 \right\},$$

where  $a = (a_k)_{k \in J}$  denotes a monotone non-decreasing sequence. For instance, if  $\theta$  corresponds to the sequence of Fourier coefficients of a function  $f$  and  $a_j$  is of order  $j^s$  with  $s > 0$ , then assuming that  $\theta \in \mathcal{E}_{a,2}(R)$  is equivalent to impose conditions on the  $s$ -th derivative of  $f$ . The belonging to  $\mathcal{E}_{a,2}(R)$  may be seen as a regularity assumption on our signal. The following result characterises the minimax rate of testing over  $\mathcal{E}_{a,2}(R)$ .

**Proposition 3** *Let  $\alpha, \beta$  be fixed and denote by  $\rho_2(\mathcal{E}_{a,2}(R), \alpha, \beta)$  the minimax rate of testing over  $\mathcal{E}_{a,2}(R)$  with respect to the  $l_2$  norm. Then*

$$\rho_2^2(\mathcal{E}_{a,2}(R), \alpha, \beta) \geq \sup_{D \in J} (\rho_D^2 \wedge R^2 a_D^{-2}) := \rho_{a,2,R}^2,$$

where  $\rho_D^2$  has been introduced in Proposition 1. Moreover

$$\sup_{\theta \in \mathcal{E}_{a,2}(R), \|\theta\|_2 \geq C \rho_{a,2,R}} P_\theta(\Phi_\alpha^* = 0) \leq \beta,$$

where  $C$  is a positive constant independent of  $\sigma$  and  $\Phi_\alpha^*$  denotes the test introduced in Proposition 2 with

$$D^* = \inf \{ D \in J, R^2 a_D^{-2} \leq \rho_D^2 \} \text{ if } J = \mathbb{N}^* \\ D^* = \inf \{ D \in J, R^2 a_D^{-2} \leq \rho_D^2 \} \wedge N \text{ if } J = \{1, \dots, N\}.$$

Proposition 3 ensures that the minimax rate of testing over  $\mathcal{E}_{a,2}(R)$  is of order of  $\rho_{a,2,R}^2$ . Indeed, in a first time we prove that this quantity defines a lower bound for this rate. Then we construct an explicit test that reaches, up to a constant, this lower bound. Remark that this test was introduced in Proposition 2 where only signals with a finite number of non-zero coefficients were considered. We do not use the whole sequence  $(Y_j)_{j \in J}$  in order to test the null hypothesis " $\theta = 0$ " but only the first  $D^*$  coefficients. The price to pay is to introduce some bias in the testing procedure. However, as in classical estimation problems, this bias can be controlled by taking advantage of the constraint expressed on the decay of  $\theta$ .

**PROOF OF PROPOSITION 3.** We first prove the lower bound. For all  $D \in J$ , introduce  $r_D^2 = \rho_D^2 \wedge R^2 a_D^{-2}$ . Let  $D$  be fixed. Then for all  $\theta \in S_D$  such that  $\|\theta\|_2^2 = r_D^2$

$$\sum_{j \in J} a_j^2 \theta_j^2 = \sum_{j=1}^D a_j^2 \theta_j^2 \leq a_D^2 \|\theta\|_2^2 \leq R^2.$$

Hence

$$\{\theta \in S_D, \|\theta\|_2^2 = r_D^2\} \subset \{\theta \in \mathcal{E}_{a,2}(R), \|\theta\|_2^2 \geq r_D^2\}.$$

Since  $r_D \leq \rho_D$ , we get from Proposition 1

$$\inf_{\Phi_\alpha} \sup_{\theta \in \mathcal{E}_{a,2}(R), \|\theta\|_2 \geq r_D} P_\theta(\Phi_\alpha = 0) \geq \inf_{\Phi_\alpha} \sup_{\theta \in S_D, \|\theta\|_2 = r_D} P_\theta(\Phi_\alpha = 0) \geq \beta, \quad (4.12)$$

where the infimum is taken over all possible level- $\alpha$  testing procedures. Since inequality (4.12) holds for all  $D \in J$ , we obtain  $\rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) \geq \rho_{a,2,R}^2$ . Concerning the upper bound, we know from Proposition 2 that the test  $\Phi_\alpha^*$  is powerful as soon as:

$$\sum_{j=1}^{D^*} \theta_j^2 \geq C(\alpha, \beta) \rho_{D^*}^2 \Leftrightarrow \|\theta\|_2^2 \geq C(\alpha, \beta) \rho_{D^*}^2 + \sum_{k>D^*} \theta_k^2,$$

where  $C(\alpha, \beta)$  denotes a positive constant. Since  $\theta \in \mathcal{E}_{a,2}(R)$  and thanks to the definition of  $D^*$ , we get

$$\sum_{k>D^*} \theta_k^2 \leq a_{D^*}^{-2} R^2 \leq \rho_{D^*}^2 \text{ and } \sup_{\theta \in \mathcal{E}_{a,2}(R), \|\theta\|_2 \geq C \rho_{D^*}^2} P_\theta(\Phi_\alpha^* = 0) \leq \beta,$$

where  $C = (C(\alpha, \beta) + 1)$ . In order to conclude the proof, just remark that

$$\rho_{D^*}^2 \leq C \sup_{D \in J} (\rho_D^2 \wedge R^2 a_D^{-2}),$$

for some positive constant  $C$ .

□

Theorem 3 presents the minimax rate of testing in a general setting. Several explicit rates can be obtained when introducing specific constraints on the sequences  $(a_k)_{k \in J}$  and  $(b_k)_{k \in J}$ . These rates are summarized in Corollary 1. Let  $(\nu_k)_{k \in \mathbb{N}^*}$  be a sequence real numbers. In the following, we write  $\nu_k \sim k^l$  if there exist positive constants  $c_1$  and  $c_2$  such that, for all  $k \in \mathbb{N}^*$ ,  $c_1 k^l \leq \nu_k \leq c_2 k^l$ .

**Corollary 1** *Let  $\alpha, \beta$  be fixed. We assume that  $(Z_j)_{j \in J}$  obeys to Model (1.3). The table below presents the minimax rates of testing over the ellipsoids  $\mathcal{E}_{a,2}(R)$  with respect to the  $l_2$  norm. We consider various behaviour for the sequences  $(a_k)_{k \in \mathbb{N}^*}$  and  $(b_k)_{k \in \mathbb{N}^*}$ . For each case, we give  $f(\sigma)$  such that for all  $\sigma > 0$ ,  $C_1(\alpha, \beta) f(\sigma) \leq \rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) \leq C_2(\alpha, \beta) f(\sigma)$  where  $C_1(\alpha, \beta)$  and  $C_2(\alpha, \beta)$  denote positive constants independent of  $\sigma$ .*

	<i>Mildly ill-posed</i>	<i>Severly ill-posed</i>
	$b_k \sim k^{-t}$	$b_k \sim \exp(-\gamma k^r)$
$a_k \sim k^{-s}$	$\sigma^{\frac{4s}{2s+2t+1/2}}$	$\left(\frac{\log(\sigma^{-2})}{2\gamma}\right)^{-2s/r}$
$a_k \sim \exp(\nu k^s)$	$\sigma^2 \left(\frac{\log(\sigma^{-2})}{2\nu}\right)^{(2t+1/2)/s}$	$e^{-2\nu \tilde{D}^s}$

where  $\tilde{D}$  is the solution of the equation  $2\nu D^s + 2\gamma D^r = \log(\sigma^{-2})$ .

These rates are similar to the one obtained in [5] for density estimation. For the sake of convenience, we only present the proof in the particular case of mildly ill-posed problems with polynomially increasing  $a_k$ . We leave the construction in other cases to an interested reader.

PROOF OF COROLLARY 1. Since  $a_k \sim k^s$  and  $b_k \sim k^{-t}$ , we get:

$$\rho_D^2 \simeq \sigma^2 D^{2t+1/2} \text{ and } R^2 a_D^{-2} \simeq D^{-2s}.$$

Then

$$D^* \simeq \arg \sup_{D \in J} (\rho_D^2 \wedge R^2 a_D^{-2}) \simeq \sigma^{\frac{2}{4s+4t+1}},$$

hence leading to the desired rate. □

## 4.2 Minimax rates of testing over $l_p$ -bodies with $0 < p < 2$

Ellipsoids contain essentially smooth functions. For instance, in the particular case where  $\theta$  corresponds to the Fourier coefficients of a given function  $f$ , the constraints expressed through the belonging of one of the spaces introduced above may be incompatible with the presence of discontinuities. In order to extend the covered cases, we consider in this part sequences  $\theta$  belonging to  $l_p$ -bodies  $\mathcal{E}_{a,p}(R)$  defined as

$$\mathcal{E}_{a,p}(R) = \left\{ \nu \in l_2(J), \sum_{j \in J} a_j^p \nu_j^p \leq R^p \right\},$$

where  $a = (a_k)_{k \in J}$  denotes a monotone non-decreasing sequence and  $0 < p < 2$ . The following theorem proposes a lower bound for the minimax rate of testing over such spaces.

**Theorem 3** *Let  $(Y_j)_{j \in J}$  obey to the model (1.1). Let  $\alpha, \beta$  be fixed and denote by  $\rho_2(\mathcal{E}_{a,p}(R), \alpha, \beta)$  the minimax rate of testing over  $\mathcal{E}_{a,p}(R)$  with respect to the  $l_2$  norm. For all  $D \in J$  and for all  $0 \leq l \leq D - \lceil \sqrt{D} \rceil$ , we set*

$$\rho_{\lceil \sqrt{D} \rceil, D, l}^2 = \Sigma_{l, \lceil \sqrt{D} \rceil}^2 \ln \left( 1 + \sqrt{1 - \frac{l}{D}} \right),$$

where  $\Sigma_{l, \lceil \sqrt{D} \rceil}^2$  is given in (2.5). Then

$$\rho^2(\mathcal{E}_{a,p}(R), \alpha, \beta) \geq \sup_{D \in J} (\rho_1(D) \vee \rho_2(D)),$$

where

$$\rho_1(D) = \max_{0 \leq l \leq D - \lceil \sqrt{D} \rceil} \left( \sqrt{D}^{1-2/p} R^2 a_D^{-2} \frac{\Sigma_{l, \lceil \sqrt{D} \rceil}^2}{\Sigma_{D - \lceil \sqrt{D} \rceil, \lceil \sqrt{D} \rceil}^2} \wedge \rho_{\lceil \sqrt{D} \rceil, D, l}^2 \right),$$

and

$$\rho_2(D) = \sqrt{D}^{1-2/p} R^2 a_D^{-2} \wedge \left( \sum_{j=D - \lceil \sqrt{D} \rceil + 1}^D \sigma_j^4 \right)^{1/2}.$$

To the end of this section, we assume that the sequence  $(b_j)_{j \in J}$  is polynomially or exponentially increasing, which yet correspond to the main case of interest in inverse problems. The lower bounds in these particular cases are easier to handle, as proved in the following corollary.

**Corollary 3** *Let  $(Z_j)_{j \in J}$  obey to the model (1.3) with  $b_j = \sigma \sigma_j^{-1}$  for all  $j \in J$ . Then, assuming that for all  $j \in J$ ,  $\sigma_j = j^{2\gamma}$  or that for all  $j \in J$ ,  $\sigma_j = \exp(\gamma j)$  for some  $\gamma \geq 0$ , we obtain*

$$\rho^2(\mathcal{E}_{a,p}(R), \alpha, \beta) \geq C(\gamma) \sup_{D \in J} (\rho_{[\sqrt{D}], D}^2 \wedge \sqrt{D}^{1-2/p} R^2 a_D^{-2}) := \rho_{a,p,R}^2,$$

where  $\rho_{[\sqrt{D}], D}^2$  is defined in (2.6).

The proof of Corollary 3 derives easily from the comments of Theorem 1.

In order to attain the lower bound presented above, a test similar to the one introduced in Theorem 3 with

$$D^\dagger = \inf \left\{ D \in J, R^2 a_D^{-2} \sqrt{D}^{1-2/p} \leq \rho_{[\sqrt{D}], D}^2 \right\}.$$

is not sufficient. On  $l_p$ -bodies, the bias after the rank  $D^\dagger$  is indeed more difficult to control than for ellipsoids. Some significant coefficients (in a sense which will be precised in the proof) may be contained in the sequence  $\theta$  after the rank  $D^\dagger$ . Hence, we have to introduce specific tests in order to detect these coefficients on the subset  $\{D^\dagger, \dots, N\}$ . More precisely, for all  $j \in J$  and  $\alpha \in (0, 1)$ , introduce

$$\Phi_{\{j\}, \alpha} = \mathbf{1}_{\{|Y_j| \geq q_{j, \alpha}\}},$$

wher  $q_{j, \alpha}$  denotes the  $\alpha$  quantile of a Gaussian random variable with mean 0 and variance  $\sigma_j^2$ . Then define

$$\Phi_\alpha^\dagger = \Phi_{\text{loc}, \alpha/2} \wedge \Phi_{D^\dagger, \alpha/2} \text{ with } \Phi_{\text{loc}, \alpha/2} = \sup_{j \in \{D^\dagger, \dots, N\}} \Phi_{\{j\}, 2\alpha/\pi^2(j-D^\dagger)^2},$$

and where  $\Phi_{D^\dagger, \alpha/2}$  denotes the test constructed in Proposition 2. The following theorem emphasizes the performances of the test  $\Phi_\alpha^\dagger$ .

**Proposition 4** *Let  $\alpha, \beta$  be fixed. We assume that the sequence  $\left(a_j^{-p} b_j^{-(2-p)}\right)_{j \in \mathbb{N}}$  is monotone non-increasing. Suppose that  $J = \{1, \dots, N\}$ . Then*

$$\sup_{\theta \in \mathcal{E}_{a,p}(R), \|\theta\|^2 \geq \lambda_\sigma \rho_{a,2,R}^2} P_\theta(\Phi_\alpha^\dagger = 0) \leq \beta,$$

with

- $\lambda_\sigma = C \log(N)$  for mildly ill-posed problems,
- $\lambda_\sigma = C \log(N) \sqrt{D^\dagger}^{1-p/2}$  for severely ill-posed inverse problems,

where  $C$  denotes a positive constant independent of  $\sigma$ .

Our test reaches the lower bound established in Proposition 3 up to a log term: it is not sharp. Remark that a similar property occurs in the homoscedastic case: see [2] for more details. Hence, the lower bound established in Proposition 3 corresponds certainly to the minimax rate on  $\mathcal{E}_{a,p}(R)$ .

For the sake of convenience, the upper bound is presented for  $J = \{1, \dots, N\}$  which more or less corresponds to the regression setting. Nevertheless, our result can be easily extended to the case where  $J = \mathbb{N}^*$ . In such a situation, our test will be performed on  $\{1, \dots, \tilde{N}\}$ , where  $\tilde{N}$  is a trade-off between the bias after the rank  $\tilde{N}$  on  $\mathcal{E}_{a,p}(R)$  and the growth of  $\log(N)$ . A good candidate for  $\tilde{N}$  is a power of  $\sigma^{-2}$ .

In order to conclude this discussion, we point out that we impose a condition on the sequence  $\left(a_j^{-p} b_j^{-(2-p)}\right)_{j \in \mathbb{N}}$ . This condition is necessary in order to control the bias after the rank  $D^\dagger$ . It always hold when  $p = 2$  since  $(a_j)_{j \in \mathbb{N}}$  is an increasing sequence. When  $p < 2$ , the considered function has to be sufficiently smooth with respect to the ill-posedness of the problem. A similar condition can be found for instance in [7].

The corollary bellow considers the particular case of mildly ill-posed problems with polynomial  $l_p$ -bodies, where the situation is easier to handle.

**Corollary 2** *Assume that  $a_k \sim k^s$  and  $b_k \sim k^{-t}$  for all  $k \in \mathbb{N}$  where  $s, t$  denote positive constants such that  $s > t(2/p - 1)$ . Then*

$$\rho^2(\mathcal{E}_{a,p}(R), \alpha, \beta) \simeq \sigma^{\frac{4s+2/p-1}{2s+2t+1/p}}$$

Remark that the sequence  $\left(a_j^{-p} b_j^{-(2-p)}\right)_{j \in \mathbb{N}}$  is monotone non-increasing as soon as  $s > t(2/p - 1)$ . Hence the conditions of Proposition 4 are satisfied. The proof follows the same argument as in Corollary 1.

## 5 Proofs

### 5.1 Proof of the lower bounds

The proofs of the lower bounds use a Bayesian approach following the methods developed in the papers by Ingster [10] and by Baraud [2]. We use the following lemma :

**Lemma 1** *Let  $\mathcal{F}$  be some subset of  $l_2(J)$ . Let  $\mu_\rho$  be some probability measure on*

$$\mathcal{F}_\rho = \{\theta \in \mathcal{F}, \|\theta\|_2 \geq \rho\}$$

and let

$$\mathbb{P}_{\mu_\rho} = \int \mathbb{P}_\theta d\mu_\rho(\theta).$$

Assuming that  $\mathbb{P}_{\mu_\rho}$  is absolutely continuous with respect to  $\mathbb{P}_0$ , we define

$$L_{\mu_\rho}(y) = \frac{d\mathbb{P}_{\mu_\rho}}{d\mathbb{P}_0}(y).$$

For all  $\alpha > 0$ ,  $\beta \in ]0, 1 - \alpha[$ , if

$$\mathbb{E}_0 \left( L_{\mu_{\rho^*}}^2(Y) \right) \leq 1 + 4(1 - \alpha - \beta)^2,$$

then

$$\forall \rho \leq \rho^*, \quad \inf_{\Phi_\alpha} \sup_{\theta \in \mathcal{F}_\rho} \mathbb{P}_\theta(\Phi_\alpha = 0) \geq \beta.$$

This implies that

$$\rho(\mathcal{F}, \alpha, \beta) \geq \rho^*.$$

For the proof of this lemma, we refer to Baraud [2], Section 7.1.

### 5.1.1 Proof of Proposition 1

Let  $\rho > 0$ , we set for  $1 \leq j \leq D$ ,

$$\theta_j = \omega_j \sigma_j^2 \rho \left( \sum_{j=1}^D \sigma_j^4 \right)^{-1/2}$$

where  $(\omega_j, 1 \leq j \leq D)$  are i.i.d. Rademacher random variables :  $\mathbb{P}(\omega_j = 1) = \mathbb{P}(\omega_j = -1) = 1/2$ . Let  $\mu_\rho$  be the distribution of  $(\theta_1, \dots, \theta_D)$ .  $\mu_\rho$  is a probability measure on

$$\{\theta \in S_D, \|\theta\|_2 = \rho\}.$$

Let us now evaluate the likelihood ratio  $L_{\mu_\rho}(Y) = \frac{d\mathbb{P}_{\mu_\rho}}{d\mathbb{P}_0}(Y)$ .

$$\begin{aligned} L_{\mu_\rho}(Y) &= \mathbb{E}_\omega \left[ \exp \left( -\frac{1}{2} \sum_{j=1}^D \frac{1}{\sigma_j^2} \left( Y_j - \frac{\sigma_j^2 \omega_j \rho}{\sqrt{\sum_{j=1}^D \sigma_j^4}} \right)^2 \right) \exp \left( \frac{1}{2} \sum_{j=1}^D \frac{Y_j^2}{\sigma_j^2} \right) \right] \\ &= \exp \left( -\frac{\rho^2 \sum_{j=1}^D \sigma_j^2}{2 \sum_{j=1}^D \sigma_j^4} \right) \prod_{j=1}^D \cosh \left( \frac{\rho Y_j}{\sqrt{\sum_{j=1}^D \sigma_j^4}} \right). \end{aligned}$$

Let  $Z$  be some standard normal variable. For all  $\lambda \in \mathbb{R}$ ,

$$\mathbb{E}(\cosh^2(\lambda Z)) = \exp(\lambda^2) \cosh(\lambda^2). \quad (5.13)$$

Hence, since  $Y_j/\sigma_j \sim \mathcal{N}(0, 1)$ ,

$$\mathbb{E}_0 \left( L_{\mu_\rho}^2(Y) \right) = \prod_{j=1}^D \cosh \left( \frac{\rho^2 \sigma_j^2}{\sum_{j=1}^D \sigma_j^4} \right).$$

Since for all  $x \in \mathbb{R}$ ,  $\cosh(x) \leq \exp(x^2/2)$ , we obtain

$$\mathbb{E}_0 \left( L_{\mu_\rho}^2(Y) \right) \leq \exp \left( \frac{\rho^4}{2 \sum_{j=1}^D \sigma_j^4} \right).$$

For  $\rho = \rho_D$  we obtain :

$$\mathbb{E}_0 \left( L_{\mu_\rho}^2(Y) \right) \leq 1 + 4(1 - \alpha - \beta)^2,$$

which implies that  $\rho(S_D, \alpha, \beta) \geq \rho_D$  by Lemma 1.

### 5.1.2 Proof of Theorem 1

Without loss of generality, we can assume that the sequence  $(\sigma_j)_{j \in J}$  is non decreasing (if this is not the case, we can reorder the observations  $Y_j$ ). We fix some  $l \in \{0, 1, \dots, n - k\}$ . Let  $\mathcal{M}_{k,l,n}$  denote the set of all subsets of  $\{l + 1, \dots, n\}$  with cardinality  $k$ . Let  $\hat{m}$  be a random set of  $\{l + 1, \dots, n\}$ , which is uniformly distributed on  $\mathcal{M}_{k,l,n}$ . This means that for all  $m \in \mathcal{M}_{k,l,n}$ ,  $\mathbb{P}(\hat{m} = m) = 1/C_{n-l}^k$ . Let  $(\omega_j, 1 \leq j \leq n)$  be i.i.d. Rademacher random variables, independent of  $\hat{m}$ . Let us recall that

$$\Sigma_{l,k}^2 = \sum_{j=l+1}^{l+k} \sigma_j^2.$$

We set

$$\theta_j = (\rho \omega_j \sigma_j / \Sigma_{l,k}) \mathbb{1}_{j \in \hat{m}} \quad (5.14)$$

Note that  $\theta = (\theta_j)_{j \in J} \in \mathcal{S}_{k,n}$  and that, since  $(\sigma_j)_{j \in J}$  is non decreasing,

$$\|\theta\|_2^2 = \rho^2 \frac{\sum_{j \in \hat{m}} \sigma_j^2}{\Sigma_{l,k}^2} \geq \rho^2.$$

$$\begin{aligned} L_{\mu_\rho}(Y) &= \mathbb{E}_{\hat{m}, \omega} \left[ \exp \left( -\frac{1}{2} \sum_{j \in J} \frac{1}{\sigma_j^2} (Y_j - \theta_j)^2 \right) \exp \left( -\frac{1}{2} \sum_{j \in J} \frac{Y_j^2}{\sigma_j^2} \right) \right] \\ &= \mathbb{E}_{\hat{m}, \omega} \left[ \exp \left( \sum_{j \in \hat{m}} \frac{Y_j \theta_j}{\sigma_j^2} \right) \exp \left( -\frac{1}{2} \sum_{j \in \hat{m}} \frac{\theta_j^2}{\sigma_j^2} \right) \right] \\ &= \mathbb{E}_{\hat{m}, \omega} \left[ \exp \left( \sum_{j \in \hat{m}} \frac{Y_j \omega_j \rho}{\sigma_j \Sigma_{l,k}} \exp \left( -\frac{k \rho^2}{2 \Sigma_{l,k}^2} \right) \right) \right]. \end{aligned}$$

$$\begin{aligned} L_{\mu_\rho}(Y) &= \frac{1}{C_{n-l}^k} \sum_{m \in \mathcal{M}_{k,l,n}} \mathbb{E}_\omega \left[ \exp \left( \sum_{j \in m} \frac{Y_j \omega_j \rho}{\sigma_j \Sigma_{l,k}} \right) \exp \left( -\frac{k \rho^2}{2 \Sigma_{l,k}^2} \right) \right] \\ &= \exp \left( -\frac{k \rho^2}{2 \Sigma_{l,k}^2} \right) \frac{1}{C_{n-l}^k} \sum_{m \in \mathcal{M}_{k,l,n}} \prod_{j \in m} \cosh \left( \frac{\rho Y_j}{\sigma_j \Sigma_{l,k}} \right). \end{aligned}$$

We use (5.13) together with  $\mathbb{E}(\cosh(\lambda Z)) = \exp(\lambda^2/2)$  for a standard Gaussian variable  $Z$ . Since  $Y_j/\sigma_j$  is a standard normal variable, we obtain that

$$\begin{aligned} \mathbb{E}_0 \left( L_{\mu_\rho}^2(Y) \right) &= \exp \left( -\frac{k \rho^2}{\Sigma_{l,k}^2} \right) \frac{1}{(C_{n-l}^k)^2} \sum_{m, m' \in \mathcal{M}_{k,l,n}} \prod_{j \in m \setminus m'} \exp \left( \frac{\rho^2}{2 \Sigma_{l,k}^2} \right) \\ &\quad \times \prod_{j \in m' \setminus m} \exp \left( \frac{\rho^2}{2 \Sigma_{l,k}^2} \right) \prod_{j \in m \cap m'} \exp \left( \frac{\rho^2}{\Sigma_{l,k}^2} \right) \cosh \left( \frac{\rho^2}{\Sigma_{l,k}^2} \right). \end{aligned}$$



Since, for all  $m, m' \in \mathcal{M}_{k,l,n}$ ,

$$|m \setminus m'| + |m' \setminus m| + 2|m \cap m'| = |m| + |m'| = 2k,$$

we obtain

$$\mathbb{E}_0 \left( L_{\mu_\rho}^2(Y) \right) = \frac{1}{(C_{n-l}^k)^2} \sum_{m, m' \in \mathcal{M}_{k,l,n}} \left( \cosh \left( \frac{\rho^2}{\Sigma_{l,k}^2} \right) \right)^{|m \cap m'|}.$$

The end of the proof is similar to the proof of Theorem 1 in Baraud [2], similar arguments are also given in Fromont et al. [8]. Let us recall these arguments.

$$\mathbb{E}_0 \left( L_{\mu_\rho}^2(Y) \right) = \mathbb{E} \left[ \exp \left( |\hat{m} \cap \hat{m}'| \ln \cosh \left( \frac{\rho^2}{\Sigma_{l,k}^2} \right) \right) \right],$$

where  $\hat{m}, \hat{m}'$  are independent random subsets with uniform distribution on  $\mathcal{M}_{k,l,n}$ . For fixed  $\hat{m}$ ,  $|\hat{m} \cap \hat{m}'|$  is a hypergeometric variable with parameters  $(n-l, k, k/(n-l))$ . We know from Aldous [1] that there exists a binomial variable  $B$  with parameters  $(k, k/(n-l))$  and a  $\sigma$ -algebra  $\mathcal{B}$  such that  $\mathbb{E}(B/\mathcal{B}) = |\hat{m} \cap \hat{m}'|$ . By Jensen's inequality,

$$\mathbb{E}_0 \left( L_{\mu_\rho}^2(Y) \right) \leq \mathbb{E} \left[ \exp \left( B \ln \cosh \left( \frac{\rho^2}{\Sigma_{l,k}^2} \right) \right) \right].$$

Since  $B$  is a binomial variable with parameters  $(k, k/(n-l))$ ,

$$\mathbb{E} \left[ \exp \left( B \ln \cosh \left( \frac{\rho^2}{\Sigma_{l,k}^2} \right) \right) \right] = \exp \left[ k \ln \left( 1 + \frac{k}{n-l} \left( \cosh \left( \frac{\rho^2}{\Sigma_{l,k}^2} \right) - 1 \right) \right) \right].$$

Let  $c = 1 + 4(1 - \alpha - \beta)^2$ , and  $A = \frac{n-l}{k^2} \ln(c)$ . Since the function  $\cosh$  is increasing on  $\mathbb{R}^+$ , we obtain that if

$$\frac{\rho^2}{\Sigma_{l,k}^2} \leq \ln \left( 1 + A + \sqrt{2A + A^2} \right),$$

then

$$\cosh \left( \frac{\rho^2}{\Sigma_{l,k}^2} \right) - 1 \leq \frac{1}{2} \left( A + \sqrt{2A + A^2} - 1 \right) + \frac{1}{2} \left( A + \sqrt{2A + A^2} + 1 \right)^{-1} = A.$$

We finally obtain that

$$\mathbb{E}_0 \left( L_{\mu_\rho}^2(Y) \right) \leq \exp \left[ k \ln \left( 1 + \frac{k}{n-l} A \right) \right] \leq c.$$

By Lemma 1, this implies that

$$\begin{aligned} \rho(\mathcal{S}_{k,n}, \alpha, \beta) &\geq \Sigma_{l,k}^2 \ln \left( 1 + A + \sqrt{2A + A^2} \right) \\ &\geq \Sigma_{l,k}^2 \ln \left( 1 + 2A \vee \sqrt{2A} \right). \end{aligned}$$

If  $\alpha + \beta \leq 0.59$ ,  $\ln(c) \geq 1/2$ , which implies that

$$\rho(\mathcal{S}_{k,n}, \alpha, \beta) \geq \sum_{l,k}^2 \ln \left( 1 + \frac{n-l}{k^2} \vee \sqrt{\frac{n-l}{k^2}} \right).$$

Since this result holds for all  $l \in \{0, n-k\}$ , we get

$$\rho(\mathcal{S}_{k,n}, \alpha, \beta) \geq \max_{0 \leq l \leq n-k} \sum_{l,k}^2 \ln \left( 1 + \frac{n-l}{k^2} \vee \sqrt{\frac{n-l}{k^2}} \right).$$

In order to prove that

$$\rho(\mathcal{S}_{k,n}, \alpha, \beta) \geq \left( \sum_{j=n-k+1}^n \sigma_j^4 \right)^{1/2},$$

we define, as in the proof of Proposition 1,

$$\begin{aligned} \theta_j &= \omega_j \sigma_j^2 \rho \left( \sum_{j=n-k+1}^n \sigma_j^4 \right)^{-1/2} & \forall j \in \{n-k+1, \dots, n\}, \\ &= 0 & \forall j \notin \{n-k+1, \dots, n\}, \end{aligned}$$

where  $(\omega_j, n-k+1 \leq j \leq n)$  are i.i.d. Rademacher random variables. Note that  $(\theta_j)_{j \in J} \in \mathcal{S}_{k,n}$  and that  $\|\theta\|_2^2 = \rho^2$ . We now conclude as in the proof of Proposition 1, using that  $C(\alpha, \beta) = \sqrt{2 \ln(c)} \geq 1$ .

## 5.2 Proof of the upper bounds

### 5.2.1 Proof of Proposition 2

In order to prove Proposition 2, we have to show that for all  $\theta \in S_D$  such that  $\|\theta\|_2^2 \geq C(\alpha, \beta) \left( \sum_{j=1}^D \sigma_j^4 \right)^{1/2}$ ,

$$\mathbb{P}_\theta \left( \sum_{j=1}^D Y_j^2 \leq t_{D,1-\alpha}(\sigma) \right) < \beta. \quad (5.15)$$

We denote by  $t_{D,\beta}(\theta, \sigma)$  the  $\beta$  quantile of  $\sum_{j=1}^D Y_j^2$ , when  $Y = (Y_j)_{j \in J}$  obeys to Model (1.1). In order to prove (5.15), it suffices to show that

$$t_{D,1-\alpha}(\sigma) < t_{D,\beta}(\theta, \sigma).$$

To prove this inequality, we will give an upper bound for  $t_{D,1-\alpha}(\sigma)$  and a lower bound for  $t_{D,\beta}(\theta, \sigma)$ .

**Upper bound for  $t_{D,1-\alpha}(\sigma)$  :**

We use an exponential inequality for chi-square distributions due to Laurent and Massart [12] (see Lemma 1). It follows from this inequality that for all  $x \geq 0$ ,

$$\mathbb{P} \left( \sum_{j=1}^D \sigma_j^2 (\epsilon_j^2 - 1) \geq 2\sqrt{x} \left( \sum_{j=1}^D \sigma_j^4 \right)^{1/2} + 2x \sup_{1 \leq j \leq D} (\sigma_j^2) \right) \leq \exp(-x).$$

Setting  $x_\alpha = \ln(1/\alpha)$ , we obtain that

$$t_{D,1-\alpha}(\sigma) \leq \sum_{j=1}^D \sigma_j^2 + 2\sqrt{x_\alpha} \left( \sum_{j=1}^D \sigma_j^4 \right)^{1/2} + 2x_\alpha \sup_{1 \leq j \leq D} (\sigma_j^2).$$

Since  $\sup_{1 \leq j \leq D} \sigma_j^2 \leq \left( \sum_{j=1}^D \sigma_j^4 \right)^{1/2}$ ,

$$t_{D,1-\alpha}(\sigma) \leq \sum_{j=1}^D \sigma_j^2 + C(\alpha) \left( \sum_{j=1}^D \sigma_j^4 \right)^{1/2}. \quad (5.16)$$

**Lower bound for  $t_{D,\beta}(\theta, \sigma)$  :**

We prove the following lemma, which generalizes the results obtained by Birgé [4] to the heteroscedastic framework :

**Lemma 2** *Let*

$$Y_j = \theta_j + \sigma_j \epsilon_j, \quad 1 \leq j \leq D,$$

where  $\epsilon_1, \dots, \epsilon_D$  are i.i.d. Gaussian variables with mean 0 and variance 1.

We define  $\hat{T} = \sum_{j=1}^D Y_j^2$  and

$$\Sigma = \sum_{j=1}^D \sigma_j^4 + 2 \sum_{j=1}^D \sigma_j^2 \theta_j^2.$$

The following inequalities hold for all  $x \geq 0$  :

$$\mathbb{P} \left( \hat{T} - \mathbb{E}(\hat{T}) \geq 2\sqrt{\Sigma x} + 2 \sup_{1 \leq j \leq D} (\sigma_j^2) x \right) \leq \exp(-x). \quad (5.17)$$

$$\mathbb{P} \left( \hat{T} - \mathbb{E}(\hat{T}) \leq -2\sqrt{\Sigma x} \right) \leq \exp(-x). \quad (5.18)$$

The proof of this lemma is given in the Appendix.

Inequality (5.18) provides a lower bound for  $t_{D,\beta}(\theta, \sigma)$ . Indeed, setting  $x_\beta = \log(1/\beta)$ , we obtain that

$$\mathbb{P} \left( \hat{T} - \mathbb{E}(\hat{T}) \leq -2\sqrt{\Sigma x_\beta} \right) \leq \beta.$$

Hence,  $t_{D,\beta}(\theta, \sigma) \geq \sum_{j=1}^D (\theta_j^2 + \sigma_j^2) - 2\sqrt{\Sigma x_\beta}$ . (5.15) is satisfied if  $t_{D,1-\alpha}(\sigma) < t_{D,\beta}(\theta, \sigma)$ , which holds as soon as

$$\sum_{j=1}^D \theta_j^2 - 2\sqrt{\Sigma x_\beta} > 2\sqrt{x_\alpha} \sqrt{\sum_{j=1}^D \sigma_j^4 + 2x_\alpha \sup_{1 \leq j \leq D} (\sigma_j^2)}. \quad (5.19)$$

Let us note that

$$\begin{aligned} \sqrt{\Sigma} &= \sqrt{\sum_{j=1}^D \sigma_j^4 + 2\sigma_j^2 \theta_j^2} \leq \sqrt{\sum_{j=1}^D \sigma_j^4} + \sqrt{2} \sqrt{\sum_{j=1}^D \sigma_j^2 \theta_j^2} \\ &\leq \sqrt{\sum_{j=1}^D \sigma_j^4} + \sqrt{2} \sup_{1 \leq j \leq D} (\sigma_j) \sqrt{\sum_{j=1}^D \theta_j^2} \end{aligned}$$

Hence, the following inequality implies (5.19) :

$$\sum_{j=1}^D \theta_j^2 - 2\sqrt{2} \sup_{1 \leq j \leq D} (\sigma_j) \sqrt{x_\beta} \sqrt{\sum_{j=1}^D \theta_j^2} - 2 \sqrt{\sum_{j=1}^D \sigma_j^4 (\sqrt{x_\beta} + \sqrt{x_\alpha})} - 2 \sup_{1 \leq j \leq D} (\sigma_j^2) x_\alpha > 0.$$

This inequality holds if

$$\sum_{j=1}^D \theta_j^2 \geq C \left[ \sup_{1 \leq j \leq D} (\sigma_j^2) (x_\beta + x_\alpha) + \sqrt{\sum_{j=1}^D \sigma_j^4 (\sqrt{x_\beta} + \sqrt{x_\alpha})} \right],$$

where  $C$  is an absolute constant (which can be taken equal to 8). Hence, we have proved that

$$\rho(S_D, \alpha, \beta) \leq C(\alpha, \beta) \sqrt{\sum_{j=1}^D \sigma_j^4},$$

which concludes the proof of Proposition 2.

### 5.3 Proof of minimax rates on ellipsoids and $l_p$ -bodies

#### 5.3.1 Proof of Theorem 3

The proof will use the one of Theorem 1. Let us first establish a relation between the  $l_p$  ball  $\mathcal{E}_{a,p}(R)$  and the sets  $\mathcal{S}_{k,n}$ . For all  $D \in J$ , for all  $\theta \in \mathcal{S}_{\lceil \sqrt{D} \rceil, D}$  such that  $\|\theta\|_2^2 \leq \sqrt{D}^{1-2/p} R^2 a_D^{-2}$ , we have  $\theta \in \mathcal{E}_{a,p}(R)$ . Indeed, using a Hölder's inequality

$$\sum_{j=1}^{+\infty} a_j^p \theta_j^p = \sum_{j:\theta_j \neq 0} a_j^p \theta_j^p \leq \sqrt{D}^{1-p/2} (\|\theta\|_2^2)^{p/2} a_D^p \leq R^p.$$

We set  $k = \lceil \sqrt{D} \rceil$ ,  $n = D$  and for all  $l \in \{0, 1, \dots, n-k\}$ , we define  $\theta = (\theta_j, j \in J)$  by (5.14). As noticed in the proof of Theorem 1,  $\theta \in \mathcal{S}_{k,n}$  and  $\|\theta\|_2^2 \geq \rho^2$ . We also have  $\|\theta\|_2^2 \leq \rho^2 \Sigma_{n-k,k}^2 / \Sigma_{l,k}^2$ . This implies that if

$$\rho^2 \frac{\Sigma_{n-k,k}^2}{\Sigma_{l,k}^2} \leq \sqrt{D}^{1-2/p} R^2 a_D^{-2},$$

then  $\theta \in \mathcal{E}_{a,p}(R)$ .

Moreover, in the proof of Theorem 1, we proved that if

$$\rho^2 \leq \Sigma_{l,k}^2 \ln \left( 1 + \frac{n-l}{k^2} \vee \sqrt{\frac{n-l}{k^2}} \right),$$

then

$$\mathbb{E}_0(L_{\mu_\rho}^2(Y)) \leq 1 + 4(1 - \alpha - \beta)^2.$$

This implies by Lemma 1 that  $\rho_2^2(\mathcal{E}_{a,p}(R)) \geq \rho^2$ . We finally get

$$\rho_2^2(\mathcal{E}_{a,p}(R)) \geq \Sigma_{l,k}^2 \ln \left( 1 + \frac{n-l}{k^2} \vee \sqrt{\frac{n-l}{k^2}} \right) \wedge \sqrt{D}^{1-2/p} R^2 a_D^{-2} \frac{\Sigma_{l,k}^2}{\Sigma_{n-k,k}^2}.$$

Since the result holds for all  $l \in \{0, 1, \dots, n-k\}$ , we obtain that  $\rho_2^2(\mathcal{E}_{a,p}(R)) \geq \rho_1(D)$ . To obtain that  $\rho_2^2(\mathcal{E}_{a,p}(R)) \geq \rho_2(D)$ , we consider, as in the proof of Theorem 1, for  $k = \lceil \sqrt{D} \rceil$  and  $n = D$

$$\begin{aligned}\theta_j &= \omega_j \sigma_j^2 \rho \left( \sum_{j=n-k+1}^n \sigma_j^4 \right)^{-1/2} & \forall j \in \{n-k+1, \dots, n\}, \\ &= 0 & \forall j \notin \{n-k+1, \dots, n\}.\end{aligned}$$

Since  $\rho_2^2(\mathcal{E}_{a,p}(R)) \geq \rho_1(D) \vee \rho_2(D)$  for all  $D \in J$ , the result follows.

### 5.3.2 Proof of Proposition 4

Introduce

$$A = \left\{ D \in J, R^2 a_D^{-2} \sqrt{D}^{1-p/2} \leq \rho_{\lfloor \sqrt{D} \rfloor, D}^2 \right\}.$$

In a first time, we suppose that  $A$  is empty. From the definition of  $D^\dagger$ , we get  $D^\dagger = N$  and

$$P_\theta(\Phi_\alpha^\dagger = 0) \leq P_\theta(\Phi_{D^\dagger, \alpha/2} = 0) = P_\theta(\Phi_{N, \alpha/2} = 0) \leq \beta,$$

for all function  $f$  satisfying

$$\sum_{j \in J} \theta_j^2 = \|\theta\|^2 \geq C \rho_N^2,$$

For some positive constant  $C$ . Since  $(b_k)_{k \in \mathbb{N}} \sim (k^{-t})_{k \in \mathbb{N}}$  for some  $t > 0$ ,  $\rho_D^2 \simeq \rho_{\lfloor \sqrt{D} \rfloor, D}^2 \simeq D^{2t+1/2}$  for all  $D \in \mathbb{N}$ . Then

$$\rho_N^2 \leq (\rho_N^2 \wedge \sqrt{N}^{1-2/p} a_N^{-2} R^2) \leq \sup_D (\rho_{\lfloor \sqrt{D} \rfloor, D}^2 \wedge \sqrt{D}^{1-2/p} a_D^{-2} R^2) \leq \|\theta\|^2.$$

Hence, our test is powerful in such a situation. From now on, we assume that the set  $A$  is not empty:  $D^\dagger \leq N$ . For all  $j \in J$ , set

$$\mu_j = \ln \left( \frac{\pi^2 (j - D^\dagger)^2}{2\alpha} \right).$$

Two different situations can occur:

- 1/ For all  $j > D^\dagger$ ,  $b_j^2 \theta_j^2 \leq \epsilon^2 \mu_j^2$  for some sequence  $\mu_j$ , i.e. all the coefficients  $\theta_k$  have poor importance after the rank  $D^\dagger$ .
- 2/ There exists at least  $j \in \{D^\dagger, \dots, N\}$  such that  $b_j^2 \theta_j^2 \geq \epsilon^2 \mu_j^2$ , i.e. there exist significant coefficients after the rank  $D^\dagger$ .

First consider the case 1/. Recall that in this case, the set  $A$  is not empty and there exists  $j \in \{D^\dagger, \dots, N\}$  such that  $b_j^2 \theta_j^2 > \sigma^2 \mu_j^2$ . In this particular setting, we have to use the threshold test in order to detect these coefficients. More precisely,

$$P_\theta(\Phi_\alpha^\dagger = 0) \leq P_\theta \left( \sup_{j > D^\dagger} \Phi_{\{j\}, 2\alpha/\pi^2(j-D^\dagger)^2} = 0 \right) \leq P_\theta \left( \Phi_{\{j'\}, 2\alpha/\pi^2(j'-D^\dagger)^2} = 0 \right),$$

where  $j'$  is such that  $\theta_{j'}^2 > \sigma^2 \mu_{j'}^2$ . Thanks to the results established in a previous section for threshold tests, we know that our test is powerful as soon as:

$$\theta_{j'}^2 > \sigma^2 b_{j'}^{-2} \ln \left( \frac{\pi^2 (j - D^\dagger)^2}{2\alpha} \right).$$

This is exactly the assumption made in case 1/. Now, we consider the second point. Let  $j > D^\dagger$ ,

$$\begin{aligned}\theta_j^2 &= \theta_j^p \theta_j^{2-p}, \\ &= \theta_j^{2-p} b_j^{2-p} \theta_j^p b_j^{-(2-p)}, \\ &\leq \sigma^{2-p} \mu_j^{2-p} \theta_j^p b_j^{-(2-p)}.\end{aligned}$$

Then, we get

$$\begin{aligned}\sum_{k>D^\dagger} \theta_j^2 &\leq \sigma^{2-p} \sum_{j>D^\star} \theta_j^p b_j^{-(2-p)} \mu_j^{2-p}, \\ &= \sigma^{2-p} \sum_{j>D^\star} a_j^p \theta_j^p a_j^{-p} b_j^{-(2-p)} \mu_N^{2-p}, \\ &\leq \sigma^{2-p} R^p \max_{j>D^\dagger} a_j^{-p} b_j^{-(2-p)} \mu_N^{2-p}.\end{aligned}$$

Since the sequence  $(a_j^{-p} b_j^{-(2-p)})_{j \in \mathbb{N}}$  is assumed to be monotone non increasing, we can control the bias as follows

$$\sum_{k>D^\dagger} \theta_j^2 \leq \sigma^{2-p} R^p a_{D^\dagger}^{-p} b_{D^\dagger}^{-(2-p)} \mu_N^{2-p}.$$

In order to conclude the proof, we have to bound the right hand side of the above inequality. First assume that the problem is mildly ill-posed, i.e.  $(b_k)_{k \in \mathbb{N}^\star} \sim (k^{-t})_{k \in \mathbb{N}}$  for some  $t > 0$ . Then

$$\begin{aligned}D^\dagger &= \inf \left\{ D \in J, R^2 a_D^{-2} (\sqrt{D})^{1-2/p} \leq \sigma^2 D^{2t+1/2} \right\}, \\ &= \inf \left\{ D \in J, R^2 a_D^{-2} \leq \sigma^2 D^{2t+1/p} \right\}.\end{aligned}$$

Thus

$$\sum_{k>D^\dagger} \theta_j^2 \leq \sigma^{2-p} \sigma^p (D^\dagger)^{tp+1/2} (D^\dagger)^{2t-tp} \mu_N^{2-p} \leq \sigma^2 (D^\dagger)^{2t+1/2} \mu_N^{2-p}. \quad (5.20)$$

Hence

$$P_f(\Phi_\alpha^\dagger = 0) \leq P_\theta(\Phi_{D^\dagger, \alpha/2} = 0) \leq \beta,$$

for all sequence  $\theta$  satisfying

$$\sum_{k=1}^{D^\dagger} \theta_k^2 \geq C_{\alpha, \beta} \sigma^2 (D^\dagger)^{2t+1/2} \Leftrightarrow \|\theta\|^2 \geq C_{\alpha, \beta} \sigma^2 (D^\dagger)^{2t+1/2} + \sum_{k>D^\dagger} \theta_j^2. \quad (5.21)$$

The first point of Proposition 4 follows from (5.20) and (5.21). Now assume that the problem is severly ill-posed, i.e.  $(b_k)_{k \in \mathbb{N}^\star} \sim (e^{-\gamma k^r})_{k \in \mathbb{N}^\star}$  for some positive constants  $\gamma$  and  $r$ . In this setting,

$$D^\dagger = \inf \left\{ D \in J, R^2 a_D^{-2} \sqrt{D}^{1-p/2} \leq \sigma^2 e^{2\gamma D^r} \right\}.$$

Hence,

$$\sum_{j>D^\dagger} \theta_j^2 \leq \sigma^2 e^{2\gamma D^r} \sqrt{D^\dagger}^{1-p/2} \mu_N^{2-p}.$$

An inequality similar to (5.21) holds, which concludes the second point.

## 6 Proof of Theorem 2

The test  $\Phi_\alpha$  is obviously of level  $\alpha$  thanks to Bonferroni's inequality :

$$\begin{aligned}\mathbb{P}_0(\Phi_\alpha = 1) &\leq \mathbb{P}_0(\Phi_{\alpha/2}^{(1)} = 1) + \mathbb{P}_0(\Phi_{\alpha/2}^{(2)} = 1) \\ &\leq \frac{\alpha}{2} + \frac{\alpha}{2} \leq \alpha.\end{aligned}$$

Let us now evaluate the power of the test.

$$\mathbb{P}_\theta(\Phi_\alpha = 1) \geq \max\left(\mathbb{P}_\theta(\Phi_{\alpha/2}^{(1)} = 1) + \mathbb{P}_\theta(\Phi_{\alpha/2}^{(2)} = 1)\right).$$

It follows from Proposition 2 that for all  $\theta \in \mathcal{S}_{k,n}$  such that

$$\|\theta\|_2^2 \geq C(\alpha, \beta) \left( \sum_{j=1}^n \sigma_j^4 \right)^{1/2},$$

we have  $\mathbb{P}_\theta(\Phi_\alpha = 1) \geq 1 - \beta$ . It remains to evaluate the power of the test  $\Phi_{\alpha/2}^{(2)}$ .

$$\begin{aligned}\mathbb{P}_\theta(\Phi_\alpha^{(2)} = 0) &= \mathbb{P}_\theta(\forall j \in \{1, \dots, n\}, \frac{Y_j^2}{\sigma_j^2} \leq q_{n,1-\alpha}) \\ &\leq \inf_{1 \leq j \leq n} \mathbb{P}_\theta\left(\frac{Y_j^2}{\sigma_j^2} \leq q_{n,1-\alpha}\right).\end{aligned}$$

$$\begin{aligned}\mathbb{P}_\theta\left(\frac{Y_j^2}{\sigma_j^2} \leq q_{n,1-\alpha}\right) &= \mathbb{P}\left(|\theta_j + \sigma_j \epsilon_j| \leq \sigma_j \sqrt{q_{n,1-\alpha}}\right) \\ &\leq \mathbb{P}\left(|\theta_j| - \sigma_j |\epsilon_j| \leq \sigma_j \sqrt{q_{n,1-\alpha}}\right) \\ &\leq \mathbb{P}\left(\sigma_j |\epsilon_j| \geq |\theta_j| - \sigma_j \sqrt{q_{n,1-\alpha}}\right).\end{aligned}$$

Let  $q_\beta$  denote the  $1 - \beta$  quantile of  $|\epsilon_j|$ . We obtain that if

$$\exists j \in \{1, \dots, n\}, \quad |\theta_j| \geq \sigma_j (q_\beta + \sqrt{q_{n,1-\alpha}}), \tag{6.22}$$

then

$$\mathbb{P}_\theta(\Phi_\alpha^{(2)} = 0) \leq \beta.$$

Condition (6.22) is equivalent to

$$\exists m \in \mathcal{M}_{k,n}, \quad \sum_{j \in m} \theta_j^2 \geq \sum_{j \in m} \sigma_j^2 (q_\beta + \sqrt{q_{n,1-\alpha}})^2.$$

In particular, if

$$\|\theta\|_2^2 \geq \left( \sum_{j, \theta_j \neq 0} \sigma_j^2 \right) (q_\beta + \sqrt{q_{n,1-\alpha}})^2,$$

then (6.22) holds. This implies that for all  $\theta \in \mathcal{S}_{k,n}$  such that

$$\|\theta\|_2^2 \geq \max_{m \in \mathcal{M}_{k,n}} \left( \sum_{j \in m} \sigma_j^2 \right) (q_\beta + \sqrt{q_{n,1-\alpha}})^2,$$

we have  $\mathbb{P}_\theta(\Phi_\alpha^{(2)} = 0) \leq \beta$ . It remains to give an upper bound for  $q_{n,1-\alpha}$ . We use the inequality  $\mathbb{P}(|\epsilon_1| \geq x) \leq \exp(-x^2/2)$ . This leads to

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq j \leq n} \epsilon_j^2 \geq 2 \ln(n/\alpha)\right) &\leq n\mathbb{P}(|\epsilon_1| \geq \sqrt{2 \ln(n/\alpha)}) \\ &\leq \alpha. \end{aligned}$$

Hence,  $q_{n,1-\alpha} \leq 2 \ln(n/\alpha)$ , which concludes the proof of Theorem 2.

The proof of Theorem 2 follows directly from (6.22).

## 7 Appendix

### Proof of Lemma 2 :

We first compute the Laplace transform of  $\hat{T}$ . Easy computations show that for  $t < 1/(2\sigma_j^2)$ ,

$$\mathbb{E} \left[ \exp(t(\theta_j + \sigma_j \epsilon_j)^2) \right] = \frac{1}{\sqrt{1 - 2t\sigma_j^2}} \exp\left(\frac{t\theta_j^2}{1 - 2t\sigma_j^2}\right).$$

This implies that for  $t < \min_{1 \leq j \leq D} 1/(2\sigma_j^2)$ ,

$$\mathbb{E} \left[ \exp(t\hat{T}) \right] = \exp\left(\sum_{j=1}^D \frac{t\theta_j^2}{1 - 2t\sigma_j^2}\right) \prod_{j=1}^D \frac{1}{\sqrt{1 - 2t\sigma_j^2}}.$$

Moreover,

$$\mathbb{E}(\hat{T}) = \sum_{j=1}^D \theta_j^2 + \sum_{j=1}^D \sigma_j^2.$$

This leads to

$$\begin{aligned} \mathbb{E} \left[ \exp(t(\hat{T} - \mathbb{E}(\hat{T}))) \right] &= \exp\left(\sum_{j=1}^D \frac{2t^2\theta_j^2\sigma_j^2}{1 - 2t\sigma_j^2} - t\sigma_j^2\right) \prod_{j=1}^D \frac{1}{\sqrt{1 - 2t\sigma_j^2}} \\ &= \exp\left(\sum_{j=1}^D \frac{2t^2\theta_j^2\sigma_j^2}{1 - 2t\sigma_j^2} - t\sigma_j^2\right) \exp\left(-\frac{1}{2} \sum_{j=1}^D \log(1 - 2t\sigma_j^2)\right). \end{aligned}$$

We use the following inequality which holds for  $x < 1/2$  :

$$x \left[ \frac{1}{2} \log(1 - 2x) + x + \frac{x^2}{1 - 2x} \right] \geq 0. \quad (7.23)$$



This inequality implies that for all  $t < \min_{1 \leq j \leq D} 1/2\sigma_j^2$ ,

$$\log \mathbb{E} \left[ \exp(t(\hat{T} - \mathbb{E}(\hat{T}))) \right] \leq \sum_{j=1}^D \frac{t^2 \sigma_j^4}{1 - 2t\sigma_j^2} + 2t^2 \sum_{j=1}^D \frac{\theta_j^2 \sigma_j^2}{1 - 2t\sigma_j^2}.$$

This leads to

$$\log \mathbb{E} \left[ \exp(t(\hat{T} - \mathbb{E}(\hat{T}))) \right] \leq \frac{t^2 \Sigma}{1 - 2t \sup_{1 \leq j \leq D} (\sigma_j^2)}.$$

We now use the following lemma which is proved in Birgé [4] (see Lemma 8.2) :

**Lemma 3** *Let  $X$  be a random variable such that*

$$\log (\mathbb{E} [\exp(tX)]) \leq \frac{(at)^2}{1 - bt} \quad \text{for } 0 < t < 1/b$$

where  $a$  and  $b$  are positive constants. Then

$$\mathbb{P} (X \geq 2a\sqrt{x} + bx) \leq \exp(-x) \quad \text{for all } x > 0.$$

Hence, inequality (5.17) is proved. Let us now prove inequality (5.18).

For all  $z \in \mathbb{R}$ ,

$$\begin{aligned} \mathbb{P} (\hat{T} - \mathbb{E}(\hat{T}) \leq -z) &= \mathbb{P} (-\hat{T} + \mathbb{E}(\hat{T}) - z \geq 0) \\ &\leq \inf_{t > 0} \mathbb{E} \left( e^{t(-\hat{T} + \mathbb{E}(\hat{T}) - z)} \right) \\ &\leq \inf_{t < 0} \mathbb{E} \left( e^{t(\hat{T} - \mathbb{E}(\hat{T}) + z)} \right). \end{aligned}$$

We have, from the above computations

$$\ln \left( \mathbb{E} \left( e^{t(\hat{T} - \mathbb{E}(\hat{T}) + z)} \right) \right) = \sum_{j=1}^D \left[ \frac{2t^2 \theta_j^2 \sigma_j^2}{1 - 2t\sigma_j^2} - t\sigma_j^2 - \frac{1}{2} \ln(1 - 2t\sigma_j^2) \right] + tz.$$

We now use (7.23) for  $x = t\sigma_j^2$  with  $t < 0$ . We obtain

$$\frac{1}{2} \ln(1 - 2t\sigma_j^2) + t\sigma_j^2 + \frac{t^2 \sigma_j^4}{1 - 2t\sigma_j^2} \leq 0.$$

This implies that

$$\frac{2t^2 \theta_j^2 \sigma_j^2}{1 - 2t\sigma_j^2} \leq -2t\theta_j^2 - \frac{\theta_j^2}{\sigma_j^2} \ln(1 - 2t\sigma_j^2).$$

Hence, for all  $t < 0$ ,  $z \in \mathbb{R}$ ,

$$\mathbb{E} \left( e^{t(\hat{T} - \mathbb{E}(\hat{T}) + z)} \right) \leq \exp \left[ - \sum_{j=1}^D \left( \frac{1}{2} \log(1 - 2t\sigma_j^2) + t\sigma_j^2 \right) \left( 1 + 2\frac{\theta_j^2}{\sigma_j^2} \right) + tz \right].$$

We use this inequality with  $z = 2\sqrt{\Sigma x}$ , and  $t_x = -\sqrt{x}/\sqrt{\Sigma}$ .

$$\mathbb{P}\left(\hat{T} - \mathbb{E}(\hat{T}) \leq -2\sqrt{\Sigma x}\right) \leq \mathbb{E}\left(e^{t_x(\hat{T} - \mathbb{E}(\hat{T}) + 2\sqrt{\Sigma x})}\right).$$

Moreover,

$$\mathbb{E}\left(e^{t_x(\hat{T} - \mathbb{E}(\hat{T}) + 2\sqrt{\Sigma x})}\right) = \exp\left[-\sum_{j=1}^D \left(\frac{1}{2} \log\left(1 - 2\frac{\sqrt{x}}{\sqrt{\Sigma}}\sigma_j^2\right) - \frac{\sqrt{x}}{\sqrt{\Sigma}}\sigma_j^2\right) \left(1 + 2\frac{\theta_j^2}{\sigma_j^2}\right) - 2x\right].$$

We use the following inequality which holds for all  $u \geq 0$  :

$$\frac{1}{2} \log(1 + 2u) - u \geq -u^2,$$

and we apply this inequality to  $u = -t_x\sigma_j^2$ . We obtain that for all  $x \geq 0$ ,

$$\mathbb{P}\left(\hat{T} - \mathbb{E}(\hat{T}) \leq -2\sqrt{\Sigma x}\right) \leq \exp(-x).$$

This concludes the proof of Lemma 2.

## References

- [1] Aldous, D. J. (1985) *Exchangeability and related topics*. Ecole d'été de probabilités de Saint-Flour XIII, Lect. Notes Math. 11117, 1-198.
- [2] Baraud, Y. (2002) *Non asymptotic minimax rates of testing in signal detection*, Bernoulli, **8**, 577-606.
- [3] Baraud, Y., Huet, S., and Laurent, B. (2003) *Adaptive tests of linear hypotheses by model selection*, Ann. Statist., **31**, no. 1, 225-251.
- [4] Birgé, L. (2001) *An alternative point of view on Lepski's method*, State of the art in probability and statistics (Leiden, 1999) (ed. Monogr., IMS Lecture Notes, **36**, 113-133.
- [5] Butucea, C. (2007) *Goodness-of-fit testing and quadratic functional estimation from indirect observations*, Ann. Statist., **35**, no. 5, 1907-1930.
- [6] Cavalier, L. (2008) *Nonparametric statistical inverse problems*, Inverse Problems, 24(3).
- [7] Donoho, D. (1995), *Nonlinear solution of linear inverse problems by Wavelet-Vaguelette decomposition*, Applied and computational harmonic analysis, **2**, 101-126.
- [8] Fromont, M., Laurent, B. and Reynaud-Bouret, P. (2009) *Adaptive test of homogeneity for a Poisson process*, ArXiv:0905.0989v1, Submitted.
- [9] Ermakov, M. S. (2006), *Minimax detection of a signal in the heteroscedastic Gaussian white noise*, J. Math. Sci., **137**, No. 1, 4516-4524.
- [10] Ingster, Yu.I. (1993) *Asymptotically minimax testing for nonparametric alternatives I-II-III*, Math. Methods Statist., **2**, 85-114, 171-189, 249-268.

- [11] Ingster, Yu.I. and Suslina, I.A. (1998) *Minimax detection of a signal for Besov bodies and balls*, Problems Inform. Transmission, **34**, 48-59.
- [12] Laurent, B. and Massart, P. (2000) *Adaptive estimation of a quadratic functional by model selection*, Ann. Statist., **28**, no. 5, 1302-1338.
- [13] Lepski, O. V., and Spokoiny, V. G. (1999) *Minimax nonparametric hypothesis testing: the case of inhomogeneous alternative*, Bernoulli, **5**, 333-358.