Goodness-of-fit Tests for high-dimensional Gaussian linear models

Nicolas VERZELEN[1]

Fanny VILLERS[2]

Rapport technique 2007-4, 44 pp.

Unité Mathématiques et Informatique Appliquées
INRA
Domaine de Vilvert
78352 Jouy-en-Josas Cedex
France

[1] Université Paris-Sud, Laboratoire de Mathématique d'Orsay, INRIA Futurs, Equipe SELECT, Université Paris-Sud 91405 Orsay Cedex France
[2] INRA, unité MIA, fanny.villers@jouy.inra.fr

Goodness-of-fit Tests for high-dimensional Gaussian linear models

November 13, 2007

Abstract

Let $(Y, (X_i)_{i\in\mathcal{I}})$ be a zero mean Gaussian vector and V be a subset of $\mathcal{I}$. Suppose we are given n i.i.d. replications of the vector (Y, X). We propose a new test for testing that Y is independent of $(X_i)_{i\in\mathcal{I}\setminus V}$ 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 neighbourhood of a Gaussian graphical model. The procedure is based on a model of Gaussian regression with random Gaussian covariates. We give non asymptotic 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. Besides, it allows us to derive non asymptotic minimax rates of testing in this 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

$$Y = \sum_{i\in\mathcal{I}} \theta_i X_i + \epsilon \tag{1}$$

where θ is an unknown vector of $\mathbb{R}^{\mathcal{I}}$. The vector X follows a zero mean Gaussian distribution with non singular covariance matrix Σ and ϵ is a zero mean Gaussian random variable independent of X. We note p the cardinal of $\mathcal{I}$ and $\text{var}(Y)$ the variance of Y. Straightforwardly, the variance of ϵ corresponds to the conditional variance of Y given X, $\text{var}(Y|X)$.

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 (2006), Candès and Tao (2007), Zhao and Yu (2006), Zou and Hastie (2005), or Bühlmann and Kalisch (2007). Our issue is the natural testing counterpart of this variable selection problem: we aim at defining a computationally feasible testing procedure which achieves an optimal rate in some sense.

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 $(\theta_i)_{i\in V}$. The first purpose of this paper is to propose a test of the null hypothesis "$\theta_{\mathcal{I}\setminus V} = 0$" against the general alternative "$\theta_{\mathcal{I}\setminus V} \neq 0$"

under 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". Moreover, we want to be 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. From a theoretical point of view, our second purpose is to derive non asymptotic minimax rates of testing for this model over various alternatives.

1.1 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 non singular. We observe n i.i.d. replications of this vector Z and we are given a graph $\mathcal{G} = (\Gamma, E)$ where $\Gamma = \{1, \ldots |\mathcal{J}|\}$ 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 (1996) 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 (Cressie, 1993; Rue and Held, 2005) or probabilistic expert systems (Cowell et al., 1999). 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 (2005), Kishino and Waddell (2000) or Wille et al. (2004). 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 (2005) or Wille and Bühlmann (2006). Other methods are based on variable selection for high-dimensional data we previously mentioned. For instance, Meinshausen and Bühlmann (2006) proposed a computationally feasible model selection algorithm using Lasso penalisation. Huang et al. (2006) and Yuan and Lin (2007) extend this method to infer directly the inverse covariance matrix Σ'^{-1} by minimizing the log-likehood penalised by the l^1 norm.

While the issue of graph and covariance estimation is extensively studied, it seems that the problem of hypothesis testing of GGM in a high-dimensional setting has not yet raised much interest. 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 reveal 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 (2007) and Schäfer and Strimmer (2005). 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 non asymptotic way.

In a subsequent paper (Verzelen and Villers, 2007) we define a test of graph based on the present work. It benefits the ability of handling high dimensional GGM and has minimax properties. Besides we show numerical evidence of its efficiency; see Verzelen and Villers (2007) 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 neighbours 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 conditionnally 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 (2006).

1.2 Connection with tests in fixed design regression

Our work is directly inspired by the testing procedure of Baraud et al. (2003) in fixed design regression framework. Contrary to model (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). Let us define this second regression model:

$$Y_i = f_i + \sigma \epsilon_i, \quad i \in \{1, \ldots, N\}, \tag{2}$$

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 Baraud et al. (2003) for a short review of non parametric tests in this framework. Besides, we are interested in the performance of the procedures from a minimax perspective. To our knowledge, there has been no results in model (1). On the other hand, there are numerous papers on this issue in the fixed design regression model. First, we refer to the seminal work of Ingster (1993,a,b,c) which gives asymptotic minimax rates over non parametric alternatives. Our work is closely related to the results of Baraud (2002) where he gives non asymptotic 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.

1.3 Principle of our testing procedure

Let us briefly describe the idea underlying our testing procedure. 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 $\mathbb{R}^{\mathcal{I}}$ 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

$$\phi_m(\mathbf{Y}, \mathbf{X}) := \frac{N_m \|\Pi_{V \cup m} \mathbf{Y} - \Pi_V \mathbf{Y}\|_n^2}{D_m \|\mathbf{Y} - \Pi_{V \cup m} \mathbf{Y}\|_n^2}, \tag{3}$$

where Π_V refers to the orthogonal projection onto the space generated by the vectors $(\mathbf{X}_i)_{i \in V}$ and $\|.\|_n$ is the canonical norm in $\mathbb{R}^n$.

Let us consider a finite collection $\mathcal{M}$ of non empty 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:

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

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, 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.

1.4 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 write $\mathbb{R}^{\mathcal{I}}$ for $\mathbb{R}^{\mathcal{I}}$ endowed with the euclidean norm

$$\|\theta\|^2 := \theta^t \Sigma \theta = \mathrm{var}\left(\sum_{i\in\mathcal{I}} \theta_i X_i\right) . \tag{4}$$

As ϵ and X are independent, we derive from the definition of $\|.\|^2$ that $\mathrm{var}(Y) = \|\theta\|^2 + \mathrm{var}(Y|X)$. Thus, if we have $\|\theta\|$ vary, either the quantity $\mathrm{var}(Y)$ or $\mathrm{var}(Y|X)$ has to vary. In the sequel, we suppose that $\mathrm{var}(Y)$ is fixed. We briefly justify this choice in Section 4.2. Consequently, if $\|\theta\|^2$ is increasing, then $\mathrm{var}(Y|X)$ 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 $\mathrm{var}(Y)$, we denote $\mathbb{P}_\theta$ the joint distribution of $(\mathbf{Y}, \mathbf{X})$. For the sake of simplicity, we do not emphasize the dependence of $\mathbb{P}_\theta$ on $\mathrm{var}(Y)$ 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 $\rho(\psi_\alpha, \Theta, \delta, \mathrm{var}(Y), \Sigma)$ defined by:

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

where the quantity

$$r_{s/n}(\theta) := \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \tag{5}$$

appears naturally as it corresponds to the ratio $\|\theta\|^2/\mathrm{var}(Y|X)$ 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

$$\inf_{\psi_\alpha} \rho(\psi_\alpha, \Theta, \delta, \mathrm{var}(Y), \Sigma) := \rho(\Theta, \alpha, \delta, \mathrm{var}(Y), \Sigma), \tag{6}$$

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 $\Theta \subset \mathbb{R}^{\mathcal{I}}$ and $\alpha \in]0,1[$, we denote $\beta_\Sigma(\Theta)$ the quantity

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

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.

1.5 Organization of the Paper

We present the procedure in Section 2. In Section 3, we give a general theorem which characterizes a set of vectors θ over which the test is powerful in a non asymptotic setting. In Section 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 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 exhibits the same ranges in our statistical model (1) and in fixed design regression model (2). In section 5, we define two procedures which achieve the simultaneous minimax rates of testing over large classes of ellipsoids (to sometimes the price of a $\log(p)$ factor). Besides, we show that the problem of adaptation over classes of ellipsoids is impossible without a loss in efficiency. This was previously pointed out by Spokoiny (1996) in fixed design regression framework. The simulation studies are presented in Section 6. Finally, Sections 7 and 8 contain the proofs.

Let us now introduce some notations that will be used throughout the paper. For $x, y \in \mathbb{R}$, we set

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

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.

2 The Testing procedure

In this section, we adapt the testing procedure of Baraud et al. (2003) in our random design model (1).

2.1 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 non empty subsets of $\mathcal{I} \setminus V$ such that for all $m \in \mathcal{M}$, $1 \leq D_m \leq n-d-1$. Most of the notations used in this definition were defined in Section 1.3. We introduce the following test of level α. We reject H_0:"$\theta \in S_V$" when the statistic

$$T_\alpha := \sup_{m \in \mathcal{M}} \left\{\phi_m(\mathbf{Y}, \mathbf{X}) - \bar{F}^{-1}_{D_m, N_m}(\alpha_m(\mathbf{X}))\right\} \tag{7}$$

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_m 's do not depend on $\mathbf{X}$ and satisfy the equality :

$$\sum_{m\in\mathcal{M}} \alpha_m = \alpha \ . \tag{8}$$

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

$$\inf_{m\in\mathcal{M}} \bar{F}_{D_m,N_m}\left(\frac{\|\Pi_{V\cup m}(\boldsymbol{\epsilon}) - \Pi_V(\boldsymbol{\epsilon})\|_n^2/D_m}{\|\boldsymbol{\epsilon} - \Pi_{V\cup m}(\boldsymbol{\epsilon})\|_n^2/N_m}\right) \tag{9}$$

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 (9) 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 $\mathbf{X}$. Thus, we can easily work out its quantile using Monte-Carlo method.

It is clear that 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 Verzelen and Villers (2007), we favour collections $\mathcal{M}$ whose size is linear with respect to the number of covariates p.

2.2 Behavior of the test under the null hypothesis

The test associated with Procedure P_1 corresponds to a Bonferroni test and is of size less than α by arguing as follows: let θ be an element of S_V (defined in Section 1.3),

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

where $\phi_m(\mathbf{X},\mathbf{Y})$ is defined in (3). We now define the test statistic $\phi_{m,\alpha_m}(\mathbf{X},\mathbf{Y})$ as

$$\phi_{m,\alpha_m}(\mathbf{X},\mathbf{Y}) = \phi_m(\mathbf{X},\mathbf{Y}) - \bar{F}^{-1}_{D_m,N_m}(\alpha_m). \tag{10}$$

The test is rejected if $\phi_{m,\alpha_m}(\mathbf{X},\mathbf{Y})$ is positive. 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 $\phi_m(\mathbf{X},\mathbf{Y})$ is equal to

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

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

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

Procedure P_1 is therefore conservative.

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

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

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 $\mathbf{Y} - \Pi_{V\cup m}\mathbf{Y} = \boldsymbol{\epsilon} - \Pi_{V\cup m}\boldsymbol{\epsilon}$.

2.3 Comparison of Procedures P_1 and P_2

We show in this section that the test (7) with Procedure P_2 is more powerful than the corresponding test defined with Procedure P_1 with weights $\alpha_m = \alpha/|\mathcal{M}|$. We respectively refer to T^1_α and T^2_α for these two tests associated with Procedure P_1 and P_2. More precisely, let us prove that

$$\forall\theta \notin S_V,\ \mathbb{P}_\theta\left(T^2_\alpha(\mathbf{X},\mathbf{Y}) > 0\middle|\mathbf{X}\right) \geq \mathbb{P}_\theta\left(T^1_\alpha(\mathbf{X},\mathbf{Y}) > 0\,|\,\mathbf{X}\right) \quad \mathbf{X}\ a.s.\ . \tag{11}$$

In fact this previous inequality is straighforward when considering the definitions of T^1_α and T^2_α:

$$\begin{aligned} T^1_\alpha(\mathbf{X},\mathbf{Y}) &= \sup_{m\in\mathcal{M}}\left\{\phi_m(\mathbf{X},\mathbf{Y}) - \bar{F}^{-1}_{D_m,N_m}(\alpha/|\mathcal{M}|)\right\} \\ T^2_\alpha(\mathbf{X},\mathbf{Y}) &= \sup_{m\in\mathcal{M}}\left\{\phi_m(\mathbf{X},\mathbf{Y}) - \bar{F}^{-1}_{D_m,N_m}(q_{\mathbf{X},\alpha})\right\} \end{aligned}$$

Conditionally on $\mathbf{X}$, the size of T^1_α is smaller than α, whereas the size T^2_α is exactly α. As a consequence $q_{\mathbf{X},\alpha} \geq \alpha/|\mathcal{M}|$ as the statistics T^1_α and T^2_α differ only through these quantities. Thus, $T^2_\alpha(\mathbf{X},\mathbf{Y}) \geq T^1_\alpha(\mathbf{X},\mathbf{Y})$, $(\mathbf{X},\mathbf{Y})$ almost surely and the result (11) follows.

The choice of Procedure P_1 allows to avoid the computation of the quantile $q_{\mathbf{X},\alpha}$ and possibly permits 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. This will be illustrated in Section 6. In the next three sections we study the power and rates of testing of T_α with Procedure P_1.

3 Power of the Test

In this section, we aim at describing a set of vectors θ in $\mathbb{R}^{\mathcal{I}}$ over which the test defined in Section 2 with Procedure P_1 is powerful. We note that 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 introduce some quantities that depend on α_m, δ, D_m, and N_m. We set $L = \log\left(\frac{1}{\delta}\right)$ and for any $m \in \mathcal{M}$, we define $L_m = \log\left(\frac{1}{\alpha_m}\right)$, $k_m = 2\exp(4L_m/N_m)$, and $l_m = \left(1 + 2\sqrt{\frac{L}{N_m}} + 2\frac{L}{N_m}\right)$.

Under the following condition, k_m and l_m behave like constants:

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

For typical choices of the collections $\mathcal{M}$ and $\{\alpha_m, m \in \mathcal{M}\}$, these conditions are fulfilled. In Sections 4 and 5, we discuss these assumptions for various settings. Let us now turn to the main result.

Theorem 1. *Let T_α be the test procedure defined by (7). We assume that $n > d+2$. Then $\mathbb{P}_\theta(T_\alpha > 0) \geq 1-\delta$ for all θ belonging to the set*

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

where

$$\Delta(m) := \frac{4l_m\sqrt{D_m \log\left(\frac{1}{\alpha_m\delta}\right)}\left(1+\sqrt{\frac{D_m}{N_m}}\right) + \log\left(\frac{1}{\alpha_m\delta}\right)\left[8 \vee k_m l_m\left(1+2\frac{D_m}{N_m}\right)\right]}{(n-d)\left(1-1/2 \vee 2\sqrt{\frac{2L}{N_m}}\right)}. \tag{12}$$

Under the hypothesis $H_{\mathcal{M}}$, for any $m \in \mathcal{M}$,

$$\Delta(m) \leq \frac{C_1\sqrt{D_m \log\left(\frac{1}{\alpha_m\delta}\right)}\left(1+\sqrt{\frac{D_m}{N_m}}\right) + C_2\left(1+2\frac{D_m}{N_m}\right)\log\left(\frac{1}{\alpha_m\delta}\right)}{n-d}, \tag{13}$$

where C_1 and C_2 are universal constants.

This result is similar to Theorem 1 in Baraud et al. (2003) 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 (7) 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 1, we know that under the 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

$$\mathcal{F}'_m(\delta,\alpha_m) = \left\{\theta \in \Theta, \frac{\mathrm{var}(Y) - \mathrm{var}(Y|X_m)}{\mathrm{var}(Y|X_m)} \geq \frac{C'_1(D_m, N_m)}{n}\left(\sqrt{D_m \log\left(\frac{1}{\alpha_m\delta}\right)} + \log\left(\frac{1}{\alpha_m\delta}\right)\right)\right\}. \tag{14}$$

Besides, $C'_1(D_m, N_m)$ 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 1, we know that it contains $\mathcal{F}'_m(\delta,\alpha)$. Moreover, the following Proposition shows that this set is not much larger than $\mathcal{F}'_m(\delta,\alpha)$:

Proposition 2. *Let $\delta \in]0, 1-\alpha[$ and*

$$t(\alpha,\delta) := \sqrt{\log\left(1+8\left(1-\alpha-\delta\right)^2\right)}\left[1 \wedge \sqrt{\log\left(1+8\left(1-\alpha-\delta\right)^2\right)/(2\log 2)}\right].$$

If

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

then $\mathbb{P}_\theta\left(\phi_m > 0\right) \leq 1-\delta$.

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 (14) differ from the fact that $\log(1/\alpha)$ is replaced by $\log(1/\alpha_m)$. For the main applications that we will study in section 4, 5, and 6, the difference $\log(1/\alpha_m)-\log(1/\alpha)$ is of order $k\log(ep/k)$ where k is a "small" integer of the order $\log(n)$ or $\log\log n$. 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 (7) applies in the same way when one considers 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 non-zero coordinates

Let us fix a number k between 1 and p. In this section, we are interested in testing $\theta=0$ against the class of θ with a most k non zero components. This typically corresponds to the situation encountered when considering tests of neighbourhood for large sparse graphs. As the graph is assumed to be sparse, only a small number of neighbours 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 $\mathcal{I}=\{1,\ldots,p\}$ of cardinality k. The set $\Theta[k,p]$ stands for the subset of $\theta\in\mathbb{R}^{\mathcal{I}}$, such that at most k coordinates of θ are non-zero.

First, we define a test T_α of the form (7) 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 matrices Σ.

4.1 Rate of testing of T_α

Proposition 3. *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:*

$$n\geq k+\left[10\left[\log\left(\frac{1}{\alpha}\right)+k\log\left(\frac{ep}{k}\right)\right]\vee 21\log(1/\delta)\right].$$

Let us set the quantity

$$\rho'^2_{k,n,p}:=\frac{C_3 k\log\left(\frac{ep}{k}\right)+C_4\left[\sqrt{k\log\left(\frac{1}{\alpha\delta}\right)}\vee\log\left(\frac{1}{\alpha\delta}\right)\right]}{n},\qquad(15)$$

where C_3 and C_4 are universal constants. For any θ in $\Theta[k,p]$, such that $\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2}\geq\rho'^2_{k,n,p}$, $\mathbb{P}_\theta(T_\alpha>0)\geq 1-\delta$.

This Proposition follows easily from Theorem 1 and its proof is given in Section 7. Let us note that this upper bound does not directly depend on the covariance matrix of the vector X. We will further discuss this result after deriving lower bounds for the minimax rate of testing in this setting.

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. That 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 non asymptotic 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. As these bounds coincide with the upper bound obtained in Section 4.1, this will show that our test T_α is rate optimal.

In order to simplify the notations, we set $\eta = 2(1-\alpha-\delta)$ and $\mathcal{L}(\eta) = \frac{\log(1+2\eta^2)}{2}$. We first give a lower bound for the (α,δ)-minimax rate of detection of all p non-zero coordinates, as we will need it later.

Proposition 4. *Let us suppose that* $\mathrm{var}(Y)$ *is known. Let us set* $\rho^2_{p,n}$ *such that:*

$$\rho^2_{p,n}(\eta) := \sqrt{2}\left[\sqrt{\mathcal{L}(\eta)} \wedge \frac{\mathcal{L}(\eta)}{\sqrt{\log(2)}}\right]\frac{\sqrt{p}}{n}. \tag{16}$$

Then for all $\rho < \rho_{p,n}(\eta)$*,*

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

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

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

Theorem 5. *Let us set* $\rho^2_{k,p,n}$ *such that*

$$\rho^2_{k,p,n} := \frac{k(\mathcal{L}(\eta)\wedge 1)}{n}\log\left(1+\frac{p}{k^2}+\sqrt{2\frac{p}{k^2}}\right). \tag{17}$$

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

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

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

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

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

This result implies the lower bound

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

The proof is given in Section 8. To the price of more technicity, it is possible to prove that the lower bound still holds if the variables (X_i) are assumed independent with known variances possibly different. This theorem recovers approximately the lower bounds for the minimax rates of testing in signal detection framework obtained by Baraud (2002). The main difference lies in the fact that we suppose $\mathrm{var}(Y)$ 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 3 and Theorem 5. 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 3, the quantities $\rho^2_{k,n,p}$ and $\rho'^2_{k,n,p}$ 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 5 is sharp. Additionally, in this case, the procedure T_α defined in Proposition 3 follows approximately the minimax rate of testing. We recall that our procedure T_α does not depend on the knowledge of var(Y) and corr(X). In applications, this choice of 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 $\sqrt{p} \leq k \leq p$, the lower bound and the upper bound do not coincide anymore. Nevertheless, if $n \geq (1+\gamma)p$ for some $\gamma > 0$, Theorem 1 shows that the test $\phi_{\mathcal{I},\alpha}$ defined in (10) has power greater than δ over the vectors θ which satisfy

$$\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq C(\gamma,\alpha,\delta)\frac{\sqrt{p}}{n}. \tag{18}$$

  This upper bound and the lower bound do not depend on k. Here again, the lower bound obtained in Theorem 5 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 Baraud (2002).

- When $k < \sqrt{p}$ and k is close to $\sqrt{p}$, the lower bound and the upper bound given by Proposition 3 differ from at most a $\log(p)$ factor. For instance, if k is of order $\sqrt{p}/\log p$, the lower bound in Theorem 5 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 and if the minimax rates of testing coincide when 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 (Baraud, 2002) when the covariates are independent. Moreover, our result shows that the minimax rate of testing is slower when the $(X_i)_{i\in\mathcal{I}}$ are independent than for any form of dependence. Indeed, the upper bounds obtained in Proposition 3 and in (18) 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 only 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 non-zero, and $\frac{\log(p)}{n}$, the minimax rate of testing for independent covariates.

Proposition 6. *Let us suppose that there exists a positive number c such that for any $i \neq j$,*

$$|corr(X_i,X_j)| \leq c.$$

We define $\rho^2_{1,p,n,c}$ as

$$\rho^2_{1,p,n,c} := \frac{1}{n}\left[\log(1+\eta^2 p) \wedge \frac{1}{c}\log\left(1+\eta^2\right)\right]. \tag{19}$$

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

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

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

$$n \geq \left[C_5\left(1+\log\left(\frac{2p}{\alpha}\right)\right)\right] \vee \left[C_6 \log\left(\frac{1}{\delta}\right)\right] \tag{20}$$

If $\alpha < 60\%$ *and* $\delta < 60\%$ *the test* T_α *defined by*

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

satisfies

$$\mathbb{P}_0\left(T_\alpha > 0\right) \leq \alpha \text{ and } \mathbb{P}_\theta\left(T_\alpha > 0\right) \geq 1-\delta,$$

for any θ *in* $\Theta[1,p]$ *such that*

$$\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq \rho'^2_{1,n,p,c},$$

where

$$\rho'^2_{1,n,p,c} := \frac{C_7}{n}\left(\log\left(\frac{2p}{\alpha\delta}\right) \bigwedge \frac{1}{c}\log\left(\frac{2}{\alpha\delta}\right)\right), \tag{21}$$

and C_5, C_6, *and* C_7 *are universal constants.*

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 3 is not rate optimal when the correlation is known and is large. Indeed, when the correlation between the covariates is large, all the tests statistics ϕ_{m,α_m} defining T_α are highly correlated. The choice of the weights α_m in Procedure P_1 corresponds to a Bonferroni procedure. The loss due to a Bonferroni procedure is precisely large when the tests are positively correlated.

This example shows 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 neighbours in the graph and smaller if they are far (w.r.t. the graph distance). That is why we derive lower bounds of the rate of testing for other kind of correlation matrices often used to model stationary processes.

Proposition 8. *Let* $X_1,\ldots,X_p$ *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* $|.|_p$ *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

$$\begin{aligned}\mathrm{corr}(X_i,X_j) &= \exp(-w|i-j|_p) \text{ where } w>0,\\ \mathrm{corr}(X_i,X_j) &= (1+|i-j|_p)^{-t} \text{ where } t>0.\end{aligned}$$

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

$$\begin{aligned}\rho^2_{1,p,n,\Sigma_1}(w) &:= \frac{1}{n}\log\left(1+2p\eta^2\frac{1-e^{-w}}{1+e^{-w}}\right)\\ \rho^2_{1,p,n,\Sigma_2}(t) &:= \begin{cases}\frac{1}{n}\log\left(1+\frac{2p(t-1)\eta^2}{t+1}\right) & \text{if} \quad t>1\\ \frac{1}{n}\log\left(1+\frac{2p\eta^2}{1+2\log(p-1)}\right) & \text{if} \quad t=1\\ \frac{1}{n}\log\left(1+p^t2^{1-t}(1-t)\eta^2\right) & \text{if} \quad 0<t<1.\end{cases}\end{aligned}$$

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

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

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

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

All in all, 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 3 is rate-optimal for these correlations matrices.

To conclude, when $k\leq p^\gamma$ (for $\gamma\leq 1/2$), the test T_α defined in Proposition 3 is approximately (α,δ)-minimax against the alternative $\theta\in\Theta[k,p]$, when neither $\mathrm{var}(Y)$ nor the covariance matrix of X is fixed. Indeed, the rate of testing of T_α coincide (up to a constant) with the following quantity:

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

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 (18) has the same behavior.

However, our procedure does not adapt to Σ: for some correlation matrices (as for instance in Proposition 7), T_α with Procedure P_1 is not rate optimal. Nevertheless, we believe and this will be illustrated in Section 6 that Procedure P_2 slightly improves the power of the test in practice.

5 Rates of testing on "ellipsoids" and adaptation

In this section, we define tests T_α of the form (7) in order to test simultaneously $\theta=0$ against θ belongs 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. In this section, $\mathcal{I}$ is supposed again to be $\{1,\ldots,p\}$.

In the sequel for any non increasing 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

$$\mathcal{E}_a(R):=\left\{\theta\in\mathbb{R}^{\mathcal{I}},\sum_{i=1}^{p}\frac{\mathrm{var}(Y|X_{m_{i-1}})-\mathrm{var}(Y|X_{m_i})}{a_i^2}\leq R^2\mathrm{var}(Y|X)\right\}, \tag{22}$$

where for any $1 \leq i \leq p$, m_i refers to the set $\{1, \ldots, i\}$ and $m_0 = \emptyset$.

Let us explain why we call this set an ellipsoid. For instance, let us suppose 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

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

The main difference between this definition and the classical definition of an ellipsoid in the fixed design regression framework (as for instance in Baraud (2002)) 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 Gram-Schmidt process. If we call θ' the vector in $\mathbb{R}^{\mathcal{I}}$ 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:

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

The main advantage of definition (22) is that it does not depend on the covariance of X.

In the sequel we also consider the special case of ellipsoids with polynomial decay,

$$\mathcal{E}'_s(R) := \left\{ \theta \in \mathbb{R}^{\mathcal{I}}, \sum_{i=1}^{p} \frac{\mathrm{var}(Y|X_{m_{i-1}}) - \mathrm{var}(Y|X_{m_i})}{i^{-2s} \mathrm{var}(Y|X)} \leq R^2 \right\}, \tag{23}$$

where $s > 0$ and $R > 0$. First, we define two tests procedures of the form (7) 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 (2002).

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. (2004). We apply this last method to build a collection of models for our testing procedure (7) in Verzelen and Villers (2007).

5.1 Simultaneous Rates of testing of T_α over classes of ellipsoids

First, we define a test of the form (7) in order to test $\theta = 0$ against θ belongs to any of the ellipsoids $\mathcal{E}_a(R)$. For any $x > 0$, $[x]$ denotes the integer part of x.

The class of models $\mathcal{M}$ and the weights α_m depend on n and p:

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

Proposition 9. *Let us assume that*

$$n \geq 42 \left(\log\left(\frac{40}{\alpha}\right) \vee \log\left(\frac{1}{\delta}\right) \right) \tag{24}$$

For any ellipsoid $\mathcal{E}_a(R)$, the test T_α defined by (7) with Procedure P_1 and with the class of models given just above satisfies

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

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

$$\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq C_8 \left(\inf_{1\leq i\leq [n/2]} \left[a_{i+1}^2 R^2 + \frac{\sqrt{i\log\left(\frac{n/2}{\alpha\delta}\right)}}{n} \right] + \frac{1}{n}\log\left(\frac{n/2}{\alpha\delta}\right) \right) \tag{25}$$

if $n < 2p$, or

$$\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq C_8 \left[\inf_{1\leq i\leq p-1} \left[a_{i+1}^2 R^2 + \frac{\sqrt{i\log\left(\frac{2(p-1)}{\alpha\delta}\right)}}{n} \right] + \frac{\log\left(\frac{2(p-1)}{\alpha\delta}\right)}{n} \right] \bigwedge \left[\frac{\sqrt{p\log\left(\frac{2}{\alpha\delta}\right)} + \log\left(\frac{2}{\alpha\delta}\right)}{n} \right] \tag{26}$$

if $n \geq 2p$.

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

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

Besides, we will show in Proposition 14 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. 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.

We now turn to define a procedure of the form (7) 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 $\cup m_k$ where k belongs to $\left\{2^j, j\geq 0\right\} \cap \{1,\ldots,[n/2]\}$ 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 $\cup m_k$ where k belongs to $\left(\left\{2^j, j\geq 0\right\} \cap \{1,\ldots,p\}\right) \cup \{p\}$, 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 10. *Let us assume that*

$$n \geq 42\left(\log\left(\frac{40}{\alpha}\right) \vee \log\left(\frac{1}{\delta}\right)\right) \tag{27}$$

and that $R^2 \geq \sqrt{\log\log n}/n$. For any $s > 0$, The test procedure T_α defined by (7) with Procedure P_1 and with a class of models given just above satisfies:

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

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

$$\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq C_9(\alpha,\delta)\left[R^{2/(1+4s)}\left(\frac{\sqrt{\log\log n}}{n}\right)^{4s/(1+4s)} + R^2(n/2)^{-2s} + \frac{\log\log n}{n}\right] \quad (28)$$

if $n < 2p$ *or*

$$\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq C_9(\alpha,\delta)\left(\left[R^{2/(1+4s)}\left(\frac{\sqrt{\log\log p}}{n}\right)^{4s/(1+4s)} + \frac{\log\log p}{n}\right]\bigwedge\frac{\sqrt{p}}{n}\right) \quad (29)$$

if $n \geq 2p$. $C_9(\alpha,\delta)$ *is a constant which only depends on* α *and* δ.

Again, we retrieve similar results to those of Corollary 2 in Baraud et al. (2003) in the fixed design regression framework. For $s > 1/4$ and $n < 2p$, the rate of testing is of order $\left(\frac{\sqrt{\log\log n}}{n}\right)^{4s/(1+4s)}$. We show in the next subsection that this logarithmic factor is due to the adaptative 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 $\left(\frac{\sqrt{\log\log p}}{n}\right)^{4s/(1+4s)} \wedge \left(\frac{\sqrt{p}}{n}\right)$, 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 11. *Let us set the sequence* $(a_i)_{1\leq i\leq p+1}$ *and the positive number* R. *We introduce*

$$\rho^2_{a,n}(R) := \sup_{1\leq i\leq p}\left[\rho^2_{i,n} \wedge a_i^2R^2\right], \quad (30)$$

where $\rho^2_{i,n}$ *is defined by (16), then for any non singular 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 var(Y) *is fixed. If* $\alpha+\delta \leq 47\%$ *then*

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

This lower bound is once more analogous to the one in the fixed design regression framework. Contrary to the lower bounds in previous section, this bound 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:

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

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

We get the corresponding upper bound only if D^* is not too large compared to n, as shown by the following proposition.

Proposition 12. *Let us assume that* $n \geq 20\log\left(\frac{1}{\alpha}\right) \vee 41\log\left(\frac{2}{\delta}\right)$. *If* $R^2 > \frac{1}{n}$ *and* $D^* \leq n/2$, *the test* $\phi_{m_{D^*},\alpha}$ *defined by (10) satisfies*

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

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

$$\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq C_{10}(\alpha,\delta) \sup_{1\leq i\leq d}\left[\frac{\sqrt{i}}{n} \wedge a_i^2 R\right],$$

where $C_{10}(\alpha,\delta)$ *is a constant which only depends on* α *and* δ.

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 Baraud (2002)). If D^* is large and n is small, the bounds in Proposition 11 and 12 do not coincide. In this case, we do not know if this comes from the fact that the test in Proposition 12 does not depend on the knowledge of $\mathrm{var}(Y)$ or if one of the bounds in Proposition 11 and 12 is not sharp.

We are now interested in computing lower bounds of rate 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 $D \in \{1,\ldots,p\}$, S_{m_D} stands for the linear spaces of vectors θ such that only their D first coordinates are possibly non zero.

Proposition 13. *For* $D \geq 2$, *let us set*

$$\bar{\rho}^2_{D,n} := \frac{1}{2\sqrt{\log(2)}}\left(1 \wedge \log(1+2\eta^2)\right)\frac{\sqrt{\log\log(D+1)}\sqrt{D}}{n}. \tag{31}$$

Then, the following lower bound holds

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

if for all 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 14. *We fix a sequence* $(a_i)_{1\leq i\leq p+1}$. *For each* $R > 0$, *let us set*

$$\bar{\rho}^2_{a,R,n} = \sup_{1\leq D\leq p}\left[\bar{\rho}^2_{D,n} \wedge (R^2 a_D^2)\right]. \tag{32}$$

where $\bar{\rho}_{D,n}$ *is given by (31). Then, for any non singular covariance matrix* Σ *of the vector* X,

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

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 Spokoiny (1996) 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 14, 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 calculation, we find that the lower bound is of the order:

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

This shows that the power of the test T_α obtained in (29) 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 10 is rate optimal over the classes of ellipsoids $\mathcal{E}'_s(R)$. On the other hand, the test T_α defined in Proposition 9 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 2.3 that the test (7) 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) with $\mathcal{I} = \{1, \ldots, p\}$ and test the null hypothesis $''\theta = 0''$, which is equivalent to "Y is independent of X", at level $\alpha = 5\%$. Let $(X_i)_{1\leqslant i\leqslant p}$ 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 $i \neq j$. Besides, in this experiment the parameter θ is chosen such that only one of its components is possibly non zero. 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 to compare the procedure for independent, weakly and highly correlated covariates. We estimate the level 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 $r_{s/n}$ defined in (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 $(X_i)_{1\leqslant i\leqslant p}$ are independent. The number of covariates p equals 500 and the number of observations n equals 50 and 100. We set for any $i \in \{1,\ldots,p\}$, $\theta_i = Ri^{-s}$. We estimate the level 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 $(\theta_i)_{1\leqslant i\leqslant p}$. 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 (7) with different collections $\mathcal{M}$ and different choices fo 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:

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

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 2. When we are using the procedure P_1, the α_m's 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 (7) 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 test $\phi_{\{1\},\alpha}$ (10) 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 non zero.

2. The test $\phi_{\{2\},\alpha}$ of $\theta_2 = 0$ against $\theta_2 \neq 0$. This test corresponds to a single test where the model under the alternative is wrong. Adapting the proof of Proposition 7, we know that this test is approximately minimax on $\Theta[1,p]$ if the correlation between the covariates is constant and large. There is nothing special about the number 2, we could use any i between 2 and p.

Contrary to our procedures, these two tests are based on a deep knowledge of θ or $\mathrm{var}(X)$. 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 more powerful than $\phi_{\{2\},\alpha}$ and is close to the test $\phi_{\{1\},\alpha}$.

We estimate the level and the power of the testing procedures with 1000 simulations. For each simulation, we simulate the gaussian vector $(X_1,\ldots,X_p)$ 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

Null hypothesis is true, $\theta_1 = 0$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.043	0.045
15	0.044	0.049

Null hypothesis is false

$\theta_1 = 0.8$, $r_{s/n} = 1.78$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.48	0.48
15	0.81	0.81

$\theta_1 = 0.9$, $r_{s/n} = 4.26$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.86	0.86
15	0.99	0.99

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

$r_{s/n}$. 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 obervations.

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 level and the power of the test for weakly correlated covariates are similar to the level and the power obtained in the independent case. Hence, we recover the remark following Proposition 6: 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 following Theorem 5: 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 procedure and of the two tests $\phi_{\{1\},\alpha}$ and $\phi_{\{2\},\alpha}$ 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 seconds with Procedure P_2. As expected, because the collection of models $\mathcal{M}^1$ depends on p, Procedure P_1 is too conservative when p increases. For $p = 100$, the power of the test based on Procedure P_1 is similar to the power of the test $\phi_{\{2\},\alpha}$, while when p is larger, $T_{\mathcal{M}^1,P_1}$ is less powerful than $\phi_{\{2\},\alpha}$. Procedure P_2 is therefore recommanded in case of a large number of highly correlated covariates. The test based on Procedure P_2 is indeed more powerful than $\phi_{\{2\},\alpha}$, 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 non-zero component of θ contrary to ours. In practice, we advise to use Procedure P_2 if the number of covariates p is large, as 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.

Null hypothesis is true, $\theta_1 = 0$

$c = 0$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.043	0.045
15	0.044	0.049

$c = 0.1$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.042	0.04
15	0.058	0.06

$c = 0.8$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.018	0.045
15	0.019	0.052

Null hypothesis is false, $\theta_1 = 0.8$

$c = 0$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.48	0.48
15	0.81	0.81

$c = 0.1$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.49	0.49
15	0.81	0.82

$c = 0.8$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$
10	0.64	0.77
15	0.89	0.94

Table 2: First simulation study, independent and dependent case. $p = 30$ Percentages of rejection.

Null hypothesis is true, $\theta_1 = 0$

$p = 100$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$	$\phi_{\{1\},\alpha}$	$\phi_{\{2\},\alpha}$
10	0.01	0.056	0.051	0.050
15	0.016	0.053	0.047	0.050

$p = 500$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$	$\phi_{\{1\},\alpha}$	$\phi_{\{2\},\alpha}$
10	0.009	0.044	0.040	0.043
15	0.011	0.040	0.042	0.039

Null hypothesis is false, $\theta_1 = 0.8$

$p = 100$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$	$\phi_{\{1\},\alpha}$	$\phi_{\{2\},\alpha}$
10	0.60	0.77	0.91	0.62
15	0.85	0.92	0.99	0.82

$p = 500$

n	$T_{\mathcal{M}^1,P_1}$	$T_{\mathcal{M}^1,P_2}$	$\phi_{\{1\},\alpha}$	$\phi_{\{2\},\alpha}$
10	0.52	0.76	0.91	0.63
15	0.77	0.94	0.99	0.83

Table 3: First simulation study, dependent case: $c = 0.8$. Percentages of rejection.

Null hypothesis is true, $R = 0$

n	$T_{\mathcal{M}^2,P_1}$	$T_{\mathcal{M}^2,P_2}$	$T_{\mathcal{M}^3,P_1}$	$T_{\mathcal{M}^3,P_2}$
50	0.013	0.052	0.036	0.059
100	0.009	0.059	0.042	0.059

Null hypothesis is false, $R = 0.2, s = 0.5$

n	$T_{\mathcal{M}^2,P_1}$	$T_{\mathcal{M}^2,P_2}$	$T_{\mathcal{M}^3,P_1}$	$T_{\mathcal{M}^3,P_2}$
50	0.17	0.33	0.31	0.38
100	0.42	0.66	0.62	0.69

Table 4: Second simulation study. Percentages of rejection.

7 Proofs of Theorem 1, Proposition 3, 7, 9, 10, and 12

7.1 Proof of Theorem 1

This proof follows the same approach as the proof of Theorem 1 in Baraud et al. (2003). The main differences and difficulties come from the fact that the design is now random.

Using the definition of T_α, we notice that $\mathbb{P}_\theta(T_\alpha \leq 0) \leq \inf_{m\in\mathcal{M}} P_\theta(m)$ where

$$P_\theta(m) = \mathbb{P}_\theta\left(\frac{N_m\|\Pi_{V\cup m}\mathbf{Y} - \Pi_V\mathbf{Y}\|_n^2}{D_m\|\mathbf{Y} - \Pi_{V\cup m}\mathbf{Y}\|_n^2} \leq \overline{F}^{-1}_{D_m,N_m}(\alpha_m)\right). \tag{33}$$

First, we derive the distribution of the test statistic $\phi_m(\mathbf{X},\mathbf{Y})$ under $\mathbb{P}_\theta$, then we give an upper bound for $P_\theta(m)$ and finally we shall gather the results in order to find a subset of $\mathbb{R}^{\mathcal{I}}$ over which the power of T_α is larger than δ.

The distribution of Y conditionally to the set of variables $(X_{V\cup m})$ is of the form

$$Y = \sum_{i\in V\cup m} \theta_i^{V\cup m} X_i + \epsilon^{V\cup m}, \tag{34}$$

where the vector $\theta^{V\cup m}$ is a constant and $\epsilon^{V\cup m}$ is a zero mean gaussian variable independent of $X_{V\cup m}$, whose variance is $\text{var}(Y|X_{V\cup m})$. 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 $\phi_m(\mathbf{X},\mathbf{Y})$:

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

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 $\mathbf{X}_V$,

$$\mathbf{X}_i = \sum_{j\in V} \theta_j^{V,i}\mathbf{X}_j + \boldsymbol{\epsilon}_i^V. \tag{35}$$

where $\theta_j^{V,i}$ are constants and $\boldsymbol{\epsilon}_i^V$ is a zero-mean normal gaussian random variable whose variance is $\text{var}(X_i|X_V)$ and which is independent of $\mathbf{X}_V$. This enables us to express

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

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

$$\phi_m(\mathbf{X}, \mathbf{Y}) = \frac{N_m \|\Pi_{V^\perp \cap (V\cup m)} \left(\sum_{i\in m} \theta_i^{V\cup m} \boldsymbol{\epsilon}_i^V + \boldsymbol{\epsilon}^{V\cup m} \right) \|_n^2}{D_m \|\Pi_{(V\cup m)^\perp} \boldsymbol{\epsilon}^{V\cup m}\|_n^2} . \tag{36}$$

Let us define the random variable $Z_m^{(1)}$ and $Z_m^{(2)}$ where $Z_m^{(1)}$ refers to the numerator of (36) 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 $\text{var}(Y|X_{V\cup m})I_n$ 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. Additionally, $Z_m^{(2)}$ is independent of $\mathbf{X}_{V\cup m}$. Indeed, almost surely conditionally to $\mathbf{X}_{V\cup m}$, $Z_m^{(2)}/\text{var}(Y|X_{V\cup m})$ follows a χ^2 distribution with N_m degrees of freedom. This distribution does not depend on $\mathbf{X}_{V\cup m}$. As $Z_m^{(1)}$ and $Z_m^{(2)}$ are independent conditionally to $\mathbf{X}_{V\cup m}$ and as $Z_m^{(2)}$ is independent of $X_{V\cup m}$, $Z_m^{(1)}$ and $Z_m^{(2)}$ are independent.

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

$$(Z_m^{(1)}|X_{V\cup m}) \sim \text{var}(Y|X_{V\cup m})\chi^2 \left(\frac{\left\| \sum_{j\in m} \theta_j^{V\cup m} \Pi_{(V\cup m)\cap V^\perp} \boldsymbol{\epsilon}_j^V \right\|_n^2}{\text{var}(Y|X_{V\cup m})}, D_m \right) .$$

Let us derive the distribution of the non-central parameter. First, we simplify the projection term as $\boldsymbol{\epsilon}_j^V$ is a linear combinations 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

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

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 are able to derive the distribution of the non-central parameter:

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

To sum up, let us express simply the distribution of $Z_m^{(1)}$. Let U, V and W be three independent random variables which respectively follow a χ^2 distribution with $n-d$ degrees of freedom, a standard normal distribution and a χ^2 distribution with $D_m - 1$ degrees of freedom. Then,

$$Z_m^{(1)} \sim \mathrm{var}(Y|X_{V\cup m})\left[\left(\kappa_m\sqrt{U}+V\right)^2+W\right]. \tag{37}$$

In fact, κ_m^2 easily simplifies in a quotient of conditional variances. Let us first express $\mathrm{var}(Y|X_V)$ in term of $\mathrm{var}(Y|X_{m\cup V})$ using the decomposition (34) of Y.

$$\begin{aligned}
\mathrm{var}(Y|X_V) &= \mathrm{var}\left(\sum_{j\in V\cup m}\theta_j^{V\cup m}X_j + \epsilon^{V\cup m}\,|X_V\right)\\
&= \mathrm{var}\left(\sum_{j\in V\cup m}\theta_j^{V\cup m}X_j\,|X_V\right) + \mathrm{var}\left(\epsilon^{V\cup m}\,|X_V\right)\\
&= \mathrm{var}\left(\sum_{j\in V\cup m}\theta_j^{V\cup m}X_j\,|X_V\right) + \mathrm{var}\left(Y\,|X_{V\cup m}\right),
\end{aligned} \tag{38}$$

as $\epsilon^{V\cup m}$ is independent of $X_{V\cup m}$. Now using the definition of ϵ_j^V in (35), it turns out that

$$\begin{aligned}
\mathrm{var}\left(\sum_{j\in V\cup m}\theta_j^{V\cup m}X_j|X_V\right) &= \mathrm{var}\left(\sum_{j\in m}\theta_j^{V\cup m}X_j|X_V\right)\\
&= \mathrm{var}\left(\sum_{j\in m}\theta_j^{V\cup m}\epsilon_j^V\,|X_V\right)\\
&= \mathrm{var}\left(\sum_{j\in m}\theta_j^{V\cup m}\epsilon_j^V\right),
\end{aligned} \tag{39}$$

as the $(\epsilon_j^V)_{j\in m}$ are independent of X_V. Gathering formulae (38) and (39), we get

$$\kappa_m^2 = \frac{\mathrm{var}(Y|X_V) - \mathrm{var}(Y|X_{V\cup m})}{\mathrm{var}(Y|X_{V\cup m})}. \tag{40}$$

As we know the distribution of $\phi_m(\mathbf{X},\mathbf{Y})$ under the distribution $\mathbb{P}_\theta$, we are now in position to work out precise upper bounds for $P_\theta(m)$.

$$\begin{aligned}
P_\theta(m) &= \mathbb{P}_\theta\left(\frac{N_m}{D_m}Z_m^{(1)} \leq \overline{F}^{-1}_{D_m,N_m}(\alpha_m)Z_m^{(2)}\right)\\
&= \mathbb{P}_\theta\left(\frac{1}{\mathrm{var}(Y|X_{V\cup m})}\left(\frac{D_m}{N_m}\overline{F}^{-1}_{D_m,N_m}(\alpha_m)Z_m^{(2)} - Z_m^{(1)}\right)\geq 0\right).
\end{aligned} \tag{41}$$

Let us call Z_m the random variable defined by

$$Z_m := \frac{1}{\mathrm{var}(Y|X_{V\cup m})}\left(\frac{D_m}{N_m}\overline{F}^{-1}_{D_m,N_m}(\alpha_m)Z_m^{(2)} - Z_m^{(1)}\right).$$

It is possible to control the quantity $P_\theta(m)$ by bounding the deviations of Z_m.

Lemma 15. *For any $x > 0$, the random variable Z_m defined above satisfies the inequality:*

$$\mathbb{P}_\theta(Z_m - \mathbb{E}(Z_m) \geq c_m x + 2\sqrt{v_m x}) \leq \exp(-x)\,, \tag{42}$$

where c_m and v_m refer to:

$$\begin{aligned} c_m &:= \frac{2D_m}{N_m}\overline{F}^{-1}_{D_m,N_m}(\alpha_m), \\ v_m &:= D_m + (n-d)\left(2\kappa_m^2 + \kappa_m^4\right) + \frac{D_m^2}{N_m}\left(\overline{F}^{-1}_{D_m,N_m}(\alpha_m)\right)^2. \end{aligned}$$

We now apply this lemma choosing $x = L$,

$$\mathbb{P}_\theta\left(Z_m \geq \mathbb{E}(Z_m) + c_m L + 2\sqrt{v_m L}\right) \leq \delta.$$

Therefore, $\mathbb{P}_\theta(T_\alpha \leq 0) \leq \delta$ if for some $m \in \mathcal{M}$,

$$\mathbb{E}(Z_m) + c_m L + 2\sqrt{v_m L} \leq 0. \tag{43}$$

It is straightforward to compute the expectation of Z_m:

$$\mathbb{E}(Z_m) = -\kappa_m^2(n-d) - D_m + D_m\overline{F}^{-1}_{D_m,N_m}(\alpha_m).$$

Using this last equality, condition (43) is equivalent to the following inequality:

$$\kappa_m^2(n-d) \geq D_m\left(\overline{F}^{-1}_{D_m,N_m}(\alpha_m) - 1\right) + c_m L + 2\sqrt{L v_m}. \tag{44}$$

Thanks to the definition of v_m in Lemma 15, we now bound the term $2\sqrt{Lv_m}$. If $\kappa_m^2 \geq 2$, then

$$2\sqrt{Lv_m} \leq 2\sqrt{LD_m} + 2\kappa_m^2\sqrt{2L(n-d)} + 2\overline{F}^{-1}_{D_m,N_m}(\alpha_m)D_m\sqrt{\frac{L}{N_m}}.$$

On the other hand, if $\kappa_m^2 \leq 2$, we obtain an alternative upper bound using the inequality $2uv \leq 4u^2 + v^2/4$,

$$2\sqrt{Lv_m} \quad\leq\quad 2\sqrt{LD_m} + (n-d)\kappa_m^2/2 + 8L + 2\overline{F}^{-1}_{D_m,N_m}(\alpha_m)D_m\sqrt{\frac{L}{N_m}}.$$

Gathering these two inequalities, whatever the value of κ_m^2,

$$2\sqrt{Lv_m} \leq 2\sqrt{LD_m} + (n-d)\kappa_m^2\left(1/2 \vee 2\sqrt{\frac{2L}{n-d}}\right) + 2\overline{F}^{-1}_{D_m,N_m}(\alpha_m)D_m\sqrt{\frac{L}{N_m}} + 8L. \tag{45}$$

Combining the upper bound (45) with condition (44) enables to give a condition in term of κ_m^2. Indeed, $P_\theta(m) \leq \delta$ if

$$\kappa_m^2 \geq \frac{\frac{D_m}{N_m}\overline{F}^{-1}_{D_m,N_m}(\alpha_m)\left[N_m + 2\sqrt{N_m L} + 2L\right] + 2\sqrt{D_m L} - D_m + 8L}{(n-d)\left(1 - \left(\frac{1}{2} \vee 2\sqrt{\frac{2L}{n-d}}\right)\right)}. \tag{46}$$

To bound $\overline{F}^{-1}_{D_m,N_m}(\alpha_m)$, we use Lemma 1 in Baraud et al. (2003):

Lemma 16. *Let $u \in]0,1[$ and $\overline{F}^{-1}_{D,N}(u)$ be the $1-u$ quantile of a Fisher random variable with D and N degrees of freedom. Then we have*

$$\begin{aligned} D\overline{F}^{-1}_{D,N}(u) &\leq D + 2\sqrt{D\left(1+\frac{D}{N}\right)\log\left(\frac{1}{u}\right)} \\ &+ \left(1+2\frac{D}{N}\right)\frac{N}{2}\left[\exp\left(\frac{4}{N}\log\left(\frac{1}{u}\right)\right)-1\right]. \end{aligned} \tag{47}$$

Gathering inequalities (46) and (47), $P_\theta(m) \leq \delta$ if

$$\kappa_m^2 \geq \frac{A+B}{(n-d)\left(1-\left(\frac{1}{2}\vee 2\sqrt{\frac{2L}{n-d}}\right)\right)}, \tag{48}$$

where

$$\begin{aligned} A &:= 2\sqrt{D_m\left(1+\frac{D_m}{N_m}\right)\log\left(\frac{1}{\alpha_m}\right)}\left[1+2\sqrt{\frac{L}{N_m}}+2\frac{L}{N_m}\right]+2D_m\left[\sqrt{\frac{L}{N_m}}+\frac{L}{N_m}\right]+2\sqrt{D_mL}, \\ B &:= \left(1+2\frac{D_m}{N_m}\right)\frac{N_m}{2}\left[\exp\left(\frac{4}{N}\log\left(\frac{1}{\alpha_m}\right)\right)-1\right]\left(1+2\sqrt{\frac{L}{N_m}}+2\frac{L}{N_m}\right)+8L. \end{aligned}$$

By factorizing and bounding the last two terms of A, we get

$$\begin{aligned} 2D_m\left[\sqrt{\frac{L}{N_m}}+\frac{L}{N_m}\right]+2\sqrt{D_mL} &= 2\sqrt{D_mL}\left(1+\sqrt{\frac{D_m}{N_m}}+\sqrt{\frac{D_m}{N_m}}\sqrt{\frac{L}{N_m}}\right) \\ &\leq 2\sqrt{D_mL}\left[1+\sqrt{\frac{D_m}{N_m}}\right]\left[1+2\sqrt{\frac{L}{N_m}}+2\frac{L}{N_m}\right]. \end{aligned}$$

It follows that

$$\begin{aligned} A &\leq 2\sqrt{D_m}\left(1+\sqrt{\frac{D_m}{N_m}}\right)\left(1+2\sqrt{\frac{L}{N_m}}+2\frac{L}{N_m}\right)\left[\sqrt{L}+\sqrt{L_m}\right] \\ &\leq 4\sqrt{D_m}l_m\left(1+\sqrt{\frac{D_m}{N_m}}\right)\left[\sqrt{\log\left(\frac{1}{\alpha_m\delta}\right)}\right]. \end{aligned} \tag{49}$$

Using the inequality $\exp(u)-1 \leq u\exp(u)$ which holds for all $u>0$, we derive that

$$\begin{aligned} B &\leq 2\left(1+2\frac{D_m}{N_m}\right)\log\left(\frac{1}{\alpha_m}\right)\exp\left[\frac{4}{N_m}\log\left(\frac{1}{\alpha_m}\right)\right]\left[1+2\sqrt{\frac{L}{N_m}}+2\frac{L}{N_m}\right]+8L \\ &\leq \log\left(\frac{1}{\alpha_m\delta}\right)\left(8\vee k_ml_m\left(1+\frac{2D_m}{N_m}\right)\right). \end{aligned} \tag{50}$$

Combining inequalities (48), (49), and (50) we obtain the condition (12). Under assumption $H_{\mathcal{M}}$, $L_m \leq N_m/10$ for all $m \in \mathcal{M}$ and $L \leq N_m/21$. The terms L/N_m, L_m/N_m, k_m, and l_m are bounded by a constant and the second part of the theorem follows easily.

7.2 Proof of Lemma 15

We prove this deviation inequality thanks to Laplace method. First of all, one has to upper bound the Laplace transform of the variable

$$Z_m \sim \frac{D_m}{N_m}\overline{F}^{-1}_{D_m,N_m}(\alpha_m)T - \left((\kappa_m\sqrt{U}+V)^2+W\right),$$

where we recall that T, U, V, and W are independent random variables which follow respectively a χ^2 distribution with N_m degrees of freedom, a χ^2 distribution with $n-d$ degrees of freedom, a standard normal distribution and a χ^2 distribution with $D_m - 1$ degree of freedom. To keep the formulae as short as possible, λ_m will refer to $\frac{D_m}{N_m}\overline{F}^{-1}_{D_m,N_m}(\alpha_m)$.

$$\begin{aligned}\mathbb{E}\left[\exp\left(-t\left(\kappa_m\sqrt{U}+V\right)^2\right)\right] &= \int \exp\left(-t\left(\kappa_m\|x\|_{n-d}+y\right)^2\right)\frac{1}{(2\pi)^{(n-d+1)/2}}\exp\left(-\frac{\|x\|^2_{n-d}-y^2}{2}\right)dxdy \\ &= \frac{1}{\sqrt{1+2t}}\left[\frac{1+2t}{1+2t[\kappa_m^2+1]}\right]^{(n-d)/2},\end{aligned}$$

by standard Gaussian computation. After multiplication by the Laplace transform of $\frac{D_m}{N_m}\overline{F}^{-1}_{D_m,N_m}(\alpha_m)T$ and W, we get:

$$\mathbb{E}\left[\exp\left(tZ_m\right)\right] = \frac{(1+2t)^{N_m/2}}{(1+2t[\kappa_m^2+1])^{(n-d)/2}\left(1-2t\lambda_m\right)^{N_m/2}}.$$

Clearly, the expectation of Z_m is

$$\mathbb{E}(Z_m) = \lambda_m N_m - (\kappa_m^2(n-d)+D_m).$$

One then obtains $\Psi_m(t)$ the log-Laplace transform of $Z_m - \mathbb{E}(Z_m)$:

$$\begin{aligned}\Psi_m(t) &= \frac{N_m}{2}\log\left(\frac{1+2t}{1-2t\lambda_m}\right) - \frac{n-d}{2}\log\left(1+2t[\kappa_m^2+1]\right) - t\mathbb{E}(Z_m) \\ &= -\frac{D_m}{2}\log(1+2t) - \frac{n-d}{2}\log\left(1+\frac{2t\kappa_m^2}{1+2t}\right) - \frac{N_m}{2}\log(1-2t\lambda_m) - t\mathbb{E}(Z_m).\end{aligned}$$

Using the inequality $\log(1+u) \geq u - u^2/2$ which holds for all $u>0$, we derive that for any $t \geq 0$,

$$\begin{aligned}\Psi_m(t) &\leq D_m t^2 + (n-d)\left[-\frac{t\kappa_m^2}{1+2t} + t\kappa_m^2 + \frac{t^2\kappa_m^4}{(1+2t)^2}\right] - \frac{N_m}{2}\log(1-2t\lambda_m) - t\lambda_m N_m \\ &\leq D_m t^2 + (n-d)\left[\frac{2t^2\kappa_m^2}{1+2t} + \frac{t^2\kappa_m^4}{(1+2t)^2}\right] - \frac{N_m}{2}\log(1-2t\lambda_m) - t\lambda_m N_m \\ &\leq t^2\left[D_m + (n-d)(2\kappa_m^2+\kappa_m^4)\right] - \frac{N_m}{2}\log(1-2t\lambda_m) - t\lambda_m N_m.\end{aligned}$$

For any $0 \leq u \leq 1/2$, it holds that $-u - 1/2\log(1-2u) \leq \frac{u^2}{1-2u}$ (compare the power series). As a consequence, for any $0 \leq t \leq \frac{\lambda_m}{2}$,

$$\begin{aligned}\Psi_m(t) &\leq t^2\left[D_m + (n-d)(2\kappa_m^2+\kappa_m^4)\right] + N_m\frac{\lambda_m^2 t^2}{1-2t\lambda_m} \\ &\leq \frac{t^2}{1-2\lambda_m t}\left(D_m + (n-d)(2\kappa_m^2+\kappa_m^4) + N_m\lambda_m^2\right). \qquad (51)\end{aligned}$$

We now refer to Birgé and Massart (1998), where it is proved that if

$$\log\left(\mathbb{E}\left[e^{tZ}\right]\right) \leq \frac{vt^2}{2(1-ct)},$$

then for any positive x,

$$\mathbb{P}\left(Z \geq cx + \sqrt{2vx}\right) \leq e^{-x}.$$

Applying this property to the upper bound (51) and replacing λ_m by its value enable to prove (42).

7.3 Proof of Proposition 3

We first recall the classical upper bound for the binomial coefficient (see for instance (2.9) in Massart (2007)).

$$\log|\mathcal{M}(k,p)| = \log\tbinom{p}{k} \leq k\log\left(\frac{ep}{k}\right).$$

As a consequence, $\log(1/\alpha_m) \leq \log(1/\alpha) + k\log\left(\frac{ep}{k}\right)$. The assumption on n in Proposition 3 therefore implies hypothesis $H_{\mathcal{M}}$ applied to this class of models. Thus, we are in position to apply the second result of Theorem 1. Moreover, the assumption on n implies that $n \geq 11k$ and D_m/N_m is thus smaller than $1/10$ for any model m in $\mathcal{M}(k,p)$. Formula (13) in Theorem 1 then translates into

$$\triangle(m) \leq \frac{(1+\sqrt{0.1})C_1\left(\sqrt{k^2\log\left(\frac{ep}{k}\right)} + \sqrt{k\log\left(\frac{1}{\alpha\delta}\right)}\right) + 1.2C_2\left(k\log\left(\frac{ep}{k}\right) + \log\left(\frac{1}{\alpha\delta}\right)\right)}{n},$$

and it follows that Proposition 3 holds.

7.4 Proof of Proposition 7

We fix the constant C_5 to be $10 \vee 2C_4'$ where C_4' is defined below and C_6 to be 21. This choice of constants allows the procedure T_α to satisfy Hypothesis $H_{\mathcal{M}}$. An argument similar to the proof of Proposition 3 allows to show easily that there exists a universal constant C_3' such that if we set

$$\rho_1'^2 \quad = \quad \frac{C_3'\left(\log(p) + \log\left(\frac{2}{\alpha\delta}\right)\right)}{n} = \frac{C_3'}{n}\log\left(\frac{2p}{\alpha\delta}\right), \tag{52}$$

then $\frac{\|\theta\|^2}{\text{var}(Y)-\|\theta\|^2} \geq \rho_1'^2$ implies that $\mathbb{P}_\theta\left(T_\alpha > 0\right) \geq 1-\delta$. Here, the factor 2 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 $\frac{\lambda^2}{\text{var}(Y)-\lambda^2} = \rho^2$ and let $\theta \in \Theta[1,p]$ such that $\|\theta\|^2 = \lambda^2$. As $\text{corr}(X_1, X_i) = c$ for any i in $\{2\ldots p\}$,

$$\frac{\text{var}(Y) - \text{var}(Y|X_1)}{\text{var}(Y|X_1)} \geq \frac{c\lambda^2}{\text{var}(Y) - c\lambda^2}.$$

We now apply Theorem 1 to $\phi_{\{1\},\alpha/2}$ under $H_{\mathcal{M}}$. There exists a universal constant C_4' such that $\mathbb{P}_\theta\left(\phi_{\{1\},\alpha/2} > 0\right) \geq 1-\delta$ if

$$\frac{c\lambda^2}{\text{var}(Y) - c\lambda^2} \geq \frac{C_4'}{n}\log\left(\frac{2}{\alpha\delta}\right).$$

This last condition is equivalent to

$$\frac{\lambda^2}{\mathrm{var}(Y)} \geq \frac{C_4'}{nc + cC_4' \log\left(\frac{2}{\alpha\delta}\right)} \log\left(\frac{2}{\alpha\delta}\right). \tag{53}$$

Let us assume that $c \geq \log\left(\frac{2}{\alpha\delta}\right) / \log\left(\frac{2p}{\alpha\delta}\right)$. As $n \geq 2C_4' \log\left(\frac{2p}{\alpha\delta}\right)$ (hypothesis (20) and definition of C_5), $nc \geq 2C_4' \log\left(\frac{2}{\alpha\delta}\right)$. As a consequence, condition (53) is implied by:

$$\rho^2 \geq \frac{2C_4'}{nc} \log\left(\frac{2}{\alpha\delta}\right). \tag{54}$$

Let us define C_7 as the supremum of C_3' and $2C_4'$. Combining (52) and (54) allows to conclude that $\mathbb{P}_\theta\left(T_\alpha > 0\right) \geq 1 - \delta$ if

$$\rho^2 \geq \frac{C_7}{n}\left(\log\left(\frac{2p}{\alpha\delta}\right) \bigwedge \frac{1}{c}\log\left(\frac{2}{\alpha\delta}\right)\right).$$

If c is smaller than $\log\left(\frac{2}{\alpha\delta}\right) / \log\left(\frac{2p}{\alpha\delta}\right)$, this last result also holds by (52).

7.5 Proof of Proposition 9

First, we have to 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:

$$n \geq 20 \log\left(\frac{n}{\alpha}\right). \tag{55}$$

Hypothesis (24) implies that $n/2 \geq 20\log\left(\frac{40}{\alpha}\right)$. Besides, for any $n > 0$ it holds that $n/2 \geq 20\log\left(\frac{n}{40}\right)$. Combining these two lower bounds enables to obtain (55). The second condition of $H_{\mathcal{M}}$ holds if $n \geq 42\log\left(\frac{1}{\delta}\right)$ which is a consequence of hypothesis (24).

Let us first consider the case $n < 2p$. Let us apply Theorem 1 under hypothesis $H_{\mathcal{M}}$ to T_α. $\mathbb{P}_\theta\left(T_\alpha > 0\right) \geq 1 - \delta$ for all $\theta \in \Theta$ such that

$$\exists i \in \{1, \ldots, [n/2]\}, \frac{\mathrm{var}(Y) - \mathrm{var}\left(Y|X_{m_i}\right)}{\mathrm{var}\left(Y|X_{m_i}\right)} \geq C_5' \frac{\sqrt{i \log\left(\frac{[n/2]}{\alpha\delta}\right)} + \log\left(\frac{[n/2]}{\alpha\delta}\right)}{n}, \tag{56}$$

where C_5' is universal constant (equals $2C_1 \vee 4C_2$).

Let θ be an element of $\mathcal{E}_a(R)$ which satisfies

$$\|\theta\|^2 \geq (1 + C_5')\left(\mathrm{var}(Y|X_{m_i}) - \mathrm{var}(Y|X)\right) + (1 + C_5')\mathrm{var}(Y|X)\frac{\sqrt{i \log\left(\frac{[n/2]}{\alpha\delta}\right)} + \log\left(\frac{[n/2]}{\alpha\delta}\right)}{n}.$$

Using hypothesis (24), we show that, for any i between 1 and $[n/2]$, $\frac{\sqrt{i \log\left(\frac{[n/2]}{\alpha\delta}\right)} + \log\left(\frac{[n/2]}{\alpha\delta}\right)}{n} \leq 1$. It is then straighforward to check that θ satisfies (56).

As θ belongs to the set $\mathcal{E}_a(R)$,

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

As a consequence if θ belong to $\mathcal{E}_a(R)$ and satifies

$$\|\theta\|^2 \geq (1+C_5')\mathrm{var}(Y|X)\left[\left(a_{i+1}^2 R^2 + \frac{\sqrt{i \log\left(\frac{[n/2]}{\alpha\delta}\right)}}{n}\right) + \frac{1}{n}\log\left(\frac{[n/2]}{\alpha\delta}\right)\right], \tag{57}$$

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

$$\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq (1+C_5')\left[\inf_{1\leq i\leq [n/2]}\left(a_{i+1}^2 R^2 + \frac{\sqrt{i \log\left(\frac{n/2}{\alpha\delta}\right)}}{n}\right) + \frac{1}{n}\log\left(\frac{n/2}{\alpha\delta}\right)\right], \tag{58}$$

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 $p-1$ firsts tests, we obtain as in the previous case that $\mathbb{P}_\theta(T_\alpha \leq 0) \leq \delta$ if

$$\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq (1+C_5')\left[\inf_{1\leq i\leq (p-1)}\left(a_{i+1}^2 R^2 + \frac{\sqrt{i \log\left(\frac{(p-1)/2}{\alpha\delta}\right)}}{n}\right) + \frac{1}{n}\log\left(\frac{(p-1)/2}{\alpha\delta}\right)\right].$$

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

$$\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq C_5' \frac{\sqrt{p \log\left(\frac{2}{\alpha\delta}\right)} + \log\left(\frac{2}{\alpha\delta}\right)}{n}.$$

Gathering these two conditions allows to prove (26).

7.6 Proof of Proposition 10

The approach behind this proof is similar to the one for Proposition 9. First, we check that our class of models $\mathcal{M}$ and weights α_m satisfy hypothesis $H_{\mathcal{M}}$ as in the previous proof.

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

$$|\mathcal{M}| \leq 1 + \log(n/2 \wedge p)/\log(2) \leq \log(n \wedge 2p)/\log(2). \tag{59}$$

We deduce from (59) that there exists a constant $C(\alpha,\delta)$ only depending on α and δ such that for all $m \in \mathcal{M}$,

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

First, let us consider the case $n < 2p$. We apply Theorem 1 under the assumption $H_{\mathcal{M}}$. As in the proof of Proposition 9, we obtain that $\mathbb{P}_\theta(T_\alpha > 0) \geq 1-\delta$ if

$$\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2} \geq C'(\alpha,\delta)\left[\inf_{i\in\{2^j,j\geq 0\}\cap\{1,\ldots,[n/2]\}}\left(R^2(i+1)^{-2s}+\frac{\sqrt{i\log\log n}}{n}\right)+\frac{\log\log n}{n}\right],$$

where $C'(\alpha,\delta)$ is a constant which only depends on α and δ. It is worth noting that $R^2 i^{-2s} \leq \frac{\sqrt{i\log\log n}}{n}$ if and only if

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

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

$$\begin{aligned}\inf_{i\in\{2^j,j\geq 0\}\cap\{1,\ldots,[n/2]\}}\left(R^2 i^{-2s}+\frac{\sqrt{i\log\log n}}{n}\right) &\leq 2\frac{\sqrt{i'\log\log n}}{n}\\ &\leq 2\sqrt{2}R^{2/(1+4s)}\left(\frac{\sqrt{\log\log n}}{n}\right)^{4s/(1+4s)}. \qquad (60)\end{aligned}$$

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

$$\inf_{i\in\{2^j,j\geq 0\}\cap\{1,\ldots,[n/2]\}}\left(R^2 i^{-2s}+\frac{\sqrt{i\log\log n}}{n}\right)\leq 2R^2 i'^{-2s}\leq 2R^2\left(\frac{n}{2}\right)^{-2s}. \qquad (61)$$

Gathering inequalities (60) and (61) allows to prove (28).

We now turn to the case $n\geq 2p$. As in the proof of Proposition 9, 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 (29).

7.7 Proof of Proposition 12

We first note that the assumption about R^2 implies that $D^*\geq 2$. As N_m is larger than $n/2$, it is straightforward to show that this test satisfies condition $H_{\mathcal{M}}$. As a consequence, we can apply the second part of Theorem 1. $\mathbb{P}_\theta(T^*_\alpha\leq 0)\leq\delta$ for any θ such that

$$\frac{\mathrm{var}(Y)-\mathrm{var}(Y|X_{m_{D^*}})}{\mathrm{var}(Y|X_{m_{D^*}})}\geq C_2'(\alpha,\delta)\frac{\sqrt{D^*}}{n}, \qquad (62)$$

where $C_2'(\alpha,\delta)$ only depends on α and δ. Now, we use the same sketch as in the proof of Proposition 9. For any $\theta\in\mathcal{E}_a(R)$, condition (62) is equivalent to:

$$\|\theta\|^2\geq(\mathrm{var}(Y|X_{m_{D^*}})-\mathrm{var}(Y|X))\left(1+C_2'(\alpha,\delta)\frac{\sqrt{D^*}}{n}\right)+\mathrm{var}(Y|X)C_2'(\alpha,\delta)\frac{\sqrt{D^*}}{n}. \qquad (63)$$

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

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

As $\sqrt{D^*}/n$ is smaller than one, condition (63) is implied by

$$\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq (1 + C_2'(\alpha,\delta)) \left(a_{D^*}^2 R^2 + \frac{\sqrt{D^*}}{n} \right).$$

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

$$\frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} \geq 2(1 + C_2'(\alpha,\delta)) \sup_{1\leq i\leq p} \left[\frac{\sqrt{i}}{n} \wedge a_i^2 R^2 \right].$$

8 Proofs of Theorem 5, Proposition 2, 4, 6, 8, 11, 13, and 14

8.1 Proof of Theorem 5

This proof follows the general method for obtaining lower bounds described in Section 7.1 in Baraud (2002). 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

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

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

$$\begin{aligned} \beta_I(\Theta[k,p,\rho]) &\geq \inf_{\phi_\alpha \in \Phi_\alpha} \mathbb{P}_{\mu_\rho}[\phi_\alpha = 0] \\ &\geq 1 - \alpha - \sup_{A,\ \mathbb{P}_0(A)\leq\alpha} |\mathbb{P}_{\mu_\rho}(A) - \mathbb{P}_0(A)| \\ &\geq 1 - \alpha - \frac{1}{2}\|\mathbb{P}_{\mu_\rho} - \mathbb{P}_0\|_{TV}, \end{aligned} \tag{64}$$

where $\|\mathbb{P}_{\mu_\rho} - \mathbb{P}_0\|_{TV}$ 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

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

Then, we get the upper bound

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

Thus, we deduce from (64) that

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

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

$$\log\left(\mathbb{E}_0\left[L^2_{\mu_{\rho^*}}(\mathbf{Y},\mathbf{X})\right]\right) \leq \mathcal{L}(\eta), \tag{65}$$

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

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

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 $\mathbb{R}^{\mathcal{I}}$ defined by

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

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

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

We aim at bounding the quantity $\mathbb{E}_0(L^2_{\mu_\rho})$ and obtaining an inequality of the form (65). First, we work out $L_{\mu_{m,\zeta^m,\rho}}$:

$$\begin{aligned} L_{\mu_{m,\zeta^m,\rho}}(\mathbf{Y},\mathbf{X}) &= \Bigg[\left(\frac{1}{1-\frac{\lambda^2 k}{\mathrm{var}(Y)}}\right)^{n/2}\exp\Bigg(-\frac{\|\mathbf{Y}\|_n^2}{2}\frac{\lambda^2 k}{\mathrm{var}(Y)(\mathrm{var}(Y)-\lambda^2 k)} \\ &+ \lambda\sum_{j\in m}\zeta^m_j\frac{<\mathbf{Y},\mathbf{X}_j>_n}{\mathrm{var}(Y)-\lambda^2 k} - \lambda^2\sum_{j,j'\in m}\zeta^m_j\zeta^m_{j'}\frac{<\mathbf{X}_j,\mathbf{X}_{j'}>_n}{2(\mathrm{var}(Y)-\lambda^2 k)}\Bigg)\Bigg], \end{aligned} \tag{66}$$

where $< . >_n$ 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 $\mathbb{E}_0\left(L_{\mu_{m_1,\zeta^1,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{m_2,\zeta^2,\rho}}(\mathbf{Y},\mathbf{X})\right)$. 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:

$$\begin{aligned} m_3 &= \left\{j\in m_1\cap m_2 | \zeta^1_j = \zeta^2_j\right\} \\ m_4 &= \left\{j\in m_1\cap m_2 | \zeta^1_j = -\zeta^2_j\right\}. \end{aligned}$$

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 $\zeta_j = \zeta^1_j$ if $j \in m_1$ and $\zeta_j = \zeta^2_j$ if $j \in m_2 \setminus m_1$. Using these notations, we compute the expectation of $L_{m_1,\zeta^1,\rho}(\mathbf{Y},\mathbf{X}) L_{m_2,\zeta^2,\rho}(\mathbf{Y},\mathbf{X})$.

$$\mathbb{E}_0\left(L_{\mu_{m_1,\zeta^1,\rho}}(\mathbf{Y},\mathbf{X}) L_{\mu_{m_2,\zeta^2,\rho}}(\mathbf{Y},\mathbf{X})\right) = \left(\frac{1}{\mathrm{var}(Y)(1-\frac{\lambda^2 k}{\mathrm{var}(Y)})^2}\right)^{n/2} |A|^{-n/2} , \qquad (67)$$

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

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

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

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

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

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

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

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

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 non-zero eigenvalues, it is then straightforward to derive the determinant of A

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

Gathering this equality with (67) yields

$$\mathbb{E}_0\left(L_{\mu_{m_1,\zeta^1,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{m_2,\zeta^2,\rho}}(\mathbf{Y},\mathbf{X})\right) = \left[\frac{1}{1-\frac{\lambda^2(2|m_3|-|m_1\cap m_2|)}{\mathrm{var}(Y)}}\right]^n . \qquad (68)$$

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 (68) depends on ζ^1 and ζ^2 only towards 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

$$\mathbb{E}_0(L_{\mu_{m_1,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{m_2,\rho}}(\mathbf{Y},\mathbf{X})) \;=\; \mathbb{E}_0\left[\frac{1}{1-\frac{\lambda^2 Z}{\mathrm{var}(Y)}}\right]^n . \qquad (69)$$

When Z is non-positive, this expression is smaller than one. On the other hand, when Z is non negative:

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

as $\log(1+x)\leq x$ and as Z is smaller than k. We define an event $\mathbb{A}$ such that $\{Z>0\}\subset\mathbb{A}\subset\{Z\geq 0\}$ 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 (69) is smaller or equal to one. All in all, we bound (69) by:

$$\mathbb{E}_0(L_{\mu_{m_1,\rho}}(\mathbf{Y},\mathbf{X})L_{\mu_{m_2,\rho}}(\mathbf{Y},\mathbf{X})) \leq \frac{1}{2}+\mathbb{E}_0\left[\mathbf{1}_{\mathbb{A}}\exp\left[n\frac{\frac{\lambda^2 Z}{\mathrm{var}(Y)}}{1-\frac{\lambda^2 k}{\mathrm{var}(Y)}}\right]\right], \qquad (70)$$

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.

$$\begin{aligned}\mathbb{E}_0\left[\mathbf{1}_{\mathbb{A}}\exp\left[n\frac{\frac{\lambda^2 Z}{\mathrm{var}(Y)}}{1-\frac{\lambda^2 k}{\mathrm{var}(Y)}}\right]\right] &\leq \mathbb{P}(\mathbb{A})^{1-v}\left[\mathbb{E}_0\exp\left(\frac{n}{v}\frac{\frac{\lambda^2 Z}{\mathrm{var}(Y)}}{1-\frac{\lambda^2 k}{\mathrm{var}(Y)}}\right)\right]^v\\ &\leq \left(\frac{1}{2}\right)^{1-v}\left[\cosh\left(\frac{n\lambda^2}{v(\mathrm{var}(Y)-\lambda^2 k)}\right)\right]^{|m_1\cap m_2|v} . \qquad (71)\end{aligned}$$

Gathering inequalities (70) and (71) yields

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

Following the approach of Baraud (2002) in Section 7.2, we note that if m_1 and m_2 are taken uniformly and indepently 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

$$\mathbb{E}_0\left[L^2_{\mu_\rho}(\mathbf{Y},\mathbf{X})\right] \leq \frac{1}{2} + \left(\frac{1}{2}\right)^{1-v} \mathbb{E}\left(\cosh\left(\frac{n\lambda^2}{v(\mathrm{var}(Y)-\lambda^2 k)}\right)^{vT}\right) \tag{72}$$

where T is a random variable distributed according to a Hypergeometric distribution with parameters p, k and k/p. We know from Aldous (1985, p.173) that T has the same distribution as the random variable $\mathbb{E}(W|\mathcal{B}_p)$ where W is binomial random variable of parameters k, k/p and $\mathcal{B}_p$ some suitable σ-algebra. By a convexity argument, we then upper bound (72).

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

To get the upper bound on the total variation distance appearing in (64), we aim at constraining this last expression to be smaller than $1+\eta^2$. This is equivalent to the following inequality:

$$2^v \exp\left[k\log\left(1+\frac{k}{p}\left(\cosh\left(\frac{n\lambda^2 k}{vk(\mathrm{var}(Y)-\lambda^2 k)}\right)^{v} - 1\right)\right)\right] \leq 1+2\eta^2\ . \tag{73}$$

We now choose $v = \frac{\mathcal{L}(\eta)}{\log(2)} \wedge 1$. If v is strictly smaller than one, then (73) is equivalent to:

$$k\log\left[1+\frac{k}{p}\left(\cosh\left(\frac{n\lambda^2 k}{vk(\mathrm{var}(Y)-\lambda^2 k)}\right)^{v} - 1\right)\right] \leq \frac{\log(1+2\eta^2)}{2}. \tag{74}$$

It is straightforward to show that this last inequality also implies (73) if v equals one. We now suppose that

$$\frac{n\lambda^2}{v(\mathrm{var}(Y)-\lambda^2 k)} \leq \log\left((1+u)^{\frac{1}{v}} + \sqrt{(1+u)^{\frac{2}{v}}-1}\right), \tag{75}$$

where $u = \frac{p\mathcal{L}(\eta)}{k^2}$. Using the classical equality $\cosh\left[\log(1+x+\sqrt{2x+x^2})\right] = 1+x$ with $x = (1+u)^{\frac{1}{v}} - 1$, we deduce that inequality (75) implies (74) because

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

For any $\beta \geq 1$ and any $x > 0$, it holds that $(1+x)^\beta \geq 1+\beta x$. As $\frac{1}{v} \geq 1$, condition (75) is implied by:

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

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

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

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

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

To resume, if we take ρ^2 smaller than this last quantity, then

$$\beta_I\left(\Theta[k,p,\rho]\right)\geq\delta\ .$$

To prove the second part of the theorem, one has to observe that $\alpha+\delta\leq 53\%$ implies that $\mathcal{L}(\eta)\geq\frac{1}{2}$.

8.2 Proof of Proposition 4

Let us first assume that the covariance matrix of X is the identity. We argue as in the proof of Theorem 5 taking $k=p$. The sketch of the proof remains unchanged except that we slightly modify the last part. Inequality (74) becomes

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

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 using Gram-Schmidt process. Applying the previous argument to this new sequence of covariates allows to conclude.

8.3 Proof of Proposition 2

Let us apply proposition 4. For any $\rho\leq s(\alpha,\delta)\frac{\sqrt{D_m}}{n}$ there exists some $\theta\in S_m$ such that $\frac{\|\theta\|^2}{\mathrm{var}(Y)-\|\theta\|^2}=\rho^2$ and $\mathbb{P}_\theta(T_m\leq 0)\geq\delta$. Here, $s(\alpha,\delta)$ refers to some function only depending on α and δ. In the proof of Theorem 1, we have shown in (36) 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 $\kappa_m^2=\rho^2$. As a consequence, the distribution of ϕ_m under $\mathbb{P}_{\theta'}$ is the same as its distribution under $\mathbb{P}_\theta$, and therefore

$$\mathbb{P}_{\theta'}\left(T_m\leq 0\right)\geq\delta.$$

8.4 Proof of Proposition 6

This lower bound for dependent gaussian covariates is proved through the same approach as Theorem 5. We define the measure μ_ρ as in that proof. Under the 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 $\mathbb{E}_{0,\Sigma}\left[L^2_{\mu_\rho}(\mathbf{Y}, \mathbf{X})\right]$. 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 correspondance with the set $\{1, \ldots, p\}$.

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

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

$$\mathbb{E}_{0,\Sigma}\left(L_{\mu_{i,\rho}}(\mathbf{Y}, \mathbf{X}) L_{\mu_{j,\rho}}(\mathbf{Y}, \mathbf{X})\right) \leq \frac{1}{2} + \frac{1}{2}\left(\frac{\mathrm{var}(Y)}{\mathrm{var}(Y) - |\mathrm{corr}(i,j)|\lambda^2}\right)^n . \tag{76}$$

If $i \neq j$, $|\mathrm{corr}(i,j)|$ is smaller than c and if $i = j$, $\mathrm{corr}(i,j)$ is exactly one. As a consequence, taking the expectation of (76) with respect to i and j yields the upper bound

$$\mathbb{E}_{0,\Sigma}\left(L^2_{\mu_\rho}(\mathbf{Y}, \mathbf{X})\right) \leq \frac{1}{2} + \frac{1}{2}\left(\frac{1}{p}\left(\frac{\mathrm{var}(Y)}{\mathrm{var}(Y) - \lambda^2}\right)^n + \frac{p-1}{p}\left(\frac{\mathrm{var}(Y)}{\mathrm{var}(Y) - c\lambda^2}\right)^n\right) . \tag{77}$$

Recall that we want to constrain this quantity (77) to be smaller than $1 + \eta^2$. In particular, this holds if the two following inequalities hold:

$$\frac{1}{p}\left(\frac{\mathrm{var}(Y)}{\mathrm{var}(Y) - \lambda^2}\right)^n \leq \frac{1}{p} + \eta^2 \tag{78}$$

$$\frac{p-1}{p}\left(\frac{\mathrm{var}(Y)}{\mathrm{var}(Y) - c\lambda^2}\right)^n \leq \frac{p-1}{p} + \eta^2 . \tag{79}$$

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 (78) holds if

$$\frac{\lambda^2}{\mathrm{var}(Y) - \lambda^2} \leq \frac{1}{n}\log(1 + p\eta^2) , \tag{80}$$

whereas condition (79) is implied by

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

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

$$\frac{\lambda^2}{\mathrm{var}(Y) - \lambda^2} \leq \frac{1}{nc}\log(1 + \eta^2) . \tag{81}$$

Gathering conditions (80) and (81) allows to conclude and to obtain the desired lower bound (19).

8.5 Proof of Proposition 8

The sketch of the proof and the notations are analogous to the one in Proposition 6. The upper bound (76) still holds:

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

Using the stationarity of the covariance function, we derive from (76) the following upper bound:

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

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 $p-1$:

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

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

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

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

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

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

$$\rho^2 \quad\leq\quad \frac{1}{n}\log\left(1+\frac{2p\eta^2}{\sum_{i=0}^{p-1}|\mathrm{corr}(i,0)|}\right). \tag{82}$$

It turns out we only have to upper bound the sum of $|\mathrm{corr}(i,0)|$ for the 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 (82) simplifies as
$$\rho^2\leq\frac{1}{n}\log\left(1+2p\eta^2\frac{1-e^{-w}}{1+e^{-w}}\right).$$

2. if $\mathrm{corr}(i,j)=(1+|i-j|_p)^{-t}$ for t strictly larger than one, then $\sum_{i=0}^{p-1}|\mathrm{corr}(i,0)|\leq 1+\frac{2}{t-1}$ and condition (82) simplifies as
$$\rho^2\leq\frac{1}{n}\log\left(1+\frac{2p(t-1)\eta^2}{t+1}\right).$$

3. if $\mathrm{corr}(i,j) = (1 + |i-j|_p)^{-1}$ then $\sum_{i=0}^{p-1} |\mathrm{corr}(i,0)| \leq 1 + 2\log(p-1)$ and condition (82) simplifies as

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

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

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

and condition (82) simplifies as

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

8.6 Proof of Proposition 11

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 $D \in \{1, \ldots, p\}$. Since $r_D^2 \leq a_D^2$ and since the a_j's are non increasing,

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

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,

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

Since $r_D \leq \rho_{D,n}$, we deduce from Proposition 4 that

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

The first result of Proposition 11 follows by gathering these lower bounds for all D between 1 and p.

Moreover, $\rho_{i,n}^2$ is defined in Proposition 4 as $\rho_{i,n}^2 = \sqrt{2}\left[\sqrt{\mathcal{L}(\eta)} \wedge \frac{\mathcal{L}(\eta)}{\sqrt{\log 2}}\right]\frac{\sqrt{i}}{n}$. If $\alpha + \delta \leq 47\%$, it is straighforward to show that $\rho_{i,n}^2 \geq \frac{\sqrt{i}}{n}$.

8.7 Proof of Proposition 13

We first need the following Lemma.

Lemma 17. *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 $\theta \in \mathbb{R}^{\mathcal{I}}$ 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

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

if for all $j\in\mathcal{J}$, $r_j\leq\rho_{p(j),n}(\eta/\sqrt{k_j})$, *where the function* $\rho_{p(j),n}$ *is defined by (16).*

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

We now apply Lemma 17 with $k_j:=(1/(j+1)^2)/R(p)$ where $R(p)=\sum_{k=0}^{J}1/(k+1)^2$ to show that

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

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 equals

$$\sqrt{\log(1+2\eta^2/k_j)}\left(1\wedge\frac{\sqrt{\log(1+2\eta^2/k_j)}}{\sqrt{2\log 2}}\right)\frac{\sqrt{D}}{n}\geq\sqrt{\log(1+2\eta^2(j+1)^2R(p))}\left(1\wedge\frac{\sqrt{\log(1+2\eta^2)}}{\sqrt{2\log 2}}\right)\frac{2^{j/2}}{n}\ . \quad (83)$$

It remains to check that (83) 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

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

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

$$\begin{aligned}\sqrt{\log(1+2\eta^2(j+1)^2R(p))}\left(1\wedge\frac{\sqrt{\log(1+2\eta^2)}}{\sqrt{2\log 2}}\right)2^{j/2} &\geq \frac{1}{2\sqrt{\log(2)}}\left(1\wedge\log(1+2\eta^2)\right)\sqrt{\log\log(D+1)}\sqrt{D}\\ &= \bar{\rho}^2_{D,n}\ .\end{aligned}$$

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

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

8.8 Proof of Lemma 17

Using a similar approach to the proof of Theorem 5, we know that for each $r_j \leq \tilde{\rho}_j(\eta/\sqrt{k_j})$ there exists some measure μ_j over

$$\Theta_j[r_j] := \left\{\theta \in \Theta_j, \frac{\|\theta\|^2}{\mathrm{var}(Y) - \|\theta\|^2} = r_j^2\right\}$$

such that

$$\mathbb{E}_0\left[L^2_{\mu_j}(Y, X)\right] \leq 1 + \eta^2/k_j \ . \tag{84}$$

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,

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

and

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

Using expression (68), it is straightforward to show that if $j \neq j'$, then

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

This follows from the fact that the sets Θ_j and $\Theta_{j'}$ are orthogonal with respect to the inner product (4). Thus,

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

thanks to (84). Using the argument (65) as in the proof of Theorem 5 allows to conclude.

8.9 Proof of Proposition 14

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 Gram-Schmidt process to X and thus obtain a vector X' and a new basis for Θ 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 (2002). We use the same notations as in proof of Proposition 11. Let $D^*(R) \in \{1, \ldots, p\}$ an integer which achieves the supremum of $\bar{\rho}^2_D \wedge (R^2 a^2_D) = \bar{r}^2_D$. As in proof of Proposition 11, for any $R > 0$,

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

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

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

and the result follows from proposition 13.

Acknowledgements

We gratefully thank Sylvie Huet and Pascal Massart for many fruitful discussions.

References

Aldous, D. J., 1985. *Exchangeability and related topics*, École d'été de probabilités de Saint Flour XIII. Vol. **1117** of Lecture Notes in Mathematics. Springer-Verlag, Berlin.

Baraud, Y., 2002. Non-asymptotic rates of testing in signal detection. *Bernoulli* 8 (**5**), 577–606.

Baraud, Y., Huet, S., Laurent, B., 2003. Adaptative tests of linear hypotheses by model selection. *Ann. Statist.* **31** (1), 225–251.

Birgé, L., Massart, P., 1998. Minimum contrast estimators on sieves: exponentntial bounds and rates of convergence. *Bernoulli* **4** (3), 329–375.

Bühlmann, P., Kalisch, M., 2007. Variable selection for high-dimensional models: partial faithful distributions, strong assocations and the PC-algorithm. Tech. rep., *Seminar für Statistik, ETH Zürich.*

Candès, E., Tao, T., 2007. The dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.* (in press).

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., Spiegelhalter, D. J., 1999. *Probabilistic Networks and Expert Systems.* Statistics for Engineering and Information Science. Springer-Verlag, New-York.

Cressie, N., 1993. *Statistics for Spatial Data*, revised Edition. Wiley, New York.

Drton, M., Perlman, M., 2007. Mutiple testing and error control in Gaussian graphical model selection. *to appear in Statist. Sci.*

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. *Ann. Statist.* **32** (2), 407–499.

Huang, J., Liu, N., Pourahmadi, M., Liu, L., 2006. Covariance matrix selection and estimation via penalised normal likehood. *Biometrika* **93** (1), 85–98.

Ingster, Y. I., 1993a. Asymptotically minimax hypothesis testing for nonparametric alternatives I. *Math. Methods Statist.* **2**, 85–114.

Ingster, Y. I., 1993b. Asymptotically minimax hypothesis testing for nonparametric alternatives II. *Math. Methods Statist.* **3**, 171–189.

Ingster, Y. I., 1993c. Asymptotically minimax hypothesis testing for nonparametric alternatives III. *Math. Methods Statist.* **4**, 249–268.

Kishino, H., Waddell, P., 2000. Correspondence analysis of genes and tissue types and finding genetic links from microarray data. *Genome Informatics* **11**, 83–95.

Lauritzen, S. L., 1996. *Graphical Models.* Oxford University Press, New York.

Massart, P., 2007. *Concentration Inequalities and Model Selection*, École d'été de probabilités de Saint Flour XXXIII. Vol. **1896** of Lecture Notes in Mathematics. Springer-Verlag, Berlin.

Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the Lasso. *Ann. Statist.* **34** (3), 1436–1462.

Rue, H., Held, L., 2005. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC, London.

Schäfer, J., Strimmer, K., 2005. An empirical bayes approach to inferring large-scale gene association network. *Bioinformatics* **21**, 754–764.

Spokoiny, V. G., 1996. Adaptative hypothesis testing using wavelets. *Ann. Statist.* **24**, 2477–2498.

Verzelen, N., Villers, F., 2007. Tests for Gaussian graphical models. Tech. Rep. 2007-5, *INRA, Unité MIA*, Jouy-en-Josas, France.

Wille, A., Bühlmann, P., 2006. Low-order conditional independence graphs for inferring genetic networks. *Stat. Appl. Genet. Mol. Biol.* **5**, Art. 1.

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., Bühlmann, P., 2004. Sparse graphical Gaussian modelling of the isoprenoid gene network in *arabidopsis thaliana*. *Genome Biology* **5** (11).

Yuan, M., Lin, Y., 2007. Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**, 19–35.

Zhao, P., Yu, B., 2006. On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7**, 2541–2563.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the Elastic Net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67**, 301–320.