Electronic Journal of Statistics
Vol. 0 (2010) 1–8
ISSN: 1935-7524

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

Béatrice Laurent, Jean-Michel Loubes and Clément Marteau

Institut de Mathématiques de Toulouse,
INSA de Toulouse,
Université de Toulouse.
e-mail: beatrice.laurent@insa-toulouse.fr ;
jean-michel.loubes@math.univ-toulouse.fr ;
clement.marteau@insa-toulouse.fr

Abstract: The aim of this paper is to establish non-asymptotic minimax rates for goodness-of-fit hypotheses testing in an 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 norm, without assumption on $(\sigma_j)_{j\in J}$ and on several functions spaces. Our point of view is entirely non-asymptotic.

AMS 2000 subject classifications: 62G05, 62K20.
Keywords and phrases: Goodness-of-fit tests, heterogeneous variances, inverse problems.

1. Introduction

We consider the following heteroscedastic statistical model

$$Y_j = \theta_j + \sigma_j \epsilon_j, \quad j \in J, \tag{1.1}$$

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 θ has to be tested from the observations $(Y_j)_{j\in J}$ in order to decide whether "$\theta = 0$" or not. The particular case $\sigma_j = \sigma$ for all $j \in J$ corresponds to the classical statistical model where 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.

We point out that the model (1.1) can describe inverse problems. Indeed, for a linear operator T on an Hilbert space H with inner product $(.,.)$, consider an unknown function f indirectly observed in a Gaussian white noise model

$$Y(g) = (Tf, g) + \sigma\epsilon(g),\ g \in H, \tag{1.2}$$

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

$$T\phi_j = b_j\psi_j, \quad T^*\psi_j = b_j\phi_j,\ j\in\mathbb{N}^\star,$$

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

$$Z_j = b_j\theta_j + \sigma\epsilon_j,\ j\in\mathbb{N}^\star, \tag{1.3}$$

with $\epsilon_j = \epsilon(\psi_j)$, $(Tf,\psi_j) = b_j\theta_j$ and $\theta_j = (f,\phi_j)$. This model is often considered in the inverse problem literature, see eg [7]. Setting $Y_j = b_j^{-1}Z_j$ and $\sigma_j = \sigma b_j^{-1}$ for all $j\in\mathbb{N}^\star$, we obtain (1.1). Hence inference on observations from model (1.1) provides the same results for inverse problems. We stress that if estimation issues for inverse problem have been well studied over the past years (see for instance [20], [7] or [18, 19] for a model selection approach), tests for inverse problems have been mostly investigated only for the very specific case of the convolution problem, see in [6] and references therein. In the general case some tests are provided in [5], but our results in this paper and in [16] provide the first rates in this context, simultaneously with the work in [12].

For all $\theta\in l_2(J)$, we set $\|\theta\|^2 = \sum_{j\in J}\theta_j^2$. The purpose of this paper is to provide rates of testing for the hypothesis "$\theta = 0$" against the alternative "$\theta\in\mathcal{F},\ \|\theta\|\geq\rho$". More precisely, let us fix some level $\alpha\in]0,1[$, and consider a level α test Φ_α with values in $\{0,1\}$ in order to test the null hypothesis "$\theta = 0$" (we reject the null hypothesis when $\Phi_\alpha = 1$). Then, given $\beta\in]0,1[$, and a class of vectors $\mathcal{F}\subset l_2(J)$, we define the uniform separation rate $\rho(\Phi_\alpha,\mathcal{F},\beta)$ of the test Φ_α over the class $\mathcal{F}$ with respect to the l_2 norm as the smallest radius ρ such that the test guarantees a power greater that $1-\beta$ for all alternatives $\theta\in\mathcal{F}$ such that $\|\theta\|\geq\rho$. More formally

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

We define the (α,β) minimax rate of testing over the class $\mathcal{F}$ by

$$\rho(\mathcal{F},\alpha,\beta) = \inf_{\Phi_\alpha}\rho(\Phi_\alpha,\mathcal{F},\beta),$$

where the infimum is taken over all level α test Φ_α. The aim of the paper is to determine this minimax rate of testing over various classes of alternatives $\mathcal{F}$, for signal detection in Model (1.1) with respect to the l_2 norm.

The main reference for computing minimax rates of testing over non parametric alternatives is the series of paper due to Ingster [11], where various statistical models and a wide range of sets of alternatives are considered. Lepski and Spokoiny [17] 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 [13]. Ermakov [10] determines a family of asymptotic minimax procedures 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] considers a non asymptotic point of view, which means that the noise level σ 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 Model (1.3) leads to the so-called mildly and severely ill-posed inverse problems. When allowing the noise level to decrease towards zero, we recover asymptotic rates of testing. Note that we do not aim at providing adaptive minimax rates, which will be the core of a future work.

The paper is organized as follows. In Section 2, we provide lower bounds for the minimax separation rate over classes of vectors θ 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 and 5, we compute minimax rates of testing over ellipsoids and l_p balls. The proofs are gathered in Sections 6 and 7.

To end this introduction, let us define some notations. Assume that $Y = (Y_j)_{j\in J}$ obeys to Model (1.1). We denote by θ 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 Φ_α 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 $\Phi_\alpha = 1$. Finally, for all $x \in \mathbb{R}$, we denote by $\lfloor x \rfloor$ the greater integer smaller than x and we set $\lceil x \rceil = \lfloor x \rfloor + 1$.

2. Non asymptotic lower bounds

The bounds will be established for two classes of signals characterized by their non zeros coefficients. The first one deals with the elementary case where the coefficients are equal to zero after a certain rank. The second one concerns the so-called sparse signals which are defined by the amount of non zeroes coefficients which can be located at different scales.

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* $\beta \in]0, 1-\alpha[$, $c(\alpha,\beta) = (2\ln(1+4(1-\alpha-\beta)^2))^{1/4}$ *and*

$$\rho_D = \left(\sum_{j=1}^{D} \sigma_j^4\right)^{1/4}.$$

The following result holds:

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

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

$$\rho(S_D, \alpha, \beta) \ge c(\alpha,\beta)\rho_D.$$

This proposition can be understood as follows: whatever the α-level test chosen, for all $\rho \le c(\alpha,\beta)\rho_D$, there exists a signal $\theta \in S_D$ with norm ρ, such that the error of the second kind is greater than β. The result obtained in Proposition 1 coincides 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 $k \le n$. When $J = \{1,\ldots,N\}$, we assume that $n \le N$. We want to obtain lower bounds for the minimax separation rate of signal detection over the set $\mathcal{S}_{k,n}$ defined by

$$\mathcal{S}_{k,n} = \{\theta \in l_2(J),\ \forall j > n,\ \theta_j = 0,\ \text{Card}\,\{j \le n, \theta_j \ne 0\} \le k\}. \tag{2.1}$$

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

$$\Sigma_{l,k}^2 = \sum_{j=l+1}^{l+k} \sigma_{(j)}^2. \tag{2.2}$$

Let $\beta \in]0, 1-\alpha[$, *such that* $\alpha+\beta \le 0.59$ *and*

$$\rho_{k,n} = \left[\max_{0\le l\le n-k} \Sigma_{l,k} \ln^{1/2}\left(1 + \frac{n-l}{k^2} \vee \sqrt{\frac{n-l}{k^2}}\right) \vee \left(\sum_{j=n-k+1}^{n} \sigma_{(j)}^4\right)^{1/4}\right]. \tag{2.3}$$

The following result holds

$$\inf_{\Phi_\alpha} \sup_{\theta\in\mathcal{S}_{k,n}, \|\theta\|\ge\rho_{k,n}} \mathbb{P}_\theta(\Phi_\alpha = 0) \ge \beta.$$

This implies that the minimax rate of signal detection over $\mathcal{S}_{k,n}$ *with respect to the* l_2 *norm satisfies*

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

Comments : Let us consider three cases governing 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, $\Sigma^2_{l,k} = \sigma^2 k$ for all l and, taking $l = 0$, we obtain that
$$\rho^2_{k,n} \geq \sigma^2 k \ln\left(1 + \frac{n}{k^2} \vee \sqrt{\frac{n}{k^2}}\right),$$
which corresponds to the lower bound established by Baraud [2].
2. When $k \leq n/2$ and $\Sigma^2_{\lfloor n/2 \rfloor,k} \geq C\Sigma^2_{n-k,k}$ for some absolute constant C (independent of k and n), we obtain that
$$\rho^2_{k,n} \geq \left[C\Sigma^2_{n-k,k} \ln\left(1 + \frac{n}{2k^2} \vee \sqrt{\frac{n}{2k^2}}\right) \vee \left(\sum_{j=n-k+1}^{n} \sigma^4_{(j)}\right)^{1/2}\right]. \quad (2.4)$$
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 $\rho^2_{k,n}$ is $\Sigma^2_{n-k,k}$ which corresponds to the largest possible variance for a set of cardinality k in $\{1, \dots, n\}$, indeed
$$\Sigma^2_{n-k,k} = \max_{m\in\mathcal{M}_{k,n}} \sum_{j\in m} \sigma^2_j,$$
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, $\sigma_j = \sigma j^\gamma$ for some $\sigma > 0$ and $\gamma > 0$. Actually this corresponds to a mildly ill-posed inverse problem. In this case,
$$\Sigma^2_{n-k,k} \leq k\sigma^2 n^{2\gamma}, \quad \Sigma^2_{n/2,k} \geq k\sigma^2\sigma^2 \left(\frac{n}{2}\right)^{2\gamma} \geq \frac{1}{2^{2\gamma}} \Sigma^2_{n-k,k}.$$
3. When $(\sigma_j)_{j\in J}$ grows at an exponential rate : $\sigma_j = \sigma \exp(\gamma j)$ for some $\sigma > 0$ and $\gamma > 0$, we obtain that $\rho^2_{k,n} \geq \sigma^2_{(n)}$, providing a bound for the severely ill-posed inverse problems.

3. Non asymptotic upper bounds

In this section, we construct upper bounds for the (α, β) 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 (α, β) minimax rate of testing with respect to the l_2 norm over a set $\mathcal{F}$ is bounded from above by ρ, it suffices to define a test statistic Φ_α such that the power of the test at each point θ in $\mathcal{F}$ satisfying $\|\theta\| \geq \rho$ 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 $\Phi_{D,\alpha}$ *be the level-*α *test defined by*

$$\Phi_{D,\alpha} = \mathbb{1}_{\sum_{j=1}^{D} Y_j^2 > t_{D,1-\alpha}(\sigma)}. \tag{3.1}$$

We define $x_\alpha = \ln(\alpha^{-1})$, $x_\beta = \ln(\beta^{-1})$

$$C(\alpha,\beta) = \sqrt{2x_\beta} + \sqrt{2(x_\alpha + x_\beta)} + \sqrt{2}(\sqrt{x_\alpha} + \sqrt{x_\beta})^{1/2} \tag{3.2}$$

and $\rho_D = \left(\sum_{j=1}^{D} \sigma_j^4\right)^{1/4}$.
For all $\theta \in S_D$ *such that* $\|\theta\| \geq C(\alpha,\beta)\rho_D$ *we have* $\mathbb{P}_\theta(\Phi_{D,\alpha} = 1) > 1-\beta$. *Hence, we obtain that*

$$\rho(S_D, \alpha, \beta) \leq C(\alpha,\beta)\rho_D.$$

We deduce from Propositions 1 and 2 that

$$c(\alpha,\beta)\rho_D \leq \rho(S_D,\alpha,\beta) \leq C(\alpha,\beta)\rho_D.$$

Hence the upper and lower bounds coincide up to multiplicative constants. Note that the computation of optimal constants remains an open problem.

Let us now propose a testing procedure for sparse signal detection. This procedure will be defined by a combination of two tests. 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 [9]. The second one is the test (3.1) with $D = n$. More precisely, for all $\alpha \in]0,1[$, let $t_{n,1-\alpha}(\sigma)$ denotes the $1-\alpha$ quantile of $\sum_{j=1}^{n} \sigma_j^2 \epsilon_j^2$.
Let $\Phi_\alpha^{(1)}$ be the test defined by

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

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

$$\Phi_\alpha^{(2)} = \mathbb{1}_{\max_{1\leq j\leq n}\left(\frac{Y_j^2}{\sigma_j^2}\right) > q_{n,1-\alpha}}.$$

Then define the level-α test

$$\Phi_\alpha = \max\left(\Phi_{\alpha/2}^{(1)}, \Phi_{\alpha/2}^{(2)}\right). \tag{3.3}$$

Theorem 2. *Assume that $Y = (Y_j)_{j\in J}$ obeys to Model (1.1). Let $\alpha, \beta \in]0,1[$ and Φ_α defined in (3.3). Let $C(\alpha,\beta)$ be defined by (3.2), If $\alpha \le 1/e$ and $n \ge 3$, for all $\theta \in \mathcal{S}_{k,n}$ satisfying*

$$\|\theta\|^2 \ge 4C^2(\alpha,\beta)\left[\left(\sum_{j=1}^n \sigma_j^4\right)^{1/2} \wedge \sum_{j,\theta_j\neq 0} \sigma_j^2 \ln(n)\right], \tag{3.4}$$

we have

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

Hence, we obtain that for all $k \in \{1,\ldots,n\}$,

$$\rho^2(\mathcal{S}_{k,n},\alpha,\beta) \le 4C^2(\alpha,\beta)\left[\left(\sum_{j=1}^n \sigma_j^4\right)^{1/2} \wedge \Sigma^2_{n-k,k} \ln(n)\right], \tag{3.5}$$

where $\Sigma_{l,k}$ has been defined in (2.2).

Comments : Let us compare these results with the lower bounds obtained in Theorem 1. For the sake of simplicity, we do not compute explicit constants until the end of this section.

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 $\gamma \ge 0$ (this includes the homoscedastic case). In this case, when $k \le n/2$ there exists a constant $C > 0$ such that $\Sigma^2_{\lfloor n/2\rfloor,k} \ge C\Sigma^2_{n-k,k}$. A lower bound for the (α,β) minimax separation rate of signal detection over $\mathcal{S}_{k,n}$ is given by (2.4). This lower bound has to be compared with the upper bound (3.5).

    (a) When $k = n^l$ with $l < 1/2$, the upper and lower bounds coincide and are of order $\Sigma^2_{n-k,k}\ln(n)$.

    (b) When $k = n^l$ with $l \ge 1/2$, the lower bound is of order $\Sigma^2_{n-k,k}\sqrt{n}/k$ and $\Sigma^2_{n-k,k} \ge Ck\sigma^2 n^{2\gamma}$, which leads to a lower bound of order $C\sigma^2 n^{2\gamma+1/2}$. The upper bound is smaller that $\left(\sum_{j=1}^n \sigma_j^4\right)^{1/2}$, which is smaller than $\sigma^2 n^{2\gamma+1/2}$. Hence, the two bounds coincide.

    (c) When $k = \sqrt{n}/\phi(n)$ where $\phi(n) \to +\infty$ and $\phi(n)/n \to 0$ as $n \to +\infty$ (typically $\phi(n) = \ln(n)$), the lower bound is of order $\Sigma^2_{n-k,k}\ln(\phi(n))$ and the upper bound is of order $\Sigma^2_{n-k,k}\ln(n)$.
    In this case, the upper and lower bound only differ from a logarithmic term. This gap is also observed in the homoscedastic model (see Baraud [2]) and remains, up to our knowledge, an open problem.

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 that $4C^2(\alpha,\beta)\left(\sum_{j=1}^n \sigma_j^4\right)^{1/2}$, 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 the test

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

which achieves the lower bound for the separation rate.

3. The result stated in (3.4) is more precise than the minimax upper bound given in (3.5). 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 $\|\theta\|^2$ is greater than the right hand term in (3.5) for the test to be powerful for this value of θ. The minimax bound given in (3.5) corresponds to the worst situation, that is the case where the set J_1 corresponds to the largest values for the variances.

Hence, we have provided, for the specific problem of signal detection for inverse problems, minimax rates for the both mildly and severely ill-posed problems except for the particular case of mildly inverse problems with a number of non zero coefficients of order $\sqrt{n}/\phi(n)$ specified in 1. (c).

4. Minimax rates over ellipsoids

In the previous sections, the only constraint on the signal was expressed through the number of non-zero coefficients. In several situations, one deals instead with infinite sequences having a finite number of significant coefficients, the reminder being considered as negligible (in a sense to be made precise 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 θ_k's and the associated rate of testing. We consider in the following two different kinds of function spaces: ellipsoids and l_p-bodies.

4.1. Non asymptotic 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 θ 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 characterizes the minimax rate of testing over $\mathcal{E}_{a,2}(R)$.

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

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

where ρ_D and $c(\alpha, \beta)$ have been introduced in Proposition 1. Moreover, for all $D \in J$,

$$\sup_{\theta \in \mathcal{E}_{a,2}(R), \|\theta\|^2 \geq C^2(\alpha,\beta)\rho_D^2 + R^2 a_D^{-2}} P_\theta(\Phi_{D,\alpha} = 0) \leq \beta,$$

where $\Phi_{D,\alpha}$ and $C(\alpha, \beta)$ are respectively defined by (3.1) and (3.2). Hence,

$$\rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) \leq \inf_{D \in J} (C^2(\alpha, \beta)\rho_D^2 + R^2 a_D^{-2}).$$

Proposition 3 presents both an upper and a lower bound for the minimax rate of testing over $\mathcal{E}_{a,2}(R)$. Remark that the upper bound is attained by the test Φ_α introduced in Proposition 2 where only signals with a finite number of non-zero coefficients were considered. We point out that 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 this bias can be controlled by taking advantage of the constraint expressed on the decay of θ.

4.2. Asymptotic minimax rates of testing for inverse problems

A good characterization of $\rho(\mathcal{E}_{a,2}(R), \alpha, \beta)$ can be obtained as soon as the lower and upper bounds in Proposition 3 are of the same order. As many statistical problems encountered in the literature, one has to find a trade-off between the bias $R^2 a_D^{-2}$ and some kind of variance term ρ_D^2. This trade-off can be performed in several situations, when introducing specific constraints on the sequences $(a_k)_{k \in J}$ and $(b_k)_{k \in J}$, hence leading to explicit rates of convergence as shown below.

Let α, β be fixed. We assume that $J = \mathbb{N}^\star$ and $(Z_j)_{j \in J}$ obeys to Model (1.3) with $(\nu_k)_{k \in \mathbb{N}^\star}$ a sequence of 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 $k \in \mathbb{N}^\star$, $c_1 k^l \leq \nu_k \leq c_2 k^l$.

The Table 1 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 behaviours for the sequences $(a_k)_{k \in \mathbb{N}^\star}$ and $(b_k)_{k \in \mathbb{N}^\star}$. For each case, we give $f(\sigma)$ such that for all $1 > \sigma > 0$, $C_1(\alpha, \beta) f(\sigma) \leq \rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) \leq C_2(\alpha, \beta) f(\sigma)$ where $C_1(\alpha, \beta)$ and $C_2(\alpha, \beta)$ denote positive constants independent of σ.
In the following table, s, t, ν, γ and r denote positive constants.

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

TABLE 1
Asymptotic minimax rates of testing on ellipsoids.

Here $\tilde{D}$ denotes the integer part of the solution of $\rho_D^2 = R^2 a_D^{-2}$.

Some of these rates have already been presented in the literature. The case $a_k \sim k^s$ and $b_k \sim k^{-t}$ was first studied in [14]. More recently, [12] deals with other cases. Similar rates are also available in [6] in the context of density estimation with errors in the variables. The aim of this discussion is to show that our approach can lead to important minimax results. Our point of view in this paper is indeed entirely non-asymptotic and is not restricted to ill-posed inverse problems.

Concerning severely ill-posed problems with supersmooth functions (i.e. $b_k \sim \exp(-\gamma k^r)$ and $a_k \sim \exp(\nu k^s)$), we do not handle the general case since we assume that $s \leq 1$. When this assumption is violated, the upper and lower bounds in Proposition 3 do not coincide: our test does not attain the minimax rate of testing. This is certainly due to our approach, which in some sense is related to a rough regularization scheme. We refer to [12] for a complete study of this case.

The whole algebra leading from Proposition 3 to the rates presented in Table 1 can be found in Section 6.3.2

5. Minimax rates of testing for inverse problems over l_p-bodies

Ellipsoids contain essentially smooth functions. In the particular case where θ corresponds to the Fourier coefficients of a given function f, the constraints expressed through the belonging to 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 subsection sequences θ 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 $0 < p < 2$.

5.1. Non asymptotic lower bounds

Let $(Z_j)_{j \in J}$ obey to the model

$$Z_j = b_j \theta_j + \sigma \epsilon_j,\ j \in J.$$

To the end of this section, we assume that the sequence $(b_j)_{j\in J}$ is polynomially or exponentially decreasing, which yet correspond to the main cases of interest in inverse problems. The lower bounds are given in the following theorem.

Theorem 3. *Assuming that for all $j \in J$, $b_j = j^{-\gamma}$ for some $\gamma \geq 0$, we obtain*

$$\rho^2(\mathcal{E}_{a,p}(R), \alpha, \beta) \geq \frac{1}{2^{1+2\gamma}} \sup_{D\in J} \left[\sqrt{D}^{1-2/p} R^2 a_D^{-2} \wedge \tilde{\rho}_D^2 \right]$$

where

$$\tilde{\rho}_D^2 = \sigma^2 D^{1/2+2\gamma} \ln\left(1 + \sqrt{\left(\frac{1}{2} - \frac{1}{D}\right) \vee 0}\right).$$

Assuming that for all $j \in J$, $b_j = \exp(-\gamma j)$ for some $\gamma > 0$, we obtain

$$\rho^2(\mathcal{E}_{a,p}(R), \alpha, \beta) \geq \sup_{D\in J} \left[\sqrt{D}^{1-2/p} R^2 a_D^{-2} \wedge \sigma^2 \exp(2\gamma D) \right].$$

5.2. Non asymptotic upper bounds

In order to attain the lower bound presented above, a test similar to the one introduced in Proposition 2 is not sufficient. On l_p-bodies, the bias after a given rank D is indeed more difficult to control than for ellipsoids. Some significant coefficients (in a sense which will be made precise in the proof) may be contained in the sequence θ after the rank D. Hence, we have to introduce specific tests in order to detect these coefficients. We first define

$$D^\dagger = \inf\left\{ D \in J, R^2 a_D^{-2} \sqrt{D}^{1-2/p} \leq \sigma^2 D^{1/2+2\gamma} \ln\left(1 + \sqrt{\left(\frac{1}{2} - \frac{1}{D}\right) \vee 0}\right) \right\}, \tag{5.1}$$

if the problem is mildly ill-posed and

$$D^\dagger = \inf\left\{ D \in J, R^2 a_D^{-2} \sqrt{D}^{1-2/p} \leq \sigma^2 \exp(2\gamma D) \right\}. \tag{5.2}$$

if the problem is severely ill-posed. By convention, $D^\dagger = N$ if the set in (5.1) (or in (5.2)) is empty.

For all $j \in J$ and $\alpha \in]0,1[$, we introduce

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

where $z_{j,\alpha}$ denotes the $1-\alpha/2$ quantile of a Gaussian random variable with mean 0 and variance σ_j^2. We now define the test

$$\Phi_{\text{loc},\alpha/2} = \begin{cases} \sup_{j\in\{D^\dagger+1,..N\}} \Phi_{\{j\},\alpha/2(N-D^\dagger)} \text{ if } D^\dagger < N, \\ 0 \text{ if } D^\dagger = N. \end{cases}$$

The final test that we consider is a combination of these two procedures :

$$\Phi_\alpha^\dagger = \Phi_{\mathrm{loc},\alpha/2} \vee \Phi_{D^\dagger,\alpha/2},$$

where the testing procedure $\Phi_{D^\dagger,\alpha/2}$ is defined by (3.1).

The following proposition emphasizes the performances of the test $\Phi_\alpha^\dagger$. The constants $C_{1,p}$ and $C_{2,p}$ given below are explicitly computable. An interested reader can find the value of $C_{1,p}$ at the end of the proof of Proposition 4.

Proposition 4. *Let $\alpha, \beta \in]0,1[$. We assume that $\left(a_j^{-p} b_j^{-(2-p)}\right)_{j\in\mathbb{N}^*}$ is a monotone non-increasing sequence. One observes*

$$Z_j = b_j\theta_j + \sigma\epsilon_j,\ j \in J = \{1, \dots, N\}.$$

The following result holds

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

with

- $\nu_N = C_{1,p} \log^{1-p/2}(N)$ *when* $b_k = k^{-\gamma}$ *for all* $k \in J$ *(mildly ill-posed inverse problems).*
- $\nu_N = C_{2,p} \log^{1-p/2}(N) \sqrt{D^\dagger}^{1-p/2}$ *when* $b_k = e^{-\gamma k}$ *for all* $k \in J$ *(severely ill-posed inverse problems),*

where $C_{1,p}, C_{2,p}$ denote positive constants independent of σ.

Remark that the test $\Phi_\alpha^\dagger$ reaches the lower bound established in Corollary 3 up to some logarithmic term. Hence, the lower and upper bounds presented respectively in Theorem 3 and Proposition 4 do not coincide. This drawback is not characteristic of the heteroscedastic model since a similar property occurs in the homoscedastic case ($\gamma = 0$): see [2] for more details. In this particular homoscedastic setting, the lower bound of Theorem 3 is known to be sharp according to the results of Spokoiny (see [21]) on Besov spaces. We do not know if a similar property occurs in the heteroscedastic model. This a difficult problem that should be addressed in a separate paper.

For the sake of convenience, the upper bound is only presented for $J = \{1, \dots, N\}$ which, roughly speaking, corresponds to the regression setting. Nevertheless, our result can be easily extended to the case where $J = \mathbb{N}^\star$. 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 σ^{-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 holds 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 [8].

5.3. Asymptotic minimax rates for inverse problems

The non asymptotic study of minimax rate of testing is at the core of the present paper. Nevertheless, in some particular settings, we can obtain asymptotic results. For instance, we have presented in Section 4.2 some cases where our results lead to explicit rates on ellipsoids, hence recovering some existing properties. Such a discussion is possible on l_p bodies, although, up to our knowledge, explicit asymptotic minimax rates of testing in an inverse problem setting have never been obtained.

Here, we deal with mildly ill-posed problems with polynomial l_p-bodies, i.e. $(a_k)_{k\in\mathbb{N}} \sim (k^s)_{k\in\mathbb{N}}$ for some $s > 0$.

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

$$C_2 \log^{1-p/2}(N)\sigma^{\frac{4s+2/p-1}{2s+2\gamma+1/p}} \geq \rho^2(\mathcal{E}_{a,p}(R), \alpha, \beta) \geq C_1 \sigma^{\frac{4s+2/p-1}{2s+2\gamma+1/p}},$$

where C_1, C_2 denote positive constant independent of σ.

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

6. Proofs

6.1. Proof of the lower bounds

The proofs of the lower bounds use a Bayesian approach extending the methods developed in the papers by Ingster [11] and by Baraud [2]. We use the following lemma :

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

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

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

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

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.

6.1.1. *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 μ_ρ be the distribution of $(\theta_1, \ldots, \theta_D)$. μ_ρ is a probability measure on

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

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

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

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

$$\mathbb{E}(\cosh^2(\lambda Z)) = \exp(\lambda^2)\cosh(\lambda^2). \tag{6.1}$$

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

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

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

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

For $\rho = c(\alpha,\beta)\rho_D$ we obtain :

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

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

6.1.2. 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 $l \in \{0,1,\ldots,n-k\}$. Let $\mathcal{M}_{k,l,n}$ denote the set of all subsets of $\{l+1,\ldots,n\}$ with cardinality k. Let $\hat{m}$ be a random set of $\{l+1,\ldots,n\}$, which is uniformly distributed on $\mathcal{M}_{k,l,n}$. This means that for all $m \in \mathcal{M}_{k,l,n}$, $\mathbb{P}(\hat{m}=m) = 1/C^k_{n-l}$. Let $(\omega_j, 1 \leq j \leq n)$ be i.i.d. Rademacher random variables, independent of $\hat{m}$. Let us recall that

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

We set

$$\theta_j = (\rho\omega_j\sigma_j/\Sigma_{l,k})\ \mathbb{1}_{j\in\hat{m}} \tag{6.2}$$

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

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

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

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

We use (6.1) together with $\mathbb{E}(\cosh(\lambda Z)) = \exp(\lambda^2/2)$ for a standard Gaussian variable Z. Since Y_j/σ_j is a standard normal variable, we obtain that

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

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

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

we obtain

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

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. [9]. Let us recall these arguments.

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

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 an hypergeometric variable with parameters $(n-l, k, k/(n-l))$. This leads to

$$\mathbb{E}_0\left(L^2_{\mu_\rho}(Y)\right) = \mathbb{E}\left[\exp(H\ln(s))\right], \qquad \text{where } s = \cosh\left(\frac{\rho^2}{\Sigma^2_{l,k}}\right)$$

and H is an hypergeometric variable with parameters $(n-l, k, k/(n-l))$. The variable H can be decomposed into a sum of dependent Bernoulli variables: $H = H_1 + \ldots + H_k$. Hence, we obtain

$$\mathbb{E}\left[\exp(H\ln(s))\right] = \mathbb{E}\left[\prod_{j=1}^k(1+(s-1)H_j)\right].$$

For all $j \in \{1, \ldots, k\}$, the distribution of $(H_{k_1}, \ldots, H_{k_j})$ is independent of $\{k_1, \ldots k_j\}$ and coincides with the distribution of $(H_1, \ldots, H_j)$. Note that

$$\begin{aligned}\mathbb{E}(H_1\ldots H_j) &= \mathbb{P}(H_1 = 1, \ldots, H_j = 1) = \frac{k(k-1)\ldots(k-j+1)}{(n-l)(n-l-1)\ldots(n-l-j+1)}\\ &\le \left(\frac{k}{n-l}\right)^j.\end{aligned}$$

This implies that

$$\mathbb{E}\left[\prod_{j=1}^{k}(1+(s-1)H_j)\right] \leq \sum_{j=0}^{k} C_k^j \left(\frac{(s-1)k}{n-l}\right)^j = \left(1+\frac{(s-1)k}{n-l}\right)^k.$$

Hence,

$$\mathbb{E}_0\left(L^2_{\mu_\rho}(Y)\right) \leq \exp\left[k\ln\left(1+\frac{k}{n-l}\left(\cosh\left(\frac{\rho^2}{\Sigma^2_{l,k}}\right)-1\right)\right)\right].$$

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

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

then

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

The right hand side of the above inequality equals A. We finally obtain that

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

By Lemma 1, this implies that

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

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

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

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

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

In order to prove that

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

we define, as in the proof of Proposition 1,

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

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 $\|\theta\|^2 = \rho^2$. We now conclude as in the proof of Proposition 1, using that $c(\alpha,\beta) = (2\ln(c))^{1/4} \geq 1$ if $\alpha+\beta \leq 0.59$.

6.2. Proof of the upper bounds

6.2.1. Proof of Proposition 2

In order to prove Proposition 2, we have to show that for all $\theta \in S_D$ such that $\|\theta\| \geq C(\alpha,\beta)\rho_D$,

$$\mathbb{P}_\theta\left(\sum_{j=1}^{D} Y_j^2 \leq t_{D,1-\alpha}(\sigma)\right) < \beta. \tag{6.3}$$

We denote by $t_{D,\beta}(\theta,\sigma)$ the β 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 (6.3), it suffices to show that

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

To prove this inequality, we will first give an upper bound for $t_{D,1-\alpha}(\sigma)$ and then 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 [15] (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 $\sup_{1\leq j\leq D}\sigma_j^2 \leq \left(\sum_{j=1}^{D}\sigma_j^4\right)^{1/2}$,

$$t_{D,1-\alpha}(\sigma) \leq \sum_{j=1}^{D}\sigma_j^2 + C(\alpha)\left(\sum_{j=1}^{D}\sigma_j^4\right)^{1/2}. \tag{6.4}$$

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 $\epsilon_1, \ldots \epsilon_D$ *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 $x \geq 0$ *:*

$$\mathbb{P}\left(\hat{T} - \mathbb{E}(\hat{T}) \geq 2\sqrt{\Sigma x} + 2 \sup_{1 \leq j \leq D} (\sigma_j^2) x \right) \leq \exp(-x). \tag{6.5}$$

$$\mathbb{P}\left(\hat{T} - \mathbb{E}(\hat{T}) \leq -2\sqrt{\Sigma x} \right) \leq \exp(-x). \tag{6.6}$$

The proof of this lemma is given in the Appendix.
Inequality (6.6) 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, $t_{D,\beta}(\theta, \sigma) \geq \sum_{j=1}^{D} (\theta_j^2 + \sigma_j^2) - 2\sqrt{\Sigma x_\beta}$. (6.3) is satisfied if $t_{D,1-\alpha}(\sigma) < t_{D,\beta}(\theta, \sigma)$, which holds as soon as

$$\sum_{j=1}^{D} \theta_j^2 - 2\sqrt{\Sigma x_\beta} > 2\sqrt{x_\alpha} \sqrt{\sum_{j=1}^{D} \sigma_j^4} + 2x_\alpha \sup_{1 \leq j \leq D} (\sigma_j^2). \tag{6.7}$$

Let us note that

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

Hence, the following inequality implies (6.7) :

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

Easy computations show that this inequality holds if

$$\left(\sum_{j=1}^{D} \theta_j^2 \right)^{1/2} \geq \rho_D \left[\sqrt{2x_\beta} + \sqrt{2(x_\alpha + x_\beta)} + \sqrt{2}(\sqrt{x_\alpha} + \sqrt{x_\beta})^{1/2} \right].$$

Hence, we have proved that

$$\rho(S_D, \alpha, \beta) \leq C(\alpha, \beta)\rho_D.$$

which concludes the proof of Proposition 2.

6.2.2. *Proof of Theorem 2*

The test Φ_α is obviously of level α thanks to Bonferroni's inequality :

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

Let us now evaluate the power of the test.

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

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

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

we have $\mathbb{P}_\theta(\Phi^{(1)}_{\alpha/2} = 1) > 1-\beta$. Let us remark that $C(\alpha/2, \beta) \leq \sqrt{2}C(\alpha, \beta)$ since $\alpha \leq 1/2$.

It remains to evaluate the power of the test $\Phi^{(2)}_{\alpha/2}$.

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

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

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

$$\exists j \in \{1, \ldots, n\}, \quad |\theta_j| > \sigma_j(q_{1-\beta} + \sqrt{q_{n,1-\alpha}}), \tag{6.8}$$

then

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

Condition (6.8) is equivalent to

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

In particular, if

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

then (6.8) holds. This implies that for all $\theta \in \mathcal{S}_{k,n}$ such that

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

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

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

Hence, $q_{n,1-\alpha} \leq 2\ln(n/\alpha)$ and $q_{1-\beta} \leq \sqrt{2\ln(1/\beta)}$. By assumption, $\ln(n) \geq 1$ and $\ln(1/\alpha) \geq 1$. This implies that $\ln(n/\alpha) \leq 2\ln(n)\ln(1/\alpha)$. Let us now remark that $\sqrt{2x_\beta}+2\sqrt{2x_\alpha} \leq 2C(\alpha,\beta)$, which implies that $q_{1-\beta}+\sqrt{q_{n,1-\alpha/2}} \leq 2C(\alpha,\beta)$. This concludes the proof of Theorem 2.

6.3. Proof of minimax rates on ellipsoids and l_p-bodies

6.3.1. Proof of Proposition 3

We first prove the lower bound. For all $D \in J$, introduce $r_D^2 = c^2(\alpha,\beta)\rho_D^2 \wedge R^2 a_D^{-2}$, where $c(\alpha,\beta)$ is introduced in Proposition 1. Let D be fixed. Then for all $\theta \in S_D$ such that $\|\theta\|^2 = r_D^2$

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

Hence

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

Since $r_D \leq c(\alpha,\beta)\rho_D$, we get from Proposition 1

$$\inf_{\Phi_\alpha} \sup_{\theta\in\mathcal{E}_{a,2}(R),\|\theta\|\geq r_D} P_\theta(\Phi_\alpha = 0) \geq \inf_{\Phi_\alpha} \sup_{\theta\in S_D,\|\theta\|=r_D} P_\theta(\Phi_\alpha = 0) \geq \beta, \qquad (6.9)$$

where the infinimum is taken over all possible level-α testing procedures. Since inequality (6.9) holds for all $D \in J$, we obtain $\rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) \geq \sup_{D\in J}(c^2(\alpha,\beta)\rho_D^2 \wedge R^2 a_D^{-2})$. Concerning the upper bound, we know from Proposition 2 that the test $\Phi_{D,\alpha}$ is powerful as soon as:

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

where $C(\alpha, \beta)$ is defined by (3.2). Since $\theta \in \mathcal{E}_{a,2}(R)$, we get

$$\sum_{k>D} \theta_k^2 \leq R^2 a_D^{-2} \text{ and } \sup_{\theta\in\mathcal{E}_{a,2}(R), \|\theta\|^2 \geq C^2(\alpha,\beta)\rho_D^2 + R^2 a_D^{-2}} P_\theta(\Phi_\alpha = 0) < \beta.$$

This concludes the proof since the previous result holds for all $D \in J$.

6.3.2. Asymptotic minimax rates of testing on ellipsoids

First case: $a_k \sim k^s$ and $b_k \sim k^{-t}$. Choosing

$$\bar{D} = \left\lfloor \sigma^{\frac{2}{4s+4t+1}} \right\rfloor,$$

we can remark that $\rho_{\bar{D}}^2$ and $R^2 a_{\bar{D}}^{-2}$ are of the same order, hence leading to the desired rate.

Second case: $a_k \sim e^{\nu k^s}$ and $b_k \sim k^{-t}$. Set

$$D_0 = \left\lceil \left(\frac{1}{2\nu} \log(\sigma^{-2}) \right)^{1/s} \right\rceil.$$

Then

$$\begin{aligned} \rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) &\leq C\rho_{D_0}^2 + R^2 a_{D_0}^{-2}, \\ &\leq C\sigma^2 \left(\log(\sigma^{-2})\right)^{(2t+1/2)/s} + \sigma^2, \\ &\leq (C+1)\sigma^2 \left(\log(\sigma^{-2})\right)^{(2t+1/2)/s}, \end{aligned}$$

where C denotes a constant independent of σ. Concerning the lower bound, we set

$$D_1 = \left\lfloor \left(\frac{1}{4\nu} \log(\sigma^{-2}) \right)^{1/s} \right\rfloor.$$

Then

$$\begin{aligned} \rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) &\geq \rho_{D_1}^2 \wedge R^2 a_{D_1}^{-2}, \\ &\geq C\sigma^2 \left(\log(\sigma^{-2})\right)^{(2t+1/2)/s} \wedge \sigma = C\sigma^2 \left(\log(\sigma^{-2})\right)^{(2t+1/2)/s}, \end{aligned}$$

for some $C > 0$.

Third case: $a_k \sim k^s$ and $b_k \sim e^{-\gamma k^r}$. Set

$$D_0 = \left\lfloor \left(\frac{1}{4\gamma} \log \sigma^{-2}) \right)^{1/r} \right\rfloor .$$

Then

$$\begin{aligned} \rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) &\leq \rho^2_{D_0} + R^2 a_{D_0}^{-2}, \\ &\leq \sqrt{D_0}\sigma^2 b_{D_0}^{-2} + R^2 a_{D_0}^{-2}, \\ &\leq \sigma + C \left(\log(\sigma^{-2})\right)^{-2s/r} \leq (C+1) \left(\log(\sigma^{-2})\right)^{-2s/r}, \end{aligned}$$

for some $C > 0$. Concerning the lower bound, we set

$$D_1 = \left\lceil \left(\frac{1}{2\gamma} \log(\sigma^{-2}) \right)^{1/r} \right\rceil .$$

Then

$$\begin{aligned} \rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) &\geq \rho^2_{D_1} \wedge R^2 a_{D_1}^{-2}, \\ &\geq \sigma^2 b_{D_1}^{-2} \wedge R^2 a_{D_1}^{-2}, \\ &\geq 1 \wedge C \left(\log(\sigma^{-2})\right)^{-2s/r} = C \left(\log(\sigma^{-2})\right)^{-2s/r}, \end{aligned}$$

for some $C > 0$.

Fourth case: $a_k \sim e^{\nu k s}$ and $b_k \sim e^{-\gamma k^r}$. Denote by $\tilde{D}$ the solution of the equation

$$\rho^2_D = R^2 a_D^{-2}.$$

Remark that

$$\rho^2_{D_0} \leq R^2 a_{D_0}^{-2} \text{ where } D_0 = \lfloor \tilde{D} \rfloor,$$

since $(\rho^2_D)_{D \in \mathbb{N}^\star}$ and $(a^2_D)_{D \in \mathbb{N}^\star}$ are monotone increasing. Hence

$$\rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) \leq C\sigma^2 e^{2\gamma D_0^r} + R^2 e^{-2\nu D_0^s} \leq C' e^{-2\nu D_0^s}.$$

Then

$$\rho^2_{D_1} \geq R^2 a_{D_1}^{-2} \text{ where } D_1 = \lceil \tilde{D} \rceil.$$

We get

$$\rho^2(\mathcal{E}_{a,2}(R), \alpha, \beta) \geq C \left(\rho^2_{D_1} \wedge R^2 a_{D_1}^{-2}\right) \geq C' R^2 a_{D_1}^{-2} \geq C' R^2 e^{-2\nu D_1^s}.$$

In order to conclude the proof, we have to prove that the lower and upper bounds coincide. To this end, remark that $D_1 = D_0 + 1$. Thus

$$e^{-2\nu D_1^s} = e^{-2\nu (D_0+1)^s} = e^{-2\nu D_0^s} \times e^{2\nu \{D_0^s - (D_0+1)^s\}} \geq e^{-2\nu D_0^s} e^{-2\nu}$$

as soon as $s \leq 1$.

6.3.3. Proof of Theorem 3

The proof of this theorem will require the following proposition

Proposition 5. *Let $(Y_j)_{j\in J}$ obey to Model (1.1). Let $\alpha \in]0,1[$, $\beta \in]0, 1-\alpha[$ such that $\alpha+\beta \leq 0.59$ and denote by $\rho(\mathcal{E}_{a,p}(R), \alpha, \beta)$ 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*

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

where $\Sigma^2_{l,\lceil \sqrt{D} \rceil}$ is given in (2.2). Then

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

where

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

and

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

Proof. The proof will use the one of Theorem 1. We assume that $(\sigma_j)_{j\in J}$ is non-decreasing. 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 $\theta \in \mathcal{S}_{[\sqrt{D}],D}$ such that $\|\theta\|^2 \leq \sqrt{D}^{1-2/p} R^2 a_D^{-2}$, we have $\theta \in \mathcal{E}_{a,p}(R)$. Indeed, using Hölder's inequality

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

We set $k = \lceil \sqrt{D} \rceil$, $n = D$ and for all $l \in \{0, 1, \ldots, n-k\}$, we define $\theta = (\theta_j, j \in J)$ by (6.2). As pointed out in the proof of Theorem 1, $\theta \in \mathcal{S}_{k,n}$ and $\|\theta\|^2 \geq \rho^2$. We also have $\|\theta\|^2 \leq \rho^2 \Sigma^2_{n-k,k} / \Sigma^2_{l,k}$. This implies that if

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

then $\theta \in \mathcal{E}_{a,p}(R)$.
Moreover, in the proof of Theorem 1, we proved that if

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

then

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

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

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

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

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

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

Then, we complete the proof of Theorem 3. We now assume that $b_j = j^{-\gamma}$ which leads to $\sigma_j = \sigma j^\gamma$ for some $\gamma \geq 0$. Using the inequalities

$$\Sigma^2_{l,[\sqrt{D}]} \geq \sigma^2 l^{2\gamma} [\sqrt{D}], \quad \Sigma^2_{D-[\sqrt{D}],[\sqrt{D}]} \leq \sigma^2 D^{2\gamma+1/2},$$

one derives from Theorem 5 that

$$\rho_1(D) \geq \max_{0 \leq l \leq D-[\sqrt{D}]} \left(\sqrt{D}^{1-2/p} R^2 a_D^{-2} \frac{l^{2\gamma}}{D^{2\gamma}} \wedge \sigma^2 l^{2\gamma} [\sqrt{D}] \ln\left(1 + \sqrt{1 - \frac{l}{D}}\right)\right).$$

Taking $l = [D/2] + 1$ in the above inequality leads to the result stated in Corollary 3 for the polynomial case.
Let us now assume that $\sigma_j = \sigma \exp(j^\gamma)$. It is obvious that

$$\rho_2(D) \geq \sqrt{D}^{1-2/p} R^2 a_D^{-2} \wedge \sigma^2 \exp(2\gamma D),$$

which leads to the desired result.

6.3.4. *Proof of Proposition 4*

For the sake of convenience, we only prove this proposition in the particular case where $b_k = k^{-\gamma}$ for all $k \in \mathbb{N}^\star$. The proof for severly ill-posed inverse problems follows essentially the same algebra.

It follows from Bonferonis's inequality that $\Phi^\dagger_\alpha$ is a level-α test. Then introduce

$$A = \left\{ D \in J, R^2 a_D^{-2} \sqrt{D}^{1-p/2} \leq \sigma^2 D^{1/2+2\gamma} \lambda_D^2 \right\},$$

where
$$\lambda_D^2 = \ln\left(1+\sqrt{\left(\frac{1}{2}-\frac{1}{D}\right)\vee 0}\right).$$

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 sequence θ satisfying
$$\sum_{j\in J}\theta_j^2 = \|\theta\|^2 \geq C^2(\alpha,\beta)\rho_N^2,$$
where $C(\alpha,\beta)$ has been introduced in (3.2). Since A is empty, we get
$$\begin{aligned} C^2(\alpha,\beta)\rho_N^2 &\leq C^2(\alpha,\beta)\sigma^2\sqrt{N}N^{2\gamma}, \\ &\leq C^2(\alpha,\beta)\sigma^2 N^{1/2+2\gamma}\lambda_N^2 \ln^{-1}\left(1+\sqrt{\frac{1}{2}-\frac{1}{N}}\right), \\ &\leq 2^{2\gamma+1}C^2(\alpha,\beta)\ln^{-1}\left(1+\sqrt{\frac{1}{2}-\frac{1}{N}}\right)\rho^2(\mathcal{E}_{a,p}(R),\alpha,\beta). \end{aligned} \tag{6.10}$$

From now on, we assume that the set A is not empty and that $D^\dagger < N$. Let
$$\mu_N^2 = 2(\sqrt{5}+4)\ln\left(\frac{2(N-D^\dagger)}{\alpha\beta}\right).$$

Two different situations may occur:

1/ There exists at least $j \in \{D^\dagger,\ldots,N\}$ such that $b_j^2\theta_j^2 \geq \sigma^2\mu_N^2$, i.e. there exist significant coefficients after the rank $D^\dagger$.

2/ For all $j > D^\dagger$, $b_j^2\theta_j^2 \leq \sigma^2\mu_N^2$ i.e. all the coefficients θ_k have poor importance after the rank $D^\dagger$.

First consider the case 1/. Recall that in this case, the set A is not empty and there exists $j' \in \{D^\dagger+1,\ldots,N\}$ such that $b_{j'}^2\theta_{j'}^2 > \sigma^2\mu_N^2$. In this particular setting, we have to use the threshold test in order to detect these coefficients. More precisely,
$$P_\theta(\Phi_\alpha^\dagger = 0) \leq P_\theta\left(\sup_{j>D^\dagger}\Phi_{\{j\},\alpha/2(N-D^\dagger)} = 0\right) \leq P_\theta\left(\Phi_{\{j'\},\alpha/2(N-D^\dagger)} = 0\right).$$

Thanks to inequality (29) of [2], we know that this probability is smaller than β as soon as:
$$\theta_{j'}^2 > \sigma^2 b_{j'}^{-2}\ln\left(\frac{2(N-D^\dagger)}{\alpha\beta}\right)2(\sqrt{5}+4).$$

This inequality is implied by the assuption made in the case 1/.

Now, we consider point 2/. Let $j > D^\dagger$,

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

Then, we get

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

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

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

In order to conclude the proof, we have to bound the right hand side of the above inequality. Remark that

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

Thus

$$\begin{aligned} \sum_{j>D^\dagger} \theta_j^2 &\leq \sigma^{2-p} \sigma^p (D^\dagger)^{\gamma p+1/2} (D^\dagger)^{2\gamma-\gamma p} \mu_N^{2-p} \lambda_{D^\dagger}^p \\ &\leq \sigma^2 (D^\dagger)^{2\gamma+1/2} \mu_N^{2-p} \lambda_{D^\dagger}^p. \end{aligned}$$

Hence, we deduce from Proposition 2 that

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

for all sequence θ satisfying $\sum_{j=1}^{D^\dagger} \theta_j^2 \geq C^2(\alpha,\beta)\sigma^2 (D^\dagger)^{2\gamma+1/2}$, which is equivalent to

$$\|\theta\|^2 \geq C^2(\alpha,\beta)\sigma^2 (D^\dagger)^{2\gamma+1/2} + \sum_{j>D^\dagger} \theta_j^2. \tag{6.11}$$

In order to conclude, just remark that

$$\begin{aligned} & C^2(\alpha,\beta)\sigma^2 (D^\dagger)^{2\gamma+1/2} + \sum_{j>D^\dagger} \theta_j^2 \\ \leq\ & C^2(\alpha,\beta)\sigma^2 (D^\dagger)^{2\gamma+1/2} + \sigma^2 (D^\dagger)^{2\gamma+1/2} \mu_N^{2-p} \lambda_{D^\dagger}^p, \\ \leq\ & 2\sigma^2 (D^\dagger)^{2\gamma+1/2} (\mu_N^{2-p} \vee 1) \left(C^2(\alpha,\beta) + \lambda_N^2 \right), \\ \leq\ & C_{1,p} \rho^2(\mathcal{E}_{a,p}(R),\alpha,\beta)(\ln(N))^{1-p/2}, \end{aligned} \tag{6.12}$$

where

$$C_{1,p} = 2^{2\gamma+1} \ln^{-1}\left(1+\frac{1}{\sqrt{6}}\right)\left(C^2(\alpha,\beta)+\ln\left(1+\frac{1}{\sqrt{2}}\right)\right)\left(4(\sqrt{5}+4)\ln(2/\alpha\beta)\right)^{1-p/2}.$$

The result follows from (6.11) and (6.12). Note that when $D^\dagger = N$, $\sum_{j>D^\dagger}\theta_j^2 = 0$ and the above result also holds in this case.

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 $t < \min_{1\le j\le D} 1/(2\sigma_j^2)$,

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

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 $x < 1/2$:

$$x\left[\frac{1}{2}\log(1-2x) + x + \frac{x^2}{1-2x}\right] \ge 0. \tag{7.1}$$

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

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

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*

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

where a and b are positive constants. Then

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

Hence, inequality (6.5) is proved. Let us now prove inequality (6.6).
For all $z \in \mathbb{R}$,

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

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 (7.1) 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 $t < 0$, $z \in \mathbb{R}$,

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

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] Bissantz, N. and Claeskens, N. and Holzmann, H. and Munk, A. (2008) *Testing for lack of fit in inverse regression, with applications to biophotonic imaging*, preprint.

[6] Butucea, C. (2007) *Goodness-of-fit testing and quadratic functional estimation from indirect observations*, Ann. Statist., **35**, no. 5, 1907-1930.

[7] Cavalier, L. (2008) *Nonparametric statistical inverse problems*, Inverse Problems, 24(3).

[8] Donoho, D. (1995), *Nonlinear solution of linear inverse problems by Wavelet-Vaguelette decomposition*, Applied and computational harmonic analysis, **2**, 101-126.

[9] Fromont, M., Laurent, B. and Reynaud-Bouret, P. (2009) *Adaptive test of homogeneity for a Poisson process* . To appear in Ann. Inst. H. Poincar Probab. Statist.

[10] Ermakov, M. S. (2006), *Minimax detection of a signal in the heteroscedastic Gaussian white noise*, J. Math. Sci., **137**, No. 1, 4516-4524.

[11] Ingster, Yu.I. (1993) *Asymptotically minimax testing for nonparametric alternatives I-II-III*, Math. Methods Statist., **2**, 85-114, 171-189, 249-268.

[12] Ingster, Yu.I., Sapatinas, T. and Suslina, I.A. (2010) *Minimax signal detection in ill-posed inverse problems*. Working paper.

[13] Ingster, Yu.I. and Suslina, I.A. (1998) *Minimax detection of a signal for Besov bodies and balls*, Problems Inform. Transmission, **34**, 48-59.

[14] Ingster, Yu.I. and Suslina, I.A. (2002) *Nonparametric goodness-of-fit testing under Gaussian models*, Lecture Notes in Statistics, 169. Springer-Verlag, New York.

[15] Laurent, B. and Massart, P. (2000) *Adaptive estimation of a quadratic functional by model selection*, Ann. Statist., **28**, no. 5, 1302-1338.

[16] Laurent, B. and Loubes, J-M. and Marteau, C. (2011) *Testing inverse problems: a direct or an indirect problem*, Journal of Statistical Planning and Inference, Volume 141 (5), 1849-1861

[17] Lepski, O. V. and Spokoiny, V. G. (1999) *Minimax nonparametric hypothesis testing: the case of inhomogeneous alternative*, Bernoulli, **5**, 333-358.

[18] J-M. Loubes and C. Ludena. (2008) *Adaptive complexity regularization for inverse problems*, Electronic Journal Of Statistics, 2:661–677.

[19] J-M. Loubes and C. Ludena. (2010) *Penalized estimators for non linear inverse problems*, ESAIM Probab. Stat., Volume 14:173–191.

[20] O'Sullivan, F. (1986) *A statistical perspective on ill-posed inverse problems*, Statist. Sci., 1(**4**), 502–527.

[21] Spokoiny, V.G. (1996) *Adaptive hypothesis testing using wavelets*, Ann. Statist., 24, 2477-2498.