The Annals of Statistics
2010, Vol. 38, No. 2, 704–752
DOI: 10.1214/08-AOS629

GOODNESS-OF-FIT TESTS FOR HIGH-DIMENSIONAL GAUSSIAN LINEAR MODELS

BY NICOLAS VERZELEN AND FANNY VILLERS

Université Paris-Sud and INRIA Saclay and INRA, Laboratoire de biométrie

Let $(Y, (X_i)_{1\leq i\leq p})$ be a real zero mean Gaussian vector and V be a subset of $\{1, \dots, p\}$. Suppose we are given n i.i.d. replications of this vector. We propose a new test for testing that Y is independent of $(X_i)_{i\in\{1,\dots,p\}\setminus V}$ conditionally to $(X_i)_{i\in V}$ against the general alternative that it is not. This procedure does not depend on any prior information on the covariance of X or the variance of Y and applies in a high-dimensional setting. It straightforwardly extends to test the neighborhood of a Gaussian graphical model. The procedure is based on a model of Gaussian regression with random Gaussian covariates. We give nonasymptotic properties of the test and we prove that it is rate optimal [up to a possible $\log(n)$ factor] over various classes of alternatives under some additional assumptions. Moreover, it allows us to derive nonasymptotic minimax rates of testing in this random design setting. Finally, we carry out a simulation study in order to evaluate the performance of our procedure.

1. Introduction.

We consider the following regression model:

$$Y = \sum_{i=1}^{p} \theta_i X_i + \epsilon, \tag{1.1}$$

where θ is an unknown vector of $\mathbb{R}^p$. In the sequel, we note $\mathcal{I} := \{1, \dots, p\}$. The vector $X := (X_i)_{1\leq i\leq p}$ follows a real zero mean Gaussian distribution with nonsingular covariance matrix Σ, and ϵ is a real zero mean Gaussian random variable independent of X. Straightforwardly, the variance of ϵ corresponds to the conditional variance of Y given X, $\mathrm{var}(Y|X)$.

The variable selection problem for this model in a high-dimensional setting has recently attracted a lot of attention. A large number of papers are now devoted to the design of new algorithms and estimators which are computationally feasible and are proven to converge (see, for instance, the works of Meinshausen and Bühlmann [19], Candès and Tao [5], Zhao and Yu [29], Zou and Hastie [30], Bühlmann and Kalisch [4] or Zhao and Huang [28]). A common drawback of the previously mentioned estimation procedures is that they require restrictive conditions on the covariance matrix Σ in order to behave well. Our issue is the natural testing counterpart of this variable selection problem; we aim at defining a

Received December 2007; revised May 2008.

AMS 2000 subject classifications. Primary 62J05; secondary 62G10, 62H20.

Key words and phrases. Gaussian graphical models, linear regression, multiple testing, ellipsoid, adaptive testing, minimax hypothesis testing, minimax separation rate, goodness-of-fit.

computationally feasible testing procedure that achieves an optimal rate for any covariance matrix Σ.

1.1. *Presentation of the main results.* We are given n i.i.d. replications of the vector (Y, X). Let us respectively note $\mathbf{Y}$ and $\mathbf{X}_i$. The vectors of the n observations of Y and X_i, for any $i \in \mathcal{I}$. Let V be a subset of $\mathcal{I}$, then X_V refers to the set $\{X_i, i \in V\}$ and θ_V stands for the sequence $(\theta_i)_{i\in V}$. We first propose a collection of testing procedures T_α of the null hypothesis "$\theta_{\mathcal{I}\setminus V} = 0$" against the general alternative "$\theta_{\mathcal{I}\setminus V} \neq 0$." These procedures are based on the ideas of Baraud et al. [3] in a random design. Their definitions are very flexible as they require no prior knowledge of the covariance of X, the variance of ϵ nor the variance of Y. Note that the property "$\theta_{\mathcal{I}\setminus V} = 0$" is equivalent to "$Y$ is independent of $X_{\mathcal{I}\setminus V}$ conditionally to X_V." Hence, it also permits to test conditional independences and applies for testing the graph of Gaussian graphical model (see below). Contrary to most approaches in this setting (e.g., Drton and Pearlman [8]), we are able to consider the difficult case of tests in a high-dimensional setting: the number of covariates p is possibly much larger than the number of observations n. Such situations arise in many statistical applications like in genomics or biomedical imaging. To our knowledge, the only testing procedures (e.g., [21]) that could handle high-dimensional alternatives lack theoretical justifications. In this paper, we exhibit some tests T_α that are both computationally amenable and optimal in the minimax sense.

From a theoretical perspective, we are able to control the Family Wise Error Rate (FWER) of our testing procedures T_α. Moreover, we derive a general nonasymptotic upper bound for their power. Contrary to the various rates of convergence obtained in the estimation setting (e.g., [5] or [19]), our upper bound holds for any covariance matrix Σ. Then we derive from it nonasymptotic minimax rates of testing in the Gaussian random design framework. If the minimax rates are known for a long time in the fixed design Gaussian regression framework (e.g., [2]), they were unknown in our setting. For instance, if at most k components of θ are nonzero and if k is much smaller than p, we prove that the minimax rates of testing is of order $\frac{k\log(p)}{n}$ when the covariates X_i are independent. If the covariates are dependent, we derive faster minimax rates. To our knowledge, these are the first results for testing or estimation issues that illustrate minimax rates for dependent covariates. Afterward, we show analogous results when k is large or when the vector θ belongs to some ellipsoid or some collection of ellipsoids. For any of these alternatives, we exhibit some procedure T_α that achieves the optimal rate [at a possible $\log(n)$ factor]. Finally, we illustrate the performance of the procedure on simulated examples.

1.2. *Application to Gaussian Graphical Models* (*GGM*). Our work was originally motivated by the following question: let $(Z_j)_{j\in\mathcal{J}}$ be a random vector which follows a zero mean Gaussian distribution whose covariance matrix Σ' is nonsingular. We observe n i.i.d. replications of this vector Z and we are given a graph $\mathcal{G} = (\Gamma, E)$ where $\Gamma = \{1, \dots, |\mathcal{J}|\}$ and E is a set of edges in $\Gamma \times \Gamma$. How can we test that Z is an undirected Gaussian graphical model (GGM) with respect to the graph $\mathcal{G}$?

The random vector Z is a GGM with respect to the graph $\mathcal{G} = (\Gamma, E)$ if for any couple (i, j) which is not contained in the edge set E, Z_i and Z_j are independent, given the remaining variables. See Lauritzen [17] for definitions and main properties of GGM. Interest in these models has grown as they allow the description of dependence structure in high-dimensional data. As such, they are widely used in spatial statistics [7, 20] or probabilistic expert systems [6]. More recently, they have been applied to the analysis of microarray data. The challenge is to infer the network regulating the expression of the genes using only a small sample of data (see, for instance, Schäfer and Strimmer [21], Kishino and Waddell [15] or Wille et al. [26]). This issue has motivated the research for new estimation procedures to handle GGM in a high-dimensional setting.

It is beyond the scope of this paper to give an exhaustive review of these. Many of these graph estimation methods are based on multiple testing procedures (see, for instance, Schäfer and Strimmer [21] or Wille and Bühlmann [25]). Other methods are based on variable selection for high-dimensional data we previously mentioned. For instance, Meinshausen and Bühlmann [19] proposed a computationally feasible model selection algorithm using Lasso penalization. Huang et al. [11] and Yuan and Lin [27] extend this method to infer directly the inverse covariance matrix Σ'^{-1} by minimizing the log-likelihood penalized by the l^1 norm.

While the issue of graph and covariance estimation is extensively studied, few theoretical results are proved for the problem of hypothesis testing of GGM in a high-dimensional setting. We believe that this issue is significant for two reasons: first, when considering a gene regulation network, the biologists often have a previous knowledge of the graph and may want to test if the microarray data match with their model. Second, when applying an estimation method in a high-dimensional setting, it could be useful to test the estimated graph as some of these methods are too conservative.

Admittedly, some of the previously mentioned estimation methods are based on multiple testing. However, as they are constructed for an estimation purpose, most of them do not take into account some previous knowledge about the graph. This is, for instance, the case for the approaches of Drton and Perlman [8] and Schäfer and Strimmer [21]. Some of the other existing procedures cannot be applied in a high-dimensional setting ($|\mathcal{J}| \geq n$). Finally, most of them lack theoretical justification in a nonasymptotic way.

In a subsequent paper [23] we define a test of graph based on the present work. It benefits the ability of handling high-dimensional GGM and has minimax properties. Moreover, we show numerical evidence of its efficiency (see [23] for more details). In this article, we shall only present the idea underlying our approach.

For any $j \in \mathcal{J}$, we note $N(j)$ the set of neighbors of j in the graph $\mathcal{G}$. Testing that Z is a GGM with respect to $\mathcal{G}$ is equivalent to testing that the random variable Z_j conditionally to $(Z_l)_{l \in N(j)}$ is independent of $(Z_l)_{l \in \mathcal{J} \setminus (N(j) \cup \{j\})}$ for any $j \in \mathcal{J}$. As Z follows a Gaussian distribution, the distribution of Z_j conditionally to the other variables decomposes as follows:

$$Z_j = \sum_{k \in \mathcal{J} \setminus \{j\}} \theta_k Z_k + \epsilon_j,$$

where ϵ_j is normal and independent of $(Z_k)_{k \in \mathcal{J} \setminus \{j\}}$. Then, the statement of conditional independency is equivalent to $\theta_{\mathcal{J} \setminus \{j\} \cup N(j)} = 0$. This approach based on conditional regression is also used for estimation by Meinshausen and Bühlmann [19].

1.3. *Organization of the paper.* In Section 2, we present the approach of our procedure and connect it with the fixed design framework. Moreover, we define the notion of minimax rates of testing in this setting and gather the main notation. We define the testing procedures T_α in Section 3, and we nonasymptotically characterise the set of vectors θ over which the test T_α is powerful. In Sections 4 and 5, we apply our procedure to define tests and study their optimality for two different classes of alternatives. More precisely, in Section 4 we test $\theta = 0$ against the class of θ whose components equal 0, except at most k of them (k is supposed small). We define a test which under mild conditions achieves the minimax rate of testing. When the covariates are independent, it is interesting to note that the minimax rates exhibit the same ranges in our statistical model (1.1) and in our fixed design regression model (2.1). In Section 5, we define two procedures that achieve the simultaneous minimax rates of testing over large classes of ellipsoids [to sometimes the price of a $\log(p)$ factor]. Moreover, we show that the problem of adaptation over classes of ellipsoids is impossible without a loss in efficiency. This was previously pointed out in [22] in fixed design regression framework. The simulation studies are presented in Section 6. Finally, Sections 7, 8 and the Appendix contain the proofs.

2. Description of the approach.

2.1. *Connection with tests in fixed design regression.* Our work is directly inspired by the testing procedure of Baraud et al. [3] in fixed design regression framework. Contrary to model (1.1), the problem of hypothesis testing in fixed design regression has been extensively studied. This is why we will use the results in this framework as a benchmark for the theoretical bounds in our model (1.1). Let us define this second regression model:

$$Y_i = f_i + \sigma \epsilon_i, \qquad i \in \{1, \dots, N\}, \tag{2.1}$$

where f is an unknown vector of $\mathbb{R}^N$, σ some unknown positive number and the ϵ_i's a sequence of i.i.d. standard Gaussian random variables. The problem at hand is testing that f belongs to a linear subspace of $\mathbb{R}^N$ against the alternative that it does not. We refer to [3] for a short review of nonparametric tests in this framework. Moreover, we are interested in the performance of the procedures from a minimax perspective. To our knowledge, there have been no results in model (1.1). However, there are numerous papers on this issue in the fixed design regression model. First, we refer to the seminal work of Ingster [12–14] who gives asymptotic minimax rates over nonparametric alternatives. Our work is closely related to the results of Baraud [2] where he gives nonasymptotic minimax rates of testing over ellipsoids or sparse signals. Throughout the paper, we highlight the link between the minimax rates in fixed and in random design.

2.2. *Principle of our testing procedure.* Let us briefly describe the idea underlying our testing procedure. A formal definition will follow in Section 3.1. Let m be a subset of $\mathcal{I} \setminus V$. We respectively define S_V and $S_{V\cup m}$ as the linear subspaces of $\mathbb{R}^p$ such that $\theta_{\mathcal{I}\setminus V} = 0$, respectively $\theta_{\mathcal{I}\setminus(V\cup m)} = 0$. We note d and D_m for the cardinalities of V and m, and N_m refers to $N_m = n - d - D_m$. If $N_m > 0$, we define the Fisher statistic ϕ_m by

$$\phi_m(\mathbf{Y},\mathbf{X}) := \frac{N_m \|\Pi_{V\cup m}\mathbf{Y} - \Pi_V \mathbf{Y}\|_n^2}{D_m \|\mathbf{Y} - \Pi_{V\cup m}\mathbf{Y}\|_n^2}, \tag{2.2}$$

where Π_V refers to the orthogonal projection onto the space generated by the vectors $(\mathbf{X}_i)_{i\in V}$ and $\|\cdot\|_n$ is the canonical norm in $\mathbb{R}^n$. We define the test statistic $\phi_{m,\alpha}(\mathbf{Y},\mathbf{X})$ as

$$\phi_{m,\alpha}(\mathbf{Y},\mathbf{X}) = \phi_m(\mathbf{Y},\mathbf{X}) - \bar{F}^{-1}_{D_m,N_m}(\alpha), \tag{2.3}$$

where $\bar{F}_{D_m,N_m}(u)$ denotes the probability for a Fisher variable with D and N degrees of freedom to be larger than u. Let us consider a finite collection $\mathcal{M}$ of nonempty subsets of $\mathcal{I} \setminus V$ such that for each $m \in \mathcal{M}$, $N_m > 0$. Our testing procedure consists of doing a Fisher test for each $m \in \mathcal{M}$. We define $\{\alpha_m, m \in \mathcal{M}\}$ a suitable collection of numbers in $]0, 1[$ (which possibly depends on $\mathbf{X}$). For each $m \in \mathcal{M}$, we do the Fisher test ϕ_m of level α_m of

$$H_0\colon \theta \in S_V \quad \text{against the alternative} \quad H_{1,m}\colon \theta \in S_{V\cup m} \setminus S_V,$$

and we decide to reject the null hypothesis if one of those Fisher tests does.

The main advantage of our procedure is that it is very flexible in the choices of the model $m \in \mathcal{M}$ and in the choices of the weights $\{\alpha_m\}$. Consequently, if we choose a suitable collection $\mathcal{M}$, the test is powerful over a large class of alternatives as shown in Sections 3.3, 4 and 5.

Finally, let us mention that our procedure easily extends to the case where the expectation of the random vector (Y, X) is unknown. Let $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ denote the projections of $\mathbf{X}$ and $\mathbf{Y}$ onto the unit vector $\mathbf{1}$. Then one only has to apply the procedure to $(\mathbf{Y} - \overline{\mathbf{Y}}, \mathbf{X} - \overline{\mathbf{X}})$ and to replace d by $d + 1$. The properties of the test remain unchanged and one can adapt all the proofs to the price of more technicalities.

2.3. *Minimax rates of testing.* In order to examine the quality of our tests, we will compare their performance with the minimax rates of testing. That is why we now define precisely what we mean by the (α, δ)-minimax rate of testing over a set Θ. We endow $\mathbb{R}^p$ with the Euclidean norm,

$$\|\theta\|^2 := \theta^t \Sigma \theta = \operatorname{var}\left(\sum_{i=1}^{p} \theta_i X_i\right). \tag{2.4}$$

As ϵ and X are independent, we derive from the definition of $\|\cdot\|^2$ that $\operatorname{var}(Y) = \|\theta\|^2 + \operatorname{var}(Y|X)$. Let us remark that $\operatorname{var}(Y|X)$ does not depend on X. If $\|\theta\|$ varies, either the quantity $\operatorname{var}(Y)$ or $\operatorname{var}(Y|X)$ has to vary. In the sequel, we suppose that $\operatorname{var}(Y)$ is fixed. We briefly justify this choice in Section 4.2. Consequently, if $\|\theta\|^2$ is increasing, then $\operatorname{var}(Y|X)$ has to decrease so that the sum remains constant. Let α be a number in $]0; 1[$ and let δ be a number in $]0; 1-\alpha[$ (typically small). For a given vector θ, matrix Σ and $\operatorname{var}(Y)$, we denote by $\mathbb{P}_\theta$ the joint distribution of $(\mathbf{Y}, \mathbf{X})$. For the sake of simplicity, we do not emphasize the dependence of $\mathbb{P}_\theta$ on $\operatorname{var}(Y)$ or Σ. Let ψ_α be a test of level α of the hypothesis "$\theta = 0$" against the hypothesis "$\theta \in \Theta \setminus 0$." In our framework, it is natural to measure the performance of ψ_α using the quantity $\rho(\psi_\alpha, \Theta, \delta, \operatorname{var}(Y), \Sigma)$ defined by

$$\rho(\psi_\alpha, \Theta, \delta, \operatorname{var}(Y), \Sigma) := \inf\left\{\rho > 0, \inf\left\{\mathbb{P}_\theta(\psi_\alpha = 1), \theta \in \Theta \text{ and } \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq \rho^2\right\} \geq 1 - \delta\right\},$$

where the quantity

$$r_{\mathrm{s/n}}(\theta) := \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \tag{2.5}$$

appears naturally as it corresponds to the ratio $\|\theta\|^2 / \operatorname{var}(Y|X)$ which is the quantity of information brought by X (i.e., the signal) over the conditional variance of Y (i.e., the noise). We aim at describing the quantity

$$\inf_{\psi_\alpha} \rho(\psi_\alpha, \Theta, \delta, \operatorname{var}(Y), \Sigma) := \rho(\Theta, \alpha, \delta, \operatorname{var}(Y), \Sigma), \tag{2.6}$$

where the infimum is taken over all the level-α tests ψ_α. We call this quantity the (α, δ)-minimax rate of testing over Θ.

A dual notion of this ρ function is the function β_Σ. For any $\Theta \subset \mathbb{R}^p$ and $\alpha \in]0, 1[$, we denote by $\beta_\Sigma(\Theta)$ the quantity

$$\beta_\Sigma(\Theta) := \inf_{\psi_\alpha} \sup_{\theta \in \Theta} \mathbb{P}_\theta[\psi_\alpha = 0],$$

where the infimum is taken over all level-α tests ψ_α and where we recall that Σ refers to the covariance matrix of X.

2.4. *Notation.* We recall the main notation that we shall use throughout the paper. In the sequel, n stands for the number of independent observations, and p is the number of covariates. Moreover, X_V stands for the collection $(X_i)_{i\in V}$ of the covariates that correspond to the null hypothesis, and d is the cardinality of the set V. The models m are subsets of $\mathcal{I} \subset V$, and we note D_m, their cardinality. T_α stands for our testing procedure of level α. The statistics ϕ_m and the test $\phi_{m,\alpha}$ are respectively defined in (2.2) and (2.3). Finally, the norm $\|\cdot\|$ is introduced in (2.4).

For $x, y \in \mathbb{R}$, we set

$$x \wedge y := \inf\{x, y\}, \qquad x \vee y := \sup\{x, y\}.$$

For any $u \in \mathbb{R}$, $\bar{F}_{D,N}(u)$ denotes the probability for a Fisher variable with D and N degrees of freedom to be larger than u. In the sequel, $L, L_1, L_2, \dots$ denote constants that may vary from line to line. The notation $L(\cdot)$ specifies the dependency on some quantities. For the sake of simplicity, we only give the orders of magnitude in the results and we refer to the proofs for explicit constants.

3. The testing procedure.

3.1. *Description of the procedure.* Let us first fix some level $\alpha \in]0, 1[$. Throughout this paper, we suppose that $n \geq d + 2$. Let us consider a finite collection $\mathcal{M}$ of nonempty subsets of $\mathcal{I} \setminus V$ such that for all $m \in \mathcal{M}$, $1 \leq D_m \leq n - d - 1$. We introduce the following test of level α. We reject H_0: "$\theta \in S_V$" when the statistic

$$T_\alpha := \sup_{m\in\mathcal{M}} \{\phi_m(\mathbf{Y}, \mathbf{X}) - \bar{F}^{-1}_{D_m,N_m}(\alpha_m(\mathbf{X}))\} \tag{3.1}$$

is positive where the collection of weights $\{\alpha_m(\mathbf{X}), m \in \mathcal{M}\}$ is chosen according to one of the two following procedures:

P_1: The α_ms do not depend on $\mathbf{X}$ and satisfy the equality

$$\sum_{m\in\mathcal{M}} \alpha_m = \alpha. \tag{3.2}$$

P_2: For all $m \in \mathcal{M}$, $\alpha_m(\mathbf{X}) = q_{\mathbf{X},\alpha}$, the α-quantile of the distribution of the random variable

$$\inf_{m\in\mathcal{M}} \bar{F}_{D_m,N_m}\left(\frac{\|\Pi_{V\cup m}(\boldsymbol{\epsilon}) - \Pi_V(\boldsymbol{\epsilon})\|_n^2/D_m}{\|\boldsymbol{\epsilon} - \Pi_{V\cup m}(\boldsymbol{\epsilon})\|_n^2/N_m}\right), \tag{3.3}$$

conditionally to $\mathbf{X}$.

Note that it is easy to compute the quantity $q_{\mathbf{X},\alpha}$. Let Z be a standard Gaussian random vector of size n independent of $\mathbf{X}$. As $\boldsymbol{\epsilon}$ is independent of $\mathbf{X}$, the distribution of (3.3) conditionally to $\mathbf{X}$ is the same as the distribution of

$$\inf_{m\in\mathcal{M}} \bar{F}_{D_m,N_m}\left(\frac{\|\Pi_{V\cup m}(Z) - \Pi_V(Z)\|^2/D_m}{\|Z - \Pi_{V\cup m}(Z)\|^2/N_m}\right)$$

conditionally to $\mathbf{X}$. Hence, we can easily work out its quantile using Monte Carlo method.

Clearly, the computational complexity of the procedure is linear with respect to the size of the collection of models $\mathcal{M}$ even when using procedure P_2. Consequently, when we apply our procedure to high-dimensional data as in Section 6 or in [23], we favor collections $\mathcal{M}$ whose size is linear with respect to the number of covariates p.

3.2. *Comparison of procedures P_1 and P_2.* We respectively refer to T_α^1 and T_α^2 for the tests (3.1) associated with procedure P_1 and P_2. First, we are able to control the behavior of the test under the null hypothesis.

PROPOSITION 3.1. *The test T_α^1 corresponds to a Bonferroni procedure and therefore satisfies*

$$\mathbb{P}_\theta(T_\alpha > 0) \leq \sum_{m\in\mathcal{M}} \alpha_m \leq \alpha,$$

whereas the test T_α^2 has the property to be exactly of the size α

$$\mathbb{P}_\theta(T_\alpha > 0) = \alpha.$$

The proof is given in the Appendix. Moreover, the test T_α^2 is more powerful than the corresponding test T_α^1 defined with weights $\alpha_m = \alpha/|\mathcal{M}|$.

PROPOSITION 3.2. *For any parameter θ that does not belong to S_V, the procedure T_α^1 with weights $\alpha_m = \alpha/|\mathcal{M}|$ and the procedure T_α^2 satisfy*

$$\mathbb{P}_\theta\big(T_\alpha^2(\mathbf{X},\mathbf{Y}) > 0|\mathbf{X}\big) \geq \mathbb{P}_\theta\big(T_\alpha^1(\mathbf{X},\mathbf{Y}) > 0|\mathbf{X}\big) \qquad \mathbf{X}\ a.s. \tag{3.4}$$

Again, the proof is given in the Appendix. On the one hand, the choice of procedure P_1 allows one to avoid the computation of the quantile $q_{\mathbf{X},\alpha}$ and possibly permits one to give a Bayesian flavor to the choice of the weights. On the other hand, procedure P_2 is more powerful than the corresponding test with procedure P_1. We will illustrate these considerations in Section 6. In Sections 3.3, 4 and 5 we study the power and rates of testing of T_α with procedure P_1.

3.3. *Power of the test.* We aim at describing a set of vectors θ in $\mathbb{R}^p$ over which the test defined in Section 3 with procedure P_1 is powerful. Since procedure P_2 is more powerful than procedure P_1 with $\alpha_m = \alpha/|\mathcal{M}|$, the test with procedure P_2 will also be powerful on this set of θ.

Let α and δ be two numbers in $]0,1[$, and let $\{\alpha_m, m\in\mathcal{M}\}$ be weights such that $\sum_{m\in\mathcal{M}}\alpha_m \leq \alpha$. We define hypothesis $(H_\mathcal{M})$ as follows:

$$(H_\mathcal{M}) \quad \text{for all } m\in\mathcal{M}, \qquad \alpha_m \geq \exp(-N_m/10) \quad \text{and} \quad \delta \geq \exp 2(-N_m/21).$$

For typical choices of the collections $\mathcal{M}$ and $\{\alpha_m, m\in\mathcal{M}\}$, these conditions are fulfilled as discussed in Sections 4 and 5. Let us now turn to the main result.

THEOREM 3.3. *Let T_α be the test procedure defined by* (3.1). *We assume that $n > d+2$ and that assumption $(H_{\mathcal{M}})$ holds. Then, $\mathbb{P}_\theta(T_\alpha > 0) \geq 1-\delta$ for all θ belonging to the set*

$$\mathcal{F}_{\mathcal{M}}(\delta) := \left\{\theta \in \mathbb{R}^p, \exists m \in \mathcal{M}\colon \frac{\operatorname{var}(Y|X_V) - \operatorname{var}(Y|X_{V\cup m})}{\operatorname{var}(Y|X_{V\cup m})} \geq \Delta(m)\right\},$$

where

$$\Delta(m) := \left(L_1\sqrt{D_m \log\left(\frac{2}{\alpha_m\delta}\right)}\left(1+\sqrt{\frac{D_m}{N_m}}\right) + L_2\left(1+2\frac{D_m}{N_m}\right)\log\left(\frac{2}{\alpha_m\delta}\right)\right)\Big/(n-d). \tag{3.5}$$

This result is similar to Theorem 1 in [3] in fixed design regression framework and the same comment also holds; the test T_α under procedure P_1 has a power comparable to the best of the tests among the family $\{\phi_{m,\alpha}, m \in \mathcal{M}\}$. Indeed, let us assume, for instance, that $V = \{0\}$ and that the α_m are chosen to be equal to $\alpha/|\mathcal{M}|$. The test T_α defined by (3.1) is equivalent to doing several tests of $\theta = 0$ against $\theta \in S_m$ at level α_m for $m \in \mathcal{M}$ and it rejects the null hypothesis if one of those tests does. From Theorem 3.3, we know that under hypothesis $(H_{\mathcal{M}})$ this test has a power greater than $1-\delta$ over the set of vectors θ belonging to $\bigcup_{m\in\mathcal{M}} \mathcal{F}'_m(\delta, \alpha_m)$ where $\mathcal{F}'_m(\delta, \alpha_m)$ is the set of vectors $\theta \in \mathbb{R}^p$ such that

$$\frac{\operatorname{var}(Y) - \operatorname{var}(Y|X_m)}{\operatorname{var}(Y|X_m)} \geq \frac{L(D_m, N_m)}{n}\left(\sqrt{D_m \log\left(\frac{2}{\alpha_m\delta}\right)} + \log\left(\frac{2}{\alpha_m\delta}\right)\right). \tag{3.6}$$

The quantity, $L(D_m, N_m)$ behaves like a constant if the ratio D_m/N_m is bounded. Let us compare this result with the set of θ over which the Fisher test $\phi_{m,\alpha}$ at level α has a power greater than $1-\delta$. Applying Theorem 3.3, we know that it contains $\mathcal{F}'_m(\delta, \alpha)$. Moreover, the following proposition shows that it is not much larger than $\mathcal{F}'_m(\delta, \alpha)$:

PROPOSITION 3.4. *Let $\delta \in]0, 1-\alpha[$. If*

$$\frac{\operatorname{var}(Y) - \operatorname{var}(Y|X_m)}{\operatorname{var}(Y|X_m)} \leq L(\alpha, \delta)\frac{\sqrt{D_m}}{n},$$

then $\mathbb{P}_\theta(\phi_{m,\alpha} > 0) \leq 1-\delta$.

The proof is postponed to Section 8 and is based on a lower bound of the minimax rate of testing.

$\mathcal{F}'_m(\delta, \alpha)$ and $\mathcal{F}'_m(\delta, \alpha_m)$ defined in (3.6) differ from the fact that $\log(1/\alpha)$ is replaced by $\log(1/\alpha_m)$. For the main applications that we will study in Sections 4–6, the ratio $\log(1/\alpha_m)/\log(1/\alpha)$ is of order $\log(n)$, $\log\log n$, or $k\log(ep/k)$ where

k is a "small" integer. Thus, for each $\delta \in]0, 1-\alpha[$, the test based on T_α has a power greater than $1-\delta$ over a class of vectors which is close to $\bigcup_{m\in\mathcal{M}} \mathcal{F}'_m(\delta,\alpha)$. It follows that for each $\theta \neq 0$ the power of this test under $\mathbb{P}_\theta$ is comparable to the best of the tests among the family $\{\phi_{m,\alpha}, m \in \mathcal{M}\}$.

In the next two sections, we use this theorem to establish rates of testing against different types of alternatives. First, we give an upper bound for the rate of testing $\theta = 0$ against a class of θ for which a lot of components are equal to 0. In Section 5, we study the rates of testing and simultaneous rates of testing $\theta = 0$ against classes of ellipsoids. For the sake of simplicity, we will only consider the case $V = \{0\}$. Nevertheless, the procedure T_α defined in (3.1) applies in the same way when one considers a more complex null hypothesis and the rates of testing are unchanged except that we have to replace n by $n-d$ and $\mathrm{var}(Y)$ by $\mathrm{var}(Y|X_V)$.

4. Detecting nonzero coordinates. Let us fix an integer k between 1 and p. In this section, we are interested in testing $\theta = 0$ against the class of θ with a most k nonzero components. This typically corresponds to the situation encountered when considering tests of neighborhoods for large sparse graphs. As the graph is assumed to be sparse, only a small number of neighbors are missing under the alternative hypothesis.

For each pair of integers (k, p) with $k \leq p$, let $\mathcal{M}(k, p)$ be the class of all subsets of $\mathcal{I} = \{1, \dots, p\}$ of cardinality k. The set $\Theta[k, p]$ stands for the subset of vectors $\theta \in \mathbb{R}^p$, such that at most k coordinates of θ are nonzero.

First, we define a test T_α of the form (3.1) with procedure P_1, and we derive an upper bound for the rate of testing of T_α against the alternative $\theta \in \Theta[k, p]$. Then we show that this procedure is rate optimal when all the covariates are independent. Finally, we study the optimality of the test when $k = 1$ for some examples of covariance matrix Σ.

4.1. *Rate of testing of T_α.*

PROPOSITION 4.1. *We consider the set of models $\mathcal{M} = \mathcal{M}(k, p)$. We use the test T_α under procedure P_1, and we take the weights α_m all equal to $\alpha/|\mathcal{M}|$. Let us suppose that n satisfies*

$$n \geq L\left[\log\left(\frac{2}{\alpha\delta}\right) + k\log\left(\frac{ep}{k}\right)\right]. \tag{4.1}$$

Let us set the quantity

$$\rho'^2_{k,p,n} := L(\alpha,\delta)\frac{k\log(ep/k)}{n}. \tag{4.2}$$

For any θ in $\Theta[k, p]$, such that $\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq \rho'^2_{k,p,n}$, $\mathbb{P}_\theta(T_\alpha > 0) \geq 1-\delta$.

We recall that the norm $\|\cdot\|$ is defined in (2.4) and equals $\mathrm{var}(Y) - \mathrm{var}(Y|X)$. This proposition easily follows from Theorem 3.3 and its proof is given in Section 7. Note that the upper bound does not directly depend on the covariance matrix of the vector X. Moreover, hypothesis (4.1) corresponds to the minimal assumption needed for consistency and type-oracle inequalities in the estimation setting as pointed out by Wainwright ([24], Theorem 2) and Giraud ([10], Section 3.1). Hence, we conjecture that hypothesis (4.1) is minimal so that Proposition 4.1 holds. We will further discuss the bound (4.2) after deriving lower bounds for the minimax rate of testing.

4.2. *Minimax lower bounds for independent covariates.* In the statistical framework considered here, the problem of giving minimax rates of testing under no prior knowledge of the covariance of X and of $\mathrm{var}(Y)$ is open. This is why we shall only derive lower bounds when $\mathrm{var}(Y)$ and the covariance matrix of X are known. In this section, we give nonasymptotic lower bounds for the (α, δ)-minimax rate of testing over the set $\Theta[k, p]$ when the covariance matrix of X is the identity matrix (except Proposition 4.2). As these bounds coincide with the upper bound obtained in Section 4.1, this will show that our test T_α is rate optimal.

We first give a lower bound for the (α, δ)-minimax rate of detection of all p nonzero coordinates for any covariance matrix Σ.

PROPOSITION 4.2. *Let us suppose that* $\mathrm{var}(Y)$ *is known. Let us set* $\rho^2_{p,n}$ *such that*

$$\rho^2_{p,n} := L(\alpha, \delta)\frac{\sqrt{p}}{n}. \tag{4.3}$$

Then for all $\rho < \rho_{p,n}$,

$$\beta_\Sigma\left(\left\{\theta \in \Theta[p, p], \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} = \rho^2\right\}\right) \geq \delta,$$

where we recall that Σ *is the covariance matrix of* X.

If $n \geq (1+\gamma)p$ for some $\gamma > 0$, Theorem 3.3 shows that the test $\phi_{\mathcal{I},\alpha}$ defined in (2.3) has power greater than δ over the vectors θ that satisfy

$$\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq L(\gamma, \alpha, \delta)\frac{\sqrt{p}}{n}.$$

Hence, $\sqrt{p}/n$ is the minimax rate of testing $\Theta[p, p]$ at least when the number of observations is larger than the number of covariates. This is coherent with the minimax rate obtained in the fixed design framework (e.g., [2]). When p becomes larger we do not think that the lower bound given in Proposition 4.2 is still sharp. Note that this minimax rate of testing holds for any covariance matrix Σ contrary to Theorem 4.3.

We now turn to the lower bound for the (α, δ)-minimax rate of testing against $\theta \in \Theta[k, p]$.

THEOREM 4.3. *Let us set* $\rho^2_{k,p,n}$ *such that*

$$\rho^2_{k,p,n} := L(\alpha,\delta)\frac{k}{n}\log\left(1+\frac{p}{k^2}+\sqrt{2\frac{p}{k^2}}\right). \tag{4.4}$$

We suppose that the covariance of X *is the identity matrix* I. *Then, for all* $\rho < \rho_{k,p,n}$,

$$\beta_I\left(\left\{\theta \in \Theta[k,p], \frac{\|\theta\|^2}{\operatorname{var}(Y)-\|\theta\|^2} = \rho^2\right\}\right) > \delta,$$

where the quantity $\operatorname{var}(Y)$ *is known.*

If $\alpha+\delta \leq 53\%$, *then one has*

$$\rho^2_{k,p,n} \geq \frac{k}{2n}\log\left(1+\frac{p}{k^2}\vee\sqrt{\frac{p}{k^2}}\right).$$

This result implies the following lower bound for the minimax rate of testing:

$$\rho(\Theta[k,p],\alpha,\delta,\operatorname{var}(Y),I) \geq \rho^2_{k,p,n}.$$

The proof is given in Section 8. To the price of more technicalities, it is possible to prove that the lower bound still holds if the variables (X_i) are independent with known variances possibly different. Theorem 4.3 recovers approximately the lower bounds for the minimax rates of testing in signal detection framework obtained by Baraud [2]. The main difference lies in the fact that we suppose $\operatorname{var}(Y)$ known which in the signal detection framework translates in the fact that we would know the quantity $\|f\|^2+\sigma^2$.

We are now in position to compare the results of Proposition 4.1 and Theorem 4.3. We distinguish between the values of k:

- When $k \leq p^\gamma$ for some $\gamma < 1/2$, if n is large enough to satisfy the assumption of Proposition 4.1, the quantities $\rho^2_{k,p,n}$ and $\rho'^2_{k,p,n}$ are both of the order $\frac{k\log(p)}{n}$ times a constant (which depends on γ, α and δ). This shows that the lower bound given in Theorem 4.3 is sharp. Additionally, in this case, the procedure T_α defined in Proposition 4.1 follows approximately the minimax rate of testing. We recall that our procedure T_α does not depend on the knowledge of $\operatorname{var}(Y)$ and $\operatorname{corr}(X)$. In applications, a small k typically corresponds to testing a Gaussian graphical model with respect to a graph $\mathcal{G}$ when the number of nodes is large and the graph is supposed to be sparse. When n does not satisfy the assumption of Proposition 4.1, we believe that our lower bound is not sharp anymore.
- When $\sqrt{p} \leq k \leq p$, the lower bound and the upper bound do not coincide anymore. Nevertheless, if $n \geq (1+\gamma)p$ for some $\gamma > 0$, Theorem 3.3 shows that the test $\phi_{\mathcal{I},\alpha}$ defined in (2.3) has power greater than δ over the vectors θ that satisfy

$$\frac{\|\theta\|^2}{\operatorname{var}(Y)-\|\theta\|^2} \geq L(\gamma,\alpha,\delta)\frac{\sqrt{p}}{n}. \tag{4.5}$$

This upper bound and the lower bound do not depend on k. Here again, the lower bound obtained in Theorem 4.3 is sharp and the test $\phi_{\mathcal{I},\alpha}$ defined previously is rate optimal. The fact that the rate of testing stabilizes around $\sqrt{p}/n$ for $k > \sqrt{p}$ also appears in signal detection and there is a discussion of this phenomenon in [2].

- When $k < \sqrt{p}$ and k is close to $\sqrt{p}$, the lower bound and the upper bound given by Proposition 4.1 differ from at most a $\log(p)$ factor. For instance, if k is of order $\sqrt{p}/\log p$, the lower bound in Theorem 4.3 is of order $\sqrt{p}\log\log p/\log p$, and the upper bound is of order $\sqrt{p}$. We do not know if any of this bound is sharp or if the minimax rates of testing coincide when $\operatorname{var}(Y)$ is fixed and when it is not fixed.

All in all, the minimax rates of testing exhibit the same range of rates in our framework as in signal detection [2] when the covariates are independent. Moreover, this implies that the minimax rate of testing is slower when the $(X_i)_{i\in\mathcal{I}}$ are independent than for any other form of dependence. Indeed, the upper bounds obtained in Proposition 4.1 and in (4.5) do not depend on the covariance of X. Then a natural question arises: is the test statistic T_α rate optimal for other correlation of X? We will partially answer this question when testing against the alternative $\theta \in \Theta[1, p]$.

4.3. *Minimax rates for dependent covariates.* In this section, we look for the minimax rate of testing $\theta = 0$ against $\theta \in \Theta[1, p]$ when the covariates X_i are no longer independent. We know that this rate is between the orders $\frac{1}{n}$ which is the minimax rate of testing when we know which coordinate is nonzero and $\frac{\log(p)}{n}$, the minimax rate of testing for independent covariates.

PROPOSITION 4.4. *Let us suppose that there exists a positive number c such that for any $i \neq j$,*

$$|\operatorname{corr}(X_i, X_j)| \leq c$$

and that $\alpha + \delta \leq 53\%$. We define $\rho^2_{1,p,n,c}$ as

$$\rho^2_{1,p,n,c} := \frac{L}{n}\left(\log(p) \wedge \frac{1}{c}\right). \tag{4.6}$$

Then for any $\rho < \rho_{1,p,n,c}$,

$$\beta_\Sigma\left(\left\{\theta \in \Theta[1, p], \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} = \rho^2\right\}\right) \geq \delta,$$

where Σ refers to the covariance matrix of X.

REMARK. If the correlation between the covariates is smaller than $1/\log(p)$, then the minimax rate of testing is of the same order as in the independent case. If the correlation between the covariates is larger, we show in the following proposition that under some additional assumption, the rate is faster.

PROPOSITION 4.5. *Let us suppose that the correlation between X_i and X_j is exactly $c > 0$ for any $i \neq j$. Moreover, we assume that n satisfies the following condition*:

$$n \geq L\left[1 + \log\left(\frac{p}{\alpha\delta}\right)\right]. \tag{4.7}$$

Let introduce the random variable $X_{p+1} := \frac{1}{p}\sum_{i=1}^{p} \frac{X_i}{\sqrt{\operatorname{var}(X_i)}}$. If $\alpha < 60\%$ and $\delta < 60\%$ the test T_α defined by

$$T_\alpha = \left[\sup_{1 \leq i \leq p} \phi_{\{i\},\alpha/(2p)}\right] \vee \phi_{\{p+1\},\alpha/2}$$

satisfies

$$\mathbb{P}_0(T_\alpha > 0) \leq \alpha \quad \textit{and} \quad \mathbb{P}_\theta(T_\alpha > 0) \geq 1 - \delta$$

for any θ in $\Theta[1, p]$ such that

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq \frac{L(\alpha, \delta)}{n}\left(\log p \wedge \frac{1}{c}\right).$$

Consequently, when the correlation between X_i and X_j is a positive constant c, the minimax rate of testing is of order $\frac{\log(p)\wedge(1/c)}{n}$. When the correlation coefficient c is small, the minimax rate of testing coincides with the independent case and when c is larger those rates differ. Therefore, the test T_α defined in Proposition 4.1 is not rate optimal when the correlation is known and is large. Indeed, when the correlation between the covariates is large, the test statistics $\phi_{\{m\},\alpha_m}$ defining T_α are highly correlated. The choice of the weights α_m in procedure P_1 corresponds to a Bonferroni procedure which is precisely known to behave badly when the tests are positively correlated.

This example illustrates the limits of procedure P_1. However, it is not very realistic to suppose that the covariates have a constant correlation, for instance, when one considers a GGM. Indeed, we expect that the correlation between two covariates is large if they are neighbors in the graph and smaller if they are far (w.r.t. the graph distance). This is why we derive lower bounds of the rate of testing for other kinds of correlation matrices often used to model stationary processes.

PROPOSITION 4.6. *Let $X_1, \dots, X_p$ form a stationary process on the one-dimensional torus. More precisely, the correlation between X_i and X_j is a function of $|i - j|_p$ where $|\cdot|_p$ refers to the toroidal distance defined by*

$$|i - j|_p := (|i - j|) \wedge (p - |i - j|).$$

$\Sigma_1(w)$ and $\Sigma_2(t)$, respectively, refer to the correlation matrix of X such that

$$\operatorname{corr}(X_i, X_j) = \exp(-w|i-j|_p) \qquad \textit{where } w > 0,$$
$$\operatorname{corr}(X_i, X_j) = (1 + |i-j|_p)^{-t} \qquad \textit{where } t > 0.$$

Let us set $\rho^2_{1,p,n,\Sigma_1}(w)$ and $\rho^2_{1,p,n,\Sigma_2}(t)$ such that

$$\rho^2_{1,p,n,\Sigma_1}(w) := \frac{1}{n}\log\left(1 + L(\alpha,\delta)p\frac{1-e^{-w}}{1+e^{-w}}\right),$$
$$\rho^2_{1,p,n,\Sigma_2}(t) := \begin{cases} \frac{1}{n}\log\left(1 + L(\alpha,\delta)\frac{p(t-1)}{t+1}\right), & \textit{if } t > 1, \\ \frac{1}{n}\log\left(1 + L(\alpha,\delta)\frac{p}{1+2\log(p-1)}\right), & \textit{if } t = 1, \\ \frac{1}{n}\log\left(1 + L(\alpha,\delta)p^t 2^{-t}(1-t)\right), & \textit{if } 0 < t < 1. \end{cases}$$

Then for any $\rho^2 < \rho^2_{1,p,n,\Sigma_1}(w)$,

$$\beta_{\Sigma_1(w)}\left(\left\{\theta \in \Theta[1,p], \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} = \rho^2\right\}\right) \geq \delta,$$

and for any $\rho^2 < \rho^2_{1,p,n,\Sigma_2}(t)$,

$$\beta_{\Sigma_2(t)}\left(\left\{\theta \in \Theta[1,p], \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} = \rho^2\right\}\right) \geq \delta.$$

If the range ω is larger than $1/p^\gamma$ or if the range t is larger than γ for some $\gamma < 1$, these lower bounds are of order $\frac{\log p}{n}$. As a consequence, for any of these correlation models the minimax rate of testing is of the same order as the minimax rate of testing for independent covariates. This means that our test T_α defined in Proposition 4.1 is rate-optimal for these correlations matrices. However, if ω is smaller than $1/p$ or if t is smaller than $1/\log(p)$, we recover the parametric rates $1/n$ which is achieved by the test $\phi_{\{p+1\},\alpha}$. This comes from the fact that the correlation $\operatorname{corr}(X_1, X_i)$ does not converge to zero for such choices of ω or t. We omit the details since the arguments are similar for the proof of Proposition 4.5.

To conclude, when $k \leq p^\gamma$ (for $\gamma \leq 1/2$), the test T_α defined in Proposition 4.1 is approximately (α,δ)-minimax against the alternative $\theta \in \Theta[k,p]$ when neither $\operatorname{var}(Y)$ nor the covariance matrix of X is fixed. Indeed, the rate of testing of T_α coincides (up to a constant) with the supremum of the minimax rates of testing on $\Theta[k,p]$ over all possible covariance matrices Σ:

$$\rho(\Theta[k,p],\alpha,\delta) := \sup_{\operatorname{var}(Y)>0,\,\Sigma>0} \rho(\Theta[k,p],\alpha,\delta,\operatorname{var}(Y),\Sigma),$$

where the supremum is taken over all positive $\mathrm{var}(Y)$ and every positive definite matrix Σ. When $k \geq \sqrt{p}$ and when $n \geq (1+\gamma)p$ (for $\gamma > 0$), the test defined in (4.5) has the same behavior.

However, our procedure does not adapt to Σ; for some correlation matrices (as shown, for instance, in Proposition 4.5), T_α with procedure P_1 is not rate optimal. Nevertheless, we believe and will illustrate in Section 6 that procedure P_2 slightly improves the power of the test when the covariates are correlated.

5. Rates of testing on "ellipsoids" and adaptation. In this section, we define tests T_α of the form (3.1) in order to test simultaneously $\theta = 0$ against θ belonging to some classes of ellipsoids. We will study their rates and show that they are optimal at sometimes the price of a $\log p$ factor.

For any nonincreasing sequence $(a_i)_{1\leq i\leq p+1}$ such that $a_1 = 1$ and $a_{p+1} = 0$ and any $R > 0$, we define the ellipsoid $\mathcal{E}_a(R)$ by

$$\mathcal{E}_a(R) := \left\{\theta \in \mathbb{R}^p, \sum_{i=1}^{p} \frac{\mathrm{var}(Y|X_{m_{i-1}}) - \mathrm{var}(Y|X_{m_i})}{a_i^2} \leq R^2\, \mathrm{var}(Y|X)\right\}, \tag{5.1}$$

where m_i refers to the set $\{1,\dots,i\}$ and $m_0 = \varnothing$.

Let us explain why we call this set an ellipsoid. Assume for instance that the (X_i) are independent identically distributed with variance one. In this case, the difference $\mathrm{var}(Y|X_{m_{i-1}}) - \mathrm{var}(Y|X_{m_i})$ equals $|\theta_i|^2$, and the definition of $\mathcal{E}_a(R)$ translates in

$$\mathcal{E}_a(R) = \left\{\theta \in \mathbb{R}^p, \sum_{i=1}^{p} \frac{|\theta_i|^2}{a_i^2} \leq R^2\, \mathrm{var}(Y|X)\right\}.$$

The main difference between this definition and the classical definition of an ellipsoid in the fixed design regression framework (as, for instance, in [2]) is the presence of the term $\mathrm{var}(Y|X)$. We added this quantity in order to be able to derive lower bounds of the minimax rate. If the X_i are not i.i.d. with unit variance, it is always possible to create a sequence X_i' of i.i.d. standard Gaussian variables by orthogonalizing the X_i using the Gram–Schmidt process. If we call θ' the vector in $\mathbb{R}^p$ such that $X\theta = X'\theta'$, it is straightforward to show that $\mathrm{var}(Y|X_{m_{i-1}}) - \mathrm{var}(Y|X_{m_i}) = |\theta_i'|^2$. We can then express $\mathcal{E}_a(R)$ using the coordinates of θ' as previously;

$$\mathcal{E}_a(R) = \left\{\theta \in \mathbb{R}^p, \sum_{i=1}^{p} \frac{|\theta_i'|^2}{a_i^2} \leq R^2\, \mathrm{var}(Y|X)\right\}.$$

The main advantage of Definition 5.1 is that it does not directly depend on the covariance of X. In the sequel we also consider the special case of ellipsoids with polynomial decay,

$$\mathcal{E}_s'(R) := \left\{\theta \in \mathbb{R}^p, \sum_{i=1}^{p} \frac{\mathrm{var}(Y|X_{m_{i-1}}) - \mathrm{var}(Y|X_{m_i})}{i^{-2s}\, \mathrm{var}(Y|X)} \leq R^2\right\}, \tag{5.2}$$

where $s > 0$ and $R > 0$. First, we define two test procedures of the form (3.1) and evaluate their power respectively on the ellipsoids $\mathcal{E}_a(R)$ and on the ellipsoids $\mathcal{E}'_s(R)$. Then we give some lower bounds for the (α, δ)-simultaneous minimax rates of testing. Extensions to more general l_p balls with $0 < p < 2$ are possible to the price of more technicalities by adapting the results of Section 4 in Baraud [2].

These alternatives correspond to the situation where we are given an order of relevance on the covariates that are not in the null hypothesis. This order could either be provided by a previous knowledge of the model or by a model selection algorithm such as LARS (least angle regression) introduced by Efron et al. [9]. We apply this last method to build a collection of models for our testing procedure (3.1) in [23].

5.1. *Simultaneous rates of testing of T_α over classes of ellipsoids.* First, we define a procedure of the form (3.1) in order to test if $\theta = 0$ against θ belongs to any of the ellipsoids $\mathcal{E}_a(R)$. For any $x > 0$, $[x]$ denotes the integer part of x.

We choose the class of models $\mathcal{M}$ and the weights α_m as follows:

- If $n < 2p$, we take the set $\mathcal{M}$ to be $\bigcup_{1 \leq k \leq [n/2]} m_k$, and all the weights α_m are equal to $\alpha/|\mathcal{M}|$.
- If $n \geq 2p$, we take the set $\mathcal{M}$ to be $\bigcup_{1 \leq k \leq p} m_k$. α_{m_p} equals $\alpha/2$ and for any k between 1 and $p-1$, and α_{m_k} is chosen to be $\alpha/(2(p-1))$.

As previously, we bound the power of the tests T_α from a nonasymptotic point of view.

PROPOSITION 5.1. *Let us assume that*

$$n \geq L\left[1 + \log\left(\frac{1}{\alpha\delta}\right)\right]. \tag{5.3}$$

For any ellipsoid $\mathcal{E}_a(R)$, the test T_α defined by (3.1) *with procedure P_1 and with the class of models given just above satisfies*

$$\mathbb{P}_0(T_\alpha \leq 0) \geq 1 - \alpha,$$

and $\mathbb{P}_\theta(T_\alpha > 0) \geq 1 - \delta$ for all $\theta \in \mathcal{E}_a(R)$ such that

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq L(\alpha, \delta) \log n \inf_{1 \leq i \leq [n/2]} \left[a_{i+1}^2 R^2 + \frac{\sqrt{i}}{n}\right] \tag{5.4}$$

if $n < 2p$, or

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq L(\alpha, \delta) \left\{\left[\log p \inf_{1 \leq i \leq p-1} \left(a_{i+1}^2 R^2 + \frac{\sqrt{i}}{n}\right)\right] \wedge \frac{\sqrt{p}}{n}\right\} \tag{5.5}$$

if $n \geq 2p$.

All in all, for large values of n, the rate of testing is of order $\sup_{1\leq i\leq p}[a_i^2R^2 \wedge \frac{\sqrt{i\log(p)}}{n}]$. We show in the next subsection that the minimax rate of testing for an ellipsoid is of order

$$\sup_{1\leq i\leq p}\left[a_i^2R^2 \wedge \frac{\sqrt{i}}{n}\right].$$

Moreover, we prove in Proposition 5.6 that a loss in $\sqrt{\log\log p}$ is unavoidable if one considers the simultaneous minimax rates of testing over a family of nested ellipsoids. Nevertheless, we do not know if the term $\sqrt{\log(p)}$ is optimal for testing simultaneously against all the ellipsoids $\mathcal{E}_a(R)$ for all sequences (a_i) and all $R > 0$. When n is smaller than $2p$, we obtain comparable results except that we are unable to consider alternatives in large dimensions in the infimum (5.5).

We now turn to define a procedure of the form (3.1) in order to test simultaneously that $\theta = 0$ against θ belongs to any of the $\mathcal{E}'_s(R)$. For this, we introduce the following collection of models $\mathcal{M}$ and weights α_m:

- If $n < 2p$, we take the set $\mathcal{M}$ to be $\bigcup m_k$ where k belongs to $\{2^j, j\geq 0\}\cap\{1,\dots,[n/2]\}$, and all the weights α_m are chosen to be $\alpha/|\mathcal{M}|$.
- If $n\geq 2p$, we take the set $\mathcal{M}$ to be $\bigcup m_k$ where k belongs to $(\{2^j, j\geq 0\}\cap\{1,\dots,p\})\cup\{p\}$, α_{m_p} equals $\alpha/2$ and for any k in the model between 1 and $p-1$, α_{m_k} is chosen to be $\alpha/(2(|\mathcal{M}|-1))$.

PROPOSITION 5.2. *Let us assume that*

$$n \geq L\left[1+\log\left(\frac{1}{\alpha\delta}\right)\right] \tag{5.6}$$

and that $R^2 \geq \sqrt{\log\log n}/n$. *For any* $s > 0$, *the test procedure* T_α *defined by* (3.1) *with procedure* P_1 *and with a class of models given just above satisfies*

$$\mathbb{P}_0(T_\alpha > 0) \geq 1-\alpha,$$

and $\mathbb{P}_\theta(T_\alpha > 0)\geq 1-\delta$ *for any* $\theta\in\mathcal{E}'_s(R)$ *such that*

$$\begin{aligned}\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq L(\alpha,\delta)\Bigg[&R^{2/(1+4s)}\left(\frac{\sqrt{\log\log n}}{n}\right)^{4s/(1+4s)}\\ &+R^2(n/2)^{-2s}+\frac{\log\log n}{n}\Bigg]\end{aligned} \tag{5.7}$$

if $n < 2p$ *or*

$$\begin{aligned}&\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2}\\ &\quad\geq L(\alpha,\delta)\left(\left[R^{2/(1+4s)}\left(\frac{\sqrt{\log\log p}}{n}\right)^{4s/(1+4s)}+\frac{\log\log p}{n}\right]\wedge\frac{\sqrt{p}}{n}\right)\end{aligned} \tag{5.8}$$

if $n\geq 2p$.

Again, we retrieve similar results to those of Corollary 2 in [3] in the fixed design regression framework. For $s > 1/4$ and $n < 2p$, the rate of testing is of order $(\frac{\sqrt{\log\log n}}{n})^{4s/(1+4s)}$. We show in the next subsection that the logarithmic factor is due to the adaptive property of the test. If $s \leq 1/4$, the rate is of order n^{-2s}. When $n \geq 2p$, the rate is of order $(\frac{\sqrt{\log\log p}}{n})^{4s/(1+4s)} \wedge (\frac{\sqrt{p}}{n})$, and we mention at the end of the next subsection that it is optimal.

Here again, it is possible to define these tests with procedure P_2 in order to improve the power of the test (see Section 6 for numerical results).

5.2. *Minimax lower bounds.* We first establish the (α, δ)-minimax rate of testing over an ellipsoid when the variance of Y and the covariance matrix of X are known.

PROPOSITION 5.3. *Let us set the sequence $(a_i)_{1\leq i\leq p+1}$ and the positive number R. We introduce*

$$\rho^2_{a,n}(R) := \sup_{1\leq i\leq p} [\rho^2_{i,n} \wedge a_i^2 R^2], \tag{5.9}$$

where $\rho^2_{i,n}$ is defined by (4.3); then for any nonsingular covariance matrix Σ we have

$$\beta_\Sigma\left(\left\{\theta \in \mathcal{E}_a(R), \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq \rho^2_{a,n}(R)\right\}\right) \geq \delta,$$

where the quantity $\mathrm{var}(Y)$ *is fixed. If $\alpha + \delta \leq 47\%$, then*

$$\rho^2_{a,n}(R) \geq \sup_{1\leq i\leq p}\left[\frac{\sqrt{i}}{n} \wedge a_i^2 R^2\right].$$

This lower bound is once more analogous to the one in the fixed design regression framework. Contrary to the lower bounds obtained in the previous section, it does not depend on the covariance of the covariates. We now look for an upper bound of the minimax rate of testing over a given ellipsoid. First, we need to define the quantity D^* as

$$D^* := \inf\left\{1 \leq i \leq p, a_i^2 R^2 \leq \frac{\sqrt{i}}{n}\right\}$$

with the convention that $\inf \varnothing = p$.

PROPOSITION 5.4. *Let us assume that $n \geq L\log[1 + \log(\frac{1}{\alpha\delta})]$. If $R^2 > \frac{1}{n}$ and $D^* \leq n/2$; the test $\phi_{m_{D^*},\alpha}$ defined by (2.3) satisfies*

$$\mathbb{P}_0[\phi_{m_{D^*},\alpha} = 1] \leq \alpha \quad \textit{and} \quad \mathbb{P}_\theta[\phi_{m_{D^*},\alpha} = 0] \leq \delta$$

for all $\theta \in \mathcal{E}_a(R)$ *such that*

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq L(\alpha, \delta) \sup_{1 \leq i \leq p} \left[\frac{\sqrt{i}}{n} \wedge a_i^2 R \right].$$

If $n \geq 2D^*$, the rates of testing on an ellipsoid are analogous to the rates on an ellipsoid in fixed design regression framework (see, for instance, [2]). If D^* is large and n is small, the bounds in Propositions 5.3 and 5.4 do not coincide. In this case, we do not know if this comes from the fact that the test in Proposition 5.4 does not depend on the knowledge of $\operatorname{var}(Y)$ or if one of the bounds in Propositions 5.3 and 5.4 is not sharp.

We are now interested in computing lower bounds for rates of testing simultaneously over a family of ellipsoids in order to compare them with rates obtained in Section 5.1. First, we need a lower bound for the minimax simultaneous rate of testing over nested linear spaces. We recall that for any $D \in \{1, \dots, p\}$, S_{m_D} stands for the linear spaces of vectors θ such that only their D first coordinates are possibly nonzero.

PROPOSITION 5.5. *For* $D \geq 2$, *let us set*

$$\bar{\rho}_{D,n}^2 := L(\alpha, \delta) \frac{\sqrt{\log\log(D+1)}\sqrt{D}}{n}. \tag{5.10}$$

Then the following lower bound holds:

$$\beta_I \left(\bigcup_{1 \leq D \leq p} \left\{ \theta \in S_{m_D}, \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} = r_D^2 \right\} \right) \geq \delta$$

if for all D *between* 1 *and* p, $r_D \leq \bar{\rho}_{D,n}$.

Using this proposition, it is possible to get a lower bound for the simultaneous rate of testing over a family of nested ellipsoids.

PROPOSITION 5.6. *We fix a sequence* $(a_i)_{1 \leq i \leq p+1}$. *For each* $R > 0$, *let us set*

$$\bar{\rho}_{a,R,n}^2 := \sup_{1 \leq D \leq p} [\bar{\rho}_{D,n}^2 \wedge (R^2 a_D^2)], \tag{5.11}$$

where $\bar{\rho}_{D,n}$ *is given by* (5.10). *Then, for any non-singular covariance matrix* Σ *of the vector* X,

$$\beta_\Sigma \left(\bigcup_{R > 0} \left\{ \theta \in \mathcal{E}_a(R), \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \leq \bar{\rho}_{a,R,n}^2 \right\} \right) \geq \delta.$$

This proposition shows that the problem of adaptation is impossible in this setting: it is impossible to define a test which is simultaneously minimax over a class of nested ellipsoids (for $R > 0$). This is also the case in fixed design as proved by [22] for the case of Besov bodies. The loss of a term of the order $\sqrt{\log\log p}/n$ is unavoidable.

As a special case of Proposition 5.6, it is possible to compute a lower bound for the simultaneous minimax rate over $\mathcal{E}'_s(R)$ where R describes the positive numbers. After some calculation, we find a lower bound of order

$$\left(\frac{\sqrt{\log\log p}}{n}\right)^{4s/(1+4s)} \wedge \frac{\sqrt{p\log\log p}}{n}.$$

This shows that the power of the test T_α obtained in (5.8) for $n \geq 2p$ is optimal when $R^2 \geq \sqrt{\log\log n}/n$. However, when $n < 2p$ and $s \leq 1/4$, we do not know if the rate n^{-2s} is optimal or not.

To conclude, when $n \geq 2p$ the test T_α defined in Proposition 5.2 achieves the simultaneous minimax rate over the classes of ellipsoids $\mathcal{E}'_s(R)$. On the other hand, the test T_α defined in Proposition 5.1 is not rate optimal simultaneously over all the ellipsoids $\mathcal{E}_a(R)$ and suffers a loss of a $\sqrt{\log p}$ factor even when $n \geq 2p$.

6. Simulations studies. The purpose of this simulation study is threefold. First, we illustrate the theoretical results established in previous sections. Second, we show that our procedure is easy to implement for different choices of collections $\mathcal{M}$ and is computationally feasible even when p is large. Our third purpose is to compare the efficiency of procedures P_1 and P_2. Indeed, for a given collection $\mathcal{M}$, we know from Section 3.2 that the test (3.1) based on procedure P_2 is more powerful than the corresponding test based on P_1. However, the computation of the quantity $q_{\mathbf{X},\alpha}$ is possibly time consuming and we therefore want to know if the benefit in power is worth the computational burden.

To our knowledge, when the number of covariates p is larger than the number of observations n there is no test with which we can compare our procedure.

6.1. *Simulation experiments.* We consider the regression model (1.1) with $\mathcal{I} = \{1, \dots, p\}$ and test the null hypothesis "$\theta = 0$" which is equivalent to "Y is independent of X" at level $\alpha = 5\%$. Let $(X_i)_{1\leq i\leq p}$ be a collection of p Gaussian variables with unit variance. The random variable is defined as follows: $Y = \sum_{i=1}^p \theta_i X_i + \varepsilon$ where ε is a zero mean Gaussian variable with variance $1 - \|\theta\|^2$ independent of X.

We consider two simulation experiments described below:

1. First simulation experiment: The correlation between X_i and X_j is a constant c for any $i \neq j$. Moreover, in this experiment the parameter θ is chosen such that only one of its components is possibly nonzero. This corresponds to the situation considered in Section 4. First, the number of covariates p is fixed

equal to 30, and the number of observations n is taken equal to 10 and 15. We choose for c three different values 0, 0.1 and 0.8, allowing thus the comparison of the procedures for independent, weakly and highly correlated covariates. We estimate the size of the test by taking $\theta_1 = 0$ and the power by taking for θ_1 the values 0.8 and 0.9. Theses choices of θ lead to a small and a large signal/noise ratio $r_{\mathrm{s/n}}$ defined in (2.5) and equal in this experiment to $\theta_1^2/(1-\theta_1^2)$. Second, we examine the behavior of the tests when p increases and when the covariates are highly correlated: p equals 100 and 500, n equals 10 and 15, θ_1 is set to 0 and 0.8, and c is chosen to be 0.8.

2. Second simulation experiment: The covariates $(X_i)_{1\leq i\leq p}$ are independent. The number of covariates p equals 500 and the number of observations n equals 50 and 100. We set for any $i \in \{1,\dots,p\}$, $\theta_i = Ri^{-s}$. We estimate the size of the test by taking $R = 0$ and the power by taking for (R, s) the value $(0.2, 0.5)$ which corresponds to a slow decrease of the $(\theta_i)_{1\leq i\leq p}$. It was pointed out in the beginning of Section 5 that $|\theta_i|^2$ equals $\mathrm{var}(Y|X_{m_{i-1}}) - \mathrm{var}(Y|X_{m_i})$. Thus $|\theta_i|^2$ represents the benefit in term of conditional variance brought by the variable X_i.

We use our testing procedure defined in (3.1) with different collections $\mathcal{M}$ and different choices for the weights $\{\alpha_m, m \in \mathcal{M}\}$.

The collections $\mathcal{M}$: we define three classes. Let us set $J_{n,p} = p \wedge [\frac{n}{2}]$ where $[x]$ denotes the integer part of x, and let us define

$$\begin{aligned}
\mathcal{M}^1 &:= \{\{i\}, 1 \leq i \leq p\},\\
\mathcal{M}^2 &:= \{m_k = \{1,2,\dots,k\}, 1 \leq k \leq J_{n,p}\},\\
\mathcal{M}^3 &:= \{m_k = \{1,2,\dots,k\}, k \in \{2^j, j \geq 0\} \cap \{1,\dots,J_{n,p}\}\}.
\end{aligned}$$

We evaluate the performance of our testing procedure with $\mathcal{M} = \mathcal{M}^1$ in the first simulation experiment and $\mathcal{M} = \mathcal{M}^2$ and $\mathcal{M}^3$ in the second simulation experiment. The cardinality of these three collections is smaller than p, and the computational complexity of the testing procedures is at most linear in p.

The collections $\{\alpha_m, m \in \mathcal{M}\}$: We consider procedures P_1 and P_2 defined in Section 3. When we are using the procedure P_1, the α_ms equal $\alpha/|\mathcal{M}|$ where $|\mathcal{M}|$ denotes the cardinality of the collection $\mathcal{M}$. The quantity $q_{\mathbf{X},\alpha}$ that occurs in the procedure P_2 is computed by simulation. We use 1000 simulations for the estimation of $q_{\mathbf{X},\alpha}$. In the sequel we note $T_{\mathcal{M}^i,P_j}$, the test (3.1), with collection $\mathcal{M}^i$ and procedure P_j.

In the first experiment, when p is large, we also consider two other tests:

1. The first test is $\phi_{\{1\},\alpha}$ (2.3) of the hypothesis $\theta_1 = 0$ against the alternative $\theta_1 \neq 0$. This test corresponds to the single test when we know which coordinate is nonzero.

2. The second test is $\phi_{\{p+1\},\alpha}$ where $X_{p+1} := \frac{1}{p}\sum_{i=1}^{p} X_i$. Adapting the proof of Proposition 4.5, we know that this test is approximately minimax on $\Theta[1, p]$ if the correlation between the covariates is constant and large.

Contrary to our procedures, these two tests are based on the knowledge of $\mathrm{var}(X)$ (and eventually θ). We only use them as a benchmark to evaluate the performance of our procedure. We aim at showing that our test with procedure P_2 is as powerful as $\phi_{\{p+1\},\alpha}$ and is close to the test $\phi_{\{1\},\alpha}$.

We estimate the size and the power of the testing procedures with 1000 simulations. For each simulation, we simulate the gaussian vector $(X_1, \dots, X_p)$ and then simulate the variable Y as described in the two simulation experiments.

6.2. *Results of the simulation.* The results of the first simulation experiment for $c = 0$ are given in Table 1. As expected, the power of the tests increases with the number of observations n and with the signal/noise ratio $r_{\mathrm{s/n}}$. If the signal/noise ratio is large enough, we obtain powerful tests even if the number of covariates p is larger than the number of observations.

In Table 2 we present results of the first simulation experiment for $\theta_1 = 0.8$ when c varies.

Let us first compare the results for independent, weakly and highly correlated covariates when using procedure P_1. The size and the power of the test for weakly correlated covariates are similar to the size and the power obtained in the independent case. Hence, we recover the remark following Proposition 4.4: when the correlation coefficient between the covariates is small, the minimax rate is of the same order as in the independent case. The test for highly correlated covariates is more powerful than the test for independent covariates, recovering thus the remark

TABLE 1
First simulation study, independent case: $p = 30$, $c = 0$.
Percentages of rejection and value of the signal/noise ratio $r_{\mathrm{s/n}}$

n	$T_{\mathcal{M}^1, P_1}$	$T_{\mathcal{M}^1, P_2}$
	Null hypothesis is true, $\theta_1 = 0$	
10	0.043	0.045
15	0.044	0.049
	Null hypothesis is false, $\theta_1 = 0.8$, $r_{\mathrm{s/n}} = 1.78$	
10	0.48	0.48
15	0.81	0.81
	Null hypothesis is false, $\theta_1 = 0.9$, $r_{\mathrm{s/n}} = 4.26$	
10	0.86	0.86
15	0.99	0.99

TABLE 2
First simulation study, independent and dependent case: $p = 30$, $c = 0, 0.1, 0.8$. *Frequencies of rejection*

n	$c = 0$ $T_{\mathcal{M}^1,P_1}$	$c = 0$ $T_{\mathcal{M}^1,P_2}$	$c = 0.1$ $T_{\mathcal{M}^1,P_1}$	$c = 0.1$ $T_{\mathcal{M}^1,P_2}$	$c = 0.8$ $T_{\mathcal{M}^1,P_1}$	$c = 0.8$ $T_{\mathcal{M}^1,P_2}$
			Null hypothesis is true, $\theta_1 = 0$			
10	0.043	0.045	0.042	0.04	0.018	0.045
15	0.044	0.049	0.058	0.06	0.019	0.052
			Null hypothesis is false, $\theta_1 = 0.8$			
10	0.48	0.48	0.49	0.49	0.64	0.77
15	0.81	0.81	0.81	0.82	0.89	0.94

following Theorem 4.3: the worst case from a minimax rate perspective is the case where the covariates are independent. Let us now compare procedures P_1 and P_2. In the case of independent or weakly correlated covariates, they give similar results. For highly correlated covariates, the power of $T_{\mathcal{M}^1,P_2}$ is much larger than the one of $T_{\mathcal{M}^1,P_1}$.

In Table 3 we present results of the multiple testing procedures and of the two tests, $\phi_{\{1\},\alpha}$ and $\phi_{\{p+1\},\alpha}$, when $c = 0.8$ and the number of covariates p is large. For $p = 500$ and $n = 15$, one test takes less than one second with procedure P_1 and less than 30 s with procedure P_2. As expected, procedure P_1 is too conservative when p increases. For $p = 100$, the power of the test based on procedure P_1 is smaller than the power of the test $\phi_{\{p+1\},\alpha}$, and this difference increases when p is larger. The test based on procedure P_2 is as powerful as $\phi_{\{p+1\},\alpha}$, and its power is close to the one of $\phi_{\{1\},\alpha}$. We recall that this last test is based on the knowledge of the nonzero component of θ contrary to ours. Moreover, the test $\phi_{\{p+1\},\alpha}$ was shown in Proposition 4.5 to be optimal for this particular correlation

TABLE 3
First simulation study, dependent case: $c = 0.8$. *Frequencies of rejection*

n	$p = 100$ $T_{\mathcal{M}^1,P_1}$	$p = 100$ $T_{\mathcal{M}^1,P_2}$	$p = 100$ $\phi_{\{1\},\alpha}$	$p = 100$ $\phi_{\{p+1\},\alpha}$	$p = 500$ $T_{\mathcal{M}^1,P_1}$	$p = 500$ $T_{\mathcal{M}^1,P_2}$	$p = 500$ $\phi_{\{1\},\alpha}$	$p = 500$ $\phi_{\{p+1\},\alpha}$
				Null hypothesis is true, $\theta_1 = 0$				
10	0.01	0.056	0.051	0.045	0.009	0.044	0.040	0.040
15	0.016	0.053	0.047	0.053	0.011	0.040	0.042	0.034
				Null hypothesis is false, $\theta_1 = 0.8$				
10	0.60	0.77	0.91	0.79	0.52	0.76	0.91	0.77
15	0.85	0.92	0.99	0.92	0.77	0.94	0.99	0.94

TABLE 4
Second simulation study. Frequencies of rejection

n	$T_{\mathcal{M}^2,P_1}$	$T_{\mathcal{M}^2,P_2}$	$T_{\mathcal{M}^3,P_1}$	$T_{\mathcal{M}^3,P_2}$
		Null hypothesis is true, $R = 0$		
50	0.013	0.052	0.036	0.059
100	0.009	0.059	0.042	0.059
		Null hypothesis is false, $R = 0.2, s = 0.5$		
50	0.17	0.33	0.31	0.38
100	0.42	0.66	0.62	0.69

setting. Hence, procedure P_2 seems to achieve the optimal rate in this situation. Thus, we advise to use in practice procedure P_2 if the number of covariates p is large because procedure P_1 becomes too conservative, especially if the covariates are correlated.

The results of the second simulation experiment are given in Table 4. As expected, procedure P_2 improves the power of the test and the test $T_{\mathcal{M}^3,P_2}$ has the greatest power. In this setting, one should prefer the collection $\mathcal{M}^3$ to $\mathcal{M}^2$. This was previously pointed out in Section 5 from a theoretical point of view. Although $T_{\mathcal{M}^3,P_1}$ is conservative, it is a good compromise for practical issues: it is very easy and fast to implement, and its performances are good.

7. Proofs of Theorem 3.3, Propositions 4.1, 4.5, 5.1, 5.2 and 5.4.

PROOF OF THEOREM 3.3. In a nutshell, we shall prove that conditionally to the design $\mathbf{X}$ the distribution of the test T_α is the same as the test introduced by Baraud et al. [3]. Hence, we may apply their non asymptotic upper bound for the power.

Distribution of $\phi_m(\mathbf{Y},\mathbf{X})$. First, we derive the distribution of the test statistic $\phi_m(\mathbf{Y},\mathbf{X})$ under $\mathbb{P}_\theta$. The distribution of Y conditionally to the set of variables $(X_{V\cup m})$ is of the form

$$Y = \sum_{i\in V\cup m} \theta_i^{V\cup m} X_i + \epsilon^{V\cup m}, \tag{7.1}$$

where the vector $\theta^{V\cup m}$ is constant and $\epsilon^{V\cup m}$ is a zero mean Gaussian variable independent of $X_{V\cup m}$ whose variance is $\operatorname{var}(Y|X_{V\cup m})$. As a consequence, $\|\mathbf{Y} - \Pi_{V\cup m}\mathbf{Y}\|_n^2$ is exactly $\|\Pi_{(V\cup m)^\perp}\boldsymbol{\epsilon}^{V\cup m}\|_n^2$ where $\Pi_{(V\cup m)^\perp}$ denotes the orthogonal projection along the space generated by $(\mathbf{X}_i)_{i\in V\cup m}$.

Using the same decomposition of $\mathbf{Y}$ one simplifies the numerator of $\phi_m(\mathbf{Y},\mathbf{X})$:

$$\|\Pi_{V\cup m}\mathbf{Y} - \Pi_V\mathbf{Y}\|_n^2 = \left\| \sum_{i\in V\cup m} \theta_i^{V\cup m}(\mathbf{X}_i - \Pi_V\mathbf{X}_i) + \Pi_{V^\perp\cap(V\cup m)}\boldsymbol{\epsilon}^{V\cup m}\right\|_n^2,$$

where $\Pi_{V^\perp\cap(V\cup m)}$ is the orthogonal projection onto the intersection between the space generated by $(\mathbf{X}_i)_{i\in V\cup m}$ and the orthogonal of the space generated by $(\mathbf{X}_i)_{i\in V}$.

For any $i\in m$, let us consider the conditional distribution of X_i with respect to X_V,

$$X_i = \sum_{j\in V}\theta_j^{V,i}X_j + \epsilon_i^V, \tag{7.2}$$

where $\theta_j^{V,i}$ are constants and ϵ_i^V is a zero-mean normal Gaussian random variable whose variance is $\mathrm{var}(X_i|X_V)$ and which is independent of X_V. This enables us to express

$$\mathbf{X}_i - \Pi_V\mathbf{X}_i = \Pi_{V^\perp\cap(V\cup m)}\boldsymbol{\epsilon}_i^V \qquad \text{for all } i\in m.$$

Therefore, we decompose $\phi_m(\mathbf{Y},\mathbf{X})$ in

$$\phi_m(\mathbf{Y},\mathbf{X}) = \frac{N_m\|\Pi_{V^\perp\cap(V\cup m)}(\sum_{i\in m}\theta_i^{V\cup m}\boldsymbol{\epsilon}_i^V + \boldsymbol{\epsilon}^{V\cup m})\|_n^2}{D_m\|\Pi_{(V\cup m)^\perp}\boldsymbol{\epsilon}^{V\cup m}\|_n^2}. \tag{7.3}$$

Let us define the random variable $Z_m^{(1)}$ and $Z_m^{(2)}$ where $Z_m^{(1)}$ refers to the numerator of (7.3) divided by N_m and $Z_m^{(2)}$ to the denominator divided by D_m. We now prove that $Z_m^{(1)}$ and $Z_m^{(2)}$ are independent.

The variables $(\boldsymbol{\epsilon}_j^V)_{j\in m}$ are $\sigma(\mathbf{X}_{V\cup m})$-measurable as linear combinations of elements in $\mathbf{X}_{V\cup m}$. Moreover, $\boldsymbol{\epsilon}^{V\cup m}$ follows a zero mean normal distribution with covariance matrix $\mathrm{var}(Y|X_{V\cup m})I_n$ and is independent of $\mathbf{X}_{V\cup m}$. As a consequence, conditionally to $\mathbf{X}_{V\cup m}$, $Z_m^{(1)}$ and $Z_m^{(2)}$ are independent by Cochran's theorem as they correspond to projections onto two sets orthogonal from each other.

As $\boldsymbol{\epsilon}_j^V$ is a linear combination of the columns of $\mathbf{X}_{V\cup m}$, $Z_m^{(1)}$ follows a noncentral χ^2 distribution conditionally to $\mathbf{X}_{V\cup m}$:

$$(Z_m^{(1)}|\mathbf{X}_{V\cup m}) \sim \mathrm{var}(Y|X_{V\cup m})\chi^2\biggl(\frac{\|\sum_{j\in m}\theta_j^{V\cup m}\Pi_{(V\cup m)\cap V^\perp}\boldsymbol{\epsilon}_j^V\|_n^2}{\mathrm{var}(Y|X_{V\cup m})}, D_m\biggr).$$

We denote by $a_m^2(\mathbf{X}_{V\cup m}) := \frac{\|\sum_{j\in m}\theta_j^{V\cup m}\Pi_{(V\cup m)\cap V^\perp}\boldsymbol{\epsilon}_j^V\|_n^2}{\mathrm{var}(Y|X_{V\cup m})}$ this noncentrality parameter.

Power of T_α conditionally to $\mathbf{X}_{V\cup m}$. Conditionally to $\mathbf{X}_{V\cup m}$ our test statistic $\phi_m(\mathbf{Y},\mathbf{X})$ is the same as that proposed by Baraud et al. [3] with $n-d$ data and $\sigma^2 = \mathrm{var}(Y|X_{V\cup m})$. Arguing as in their proof of Theorem 1, there exists some quantity $\bar{\Delta}_m(\delta)$ such that the procedure accepts the hypothesis with probability not larger than $\delta/2$ if $a_m^2(\mathbf{X}_{V\cup m}) > \bar{\Delta}_m(\delta)$:

$$\begin{aligned}\bar{\Delta}_m(\delta) &:= 2.5\sqrt{1+K_m^2(U)}\sqrt{D_m\log\biggl(\frac{4}{\alpha_m\delta}\biggr)}\biggl(1+\sqrt{\frac{D_m}{N_m}}\biggr)\\ &\quad + 2.5[k_mK_m(U)\vee 5]\log\biggl(\frac{4}{\alpha_m\delta}\biggr)\biggl(1+\frac{2D_m}{N_m}\biggr),\end{aligned} \tag{7.4}$$

where $U_m := \log(1/\alpha_m)$, $U := \log(2/\delta)$, $k_m := 2\exp(4U_m/N_m)$ and

$$K_m(u) := 1 + 2\sqrt{\frac{u}{N_m}} + 2k_m\frac{u}{N_m}.$$

Consequently, we have

$$\mathbb{P}_\theta(T_\alpha \le 0|\mathbf{X}_{V\cup m})\mathbf{1}\{a_m^2(\mathbf{X}_{V\cup m}) \ge \bar{\Delta}_m(\delta)\} \le \delta/2. \tag{7.5}$$

Let us derive the distribution of the noncentral parameter $a_m(\mathbf{X}_{V\cup m})$. First, we simplify the projection term as $\boldsymbol{\epsilon}_j^V$ is a linear combination of elements of $\mathbf{X}_{V\cup m}$:

$$\Pi_{(V\cup m)\cap V^\perp}\boldsymbol{\epsilon}_j^V = \Pi_{V\cup m}\boldsymbol{\epsilon}_j^V - \Pi_V\boldsymbol{\epsilon}_j^V = \Pi_{V^\perp}\boldsymbol{\epsilon}_j^V.$$

Let us define κ_m^2 as

$$\kappa_m^2 := \frac{\operatorname{var}(\sum_{j\in m}\theta_j^{V\cup m}\epsilon_j^V)}{\operatorname{var}(Y|X_{V\cup m})}.$$

As the variable $\sum_{j\in m}\theta_j^{V\cup m}\boldsymbol{\epsilon}_j^V$ is independent of $\mathbf{X}_V$, and as almost surely the dimension of the vector space generated by $\mathbf{X}_V$ is d, we get

$$\frac{\|\sum_{j\in m}\theta_j^{V\cup m}\Pi_{V^\perp}\boldsymbol{\epsilon}_j^V\|_n^2}{\operatorname{var}(Y|X_{V\cup m})} \sim \kappa_m^2\chi^2(n-d).$$

Hence, applying for instance Lemma 1 in [16], we get

$$\mathbb{P}_\theta\left[\frac{a_m^2(\mathbf{X}_{V\cup m})}{\kappa_m^2} \ge (n-d) - 2\sqrt{(n-d)U}\right] \le \delta/2.$$

Let us gather (7.5) with this last bound. If

$$\kappa_m^2 \ge \Delta'_m(\delta) := \frac{\bar{\Delta}_m(\delta)}{(n-d)(1-2\sqrt{U/(n-d)})}, \tag{7.6}$$

then it holds that

$$\begin{aligned}
&\mathbb{P}_\theta(T_\alpha \le 0)\\
&\quad\le \mathbb{P}_\theta\big(T_\alpha \le 0, a_m^2(\mathbf{X}_{V\cup m}) > \bar{\Delta}_m(\delta)\big) + \mathbb{P}_\theta[a_m^2(\mathbf{X}_{V\cup m}) \le \bar{\Delta}_m(\delta)]\\
&\quad\le \mathbb{E}_\theta\{\mathbb{P}_\theta[T_\alpha \le 0, a_m^2(\mathbf{X}_{V\cup m}) > \bar{\Delta}_m(\delta)|\mathbf{X}_{V\cup m}]\}\\
&\qquad + \mathbb{P}_\theta\left[\frac{a_m^2(\mathbf{X}_{V\cup m})}{\kappa_m^2} \ge (n-d) - 2\sqrt{(n-d)U}\right]\\
&\quad\le \delta.
\end{aligned}$$

Computation of κ_m^2. Let us now compute the quantity κ_m^2 in order to simplify condition (7.6). Let us first express $\operatorname{var}(Y|X_V)$ in terms of $\operatorname{var}(Y|X_{m\cup V})$ using the decomposition (7.1) of Y.

$$\begin{aligned}\operatorname{var}(Y|X_V) &= \operatorname{var}\Big(\sum_{j\in V\cup m}\theta_j^{V\cup m}X_j+\epsilon^{V\cup m}|X_V\Big)\\ &= \operatorname{var}\Big(\sum_{j\in V\cup m}\theta_j^{V\cup m}X_j|X_V\Big)+\operatorname{var}(\epsilon^{V\cup m}|X_V)\\ &= \operatorname{var}\Big(\sum_{j\in V\cup m}\theta_j^{V\cup m}X_j|X_V\Big)+\operatorname{var}(Y|X_{V\cup m})\end{aligned}\tag{7.7}$$

as $\epsilon^{V\cup m}$ is independent of $X_{V\cup m}$. Now using the definition of ϵ_j^V in (7.2), it turns out that

$$\begin{aligned}\operatorname{var}\Big(\sum_{j\in V\cup m}\theta_j^{V\cup m}X_j|X_V\Big) &= \operatorname{var}\Big(\sum_{j\in m}\theta_j^{V\cup m}X_j|X_V\Big)\\ &= \operatorname{var}\Big(\sum_{j\in m}\theta_j^{V\cup m}\epsilon_j^V|X_V\Big)\\ &= \operatorname{var}\Big(\sum_{j\in m}\theta_j^{V\cup m}\epsilon_j^V\Big)\end{aligned}\tag{7.8}$$

as the $(\epsilon_j^V)_{j\in m}$ are independent of X_V. Gathering formulae (7.7) and (7.8), we get

$$\kappa_m^2=\frac{\operatorname{var}(Y|X_V)-\operatorname{var}(Y|X_{V\cup m})}{\operatorname{var}(Y|X_{V\cup m})}.\tag{7.9}$$

Under assumption $(H_{\mathcal{M}})$, $U_m\le N_m/10$ for all $m\in\mathcal{M}$ and $U\le N_m/21$. Hence, the terms U/N_m, U_m/N_m, k_m and $K_m(U)$ behave like constants and it follows from (7.6) that $\Delta'(m)\le\Delta(m)$ which completes the proof. $\square$

PROOF OF PROPOSITION 4.1. We first recall the classical upper bound for the binomial coefficient (see, for instance, (2.9) in [18]),

$$\log|\mathcal{M}(k,p)|=\log\binom{k}{p}\le k\log\Big(\frac{ep}{k}\Big).$$

As a consequence, $\log(1/\alpha_m)\le\log(1/\alpha)+k\log(\frac{ep}{k})$. Assumption (4.1) with $L=21$ therefore implies hypothesis $(H_{\mathcal{M}})$. Hence, we are in position to apply the second result of Theorem 3.3. Moreover, the assumption on n implies that $n\ge 21k$ and D_m/N_m is thus smaller than $1/20$ for any model m in $\mathcal{M}(k,p)$. Formula (3.5) in Theorem 3.3 then translates into

$$\begin{aligned}\triangle(m)\le\Bigg(&(1+\sqrt{0.05})L_1\Bigg(\sqrt{k^2\log\Big(\frac{ep}{k}\Big)}+\sqrt{k\log\Big(\frac{2}{\alpha\delta}\Big)}\Bigg)\\ &+1.1L_2\Big(k\log\Big(\frac{ep}{k}\Big)+\log\Big(\frac{2}{\alpha\delta}\Big)\Big)\Bigg)\Big/n,\end{aligned}$$

and it follows that Proposition 4.1 holds. $\square$

PROOF OF PROPOSITION 4.5. We fix the constant L in hypothesis (4.7) to be $21\log(4e) \vee C_2\log(4)$ where the universal constant C_2 is defined later in the proof. This choice of constants allows the procedure $[\sup_{1\leq i\leq p}\phi_{\{i\},\alpha/(2p)}]$ to satisfy hypothesis $(H_{\mathcal{M}})$. An argument similar to the proof of Proposition 4.1 allows to show easily that there exists a universal constant C such that if we set

$$\rho_1'^2 := \frac{C(\log(p)+\log(4/(\alpha\delta)))}{n} = \frac{C}{n}\log\left(\frac{4p}{\alpha\delta}\right), \tag{7.10}$$

then $\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq \rho_1'^2$ implies that $\mathbb{P}_\theta(T_\alpha > 0) \geq 1-\delta$. Here, the factor 4 in the logarithm comes from the fact that some weights α_m equal $\alpha/(2p)$.

Let ρ^2 and λ^2 be two positive numbers such that $\frac{\lambda^2}{\mathrm{var}(Y)-\lambda^2} = \rho^2$ and let $\theta \in \Theta[1,p]$ such that $\|\theta\|^2 = \lambda^2$. As $\mathrm{corr}(X_i, X_j) = c$ for any $i \neq j$, it follows that $\mathrm{var}(X_{p+1}) = c + \frac{1-c}{p}$ and $\mathrm{cov}(Y, X_{p+1})^2 = \|\theta\|^2[c+\frac{1-c}{p}]^2$:

$$\frac{\mathrm{var}(Y)-\mathrm{var}(Y|X_{p+1})}{\mathrm{var}(Y|X_{p+1})} = \frac{(c+(1-c)/p)\lambda^2}{\mathrm{var}(Y)-(c+(1-c)/p)\lambda^2}.$$

We now apply Theorem 3.3 to $\phi_{\{p+1\},\alpha/2}$ under hypothesis $(H_{\mathcal{M}})$. There exists a universal constant C_2 such that $\mathbb{P}_\theta(\phi_{\{p+1\},\alpha/2} > 0) \geq 1-\delta$ if

$$\frac{(c+(1-c)/p)\lambda^2}{\mathrm{var}(Y)-(c+(1-c)/p)\lambda^2} \geq \frac{C_2}{n}\log\left(\frac{4}{\alpha\delta}\right).$$

This last condition is implied by

$$\frac{c\lambda^2}{\mathrm{var}(Y)-c\lambda^2} \geq \frac{C_2}{n}\log\left(\frac{4}{\alpha\delta}\right)$$

which is equivalent to

$$\frac{\lambda^2}{\mathrm{var}(Y)} \geq \frac{C_2}{cn + cC_2\log(4/(\alpha\delta))}\log\left(\frac{4}{\alpha\delta}\right). \tag{7.11}$$

Let us assume that $c \geq \log(\frac{4}{\alpha\delta})/\log(\frac{4p}{\alpha\delta})$. As $n \geq 2C_2\log(\frac{4p}{\alpha\delta})$ (hypothesis (4.7) and definition of L), $nc \geq 2C_2\log(\frac{4}{\alpha\delta})$. As a consequence, condition (7.11) is implied by

$$\rho^2 \geq \frac{2C_2}{nc}\log\left(\frac{4}{\alpha\delta}\right). \tag{7.12}$$

Combining (7.10) and (7.12) allows us to conclude that $\mathbb{P}_\theta(T_\alpha > 0) \geq 1-\delta$ if

$$\rho^2 \geq \frac{L}{n}\left(\log\left(\frac{4p}{\alpha\delta}\right) \wedge \frac{1}{c}\log\left(\frac{4}{\alpha\delta}\right)\right).$$

$\square$

PROOF OF PROPOSITION 5.1. We fix the constant L to $42\log(80)$ in hypothesis (5.3). It follows that (5.3) implies

$$n \geq 42\left(\log\left(\frac{40}{\alpha}\right) \vee \log\left(\frac{2}{\delta}\right)\right). \tag{7.13}$$

First, we check that the test T_α satisfies condition $(H_{\mathcal{M}})$. As the dimension of each model is smaller than $n/2$, for any model m in $\mathcal{M}$, N_m is larger than $n/2$. Moreover, for any model m in $\mathcal{M}$, α_m is larger than $\alpha/(2|\mathcal{M}|)$ and $|\mathcal{M}|$ is smaller than $n/2$. As a consequence, the first condition of $(H_{\mathcal{M}})$ is implied by the inequality

$$n \geq 20\log\left(\frac{n}{\alpha}\right). \tag{7.14}$$

Hypothesis (7.13) implies that $n/2 \geq 20\log(\frac{40}{\alpha})$. Moreover, for any $n > 0$ it holds that $n/2 \geq 20\log(\frac{n}{40})$. Combining these two lower bounds enables to obtain (7.14). The second condition of $(H_{\mathcal{M}})$ holds if $n \geq 42\log(\frac{2}{\delta})$ which is a consequence of hypothesis (7.13).

We first consider the case $n < 2p$ and apply Theorem 3.3 under hypothesis $(H_{\mathcal{M}})$ to T_α. $\mathbb{P}_\theta(T_\alpha > 0) \geq 1 - \delta$ for all $\theta \in \mathbb{R}^p$ such that $\exists i \in \{1, \ldots, [n/2]\}$,

$$\begin{aligned} &\frac{\operatorname{var}(Y) - \operatorname{var}(Y|X_{m_i})}{\operatorname{var}(Y|X_{m_i})} \\ &\quad \geq C\frac{\sqrt{i\log(2[n/2]/(\alpha\delta))} + \log(2[n/2]/(\alpha\delta))}{n}, \end{aligned} \tag{7.15}$$

where C is an universal constant. Let θ be an element of $\mathcal{E}_a(R)$ that satisfies

$$\begin{aligned} \|\theta\|^2 &\geq (1+C)\big(\operatorname{var}(Y|X_{m_i}) - \operatorname{var}(Y|X)\big) \\ &\quad + (1+C)\operatorname{var}(Y|X)\frac{\sqrt{i\log(n/(\alpha\delta))} + \log(n/(\alpha\delta))}{n} \end{aligned}$$

for some $1 \leq i \leq [n/2]$. By hypothesis (5.3), it holds that

$$\frac{\sqrt{i\log(n/(\alpha\delta))} + \log(n/(\alpha\delta))}{n} \leq 1$$

for any i between 1 and $[n/2]$. It is then straightforward to check that θ satisfies (7.15). As θ belongs to the set $\mathcal{E}_a(R)$,

$$\begin{aligned} &\operatorname{var}(Y|X_{m_i}) - \operatorname{var}(Y|X) \\ &\quad = a_{i+1}^2\operatorname{var}(Y|X)\sum_{j=i+1}^{p}\frac{\operatorname{var}(Y|X_{m_{j-1}}) - \operatorname{var}(Y|X_{m_j})}{a_{i+1}^2\operatorname{var}(Y|X)} \\ &\quad \leq a_{i+1}^2\operatorname{var}(Y|X)R^2. \end{aligned}$$

Hence, if θ belongs to $\mathcal{E}_a(R)$ and satisfies

$$\|\theta\|^2 \geq (1+C)\operatorname{var}(Y|X)\left[\left(a_{i+1}^2 R^2 + \frac{\sqrt{i\log(n/(\alpha\delta))}}{n}\right) + \frac{1}{n}\log\left(\frac{n}{\alpha\delta}\right)\right],$$

then $\mathbb{P}_\theta(T_\alpha \leq 0) \leq \delta$. Gathering this condition for any i between 1 and $[n/2]$ allows us to conclude that if θ satisfies

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq (1+C)\left[\inf_{1\leq i\leq[n/2]}\left(a_{i+1}^2 R^2 + \frac{\sqrt{i\log(n/(\alpha\delta))}}{n}\right) + \frac{1}{n}\log\left(\frac{n}{\alpha\delta}\right)\right],$$

then $\mathbb{P}_\theta(T_\alpha \leq 0) \leq \delta$.

Let us now turn to the case $n \geq 2p$. Let us consider T_α as the supremum of $p-1$ tests of level $\alpha/2(p-1)$ and one test of level $\alpha/2$. By considering the $p-1$ first tests, we obtain as in the previous case that $\mathbb{P}_\theta(T_\alpha \leq 0) \leq \delta$ if

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq (1+C)\left[\inf_{1\leq i\leq(p-1)}\left(a_{i+1}^2 R^2 + \frac{\sqrt{i\log(p/(\alpha\delta))}}{n}\right) + \frac{1}{n}\log\left(\frac{p}{\alpha\delta}\right)\right].$$

On the other hand, using the last test statistic $\phi_{\mathcal{I},\alpha/2}$, $\mathbb{P}_\theta(T_\alpha \leq 0) \leq \delta$, if

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq C\frac{\sqrt{p\log(2/(\alpha\delta))} + \log(2/(\alpha\delta))}{n}.$$

Gathering these two conditions we prove (5.5). $\square$

PROOF OF PROPOSITION 5.2. The approach behind this proof is similar to the one for Proposition 5.1. We fix the constant L in assumption (5.6), as in the previous proof. Hence, the collection of models $\mathcal{M}$ and the weights α_m satisfy hypothesis $(H_{\mathcal{M}})$ as in the previous proof.

Let us give a sharper upper bound on $|\mathcal{M}|$:

$$|\mathcal{M}| \leq 1 + \log(n/2 \wedge p)/\log(2) \leq \log(n \wedge 2p)/\log(2). \tag{7.16}$$

We deduce from (7.16) that there exists a constant $L(\alpha,\delta)$ only depending on α and δ such that for all $m \in \mathcal{M}$,

$$\log\left(\frac{1}{\alpha_m\delta}\right) \leq L(\alpha,\delta)\log\log(n \wedge p).$$

First, let us consider the case $n < 2p$. We apply Theorem 3.3 under assumption $(H_{\mathcal{M}})$. As in the proof of Proposition 5.1, we obtain that $\mathbb{P}_\theta(T_\alpha > 0) \geq 1-\delta$

if

$$\frac{\|\theta\|^2}{\operatorname{var}(Y)-\|\theta\|^2} \ge L(\alpha,\delta)\left[\inf_{i\in\{2^j,j\ge 0\}\cap\{1,\dots,[n/2]\}}\left(R^2(i+1)^{-2s}+\frac{\sqrt{i\log\log n}}{n}\right)+\frac{\log\log n}{n}\right].$$

It is worth noting that $R^2 i^{-2s} \le \frac{\sqrt{i\log\log n}}{n}$ if and only if

$$i \ge i^* = \left(\frac{R^2 n}{\sqrt{\log\log n}}\right)^{2/(1+4s)}.$$

Under the assumption on R, i^* is larger than one. Let us distinguish between two cases. If there exists i' in $\{2^j, j\ge 0\}\cap\{1,\dots,[n/2]\}$ such that $i^* \le i'$, one can take $i' \le 2i^*$ and then

$$\begin{aligned}\inf_{i\in\{2^j,j\ge 0\}\cap\{1,\dots,[n/2]\}}&\left(R^2 i^{-2s}+\frac{\sqrt{i\log\log n}}{n}\right)\\ &\le 2\frac{\sqrt{i'\log\log n}}{n}\\ &\le 2\sqrt{2}R^{2/(1+4s)}\left(\frac{\sqrt{\log\log n}}{n}\right)^{4s/(1+4s)}.\end{aligned}\tag{7.17}$$

Else, we take $i' \in \{2^j, j\ge 0\}\cap\{1,\dots,[n/2]\}$ such that $n/4 \le i' \le n/2$. Since $i' \le (i^* \wedge n/2)$ we obtain that

$$\begin{aligned}\inf_{i\in\{2^j,j\ge 0\}\cap\{1,\dots,[n/2]\}}&\left(R^2 i^{-2s}+\frac{\sqrt{i\log\log n}}{n}\right)\\ &\le 2R^2 i'^{-2s} \le 2R^2\left(\frac{n}{2}\right)^{-2s}.\end{aligned}\tag{7.18}$$

Gathering inequalities (7.17) and (7.18) we prove (5.7).

We now turn to the case $n \ge 2p$. As in the proof of Proposition 5.1, we divide the proof into two parts: first we give an upper bound of the power for the $|\mathcal{M}|-1$ first tests which define T_α, and then we give an upper bound for the last test $\phi_{\mathcal{I},\alpha/2}$. Combining these two inequalities allows us to prove (5.8). $\square$

PROOF OF PROPOSITION 5.4. We fix the constant L in the assumption as in the two previous proofs. We first note that the assumption on R^2 implies that $D^* \ge 2$. As N_m is larger than $n/2$, the $\phi_{m_{D^*}}$ test clearly satisfies condition $(H_{\mathcal{M}})$. As a consequence, we may apply Theorem 3.3. Hence, $\mathbb{P}_\theta(T_\alpha^* \le 0) \le \delta$ for any θ such that

$$\frac{\operatorname{var}(Y)-\operatorname{var}(Y|X_{m_{D^*}})}{\operatorname{var}(Y|X_{m_{D^*}})} \ge L(\alpha,\delta)\frac{\sqrt{D^*}}{n}.\tag{7.19}$$

Now, we use the same sketch as in the proof of Proposition 5.1. For any $\theta \in \mathcal{E}_a(R)$, condition (7.19) is equivalent to

$$\begin{aligned}\|\theta\|^2 \geq \big(\operatorname{var}(Y|X_{m_{D^*}}) - \operatorname{var}(Y|X)\big)\left(1 + L(\alpha,\delta)\frac{\sqrt{D^*}}{n}\right)\\ + \operatorname{var}(Y|X)L(\alpha,\delta)\frac{\sqrt{D^*}}{n}.\end{aligned} \tag{7.20}$$

Moreover, as θ belongs to $\mathcal{E}_a(R)$,

$$\operatorname{var}(Y|X_{m_{D^*}}) - \operatorname{var}(Y|X) \leq a_{D^*+1}^2 R^2 \operatorname{var}(Y|X) \leq a_{D^*}^2 \operatorname{var}(Y|X) R^2.$$

As $\sqrt{D^*}/n$ is smaller than one, condition (7.20) is implied by

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq \big(1 + L(\alpha,\delta)\big)\left(a_{D^*}^2 R^2 + \frac{\sqrt{D^*}}{n}\right).$$

As $a_{D^*}^2 R^2$ is smaller than $\frac{\sqrt{D^*}}{n}$ which is smaller $\sup_{1\leq i\leq p}[\frac{\sqrt{i}}{n} \wedge a_i^2 R^2]$, it turns out that $\mathbb{P}_\theta(T_\alpha^* = 0) \leq \delta$ for any θ belonging to $\mathcal{E}_a(R)$ such that

$$\frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq 2\big(1 + L(\alpha,\delta)\big) \sup_{1\leq i\leq p}\left[\frac{\sqrt{i}}{n} \wedge a_i^2 R^2\right].$$

□

8. Proofs of Theorem 4.3, Propositions 3.4, 4.2, 4.4, 4.6, 5.3, 5.5 and 5.6. Throughout this section, we shall use the notations $\eta := 2(1-\alpha-\delta)$ and $\mathcal{L}(\eta) := \frac{\log(1+2\eta^2)}{2}$.

PROOF OF THEOREM 4.3. This proof follows the general method for obtaining lower bounds described in Section 7.1 in Baraud [2]. We first remind the reader of the main arguments of the approach applied to our model. Let ρ be some positive number and μ_ρ be some probability measure on

$$\Theta[k,p,\rho] := \left\{\theta \in \Theta[k,p], \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} = \rho\right\}.$$

We define $\mathbb{P}_{\mu_\rho} = \int \mathbb{P}_\theta \, d\mu_\rho(\theta)$ and Φ_α the set of level-α tests of hypothesis "$\theta = 0$." Then

$$\begin{aligned}\beta_I(\Theta[k,p,\rho]) &\geq \inf_{\phi_\alpha\in\Phi_\alpha} \mathbb{P}_{\mu_\rho}[\phi_\alpha = 0]\\ &\geq 1 - \alpha - \sup_{A,\mathbb{P}_0(A)\leq\alpha} |\mathbb{P}_{\mu_\rho}(A) - \mathbb{P}_0(A)|\\ &\geq 1 - \alpha - \tfrac{1}{2}\|\mathbb{P}_{\mu_\rho} - \mathbb{P}_0\|_{\mathrm{TV}},\end{aligned} \tag{8.1}$$

where $\|\mathbb{P}_{\mu_\rho} - \mathbb{P}_0\|_{\mathrm{TV}}$ denotes the total variation norm between the probabilities $\mathbb{P}_{\mu_\rho}$ and $\mathbb{P}_0$. If we suppose that $\mathbb{P}_{\mu_\rho}$ is absolutely continuous with respect to $\mathbb{P}_0$, we can upper bound the norm in total variation between these two probabilities as follows. We define

$$L_{\mu_\rho}(\mathbf{Y},\mathbf{X}) := \frac{d\mathbb{P}_{\mu_\rho}}{d\mathbb{P}_0}(\mathbf{Y},\mathbf{X}).$$

Then we get the upper bound

$$\begin{aligned}\|\mathbb{P}_{\mu_\rho} - \mathbb{P}_0\|_{\mathrm{TV}} &= \int |L_{\mu_\rho}(\mathbf{Y},\mathbf{X}) - 1|\, d\mathbb{P}_0(\mathbf{Y},\mathbf{X}) \\ &\le \big(\mathbb{E}_0[L^2_{\mu_\rho}(\mathbf{Y},\mathbf{X})] - 1\big)^{1/2}.\end{aligned}$$

Thus, we deduce from (8.1) that

$$\beta_I(\Theta[k,p,\rho]) \ge 1 - \alpha - \tfrac{1}{2}\big(\mathbb{E}_0[L^2_{\mu_\rho}(\mathbf{Y},\mathbf{X})] - 1\big)^{1/2}.$$

If we find a number $\rho^* = \rho^*(\eta)$ such that

$$\log\big(\mathbb{E}_0[L^2_{\mu_{\rho^*}}(\mathbf{Y},\mathbf{X})]\big) \le \mathcal{L}(\eta), \tag{8.2}$$

then for any $\rho \le \rho^*$,

$$\beta_I(\Theta[k,p,\rho]) \ge 1 - \alpha - \frac{\eta}{2} = \delta.$$

To apply this method, we first have to define a suitable prior μ_ρ on $\Theta[k,p,\rho]$. Let $\widehat{m}$ be some random variable uniformly distributed over $\mathcal{M}(k,p)$ and for each $m \in \mathcal{M}(k,p)$, let $\epsilon^m = (\epsilon^m_j)_{j\in m}$ be a sequence of independent Rademacher random variables. We assume that for all $m \in \mathcal{M}(k,p)$, ϵ^m and $\widehat{m}$ are independent. Let ρ be given and μ_ρ be the distribution of the random variable $\widehat{\theta} = \sum_{j\in\widehat{m}} \lambda\epsilon^{\widehat{m}}_j e_j$ where

$$\lambda^2 := \frac{\mathrm{var}(Y)\rho^2}{k(1+\rho^2)}$$

and where $(e_j)_{j\in\mathcal{I}}$ is the orthonormal family of vectors of $\mathbb{R}^p$ defined by

$$(e_j)_i = 1 \qquad \text{if } i = j \quad \text{and} \quad (e_i)_j = 0 \qquad \text{otherwise}.$$

Straightforwardly, μ_ρ is supported by $\Theta[k,p,\rho]$. For any m in $\mathcal{M}(k,p)$ and any vector $(\zeta^m_j)_{j\in m}$ with values in $\{-1;1\}$, let $\mu_{m,\zeta^m,\rho}$ be the Dirac measure on $\sum_{j\in m}\lambda\zeta^m_j e_j$. For any m in $\mathcal{M}(k,p)$, $\mu_{m,\rho}$ denotes the distribution of the random variable $\sum_{j\in m}\lambda\zeta^m_j e_j$ where (ζ^m_j) is a sequence of independent Rademacher random variables. These definitions easily imply

$$\begin{aligned}L_{\mu_\rho}(\mathbf{Y},\mathbf{X}) &= \frac{1}{\binom{p}{k}} \sum_{m\in\mathcal{M}(k,p)} L_{\mu_{m,\rho}}(\mathbf{Y},\mathbf{X}) \\ &= \frac{1}{2^k\binom{p}{k}} \sum_{m\in\mathcal{M}(k,p)} \sum_{\zeta^m\in\{-1,1\}^k} L_{\mu_{m,\zeta^m\rho}}(\mathbf{Y},\mathbf{X}).\end{aligned}$$

We aim at bounding the quantity $\mathbb{E}_0(L^2_{\mu_\rho})$ and obtaining an inequality of the form (8.2). First, we work out $L_{\mu_{m,\zeta^m,\rho}}$:

$$
\begin{aligned}
&L_{\mu_{m,\zeta^m,\rho}}(\mathbf{Y},\mathbf{X})\\
&\quad=\Bigg[\bigg(\frac{1}{1-\lambda^2k/(\mathrm{var}(Y))}\bigg)^{n/2}\exp\bigg(-\frac{\|\mathbf{Y}\|_n^2}{2}\frac{\lambda^2k}{\mathrm{var}(Y)(\mathrm{var}(Y)-\lambda^2k)}\\
&\qquad\qquad+\lambda\sum_{j\in m}\zeta_j^m\frac{\langle\mathbf{Y},\mathbf{X}_j\rangle_n}{\mathrm{var}(Y)-\lambda^2k}\\
&\qquad\qquad-\lambda^2\sum_{j,j'\in m}\zeta_j^m\zeta_{j'}^m\frac{\langle\mathbf{X}_j,\mathbf{X}_{j'}\rangle_n}{2(\mathrm{var}(Y)-\lambda^2k)}\bigg)\Bigg],
\end{aligned}\tag{8.3}
$$

where $\langle\cdot\rangle_n$ refers to the canonical inner product in $\mathbb{R}^n$.

Let us fix m_1 and m_2 in $\mathcal{M}(k,p)$ and two vectors ζ^1 and ζ^2, respectively, associated to m_1 and m_2. We aim at computing the quantity $\mathbb{E}_0(L_{\mu_{m_1,\zeta^1,\rho}}(\mathbf{Y},\mathbf{X})\times L_{\mu_{m_2,\zeta^2,\rho}}(\mathbf{Y},\mathbf{X}))$. First, we decompose the set $m_1\cup m_2$ into four sets (which possibly are empty): $m_1\setminus m_2$, $m_2\setminus m_1$, m_3 and m_4 where m_3 and m_4 are defined by

$$
\begin{aligned}
m_3&:=\{j\in m_1\cap m_2|\zeta_j^1=\zeta_j^2\},\\
m_4&:=\{j\in m_1\cap m_2|\zeta_j^1=-\zeta_j^2\}.
\end{aligned}
$$

For the sake of simplicity, we reorder the elements of $m_1\cup m_2$ from 1 to $|m_1\cup m_2|$ such that the first elements belong to $m_1\setminus m_2$, then to $m_2\setminus m_1$ and so on. Moreover, we define the vector $\zeta\in\mathbb{R}^{|m_1\cup m_2|}$ such that $\zeta_j=\zeta_j^1$ if $j\in m_1$ and $\zeta_j=\zeta_j^2$ if $j\in m_2\setminus m_1$. Using this notation, we compute the expectation of $L_{m_1,\zeta^1,\rho}(\mathbf{Y},\mathbf{X})L_{m_2,\zeta^2,\rho}(\mathbf{Y},\mathbf{X})$:

$$
\begin{aligned}
&\mathbb{E}_0(L_{\mu_{m_1,\zeta^1,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{m_2,\zeta^2,\rho}}(\mathbf{Y},\mathbf{X}))\\
&\qquad=\bigg(\frac{1}{\mathrm{var}(Y)(1-\lambda^2k/(\mathrm{var}(Y)))^2}\bigg)^{n/2}|A|^{-n/2},
\end{aligned}\tag{8.4}
$$

where $|\cdot|$ refers to the determinant and A is a symmetric square matrix of size $|m_1\cup m_2|+1$ such that

$$
A[1,j]:=\begin{cases}
\dfrac{\mathrm{var}(Y)+\lambda^2k}{\mathrm{var}(Y)(\mathrm{var}(Y)-\lambda^2k)}, & \text{if } j=1,\\
-\dfrac{\lambda\zeta_{j-1}}{\mathrm{var}(Y)-\lambda^2k}, & \text{if } (j-1)\in m_1\triangle m_2,\\
-2\dfrac{\lambda\zeta_{j-1}}{\mathrm{var}(Y)-\lambda^2k}, & \text{if } (j-1)\in m_3,\\
0, & \text{if } (j-1)\in m_4,
\end{cases}
$$

where $m_1 \triangle m_2$ refers to $(m_1 \cup m_2) \setminus (m_1 \cap m_2)$. For any $i > 1$ and $j > 1$, A satisfies

$$A[i,j] := \begin{cases} \lambda^2 \dfrac{\zeta_{i-1}\zeta_{j-1}}{\operatorname{var}(Y) - \lambda^2 k} + \delta_{i,j}, & \text{if } (i-1, j-1) \in (m_1 \setminus m_2) \times m_1, \\ \lambda^2 \dfrac{\zeta_{i-1}\zeta_{j-1}}{\operatorname{var}(Y) - \lambda^2 k} + \delta_{i,j}, & \text{if } (i-1, j-1) \in (m_2 \setminus m_1) \\ & \qquad \times (m_2 \setminus m_1 \cup m_3), \\ -\lambda^2 \dfrac{\zeta_{i-1}\zeta_{j-1}}{\operatorname{var}(Y) - \lambda^2 k}, & \text{if } (i-1, j-1) \in (m_2 \setminus m_1) \times m_4, \\ 2\lambda^2 \dfrac{\zeta_{i-1}\zeta_{j-1}}{\operatorname{var}(Y) - \lambda^2 k} + \delta_{i,j}, & \text{if } (i-1, j-1) \in [m_3 \times m_3] \\ & \qquad \cup [m_4 \times m_4], \\ 0, & \text{else}, \end{cases}$$

where $\delta_{i,j}$ is the indicator function of $i = j$.

After some linear transformation on the lines of the matrix A, it is possible to express its determinant into

$$|A| = \frac{\operatorname{var}(Y) + \lambda^2 k}{\operatorname{var}(Y)(\operatorname{var}(Y) - \lambda^2 k)} |I_{|m_1 \cup m_2|} + C|,$$

where $I_{|m_1 \cup m_2|}$ is the identity matrix of size $|m_1 \cup m_2|$. C is a symmetric matrix of size $|m_1 \cup m_2|$ such that for any (i, j),

$$C[i,j] = \zeta_i \zeta_j D[i,j]$$

and D is a block symmetric matrix defined by

$$D := \begin{bmatrix} \frac{\lambda^4 k}{\operatorname{var}^2(Y) - \lambda^4 k^2} & \frac{-\lambda^2 \operatorname{var}(Y)}{\operatorname{var}^2(Y) - \lambda^4 k^2} & \frac{-\lambda^2}{\operatorname{var}(Y) + \lambda^2 k} & \frac{\lambda^2}{\operatorname{var}(Y) - \lambda^2 k} \\ \frac{-\lambda^2 \operatorname{var}(Y)}{\operatorname{var}^2(Y) - \lambda^4 k^2} & \frac{\lambda^4 k}{\operatorname{var}^2(Y) - \lambda^4 k^2} & \frac{-\lambda^2}{\operatorname{var}(Y) + \lambda^2 k} & \frac{-\lambda^2}{\operatorname{var}(Y) - \lambda^2 k} \\ \frac{-\lambda^2}{\operatorname{var}(Y) + \lambda^2 k} & \frac{-\lambda^2}{\operatorname{var}(Y) + \lambda^2 k} & \frac{-2\lambda^2}{\operatorname{var}(Y) + \lambda^2 k} & 0 \\ \frac{\lambda^2}{\operatorname{var}(Y) - \lambda^2 k} & \frac{-\lambda^2}{\operatorname{var}(Y) - \lambda^2 k} & 0 & \frac{2\lambda^2}{\operatorname{var}(Y) - \lambda^2 k} \end{bmatrix}.$$

Each block corresponds to one of the four previously defined subsets of $m_1 \cup m_2$ (i.e., $m_1 \setminus m_2$, $m_2 \setminus m_1$, m_3 and m_4). The matrix D is of rank, at most, four. By computing its nonzero eigenvalues, it is then straightforward to derive the determinant of A,

$$|A| = \frac{[\operatorname{var}(Y) - \lambda^2(2|m_3| - |m_1 \cap m_2|)]^2}{\operatorname{var}(Y)(\operatorname{var}(Y) - \lambda^2 k)^2}.$$

Gathering this equality with (8.4) yields

$$\begin{aligned} &\mathbb{E}_0(L_{\mu_{m_1,\zeta^1,\rho}}(\mathbf{Y}, \mathbf{X}) L_{\mu_{m_2,\zeta^2,\rho}}(\mathbf{Y}, \mathbf{X})) \\ &\quad = \left[\frac{1}{1 - \lambda^2(2|m_3| - |m_1 \cap m_2|)/(\operatorname{var}(Y))}\right]^n. \end{aligned} \tag{8.5}$$

Then, we take the expectation with respect to ζ^1, ζ^2, m_1 and m_2. When m_1 and m_2 are fixed the expression (8.5) depends on ζ^1 and ζ^2 only toward the cardinality of m_3. As ζ^1 and ζ^2 correspond to independent Rademacher variables, the random variable $2|m_3| - |m_1 \cap m_2|$ follows the distribution of Z, a sum of $|m_1 \cap m_2|$ independent Rademacher variables and

$$\mathbb{E}_0(L_{\mu_{m_1,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{m_2,\rho}}(\mathbf{Y},\mathbf{X})) = \mathbb{E}_0\left[\frac{1}{1-\lambda^2 Z/\mathrm{var}(Y)}\right]^n. \tag{8.6}$$

When Z is nonpositive, this expression is smaller than one. Alternatively, when Z is nonnegative,

$$\begin{aligned}\left[\frac{1}{1-\lambda^2 Z/\mathrm{var}(Y)}\right]^n &= \exp\left(n\log\left(\frac{1}{1-\lambda^2 Z/\mathrm{var}(Y)}\right)\right)\\ &\le \exp\left[n\frac{\lambda^2 Z/\mathrm{var}(Y)}{1-\lambda^2 Z/\mathrm{var}(Y)}\right]\\ &\le \exp\left[n\frac{\lambda^2 Z/\mathrm{var}(Y)}{1-\lambda^2 k/\mathrm{var}(Y)}\right],\end{aligned}$$

as $\log(1+x) \le x$ and as Z is smaller than k. We define an event $\mathbb{A}$ such that $\{Z > 0\} \subset \mathbb{A} \subset \{Z \ge 0\}$, and $\mathbb{P}(\mathbb{A}) = \frac{1}{2}$. This is always possible as the random variable Z is symmetric. As a consequence, on the event $\mathbb{A}^c$, the quantity (8.6) is smaller or equal to one. All in all, we bound (8.6) by

$$\begin{aligned}&\mathbb{E}_0(L_{\mu_{m_1,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{m_2,\rho}}(\mathbf{Y},\mathbf{X}))\\ &\quad\le \frac{1}{2} + \mathbb{E}_0\left[\mathbf{1}_{\mathbb{A}}\exp\left[n\frac{\lambda^2 Z/\mathrm{var}(Y)}{1-\lambda^2 k/\mathrm{var}(Y)}\right]\right],\end{aligned} \tag{8.7}$$

where $\mathbf{1}_{\mathbb{A}}$ is the indicator function of the event $\mathbb{A}$. We now apply Hölder's inequality with a parameter $v \in]0;1]$ which will be fixed later:

$$\begin{aligned}&\mathbb{E}_0\left[\mathbf{1}_{\mathbb{A}}\exp\left[n\frac{\lambda^2 Z/\mathrm{var}(Y)}{1-\lambda^2 k/\mathrm{var}(Y)}\right]\right]\\ &\quad\le \mathbb{P}(\mathbb{A})^{1-v}\left[\mathbb{E}_0\exp\left(\frac{n}{v}\frac{\lambda^2 Z/\mathrm{var}(Y)}{1-\lambda^2 k/\mathrm{var}(Y)}\right)\right]^v\\ &\quad\le \left(\frac{1}{2}\right)^{1-v}\left[\cosh\left(\frac{n\lambda^2}{v(\mathrm{var}(Y)-\lambda^2 k)}\right)\right]^{|m_1\cap m_2|v}.\end{aligned} \tag{8.8}$$

Gathering inequalities (8.7) and (8.8) yields

$$\begin{aligned}&\mathbb{E}_0[L^2_{\mu_\rho}(\mathbf{Y},\mathbf{X})]\\ &\quad\le \frac{1}{2} + \left(\frac{1}{2}\right)^{1-v}\frac{1}{\binom{p}{k}^2}\sum_{m_1,m_2\in\mathcal{M}(k,p)}\cosh\left(\frac{n\lambda^2}{v(\mathrm{var}(Y)-\lambda^2 k)}\right)^{|m_1\cap m_2|v}.\end{aligned}$$

Following the approach of Baraud [2] in Section 7.2, we note that if m_1 and m_2 are taken uniformly and independently in $\mathcal{M}(k, p)$, then $|m_1 \cap m_2|$ is distributed as a hypergeometric distribution with parameters p, k and k/p. Thus we derive that

$$\mathbb{E}_0[L^2_{\mu_\rho}(\mathbf{Y}, \mathbf{X})] \le \frac{1}{2} + \left(\frac{1}{2}\right)^{1-v} \mathbb{E}\left(\cosh\left(\frac{n\lambda^2}{v(\mathrm{var}(Y) - \lambda^2 k)}\right)^{vT}\right), \tag{8.9}$$

where T is a random variable distributed according to a hypergeometric distribution with parameters p, k and k/p. We know from Aldous (see [1], page 173) that T has the same distribution as the random variable $\mathbb{E}(W|\mathcal{B}_p)$ where W is a binomial random variable of parameters k, k/p and $\mathcal{B}_p$, some suitable σ-algebra. By a convexity argument, we then upper bound (8.9):

$$\begin{aligned}
&\mathbb{E}_0[L^2_{\mu_\rho}(\mathbf{Y}, \mathbf{X})] \\
&\quad \le \frac{1}{2} + \left(\frac{1}{2}\right)^{1-v} \mathbb{E}\left(\cosh\left(\frac{n\lambda^2}{v(\mathrm{var}(Y) - \lambda^2 k)}\right)^{vW}\right) \\
&\quad = \frac{1}{2} + \left(\frac{1}{2}\right)^{1-v} \left(1 + \frac{k}{p}\left(\cosh\left(\frac{n\lambda^2}{v(\mathrm{var}(Y) - \lambda^2 k)}\right)^{v} - 1\right)\right)^k \\
&\quad = \frac{1}{2} + \left(\frac{1}{2}\right)^{1-v} \exp\left[k \log\left(1 + \frac{k}{p}\left(\cosh\left(\frac{n\lambda^2}{v(\mathrm{var}(Y) - \lambda^2 k)}\right)^{v} - 1\right)\right)\right].
\end{aligned}$$

To get the upper bound on the total variation distance appearing in (8.1), we aim at constraining this last expression to be smaller than $1 + \eta^2$. This is equivalent to the following inequality:

$$\begin{aligned}
&2^v \exp\left[k \log\left(1 + \frac{k}{p}\left(\cosh\left(\frac{n\lambda^2 k}{vk(\mathrm{var}(Y) - \lambda^2 k)}\right)^{v} - 1\right)\right)\right] \\
&\qquad \le 1 + 2\eta^2.
\end{aligned} \tag{8.10}$$

We now choose $v = \frac{\mathcal{L}(\eta)}{\log(2)} \wedge 1$. If v is strictly smaller than one, then (8.10) is equivalent to

$$k \log\left[1 + \frac{k}{p}\left(\cosh\left(\frac{n\lambda^2 k}{vk(\mathrm{var}(Y) - \lambda^2 k)}\right)^{v} - 1\right)\right] \le \frac{\log(1 + 2\eta^2)}{2}. \tag{8.11}$$

It is straightforward to show that this last inequality also implies (8.10) if v equals one. We now suppose that

$$\frac{n\lambda^2}{v(\mathrm{var}(Y) - \lambda^2 k)} \le \log\left((1 + u)^{1/v} + \sqrt{(1 + u)^{2/v} - 1}\right), \tag{8.12}$$

where $u = \frac{p\mathcal{L}(\eta)}{k^2}$. Using the classical equality $\cosh[\log(1 + x + \sqrt{2x + x^2})] = 1 + x$ with $x = (1 + u)^{1/v} - 1$, we deduce that inequality (8.12) implies (8.11) because

$$k \log\left(1+\frac{k}{p}\left(\cosh\left(\frac{n\lambda^2 k}{vk(\mathrm{var}(Y)-\lambda^2 k)}\right)^v - 1\right)\right) \le k \log\left(1+\frac{k}{p}u\right)$$
$$\le \frac{k^2}{p}u \le \mathcal{L}(\eta).$$

For any $\beta \ge 1$ and any $x > 0$, it holds that $(1+x)^\beta \ge 1+\beta x$. As $\frac{1}{v} \ge 1$, condition (8.12) is implied by

$$\frac{\lambda^2 k}{\mathrm{var}(Y)-\lambda^2 k} \le \frac{kv}{n}\log\left(1+\frac{u}{v}+\sqrt{\frac{2u}{v}}\right).$$

One then combines the previous inequality with the definitions of u and v to obtain the upper bound

$$\frac{\lambda^2 k}{\mathrm{var}(Y)-\lambda^2 k}$$
$$\le \frac{k}{n}\left(\frac{\mathcal{L}(\eta)}{\log(2)} \wedge 1\right)\log\left(1+\frac{p(\log(2)\vee\mathcal{L}(\eta))}{k^2}+\sqrt{\frac{2p(\log(2)\vee\mathcal{L}(\eta))}{k^2}}\right).$$

For any x positive and any u between 0 and 1, $\log(1+ux) \ge u\log(1+x)$. As a consequence, the previous inequality is implied by

$$\frac{\lambda^2 k}{\mathrm{var}(Y)-\lambda^2 k} \le \frac{k}{n}\left(\frac{\mathcal{L}(\eta)}{\log(2)} \wedge 1\right)\big([\mathcal{L}(\eta)\vee\log(2)]\wedge 1\big)\log\left(1+\frac{p}{k^2}+\sqrt{\frac{2p}{k^2}}\right)$$
$$= \frac{k}{n}\big(\mathcal{L}(\eta)\wedge 1\big)\log\left(1+\frac{p}{k^2}+\sqrt{\frac{2p}{k^2}}\right).$$

To resume, if we take ρ^2 smaller than (4.4), then

$$\beta_I(\Theta[k,p,\rho]) \ge \delta.$$

Moreover, the lower bound is strict if ρ^2 is strictly smaller than (4.4). To prove the second part of the theorem, one has to observe that $\alpha+\delta \le 53\%$ implies that $\mathcal{L}(\eta) \ge \frac{1}{2}$. $\square$

PROOF OF PROPOSITION 4.2. Let us first assume that the covariance matrix of X is the identity. We argue as in the proof of Theorem 4.3 taking $k = p$. The sketch of the proof remains unchanged except that we slightly modify the last part. Inequality (8.11) becomes

$$pv \log\left(\cosh\left(\frac{n\lambda^2 p}{vp(\mathrm{var}(Y)-\lambda^2 p)}\right)\right) \le \mathcal{L}(\eta),$$

where we recall that $v = \frac{\mathcal{L}(\eta)}{\log 2} \wedge 1$. For all $x \in \mathbb{R}$, $\cosh(x) \leq \exp(x^2/2)$. Consequently, the previous inequality is implied by

$$\frac{\lambda^2 p}{\mathrm{var}(Y) - \lambda^2 p} \leq \sqrt{2v\mathcal{L}(\eta)}\frac{\sqrt{p}}{n},$$

and the result follows easily.

If we no longer assume that the covariance matrix Σ is the identity, we orthogonalize the sequence X_i thanks to Gram–Schmidt process. Applying the previous argument to this new sequence of covariates completes the proof. $\square$

PROOF OF PROPOSITION 3.4. Let us define the constant $L(\alpha, \delta)$ involved in the condition:

$$L(\alpha,\delta) := \sqrt{\log\bigl(1+8(1-\alpha-\delta)^2\bigr)}\Bigl[1 \wedge \sqrt{\log\bigl(1+8(1-\alpha-\delta)^2\bigr)/(2\log 2)}\Bigr].$$

Let us apply Proposition 4.2. For any $\rho \leq L(\alpha,\delta)\frac{\sqrt{D_m}}{n}$ and any $\varsigma > 0$ there exists some $\theta \in S_m$ such that $\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} = \rho^2$ and $\mathbb{P}_\theta(\phi_{m,\alpha} \leq 0) \geq \delta - \varsigma$. In the proof of Theorem 3.3, we have shown in (7.3) and following equalities that the distribution of the test statistic ϕ_m only depends on the quantity $\kappa_m^2 = \frac{\mathrm{var}(Y)-\mathrm{var}(Y|X_m)}{\mathrm{var}(Y|X_m)}$. Let θ' be an element of S_m such that $\kappa_m^2 = \rho^2$. The distribution of ϕ_m under $\mathbb{P}_{\theta'}$ is the same as its distribution under $\mathbb{P}_\theta$, and therefore

$$\mathbb{P}_{\theta'}(\phi_{m,\alpha} \leq 0) \geq \delta - \varsigma.$$

Letting ς go to 0 completes the proof. $\square$

PROOF OF PROPOSITION 4.4. This lower bound for dependent Gaussian covariates is proved through the same approach as Theorem 4.3. We define the measure μ_ρ as in that proof. Under hypothesis (H_0), Y is independent of X. We note Σ the covariance matrix of X and $\mathbb{E}_{0,\Sigma}$ stands for the distribution of $(\mathbf{Y},\mathbf{X})$ under (H_0) in order to emphasize the dependence on Σ.

First, one has to upper bound the quantity $\mathbb{E}_{0,\Sigma}[L^2_{\mu_\rho}(\mathbf{Y},\mathbf{X})]$. For the sake of simplicity, we make the hypothesis that every covariate X_j has variance 1. If this is not the case, we only have to rescale these variables. The quantity $\mathrm{corr}(i,j)$ refers to the correlation between X_i and X_j. As we only consider the case $k=1$, the set of models m in $\mathcal{M}(1,p)$ is in correspondence with the set $\{1,\dots,p\}$:

$$\mathbb{E}_{0,\Sigma}(L_{\mu_{i,\zeta^1,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{j,\zeta^2,\rho}}(\mathbf{Y},\mathbf{X})) = \left(\frac{\mathrm{var}(Y)}{\mathrm{var}(Y)-\mathrm{corr}(i,j)\lambda^2\zeta^1\zeta^2}\right)^n.$$

When i and j are fixed, we upper bound the expectation of this quantity with respect to ζ^1 and ζ^2 by

$$\mathbb{E}_{0,\Sigma}(L_{\mu_{i,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{j,\rho}}(\mathbf{Y},\mathbf{X})) \leq \frac{1}{2} + \frac{1}{2}\left(\frac{\mathrm{var}(Y)}{\mathrm{var}(Y)-|\mathrm{corr}(i,j)|\lambda^2}\right)^n. \tag{8.13}$$

If $i \neq j$, $|\operatorname{corr}(i, j)|$ is smaller than c, and if $i = j$, $\operatorname{corr}(i, j)$ is exactly one. As a consequence, taking the expectation of (8.13) with respect to i and j yields the upper bound

$$\begin{aligned}(8.14)\qquad &\mathbb{E}_{0,\Sigma}(L^2_{\mu_\rho}(\mathbf{Y}, \mathbf{X})) \\ &\quad \leq \frac{1}{2} + \frac{1}{2}\left(\frac{1}{p}\left(\frac{\operatorname{var}(Y)}{\operatorname{var}(Y) - \lambda^2}\right)^n + \frac{p-1}{p}\left(\frac{\operatorname{var}(Y)}{\operatorname{var}(Y) - c\lambda^2}\right)^n\right).\end{aligned}$$

Recall that we want to constrain this quantity (8.14) to be smaller than $1 + \eta^2$. In particular, this holds if the two following inequalities hold:

$$\frac{1}{p}\left(\frac{\operatorname{var}(Y)}{\operatorname{var}(Y) - \lambda^2}\right)^n \leq \frac{1}{p} + \eta^2, \tag{8.15}$$

$$\frac{p-1}{p}\left(\frac{\operatorname{var}(Y)}{\operatorname{var}(Y) - c\lambda^2}\right)^n \leq \frac{p-1}{p} + \eta^2. \tag{8.16}$$

One then uses the inequality $\log(\frac{1}{1-x}) \leq \frac{x}{1-x}$ which holds for any positive x smaller than one. Condition (8.15) holds if

$$\frac{\lambda^2}{\operatorname{var}(Y) - \lambda^2} \leq \frac{1}{n}\log(1 + p\eta^2), \tag{8.17}$$

whereas condition (8.16) is implied by

$$\frac{c\lambda^2}{\operatorname{var}(Y) - c\lambda^2} \leq \frac{1}{n}\log\left(1 + \frac{p}{p-1}\eta^2\right).$$

As c is smaller than one and $\frac{p}{p-1}$ is larger than 1, this last inequality holds if

$$\frac{\lambda^2}{\operatorname{var}(Y) - \lambda^2} \leq \frac{1}{nc}\log(1 + \eta^2). \tag{8.18}$$

Gathering conditions (8.17) and (8.18) allows us to complete the proof and to obtain the desired lower bound (4.6). □

PROOF OF PROPOSITION 4.6. The sketch of the proof and the notation are analogous to the one in Proposition 4.4. The upper bound (8.13) still holds:

$$\mathbb{E}_{0,\Sigma}(L_{\mu_{i,\rho}}(\mathbf{Y}, \mathbf{X})L_{\mu_{j,\rho}}(\mathbf{Y}, \mathbf{X})) \leq \frac{1}{2} + \frac{1}{2}\left(\frac{\operatorname{var}(Y)}{\operatorname{var}(Y) - |\operatorname{corr}(i, j)|\lambda^2}\right)^n.$$

Using the stationarity of the covariance function, we derive from (8.13) the following upper bound:

$$\mathbb{E}_{0,\Sigma}(L^2_{\mu_\rho}(\mathbf{Y}, \mathbf{X})) \leq \frac{1}{2} + \frac{1}{2p}\sum_{i=0}^{p-1}\left(\frac{\operatorname{var}(Y)}{\operatorname{var}(Y) - \lambda^2|\operatorname{corr}(0, i)|}\right)^n,$$

where $\mathrm{corr}(0,i)$ equals $\mathrm{corr}(X_1, X_{i+1})$. As previously, we want to constrain this quantity to be smaller than $1+\eta^2$. In particular, this is implied if for any i between 0 and $p-1$,

$$\left(\frac{\mathrm{var}(Y)}{\mathrm{var}(Y)-\lambda^2|\mathrm{corr}(i,0)|}\right)^n \le 1+\frac{2p\eta^2|\mathrm{corr}(i,0)|}{\sum_{i=0}^{p-1}|\mathrm{corr}(i,0)|}.$$

Using the inequality $\log(1+u)\le u$, it is straightforward to show that this previous inequality holds if

$$\frac{\lambda^2}{\mathrm{var}(Y)-\lambda^2|\mathrm{corr}(i,0)|}\le \frac{1}{n|\mathrm{corr}(i,0)|}\log\left(1+\frac{2p\eta^2|\mathrm{corr}(0,i)|}{\sum_{i=0}^{p-1}|\mathrm{corr}(i,0)|}\right).$$

As $|\mathrm{corr}(i,0)|$ is smaller than one for any i between 0 and $p-1$, it follows that $\mathbb{E}_{0,\Sigma}(L^2_{\mu_\rho}(\mathbf{Y},\mathbf{X}))$ is smaller than $1+\eta^2$ if

$$\rho^2 \le \bigwedge_{i=0}^{p-1}\frac{1}{n|\mathrm{corr}(i,0)|}\log\left(1+\frac{2p\eta^2|\mathrm{corr}(0,i)|}{\sum_{i=0}^{p-1}|\mathrm{corr}(i,0)|}\right).$$

We now apply the convexity inequality $\log(1+ux)\ge u\log(1+x)$ which holds for any positive x and any u between 0 and 1 to obtain the condition

$$\rho^2 \le \frac{1}{n}\log\left(1+\frac{2p\eta^2}{\sum_{i=0}^{p-1}|\mathrm{corr}(i,0)|}\right). \tag{8.19}$$

It turns out we only have to upper bound the sum of $|\mathrm{corr}(i,0)|$ for the following different types of correlation:

1. For $\mathrm{corr}(i,j)=\exp(-w|i-j|_p)$, the sum is clearly bounded by $1+2\frac{e^{-w}}{1-e^{-w}}$ and condition (8.19) simplifies as

$$\rho^2 \le \frac{1}{n}\log\left(1+2p\eta^2\frac{1-\mathrm{e}^{-w}}{1+\mathrm{e}^{-w}}\right);$$

2. if $\mathrm{corr}(i,j)=(1+|i-j|_p)^{-t}$ for t strictly larger than one, then $\sum_{i=0}^{p-1}|\mathrm{corr}(i,0)|\le 1+\frac{2}{t-1}$ and condition (8.19) simplifies as

$$\rho^2 \le \frac{1}{n}\log\left(1+\frac{2p(t-1)\eta^2}{t+1}\right);$$

3. if $\mathrm{corr}(i,j)=(1+|i-j|_p)^{-1}$ then $\sum_{i=0}^{p-1}|\mathrm{corr}(i,0)|\le 1+2\log(p-1)$ and condition (8.19) simplifies as

$$\rho^2 \le \frac{1}{n}\log\left(1+\frac{2p\eta^2}{1+2\log(p-1)}\right);$$

4. if $\mathrm{corr}(i, j) = (1 + |i - j|_p)^{-t}$ for $0 < t < 1$, then

$$\sum_{i=0}^{p-1} |\mathrm{corr}(i, 0)| \le 1 + \frac{2}{1-t}\left[\left(\frac{p}{2}\right)^{1-t} - 1\right] \le \frac{2}{1-t}\left(\frac{p}{2}\right)^{1-t},$$

and condition (8.19) simplifies as

$$\rho^2 \le \frac{1}{n}\log(1 + p^t 2^{1-t}(1-t)\eta^2). \qquad \square$$

PROOF OF PROPOSITION 5.3. For each dimension D between 1 and p, we define $r_D^2 = \rho_{D,n}^2 \wedge a_D^2 R^2$. Let us fix some $D \in \{1, \dots, p\}$. Since $r_D^2 \le a_D^2$ and since the a_j's are nonincreasing,

$$\sum_{j=1}^{D} \frac{\mathrm{var}(Y|X_{m_{j-1}}) - \mathrm{var}(Y|X_{m_j})}{a_j^2} \le \mathrm{var}(Y|X)R^2$$

for all $\theta \in S_{m_D}$ such that $\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} = r_D^2$. Indeed, $\|\theta\|^2 = \sum_{j=1}^{D} \mathrm{var}(Y|X_{m_{j-1}}) - \mathrm{var}(Y|X_{m_j})$ and $\mathrm{var}(Y) - \|\theta\|^2 = \mathrm{var}(Y|X)$. As a consequence,

$$\left\{\theta \in S_{m_D}, \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} = r_D^2\right\} \subset \left\{\theta \in \mathcal{E}_a(R), \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \ge r_D^2\right\}.$$

Since $r_D \le \rho_{D,n}$, we deduce from Proposition 4.2 that

$$\beta_\Sigma\left(\left\{\theta \in \mathcal{E}_a(R), \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \ge r_D^2\right\}\right) \ge \delta.$$

The first result of Proposition 5.3 follows by gathering these lower bounds for all D between 1 and p.

Moreover, $\rho_{i,n}^2$ is defined in Proposition 4.2 as $\rho_{i,n}^2 = \sqrt{2}[\sqrt{\mathcal{L}(\eta)} \wedge \frac{\mathcal{L}(\eta)}{\sqrt{\log 2}}]\frac{\sqrt{i}}{n}$. If $\alpha + \delta \le 47\%$, it is straightforward to show that $\rho_{i,n}^2 \ge \frac{\sqrt{i}}{n}$. $\square$

PROOF OF PROPOSITION 5.5. We first need the following lemma.

LEMMA 8.1. *We consider $(I_j)_{j\in\mathcal{J}}$ a partition of $\mathcal{I}$. For each $j \in \mathcal{J}$ let $p(j) = |I_j|$. For any $j \in \mathcal{J}$, we define Θ_j as the set of $\theta \in \mathbb{R}^p$ such that their support is included in I_j. For any sequence of positive weights k_j such that*

$$\sum_{j\in\mathcal{J}} k_j = 1,$$

it holds that

$$\beta_I\left(\bigcup_{j\in\mathcal{J}}\left\{\theta \in \Theta_j, \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} = r_j^2\right\}\right) \ge \delta,$$

if for all $j \in \mathcal{J}$, $r_j \le \rho_{p(j),n}(\eta/\sqrt{k_j})$ where the function $\rho_{p(j),n}$ is defined by (4.3).

For all $j \geq 0$ such that $2^{j+1} - 1 \in \mathcal{I}$ [i.e., for all $j \leq J$ where $J = \log(p+1)/\log(2) - 1$], let $\bar{S}_j$ be the linear span of the e_k's for $k \in \{2^j, \ldots, 2^{j+1} - 1\}$. Then $\dim(\bar{S}_j) = 2^j$ and $\bar{S}_j \subset S_{m_D}$ for $D = D(j) = 2^{j+1} - 1$. It is straightforward to show that

$$\bigcup_{j=0}^{J} \bar{S}_j[r_{D(j)}] \subset \bigcup_{j=0}^{J} S_{m_{D(j)}}[r_{D(j)}] \subset \bigcup_{D=1}^{p} S_{m_D}[r_D],$$

where $\bar{S}_j[r_D(j)] := \{\theta \in \bar{S}_j, \frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} = r^2_{D(j)}\}$ and $S_{m_D}[r_D] := \{\theta \in S_{m_D}, \frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} = r_D^2\}$.

We choose $\mathcal{J} = \{1, \ldots, J\}$. For any $j \in \mathcal{J}$, we define $I_j = \{2^j, 2^j + 1, \ldots, 2^{j+1} - 1\}$. Applying Lemma 8.1 with $k_j := [(j+1)R(p)]^{-1}$ where $R(p) := \sum_{k=0}^{J} 1/(k+1)$, we get

$$\beta_I\left(\bigcup_{D=1}^{p}\left\{\theta \in S_{m_D}, \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} = r_D^2\right\}\right) \geq \delta,$$

if for all those $D = D(j)$,

$$r_D^2 \leq \sqrt{\log(1 + 2\eta^2/k_j)}\left(1 \wedge \frac{\sqrt{\log(1+2\eta^2/k_j)}}{\sqrt{2\log 2}}\right)\frac{\sqrt{D}}{n}.$$

For $D = D(j)$, this last quantity is lower bounded by

$$\begin{aligned}
&\sqrt{\log(1 + 2\eta^2/k_j)}\left(1 \wedge \frac{\sqrt{\log(1+2\eta^2/k_j)}}{\sqrt{2\log 2}}\right)\frac{\sqrt{D}}{n} \\
&\quad \geq \sqrt{\log\bigl(1 + 2\eta^2(j+1)R(p)\bigr)}\left(1 \wedge \frac{\sqrt{\log(1+2\eta^2)}}{\sqrt{2\log 2}}\right)\frac{2^{j/2}}{n}.
\end{aligned} \tag{8.20}$$

It remains to check that (8.20) is larger than $\bar{\rho}_{D(j),n}$. Using $j + 1 = \log(D+1)/\log(2) \geq \log(D+1)$, we get $2^{j/2} \geq \sqrt{D/2}$. Thanks to the convexity inequality $\log(1+ux) \geq u\log(1+x)$, which holds for any $x > 0$ and any $u \in]0, 1]$, we obtain

$$\begin{aligned}
&\sqrt{\log\bigl(1 + 2\eta^2(j+1)R(p)\bigr)}2^{j/2} \\
&\quad \geq \sqrt{D/2}\bigl(\eta\sqrt{2R(p)} \wedge 1\bigr)\sqrt{\log[1 + \log(D+1)]} \\
&\quad \geq \bigl((\eta\sqrt{2}) \wedge 1\bigr)\sqrt{\log\log(D+1)}\sqrt{D/2} \\
&\quad \geq \frac{1}{\sqrt{2}}\bigl(1 \wedge \sqrt{\log(1+2\eta^2)}\bigr)\sqrt{\log\log(D+1)}\sqrt{D},
\end{aligned}$$

as $R(p)$ is larger than one for any $p \geq 1$. All in all, we get the lower bound

$$\sqrt{\log\bigl(1+2\eta^2(j+1)^2R(p)\bigr)}\biggl(1\wedge\frac{\sqrt{\log(1+2\eta^2)}}{\sqrt{2\log 2}}\biggr)\frac{2^{j/2}}{n}$$
$$\geq \frac{1}{2\sqrt{\log(2)}}\bigl(1\wedge\log(1+2\eta^2)\bigr)\sqrt{\log\log(D+1)}\frac{\sqrt{D}}{n}=\bar{\rho}^2_{D,n}.$$

Thus, if for all $1 \leq D \leq p$, r_D^2 is smaller than $\bar{\rho}^2_{D,n}$, it holds that

$$\beta_I\Biggl(\bigcup_{D=1}^{p}\biggl\{\theta\in S_{m_D},\frac{\|\theta\|^2}{\operatorname{var}(Y)-\|\theta\|^2}=r_D^2\biggr\}\Biggr)\geq\delta. \qquad \square$$

PROOF OF LEMMA 8.1. Using a similar approach to the proof of Theorem 4.3, we know that for each $r_j \leq \tilde{\rho}_j(\eta/\sqrt{k_j})$ there exists some measure μ_j over

$$\Theta_j[r_j]:=\biggl\{\theta\in\Theta_j,\frac{\|\theta\|^2}{\operatorname{var}(Y)-\|\theta\|^2}=r_j^2\biggr\}$$

such that

$$\mathbb{E}_0[L^2_{\mu_j}(Y,X)]\leq 1+\eta^2/k_j. \tag{8.21}$$

We now define a probability measure $\mu=\sum_{j\in\mathcal{J}}k_j\mu_j$ over $\bigcup_{j\in\mathcal{J}}\Theta_j[r_j]$. L_{μ_j} refers to the density of $\mathbb{P}_{\mu_j}$ with respect to $\mathbb{P}_0$. Thus

$$L_\mu(Y)=\frac{d\mathbb{P}_\mu}{d\mathbb{P}_0}(\mathbf{Y},\mathbf{X})=\sum_{j\in\mathcal{J}}k_jL_{\mu_j}(\mathbf{Y},\mathbf{X})$$

and

$$\mathbb{E}_0[L^2_\mu(\mathbf{Y},\mathbf{X})]=\sum_{j,j'\in\mathcal{J}}k_jk_{j'}\mathbb{E}_0[L_{\mu_j}(\mathbf{Y},\mathbf{X})L_{\mu_{j'}}(\mathbf{Y},\mathbf{X})].$$

Using expression (8.5), it is straightforward to show that if $j\neq j'$, then

$$\mathbb{E}_0[L_{\mu_j}(\mathbf{Y},\mathbf{X})L_{\mu_{j'}}(\mathbf{Y},\mathbf{X})]=1.$$

This follows from the fact that the sets Θ_j and $\Theta_{j'}$ are orthogonal with respect to the inner product (2.4). Thus

$$\mathbb{E}_0[L_\mu\bigl(\mathbf{Y},\mathbf{X}\bigr)]=1+\sum_{j\in\mathcal{J}}k_j^2\bigl(\mathbb{E}_0[L^2_{\mu_j}(\mathbf{Y},\mathbf{X})]-1\bigr)\leq 1+\eta^2,$$

thanks to (8.21). Using argument (8.2) as in the proof of Theorem 4.3 completes the proof. $\square$

PROOF OF PROPOSITION 5.6. First of all, we only have to consider the case where the covariance matrix of X is the identity. If this is not the case, one only has to apply the Gram–Schmidt process to X and thus obtain a vector X' and a new basis for $\mathbb{R}^p$ which is orthonormal. We refer to the beginning of Section 5 for more details.

Like the previous bounds for ellipsoids, we adapt the approach of Section 6 in Baraud [2]. We use the same notation as in proof of Proposition 5.3. Let $D^*(R) \in \{1, \ldots, p\}$, an integer which achieves the supremum of $\bar{\rho}_D^2 \wedge (R^2 a_D^2) = \bar{r}_D^2$. As in proof of Proposition 5.3, for any $R > 0$,

$$\left\{\theta \in S_{m_{D^*(R)}}, \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} = r_{D^*(R)}^2\right\} \subset \left\{\theta \in \mathcal{E}_a(R), \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq r_{D^*(R)}^2\right\}.$$

When R varies, $D^*(R)$ describes $\{1, \ldots, p\}$. Thus we obtain

$$\begin{aligned} &\bigcup_{1 \leq D \leq p} \left\{\theta \in S_{m_D}, \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} = r_D^2\right\} \\ &\quad = \bigcup_{R>0} \left\{\theta \in S_{m_{D^*(R)}}, \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} = r_{D^*(R)}^2\right\} \\ &\quad \subset \bigcup_{R>0} \left\{\theta \in \mathcal{E}_a(R) \frac{\|\theta\|^2}{\operatorname{var}(Y) - \|\theta\|^2} \geq r_{D^*(R)}^2\right\}, \end{aligned}$$

and the result follows from Proposition 5.5. $\square$

APPENDIX

PROOF OF PROPOSITION 3.1. The test associated with procedure P_1 corresponds to a Bonferroni procedure. Hence we prove that its size is less than α by the following argument: let θ be an element of S_V (defined in Section 2.2),

$$\mathbb{P}_\theta(T_\alpha > 0) \leq \sum_{m \in \mathcal{M}} \mathbb{P}_\theta\left(\phi_m(\mathbf{Y}, \mathbf{X}) - \bar{F}_{D_m, N_m}^{-1}(\alpha_m) > 0\right),$$

where $\phi_m(\mathbf{Y}, \mathbf{X})$ is defined in (2.2). The test is rejected if for some model m, $\phi_m(\mathbf{Y}, \mathbf{X})$ is larger than $\bar{F}_{D_m, N_m}^{-1}(\alpha_m)$. As θ belongs to S_V, $\Pi_{V \cup m}\mathbf{Y} - \Pi_V \mathbf{Y} = \Pi_{V \cup m}\boldsymbol{\epsilon} - \Pi_V \boldsymbol{\epsilon}$ and $\mathbf{Y} - \Pi_{V \cup m}\mathbf{Y} = \boldsymbol{\epsilon} - \Pi_{V \cup m}\boldsymbol{\epsilon}$. Then the quantity $\phi_m(\mathbf{Y}, \mathbf{X})$ is equal to

$$\phi_m(\mathbf{Y}, \mathbf{X}) = \frac{N_m \|\Pi_{V \cup m}\boldsymbol{\epsilon} - \Pi_V \boldsymbol{\epsilon}\|_n^2}{D_m \|\boldsymbol{\epsilon} - \Pi_{V \cup m}\boldsymbol{\epsilon}\|_n^2}.$$

Because $\boldsymbol{\epsilon}$ is independent of $\mathbf{X}$, the distribution of $\phi_m(\mathbf{Y},\mathbf{X})$ conditionally to $\mathbf{X}$ is a Fisher distribution with D_m and N_m degrees of freedom. As a consequence, $\phi_{m,\alpha_m}(\mathbf{Y},\mathbf{X})$ is a Fisher test with D_m and N_m degrees of freedom. It follows that

$$\mathbb{P}_\theta(T_\alpha > 0) \le \sum_{m\in\mathcal{M}} \alpha_m \le \alpha.$$

The test associated with procedure P_2 has the property to be of size exactly α. More precisely, for any $\theta \in S_V$, we have that

$$\mathbb{P}_\theta(T_\alpha > 0|\mathbf{X}) = \alpha, \qquad \mathbf{X}\text{ a.s.}$$

The result follows from the fact that $q_{\mathbf{X},\alpha}$ satisfies

$$\mathbb{P}_\theta\left(\sup_{m\in\mathcal{M}}\left\{\frac{N_m\|\Pi_{V\cup m}(\boldsymbol{\epsilon}) - \Pi_V(\boldsymbol{\epsilon})\|_n^2}{D_m\|\boldsymbol{\epsilon} - \Pi_{V\cup m}(\boldsymbol{\epsilon})\|_n^2} - \bar{F}^{-1}_{D_m,N_m}(q_{\mathbf{X},\alpha})\right\} > 0\middle|\mathbf{X}\right) = \alpha,$$

and that for any $\theta \in S_V$, $\Pi_{V\cup m}\mathbf{Y} - \Pi_V\mathbf{Y} = \Pi_{V\cup m}\boldsymbol{\epsilon} - \Pi_V\boldsymbol{\epsilon}$ and $\mathbf{Y} - \Pi_{V\cup m}\mathbf{Y} = \boldsymbol{\epsilon} - \Pi_{V\cup m}\boldsymbol{\epsilon}$. □

PROOF OF PROPOSITION 3.2. We come back to the definitions of T_α^1 and T_α^2:

$$T_\alpha^1(\mathbf{X},\mathbf{Y}) = \sup_{m\in\mathcal{M}}\{\phi_m(\mathbf{Y},\mathbf{X}) - \bar{F}^{-1}_{D_m,N_m}(\alpha/|\mathcal{M}|)\},$$

$$T_\alpha^2(\mathbf{X},\mathbf{Y}) = \sup_{m\in\mathcal{M}}\{\phi_m(\mathbf{Y},\mathbf{X}) - \bar{F}^{-1}_{D_m,N_m}(q_{\mathbf{X},\alpha})\}.$$

Conditionally on $\mathbf{X}$, the size of T_α^1 is smaller than α whereas the size T_α^2 is exactly α. As a consequence, $q_{\mathbf{X},\alpha} \ge \alpha/|\mathcal{M}|$ as the statistics T_α^1 and T_α^2 differ only through these quantities. Thus $T_\alpha^2(\mathbf{X},\mathbf{Y}) \ge T_\alpha^1(\mathbf{X},\mathbf{Y})$, $(\mathbf{X},\mathbf{Y})$ almost surely, and the result (3.4) follows. □

Acknowledgments. We gratefully thank Sylvie Huet and Pascal Massart for many fruitful discussions. We also thank the two referees for their helpful comments.

REFERENCES

[1] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École d'été de probabilités de Saint Flour XIII. Lecture Notes in Math.* **1117**. Springer, Berlin. MR0883646

[2] BARAUD, Y. (2002). Non-asymptotic rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[3] BARAUD, Y., HUET, S. and LAURENT, B. (2003). Adaptative tests of linear hypotheses by model selection. *Ann. Statist.* **31** 225–251. MR1962505

[4] BÜHLMANN, P., KALISCH, M. and MAATHUIS, M. H. (2009). Variable selection for high-dimensional models: Partially faithful distributions and the PC-simple algorithm. *Biometrika*. To appear.

[5] CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

[6] COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1999). *Probabilistic Networks and Expert Systems*. Springer, New York. MR1697175

[7] CRESSIE, N. (1993). *Statistics for Spatial Data*, revised ed. Wiley, New York. MR1239641

[8] DRTON, M. and PERLMAN, M. (2007). Multiple testing and error control in Gaussian graphical model selection. *Statist. Sci.* **22** 430–449. MR2416818

[9] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166

[10] GIRAUD, C. (2008). Estimation of Gaussian graphs by model selection. *Electron. J. Stat.* **2** 542–563. MR2417393

[11] HUANG, J., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likehood. *Biometrika* **93** 85–98. MR2277742

[12] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives I. *Math. Methods Statist.* **2** 85–114. MR1257978

[13] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives II. *Math. Methods Statist.* **3** 171–189. MR1257983

[14] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives III. *Math. Methods Statist.* **4** 249–268. MR1259685

[15] KISHINO, H. and WADDELL, P. (2000). Correspondence analysis of genes and tissue types and finding genetic links from microarray data. *Genome Informatics* **11** 83–95.

[16] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic function by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785

[17] LAURITZEN, S. L. (1996). *Graphical Models*. Oxford Univ. Press, New York. MR1419991

[18] MASSART, P. (2007). Concentration inequalities and model selection. In *École d'été de probabilités de Saint Flour XXXIII*. *Lecture Notes in Math.* **1896**. Springer, Berlin. MR2319879

[19] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the Lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

[20] RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC, London. MR2130347

[21] SCHÄFER, J. and STRIMMER, K. (2005). An empirical Bayes approach to inferring large-scale gene association network. *Bioinformatics* **21** 754–764.

[22] SPOKOINY, V. G. (1996). Adaptative hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962

[23] VERZELEN, N. and VILLERS, F. (2009). Tests for Gaussian graphical models. *Comput. Statist. Data Anal.* **53** 1894–1905.

[24] WAINWRIGHT, M. J. (2007). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. Technical Report 725, Dept. Statistics, Univ. California, Berkeley.

[25] WILLE, A. and BÜHLMANN, P. (2006). Low-order conditional independence graphs for inferring genetic networks. *Stat. Appl. Genet. Mol. Biol.* **5** Art. 1 (electronic). MR2221304

[26] WILLE, A., ZIMMERMANN, P., VRANOVA, E., FÜRHOLZ, A., LAULE, O., BLEULER, S., HENNIG, L., PRELIC, A., VON ROHR, P., THIELE, L., ZITZLER, E., GRUISSEM, W. and BÜHLMANN, P. (2004). Sparse graphical Gaussian modelling of the isoprenoid gene network in *arabidopsis thaliana*. *Genome Biology* **5** 11.

[27] YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824

[28] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. MR2435448

[29] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449

[30] ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the Elastic Net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327

LABORATOIRE DE MATHÉMATIQUES D'ORSAY
UNIVERSITÉ PARIS-SUD
CNRS
ORSAY CEDEX, F-91405
FRANCE
E-MAIL: nicolas.verzelen@math.u-psud.fr

LABORATOIRE DE BIOMÉTRIE
INRA
78352 JOUY-EN-JOSAS CEDEX
FRANCE
E-MAIL: fanny.villers@jouy.inra.fr