

Minimax rate of testing in sparse linear regression

Alexandra Carpentier, Olivier Collier, Laëtitia Comminges, Alexandre Tsybakov, Yuhao Wang

► **To cite this version:**

Alexandra Carpentier, Olivier Collier, Laëtitia Comminges, Alexandre Tsybakov, Yuhao Wang. Minimax rate of testing in sparse linear regression. 2018. <hal-01770434>

HAL Id: hal-01770434

<https://hal.archives-ouvertes.fr/hal-01770434>

Submitted on 19 Apr 2018

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

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

Minimax rate of testing in sparse linear regression

Alexandra Carpentier^{*}, Olivier Collier[†], Laëticia Comminges[‡], Alexandre B. Tsybakov[§], Yuhao Wang[¶]

University of Magdeburg^{}, Modal'X, Université Paris-Nanterre[†], CEREMADE, Université Paris-Dauphine[‡], CREST, ENSAE[§], LIDS-IDSS, MIT[¶]*

Abstract: We consider the problem of testing the hypothesis that the parameter of linear regression model is 0 against an s -sparse alternative separated from 0 in the ℓ_2 -distance. We show that, in Gaussian linear regression model with $p < n$, where p is the dimension of the parameter and n is the sample size, the non-asymptotic minimax rate of testing has the form $\sqrt{(s/n) \log(1 + \sqrt{p/s})}$. We also show that this is the minimax rate of estimation of the ℓ_2 -norm of the regression parameter.

MSC 2010 subject classifications: 62J05, 62G10.

Keywords and phrases: linear regression, Gaussian sequence model, sparsity, signal detection.

1. Introduction

This paper is deals with minimax testing of hypotheses on the parameter of linear regression model under sparse alternatives. The problem of minimax testing under sparse alternatives was first studied by Ingster (1997) and Donoho and Jin (2004) who considered the Gaussian mean model. These papers were dealing with an asymptotic setting under the assumption that the sparsity index scales as a power of the dimension. Non-asymptotic setting for the Gaussian mean model was analyzed by Baraud (2002) who established bounds on the minimax rate of testing up to a logarithmic factor. Finally, the exact non-asymptotic minimax testing rate for the Gaussian mean model is derived in Collier, Comminges and Tsybakov (2017). In this paper, we present an extension of the results of Collier, Comminges and Tsybakov (2017) to linear regression model with Gaussian noise. Note that the problem of minimax testing for linear regression under sparsity was already studied in Ingster, Tsybakov and Verzelen (2010), Arias-Castro, Candès and Plan (2011), Verzelen (2012). Namely, Ingster, Tsybakov and Verzelen (2010), Arias-Castro, Candès and Plan (2011) deal with an asymptotic setting under additional assumptions on the parameters of the problem while Verzelen (2012) obtains non-asymptotic bounds up to a logarithmic factor in the spirit of Baraud (2002). Our aim here is to derive the non-asymptotic minimax rate of testing in Gaussian linear regression model with no specific assumptions on the parameters of the problem. We give a solution to this problem when $p < n$, where p is the dimension and n is the sample size.

We consider the model

$$Y = X\theta + \sigma\xi, \tag{1}$$

where $\sigma > 0$, $\xi \in \mathbb{R}^n$ is a vector of Gaussian white noise, i.e., $\xi \sim \mathcal{N}(0, I_n)$, X is a $n \times p$ matrix with real entries, I_n is the $n \times n$ identity matrix, and $\theta \in \mathbb{R}^p$ is an unknown parameter.

The following notation will be used below. For $q > 0$, we denote by $\|\cdot\|_q$ be the ℓ_q -norm, i.e., for $u = (u_1, \dots, u_p) \in \mathbb{R}^p$,

$$\|u\|_q^q = \sum_{i=1}^p |u_i|^q.$$

Let $\|\cdot\|_0$ be the ℓ_0 semi-norm, i.e., for $u \in \mathbb{R}^p$,

$$\|u\|_0 = \sum_{i=1}^p \mathbb{1}_{u_i \neq 0},$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function. We denote by $\langle u, v \rangle = u^T v$ the inner product of $u \in \mathbb{R}^p, v \in \mathbb{R}^p$. We denote by $\lambda_{\min}(M)$ and by $\text{tr}[M]$ the minimal eigenvalue and the trace of matrix $M \in \mathbb{R}^{p \times p}$. For an integer $s \in \{1, \dots, p\}$, we consider the set $B_0(s)$ of all s -sparse vectors in \mathbb{R}^p :

$$B_0(s) := \{u \in \mathbb{R}^p : \|u\|_0 \leq s\}.$$

Given the observations (X, Y) , we consider the problem of testing the hypothesis

$$H_0 : \theta = 0 \quad \text{against the alternative} \quad H_1 : \theta \in \Theta(s, \tau) \quad (2)$$

where

$$\Theta(s, \tau) = \{\theta \in B_0(s) : \|\theta\|_2 \geq \tau\}$$

for some $s \in \{1, \dots, p\}$ and $\tau > 0$. Let $\Delta = \Delta(X, Y)$ be a test statistic with values in $\{0, 1\}$. We define the risk of test Δ as the sum of the first type error and the maximum second type error:

$$\mathbf{P}_0(\Delta = 1) + \sup_{\theta \in \Theta(s, \tau)} \mathbf{P}_\theta(\Delta = 0),$$

where \mathbf{P}_θ denotes the joint distribution of (X, Y) satisfying (1). The best possible value of this risk is equal to the minimax risk

$$\mathcal{R}_{s, \tau} := \inf_{\Delta} \left\{ \mathbf{P}_0(\Delta = 1) + \sup_{\theta \in \Theta(s, \tau)} \mathbf{P}_\theta(\Delta = 0) \right\}$$

where \inf_{Δ} is the infimum over all $\{0, 1\}$ -valued statistics. We define the *minimax rate of testing on the class $B_0(s)$ with respect to the ℓ_2 -distance* as a value $\lambda > 0$, for which the following two properties hold:

(i) (*upper bound*) for any $\varepsilon \in (0, 1)$ there exists $A_\varepsilon > 0$ such that, for all $A > A_\varepsilon$,

$$\mathcal{R}_{s, A\lambda} \leq \varepsilon, \quad (3)$$

(ii) (*lower bound*) for any $\varepsilon \in (0, 1)$ there exists $a_\varepsilon > 0$ such that, for all $0 < A < a_\varepsilon$,

$$\mathcal{R}_{s, A\lambda} \geq 1 - \varepsilon. \quad (4)$$

Note that the rate λ defined in this way is a non-asymptotic minimax rate of testing as opposed to the classical asymptotic definition that can be found, for example, in Ingster and Suslina (2003). It is shown in Collier, Comminges and Tsybakov (2017) that when X is the identity matrix and $p = n$ (which corresponds to the Gaussian sequence model), the non-asymptotic minimax rate of testing on the class $B_0(s)$ with respect to the ℓ_2 -distance has the following form:

$$\lambda = \begin{cases} \sigma \sqrt{s \log(1 + p/s^2)} & \text{if } s < \sqrt{p}, \\ \sigma p^{1/4} & \text{if } s \geq \sqrt{p}. \end{cases} \quad (5)$$

For the regression model with random X and satisfying some strong assumptions, the asymptotic minimax rate of testing when n, p , and s tend to ∞ such that $s = p^a$ for some $0 < a < 1$ is studied in Ingster, Tsybakov and Verzelen (2010). In particular, it is shown in Ingster, Tsybakov and

Verzelen (2010) that for this configuration of parameters and if the matrix X has i.i.d. standard normal entries, the asymptotic rate has the form

$$\lambda = \sigma \min \left(\sqrt{\frac{s \log(p)}{n}}, n^{-1/4}, \frac{p^{1/4}}{\sqrt{n}} \right). \quad (6)$$

Similar result for a somewhat differently defined alternative H_1 is obtained in Arias-Castro, Candès and Plan (2011).

Below we show that non-asymptotically, and with no restriction on the parameters n, p and s , the lower bound (ii) for the minimax rate of testing is valid with

$$\lambda = \sigma \min \left(\sqrt{\frac{s \log(2 + p/s^2)}{n}}, n^{-1/4}, \frac{p^{1/4}}{\sqrt{n}} \right) \quad (7)$$

whenever X is a matrix with isotropic distribution and independent subgaussian rows. Furthermore, we show that the matching upper bound holds when X is a matrix with i.i.d. standard Gaussian entries and $p < n$. Note that for $p < n$ the expression (7) takes the form

$$\lambda = \sigma \min \left(\sqrt{\frac{s \log(2 + p/s^2)}{n}}, \frac{p^{1/4}}{\sqrt{n}} \right) \quad (8)$$

It will be also useful to note that, since for $s \leq \sqrt{p}$ the function $s \mapsto s \log(2 + p/s^2)$ is increasing and satisfies $\log(2 + p/s^2) \leq 2 \log(1 + p/s^2)$, the rate (8) can be equivalently (to within an absolute constant factor) written as

$$\lambda = \begin{cases} \sigma \sqrt{\frac{s \log(1 + p/s^2)}{n}} & \text{if } s < \sqrt{p}, \\ \sigma \frac{p^{1/4}}{\sqrt{n}} & \text{if } s \geq \sqrt{p}. \end{cases} \quad (9)$$

This expression is analogous to (5). Finally, note that l is also equivalent, up to a constant factor, to the more pleasant expression

$$\sqrt{\frac{s \log(1 + \sqrt{p}/s)}{n}}, \quad (10)$$

since this has a hidden elbow for $s \geq \sqrt{p}$.

2. Upper bounds on the minimax rates

In this section, we assume that X is a matrix with i.i.d. standard Gaussian entries and $p < n$ and we establish an upper bound on the minimax rate of testing in the form (9). This will be done by using a connexion between testing and estimation of functionals. We first introduce an estimator \hat{Q} of the quadratic functional $\|\theta\|_2^2$ and establish an upper bound on its risk. Then, we deduce from this result an upper bound for the risk of the estimator \hat{N} of the norm $\|\theta\|_2$ defined as follows:

$$\hat{N} = \sqrt{\max(\hat{Q}, 0)}.$$

Finally, using \hat{N} to define a test statistic we obtain an upper bound on the minimax rate of testing.

Introduce the notation

$$\alpha_s = \mathbf{E}(Z^2 | Z^2 > 2 \log(1 + p/s^2))$$

where Z is a standard normal random variable, and set

$$y_i = \{(X^T X)^{-1} X^T Y\}_i$$

where $\{(X^T X)^{-1} X^T Y\}_i$ is the i th component of the least squares estimator $(X^T X)^{-1} X^T Y$. Note that the inverse $(X^T X)^{-1}$ exists almost surely since we assume in this section that X is a matrix with i.i.d. standard Gaussian entries and $p < n$, so that X is almost surely of full rank. We consider the following estimator of the quadratic functional $\|\theta\|_2^2$:

$$\hat{Q} := \begin{cases} \sum_{i=1}^p y_i^2 - \sigma^2 \text{tr}[(X^T X)^{-1}] & \text{if } s \geq \sqrt{p}, \\ \sum_{i=1}^p [y_i^2 - \sigma^2 (X^T X)^{-1}_{ii} \alpha_s] \mathbb{1}_{y_i^2 > 2\sigma^2 (X^T X)^{-1}_{ii} \log(1+p/s^2)} & \text{if } s < \sqrt{p}. \end{cases}$$

Here and below $(X^T X)^{-1}_{ij}$ denotes the (i, j) th entry of matrix $(X^T X)^{-1}$.

For any integers n, p, s such that $s \leq p$, set

$$\psi(s, p) = \begin{cases} \frac{s \log(1+p/s^2)}{n} & \text{if } s < \sqrt{p}, \\ \frac{p^{1/2}}{n} & \text{if } s \geq \sqrt{p}. \end{cases}$$

Theorem 1 *Let n, p, s be integers such that $s \leq p, n \geq 9$, and $p \leq \min(\gamma n, n - 8)$ for some constant $0 < \gamma < 1$. Let $r > 0, \sigma > 0$. Assume that all entries of matrix X are i.i.d. standard Gaussian random variables. Then there exists a constant $c > 0$ depending only on γ such that*

$$\sup_{\theta: \|\theta\|_0 \leq s, \|\theta\|_2 \leq r} \mathbf{E}_\theta[(\hat{Q} - \|\theta\|_2^2)^2] \leq c \left(\sigma^2 \frac{r^2}{n} + \sigma^4 \psi^2(s, p) \right).$$

The proof of Theorem 1 is given in Section 5.

Arguing exactly in the same way as in the proof of Theorem 8 in Collier, Comminges and Tsybakov (2017), we deduce from Theorem 1 the following upper bound on the squared risk of the estimator \hat{N} .

Theorem 2 *Let the assumptions of Theorem 1 be satisfied. Then there exists a constant $c' > 0$ depending only on γ such that*

$$\sup_{\theta \in B_0(s)} \mathbf{E}_\theta[(\hat{N} - \|\theta\|_2)^2] \leq c' \sigma^2 \psi(s, p).$$

Theorem 2 implies that the test $\Delta^* = \mathbb{1}_{\{\hat{N} > A\lambda/2\}}$ where $\lambda = \sigma \sqrt{\psi(s, p)}$ (i.e., the same λ as in (8)) satisfies

$$\begin{aligned} & \mathbf{P}_0(\Delta^* = 1) + \sup_{\theta \in \Theta(s, A\lambda)} \mathbf{P}_\theta(\Delta^* = 0) \\ & \leq \mathbf{P}_0(\hat{N} > A\lambda/2) + \sup_{\theta \in B_0(s)} \mathbf{P}_\theta(\hat{N} - \|\theta\|_2 \leq -A\lambda/2) \\ & \leq 2 \sup_{\theta \in B_0(s)} \frac{\mathbf{E}_\theta[(\hat{N} - \|\theta\|_2)^2]}{(A/2)^2 \lambda^2} \leq C_* A^{-2} \end{aligned}$$

for some constant $C_* > 0$. Using this remark and choosing $A_\varepsilon = (C_*/\varepsilon)^{1/2}$ leads to the upper bound (i) that we have defined in the previous section. We state this conclusion in the next theorem.

Theorem 3 *Let the assumptions of Theorem 1 be satisfied and let λ be defined by (8). Then, for any $\varepsilon \in (0, 1)$ there exists $A_\varepsilon > 0$ depending only on ε and γ such that, for all $A > A_\varepsilon$,*

$$\mathcal{R}_{s, A\lambda} \leq \varepsilon.$$

3. Lower bounds on the minimax rates

In this section, we assume that the distribution of matrix X is isotropic and has independent σ_X -subgaussian rows for some $\sigma_X > 0$. The isotropy of X means that $E_X(X^T X/n) = I_p$ where E_X denotes the expectation with respect to the distribution P_X of X .

Definition 1 *Let $b > 0$. A real-valued random variable ζ is called b -subgaussian if*

$$\mathbf{E} \exp(t\zeta) \leq \exp(b^2 t^2 / 2), \quad \forall t \in \mathbb{R}.$$

A random vector η with values in \mathbb{R}^d is called b -subgaussian if all inner products $\langle \eta, v \rangle$ with vectors $v \in \mathbb{R}^d$ such that $\|v\|_2 = 1$ are b -subgaussian random variables.

The following theorem on the lower bound is non-asymptotic and holds with no restriction on the parameters n, p, s except for the inevitable condition $s \leq p$.

Theorem 4 *Let $\varepsilon \in (0, 1)$, $\sigma > 0$, and let the integers n, p, s be such that $s \leq p$. Assume that the distribution of matrix X is isotropic and X has independent σ_X -subgaussian rows for some $\sigma_X > 0$. Then, there exists $a_\varepsilon > 0$ depending only on ε and σ_X such that, for*

$$\tau = A\sigma \min \left(\sqrt{\frac{s \log(2 + p/s^2)}{n}}, n^{-1/4}, \frac{p^{1/4}}{\sqrt{n}} \right) \quad (11)$$

with any A satisfying $0 < A < a_\varepsilon$, we have

$$\mathcal{R}_{s, \tau} \geq 1 - \varepsilon.$$

The proof of Theorem 4 is given in Section 7. The next corollary is an immediate consequence of Theorems 3 and 4.

Corollary 1 *Let the assumptions of Theorem 1 be satisfied. Then the minimax rate of testing on the class $B_0(s)$ with respect to the ℓ_2 -distance is given by (8).*

In addition, from Theorem 4, we get the following lower bound on the minimax risk of estimation of the ℓ_2 -norm $\|\theta\|_2$.

Theorem 5 *Let the assumptions of Theorem 4 be satisfied, and let λ be defined in (7). Then there exists an a constant $c_* > 0$ depending only on σ_X such that*

$$\inf_{\hat{T}} \sup_{\theta \in B_0(s)} \mathbf{E}_\theta [(\hat{T} - \|\theta\|_2)^2] \geq c_* \lambda^2,$$

where $\inf_{\hat{T}}$ denotes the infimum over all estimators.

The result of Theorem 5 follows from Theorem 4 by noticing that, for τ in (11) and λ in (7) we have $\tau = A\lambda$, and for any estimator \hat{T} ,

$$\begin{aligned} \sup_{\theta \in B_0(s)} \mathbf{E}_\theta [(\hat{T} - \|\theta\|_2)^2] &\geq \frac{1}{2} \left[\mathbf{E}_0[\hat{T}^2] + \sup_{\theta \in \Theta(s, \tau)} \mathbf{E}_\theta [(\hat{T} - \|\theta\|_2)^2] \right] \\ &\geq \frac{\tau^2}{8} \left[\mathbf{P}_0(\hat{T} > \tau/2) + \sup_{\theta \in \Theta(s, \tau)} \mathbf{P}_\theta(\hat{T} \leq \tau/2) \right] \\ &\geq \frac{(A\lambda)^2}{8} \mathcal{R}_{s, \tau}. \end{aligned}$$

Corollary 2 *Let the assumptions of Theorem 1 be satisfied and let λ be defined in (8). Then the minimax rate of estimation of the norm $\|\theta\|_2$ under the mean squared risk on the class $B_0(s)$ is equal to λ^2 , that is*

$$c_*\lambda^2 \leq \inf_{\hat{T}} \sup_{\theta \in B_0(s)} \mathbf{E}_\theta[(\hat{T} - \|\theta\|_2)^2] \leq c'\lambda^2,$$

where $c_* > 0$ is an absolute constant and $c' > 0$ is a constant depending only on γ .

This corollary is an immediate consequence of Theorems 2 and 5.

Remark 1 *Inspection of the proofs reveals that the results of this section remain valid if we replace the ℓ_0 -ball $B_0(s)$ by the ℓ_0 -sphere $\bar{B}_0(s) = \{u \in \mathbb{R}^p : \|u\|_0 = s\}$.*

4. Preliminary lemmas for the proof of Theorem 1

This section treats two main technical issues for the proof of Theorem 1. The first one is to control the inverse of the smallest eigenvalue of random design matrix. The second issue is to control the errors for identifying non-zero entries in the sparse setting. For this, we need accurate bounds on the correlations between centred thresholded transformations of two correlated χ_1^2 random variables. We first recall two general facts that we will use to solve the first issue.

In what follows, we will denote by C positive constants that can vary from line to line.

Lemma 1 *[Davidson and Szarek (2001), see also Vershynin (2012).] Let $\lambda_{\min}(\hat{\Sigma})$ denote the smallest eigenvalue of the sample covariance matrix $\hat{\Sigma} = \frac{1}{n}X^T X$. Then for any $t > 0$ with probability at least $1 - 2\exp(-t^2/2)$ we have*

$$1 - \sqrt{\frac{p}{n}} - \frac{t}{\sqrt{n}} \leq \sqrt{\lambda_{\min}(\hat{\Sigma})} \leq 1 + \sqrt{\frac{p}{n}} + \frac{t}{\sqrt{n}}.$$

Lemma 2 *[(Tao and Vu, 2010, Lemma A4), see also (Bordenave and Chafaï, 2012, Lemma 4.14).] Let $1 \leq p \leq n$, let R_i be the i -th column of matrix $X \in \mathbb{R}^{n \times p}$ and $R_{-i} = \text{span}\{R_j : j \neq i\}$. If X has full rank, then*

$$(X^T X)_{ii}^{-1} = \text{dist}(R_i, R_{-i})^{-2},$$

where $\text{dist}(R_i, R_{-i})$ is the Euclidean distance of vector R_i to the space R_{-i} .

Lemma 3 *Let $n \geq 9$ and $p \leq \min(\gamma n, n - 8)$ for some constant γ such that $0 < \gamma < 1$. Assume that all entries of matrix $X \in \mathbb{R}^{n \times p}$ are i.i.d. standard Gaussian random variables. Then there exists a constant $c > 0$ depending only on γ , such that*

$$\mathbf{E}[\lambda_{\min}^{-2}(\hat{\Sigma})] \leq c. \tag{12}$$

Proof. Set $\beta = \sqrt{\gamma}$. From Lemma 1 we have

$$\mathbf{P}\left(\lambda_{\min}(\hat{\Sigma}) < \left(\frac{1-\beta}{2}\right)^2\right) \leq 2 \exp\left(-\frac{n(1-\beta)^2}{8}\right).$$

Using this inequality we obtain

$$\mathbf{E}[\lambda_{\min}^{-2}(\hat{\Sigma})] \leq \left(\frac{1-\beta}{2}\right)^{-4} + \sqrt{\mathbf{E}[\lambda_{\min}^{-4}(\hat{\Sigma})]} \sqrt{2} \exp\left(-\frac{n(1-\beta)^2}{16}\right). \tag{13}$$

We now bound the expectation $\mathbf{E}[\lambda_{\min}^{-4}(\hat{\Sigma})]$. Clearly,

$$\lambda_{\min}^{-1}(\hat{\Sigma}) \leq \text{tr}[\hat{\Sigma}^{-1}]. \quad (14)$$

Lemma 2 implies that, almost surely,

$$(\text{tr}[\hat{\Sigma}^{-1}])^4 = n^4 \left[\sum_{i=1}^p \text{dist}(R_i, R_{-i})^{-2} \right]^4 \leq n^4 p^3 \sum_{i=1}^p \text{dist}(R_i, R_{-i})^{-8}.$$

Since the random variables $\text{dist}(R_i, R_{-i})$ are identically distributed and $p \leq n$ we have

$$\mathbf{E}[(\text{tr}[\hat{\Sigma}^{-1}])^4] \leq n^8 \mathbf{E}[\text{dist}(R_1, R_{-1})^{-8}]. \quad (15)$$

Finally we only need to bound $\mathbf{E}[\text{dist}(R_1, R_{-1})^{-8}]$. If \mathcal{S} is a $p-1$ dimensional subspace of \mathbb{R}^n then the random variable $\text{dist}(R_1, \mathcal{S})^2$ has the chi-square distribution χ_{n-p+1}^2 with $n-p+1$ degrees of freedom. Hence, as R_{-1} is a span of random vectors independent of R_1 and R_{-1} is almost surely $p-1$ dimensional, we have

$$\begin{aligned} \mathbf{E}[\text{dist}(R_1, R_{-1})^{-8}] &= \mathbf{E}\left[\frac{1}{(\chi_{n-p+1}^2)^4}\right] \\ &= \frac{1}{(n-p-1)(n-p-3)(n-p-5)(n-p-7)} \leq \frac{1}{105}. \end{aligned} \quad (16)$$

Combining (13), (14), (15) and (16) we get

$$\mathbf{E}[\lambda_{\min}^{-2}(\hat{\Sigma})] \leq \left(\frac{1-\beta}{2}\right)^{-4} + \frac{n^8}{105} \sqrt{2} \exp\left(-\frac{n(1-\beta)^2}{16}\right),$$

which implies the lemma. ■

We now turn to the second issue of this section, that is bounds on the correlations. We will use the following lemma about the tails of the standard normal distribution.

Lemma 4 For $\eta \sim \mathcal{N}(0, 1)$ and any $x > 0$ we have

$$\frac{4}{\sqrt{2\pi}(x + \sqrt{x^2 + 4})} \exp(-x^2/2) \leq \mathbf{P}(|\eta| > x) \leq \frac{4}{\sqrt{2\pi}(x + \sqrt{x^2 + 2})} \exp(-x^2/2), \quad (17)$$

$$\mathbf{E}[\eta^2 \mathbf{1}_{|\eta| > x}] \leq \sqrt{\frac{2}{\pi}} \left(x + \frac{2}{x}\right) \exp(-x^2/2), \quad (18)$$

$$\mathbf{E}[\eta^4 \mathbf{1}_{|\eta| > x}] \leq \sqrt{\frac{2}{\pi}} \left(x^3 + 3x + \frac{1}{x}\right) \exp(-x^2/2). \quad (19)$$

Moreover, if $x \geq 1$, then

$$\mathbf{E}[\eta^2 \mid |\eta| > x] \leq 5x^2. \quad (20)$$

Inequalities (17) - (19) are given, e.g., in (Collier, Comminges and Tsybakov, 2017, Lemma 4) and (20) follows easily from (17) and (18).

Lemma 5 Let (η, ζ) be a Gaussian vector with mean 0 and covariance matrix $\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $0 < \rho < 1$. Set $\alpha = \mathbf{E}[\eta^2 \mid |\eta| > x]$. Then there exists an absolute constant $C > 0$ such that, for any $x \geq 1$,

$$\mathbf{E}[(\eta^2 - \alpha)(\zeta^2 - \alpha)\mathbf{1}_{|\eta|>x}\mathbf{1}_{|\zeta|>x}] \leq C\rho^2 x^4 \exp(-x^2/2).$$

Proof. From (20) we get that $\alpha \leq 5x^2$. Thus, using (19) and the fact that $x \geq 1$ we find

$$\mathbf{E}[(\zeta^2 - \alpha)^2 \mathbf{1}_{|\zeta|>x}] \leq \mathbf{E}[(\zeta^4 + \alpha^2) \mathbf{1}_{|\zeta|>x}] \leq 26\mathbf{E}[\zeta^4 \mathbf{1}_{|\zeta|>x}] \leq Cx^3 \exp(-x^2/2). \quad (21)$$

Therefore,

$$\mathbf{E}[(\eta^2 - \alpha)(\zeta^2 - \alpha)\mathbf{1}_{|\eta|>x}\mathbf{1}_{|\zeta|>x}] \leq \mathbf{E}[(\eta^2 - \alpha)^2 \mathbf{1}_{|\eta|>x}] + \mathbf{E}[(\zeta^2 - \alpha)^2 \mathbf{1}_{|\zeta|>x}] \leq Cx^3 \exp(-x^2/2).$$

This proves the lemma for $\rho > 1/2$.

Consider now the case $0 < \rho \leq 1/2$. Note that, since $\alpha \leq 5x^2$, for $0 < \rho \leq 1/2$ we also have

$$\rho \leq \frac{x}{\sqrt{\alpha}}.$$

The symmetry of the distribution of (η, ζ) implies

$$\mathbf{E}[(\eta^2 - \alpha)(\zeta^2 - \alpha)\mathbf{1}_{|\eta|>x}\mathbf{1}_{|\zeta|>x}] = 2\mathbf{E}[(\eta^2 - \alpha)(\zeta^2 - \alpha)\mathbf{1}_{|\eta|>x}\mathbf{1}_{\zeta>x}]. \quad (22)$$

Now, we use the fact that $(\eta, \zeta) \stackrel{d}{=} (\rho\zeta + \sqrt{1-\rho^2}Z, \zeta)$ where $\stackrel{d}{=}$ means equality in distribution and Z is a standard Gaussian random variable independent of ζ . Thus,

$$\begin{aligned} \mathbf{E}[(\eta^2 - \alpha)(\zeta^2 - \alpha)\mathbf{1}_{|\eta|>x}\mathbf{1}_{\zeta>x}] &= \rho^2 \mathbf{E}[(\zeta^2 - \alpha)^2 \mathbf{1}_{|\rho\zeta + \sqrt{1-\rho^2}Z|>x}\mathbf{1}_{\zeta>x}] \\ &\quad + 2\rho\sqrt{1-\rho^2} \mathbf{E}[\zeta Z (\zeta^2 - \alpha) \mathbf{1}_{|\rho\zeta + \sqrt{1-\rho^2}Z|>x}\mathbf{1}_{\zeta>x}] \\ &\quad + (1-\rho^2) \mathbf{E}[(Z^2 - \alpha)(\zeta^2 - \alpha) \mathbf{1}_{|\rho\zeta + \sqrt{1-\rho^2}Z|>x}\mathbf{1}_{\zeta>x}]. \end{aligned} \quad (23)$$

We now bound separately the three summands on the right hand side of (23). For the first summand, using (21) we get the bound

$$\rho^2 \mathbf{E}[(\zeta^2 - \alpha)^2 \mathbf{1}_{|\rho\zeta + \sqrt{1-\rho^2}Z|>x}\mathbf{1}_{\zeta>x}] \leq 26\rho^2 \mathbf{E}[\zeta^4 \mathbf{1}_{\zeta>x}] \leq C\rho^2 x^3 \exp\left(-\frac{x^2}{2}\right). \quad (24)$$

To bound the second summand, we first write

$$\mathbf{E}[\zeta Z (\zeta^2 - \alpha) \mathbf{1}_{|\rho\zeta + \sqrt{1-\rho^2}Z|>x}\mathbf{1}_{\zeta>x}] = \mathbf{E}[\zeta(\zeta^2 - \alpha)g(\zeta)\mathbf{1}_{\zeta>x}] \quad (25)$$

where $g(\zeta) := \mathbf{E}[Z \mathbf{1}_{|\rho\zeta + \sqrt{1-\rho^2}Z|>x} \mid \zeta]$. It is straightforward to check that

$$g(\zeta) = \exp\left(-\frac{(x - \rho\zeta)^2}{2(1-\rho^2)}\right) - \exp\left(-\frac{(x + \rho\zeta)^2}{2(1-\rho^2)}\right).$$

Thus $g(\zeta)$ is positive when $\zeta > x$. Therefore we have

$$\mathbf{E}[\zeta(\zeta^2 - \alpha)g(\zeta)\mathbf{1}_{\zeta>x}] \leq \mathbf{E}[\zeta^3 g(\zeta)\mathbf{1}_{\zeta>x}]. \quad (26)$$

In addition,

$$g(\zeta) = \exp\left(-\frac{(x - \rho\zeta)^2}{2(1 - \rho^2)}\right) \left(1 - \exp\left(-\frac{2x\rho\zeta}{1 - \rho^2}\right)\right) \leq 1 - \exp\left(-\frac{2x\rho\zeta}{1 - \rho^2}\right) \leq \frac{2x\rho\zeta}{1 - \rho^2}. \quad (27)$$

Combining (25) - (27) with (19) and the fact that $\rho \leq \frac{1}{2}$, we get

$$2\rho\sqrt{1 - \rho^2}\mathbf{E}[\zeta Z(\zeta^2 - \alpha)\mathbf{1}_{|\rho\zeta + \sqrt{1 - \rho^2}Z| > x}\mathbf{1}_{\zeta > x}] \leq C\rho^2 x^4 \exp\left(-\frac{x^2}{2}\right). \quad (28)$$

We now consider the third summand on the right hand side of (23). We will prove that

$$\mathbf{E}[(Z^2 - \alpha)(\zeta^2 - \alpha)\mathbf{1}_{|\rho\zeta + \sqrt{1 - \rho^2}Z| > x}\mathbf{1}_{\zeta > x}] \leq 0. \quad (29)$$

We have

$$\mathbf{E}[(Z^2 - \alpha)(\zeta^2 - \alpha)\mathbf{1}_{|\rho\zeta + \sqrt{1 - \rho^2}Z| > x}\mathbf{1}_{\zeta > x}] = \mathbf{E}[(\zeta^2 - \alpha)f(\zeta)\mathbf{1}_{\zeta > x}]$$

where

$$\begin{aligned} f(\zeta) &:= \mathbf{E}[(Z^2 - \alpha)\mathbf{1}_{|\rho\zeta + \sqrt{1 - \rho^2}Z| > x} \mid \zeta] \\ &= \int_{\frac{x - \rho\zeta}{\sqrt{1 - \rho^2}}}^{\infty} (z^2 - \alpha) \exp\left(-\frac{z^2}{2}\right) dz + \int_{-\infty}^{-\frac{x + \rho\zeta}{\sqrt{1 - \rho^2}}} (z^2 - \alpha) \exp\left(-\frac{z^2}{2}\right) dz. \end{aligned}$$

Note that $x \leq \sqrt{\alpha}$ since

$$\alpha = \frac{\mathbf{E}[\eta^2 \mathbf{1}_{|\eta| > x}]}{\mathbf{P}(|\eta| > x)} \geq x^2.$$

In order to prove (29), it is enough to show that

$$\forall \zeta \in [x, \sqrt{\alpha}], \quad f(\zeta) \geq f(\sqrt{\alpha}). \quad (30)$$

and

$$\forall \zeta \in [\sqrt{\alpha}, \infty), \quad f(\zeta) \leq f(\sqrt{\alpha}). \quad (31)$$

Indeed, assume that (30) and (31) hold. Then we have

$$\begin{aligned} \mathbf{E}[(\zeta^2 - \alpha)f(\zeta)\mathbf{1}_{x < \zeta \leq \sqrt{\alpha}}] &\leq \mathbf{E}[(\zeta^2 - \alpha)f(\sqrt{\alpha})\mathbf{1}_{x < \zeta \leq \sqrt{\alpha}}] \\ &= -\mathbf{E}[(\zeta^2 - \alpha)f(\sqrt{\alpha})\mathbf{1}_{\zeta > \sqrt{\alpha}}] \leq -\mathbf{E}[(\zeta^2 - \alpha)f(\zeta)\mathbf{1}_{\zeta > \sqrt{\alpha}}], \end{aligned}$$

where the equality is due the fact that, by the symmetry of the normal distribution and the definition of α ,

$$\mathbf{E}[(\zeta^2 - \alpha)\mathbf{1}_{\zeta > x}] = \frac{1}{2}\mathbf{E}[(\zeta^2 - \alpha)\mathbf{1}_{|\zeta| > x}] = 0.$$

Thus, to finish the proof of the lemma, it remains to prove (30) and (31). We first establish (30), for which it is sufficient to show that $f'(\zeta) < 0$ for $\zeta \in [x, \sqrt{\alpha}]$. Since $0 < \rho < x/\sqrt{\alpha}$ and $x < \sqrt{\alpha}$, we have

$$\frac{(x - \rho y)^2}{1 - \rho^2} < \alpha \quad \text{for all } y \in [x, \sqrt{\alpha}]. \quad (32)$$

Using (32) we obtain, for all $\zeta \in [x, \sqrt{\alpha}]$,

$$\begin{aligned} f'(\zeta) &= \frac{\rho}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{x+\rho\zeta}{\sqrt{1-\rho^2}}\right)^2\right) \left(\left(\frac{(x-\rho\zeta)^2}{1-\rho^2} - \alpha\right) \exp\left(\frac{2\rho x\zeta}{1-\rho^2}\right) - \left(\frac{(x+\rho\zeta)^2}{1-\rho^2} - \alpha\right)\right) \\ &\leq \frac{\rho}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{x+\rho\zeta}{\sqrt{1-\rho^2}}\right)^2\right) \left(\left(\frac{(x-\rho\zeta)^2}{1-\rho^2} - \alpha\right) - \left(\frac{(x+\rho\zeta)^2}{1-\rho^2} - \alpha\right)\right) \\ &= -\frac{\rho}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{x+\rho\zeta}{\sqrt{1-\rho^2}}\right)^2\right) \frac{4x\rho\zeta}{1-\rho^2} < 0. \end{aligned}$$

This implies (30). Finally, we prove (31). To do this, it is enough to establish the following three facts:

- (i) f' is continuous and $f'(\sqrt{\alpha}) < 0$;
- (ii) the equation $f'(y) = 0$ has at most one solution on $[\sqrt{\alpha}, +\infty)$;
- (iii) $f(\infty) = \lim_{y \rightarrow \infty} f(y) \leq f(\sqrt{\alpha})$.

Property (i) is already proved above. To prove (ii), we first observe that the solution of the equation $\frac{d}{dy}f(y) = 0$ is also solution of the equation $h(y) = 0$ where

$$h(y) := \left(\frac{(x-\rho y)^2}{1-\rho^2} - \alpha\right) \left(\exp\left(\frac{2\rho xy}{1-\rho^2}\right) - 1\right) - \frac{4\rho xy}{1-\rho^2}.$$

Next, let y_1 and y_2 be the solutions of the quadratic equation $\frac{(x-\rho y)^2}{1-\rho^2} = \alpha$:

$$y_1 = \frac{x - \sqrt{\alpha(1-\rho^2)}}{\rho} \quad \text{and} \quad y_2 = \frac{x + \sqrt{\alpha(1-\rho^2)}}{\rho}.$$

Due to (32) we have $y_1 < \sqrt{\alpha} < y_2$. Thus, $h(y) < 0$ on the interval $[\sqrt{\alpha}, y_2]$. Next, on the interval $(y_2, +\infty)$ the function h is strictly convex and $h(y_2) < 0$. It follows that h can vanish at most once on $(y_2, +\infty)$. Thus, (ii) is proved.

It remains to show that $f(\sqrt{\alpha}) \geq f(\infty) = \int_{-\infty}^{\infty} (z^2 - \alpha) \exp(-z^2/2) dz$. Rewriting $f(\sqrt{\alpha})$ as

$$f(\sqrt{\alpha}) = f(\infty) - \int_{-\frac{x+\rho\sqrt{\alpha}}{\sqrt{1-\rho^2}}}^{\frac{x-\rho\sqrt{\alpha}}{\sqrt{1-\rho^2}}} (z^2 - \alpha) \exp\left(-\frac{z^2}{2}\right) dz$$

we see that inequality $f(\infty) \leq f(\sqrt{\alpha})$ follows from (32). This proves item (iii) and thus (31). Therefore, the proof of (29) is complete. Combining (22), (23), (24), (28) and (29) yields the lemma. ■

5. Proof of Theorem 1

We consider separately the cases $s \geq \sqrt{p}$ and $s < \sqrt{p}$.

Case $s \geq \sqrt{p}$. From (1) we get that, almost surely,

$$(X^T X)^{-1} X^T Y = \theta + \tilde{\epsilon}$$

where

$$\tilde{\epsilon} = \sigma(X^T X)^{-1} X^T \xi.$$

Thus, we have

$$\begin{aligned} \mathbf{E}_\theta [(\hat{Q} - \|\theta\|_2^2)^2] &= \mathbf{E}_\theta (2\theta^T \tilde{\epsilon} + \|\tilde{\epsilon}\|_2^2 - \sigma^2 \text{tr}[(X^T X)^{-1}])^2 \\ &\leq 8 \mathbf{E}_\theta (\theta^T \tilde{\epsilon})^2 + 2 \mathbf{E}_\theta \left(\|\tilde{\epsilon}\|_2^2 - \sigma^2 \text{tr}[(X^T X)^{-1}] \right)^2. \end{aligned} \quad (33)$$

Note that, conditionally on X , the random vector $\tilde{\epsilon}$ is normal with mean 0 and covariance matrix $\sigma^2(X^T X)^{-1}$. Thus, conditionally on X , the random variable $\theta^T \tilde{\epsilon}$ is normal with mean 0 and variance $\sigma^2 \theta^T (X^T X)^{-1} \theta$. It follows that $\mathbf{E}_\theta (\theta^T \tilde{\epsilon})^2 \leq \sigma^2 r^2 \mathbf{E}[\lambda_{\min}^{-1}(X^T X)]$. Hence, applying Lemma 3 we have, for some constant C depending only on γ ,

$$\mathbf{E}_\theta (\theta^T \tilde{\epsilon})^2 \leq C \sigma^2 \frac{r^2}{n}. \quad (34)$$

Consider now the second term on the right hand side of (33). Denote by (λ_i, u_i) , $i = 1, \dots, p$, the eigenvalues and the corresponding orthonormal eigenvectors of $(X^T X)^{-1}$, respectively. Set $v_i = \sqrt{\lambda_i} u_i^T X^T \xi$. We have

$$\mathbf{E}_\theta \left(\|\tilde{\epsilon}\|_2^2 - \sigma^2 \text{tr}[(X^T X)^{-1}] \right)^2 = \sigma^4 \mathbf{E} \left(\sum_{i=1}^p \lambda_i [v_i^2 - 1] \right)^2.$$

Conditionnally on X , the random variables v_1, \dots, v_p are i.i.d. standard normal. Using this remark and Lemma 3 we get that, for some constant C depending only on γ ,

$$\mathbf{E}_\theta \left(\|\tilde{\epsilon}\|_2^2 - \sigma^2 \text{tr}[(X^T X)^{-1}] \right)^2 = 2\sigma^4 \sum_{i=1}^p \lambda_i^2 \leq 2p\sigma^4 \mathbf{E}[\lambda_{\min}^{-2}(X^T X)] \leq C \frac{\sigma^4 p}{n^2}. \quad (35)$$

Combining (33), (34) and (35) we obtain the result of the theorem for $s \geq \sqrt{p}$.

Case $s < \sqrt{p}$. Set $S = \{i : \theta_i \neq 0\}$. We have

$$\begin{aligned} \mathbf{E}_\theta (\hat{Q} - \|\theta\|_2^2)^2 &\leq 3 \mathbf{E}_\theta \left(\sum_{i \in S} (y_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s - \theta_i^2) \right)^2 \\ &\quad + 3 \mathbf{E}_\theta \left(\sum_{i \in S} [y_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s] \mathbf{1}_{y_i^2 \leq 2\sigma^2 (X^T X)_{ii}^{-1} \log(1+p/s^2)} \right)^2 \\ &\quad + 3 \mathbf{E}_\theta \left(\sum_{i \notin S} [\tilde{\epsilon}_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s] \mathbf{1}_{y_i^2 > 2\sigma^2 (X^T X)_{ii}^{-1} \log(1+p/s^2)} \right)^2, \end{aligned} \quad (36)$$

where $\tilde{\epsilon}_i$ denotes the i th component of $\tilde{\epsilon}$. We now establish upper bounds for the three terms on the right hand side of (36). For the first term, observe that

$$\mathbf{E}_\theta \left(\sum_{i \in S} (y_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s - \theta_i^2) \right)^2 \leq 8 \mathbf{E}_\theta \left(\sum_{i \in S} \theta_i \tilde{\epsilon}_i \right)^2 + 2 \mathbf{E}_\theta \left(\sum_{i \in S} (\tilde{\epsilon}_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s) \right)^2. \quad (37)$$

The second summand on the right hand side of (37) satisfies

$$\begin{aligned} \mathbf{E}_\theta \left(\sum_{i \in S} (\tilde{\epsilon}_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s) \right)^2 &\leq 2\sigma^4 (\alpha_s^2 + 3) \mathbf{E} \sum_{i \in S} \sum_{j \in S} (X^T X)_{ii}^{-1} (X^T X)_{jj}^{-1} \\ &\leq 2\sigma^4 (\alpha_s^2 + 3) s^2 \mathbf{E} [\lambda_{\min}^{-2}(X^T X)]. \end{aligned} \quad (38)$$

From (20) we obtain

$$\alpha_s \leq 10 \log(1 + p/s^2). \quad (39)$$

Thus, using (37), (38) and (39) together with Lemma 3 and (34) we find

$$\mathbf{E}_\theta \left(\sum_{i \in S} (y_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s - \theta_i^2) \right)^2 \leq C \sigma^4 s^2 \log^2(1 + p/s^2)/n^2, \quad (40)$$

where the constant C depends only on γ . For the second term on the right hand side of (36), we have immediately that it is smaller, up to an absolute constant factor, than

$$\mathbf{E} \sigma^4 \sum_{i \in S} \sum_{j \in S} (X^T X)_{ii}^{-1} (X^T X)_{jj}^{-1} (\alpha_s^2 + 4 \log^2(1 + p/s^2)).$$

Arguing as in (38) and applying Lemma 3 and (39) we get that, for some constant C depending only on γ ,

$$\mathbf{E}_\theta \left(\sum_{i \in S} [y_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s] \mathbf{1}_{y_i^2 \leq 2\sigma^2 (X^T X)_{ii}^{-1} \log(1+p/s^2)} \right)^2 \leq C \sigma^4 s^2 \log^2(1 + p/s^2)/n^2. \quad (41)$$

For the third term on the right hand side of (36), we have

$$\begin{aligned} & \mathbf{E}_\theta \left(\sum_{i \notin S} [\tilde{\epsilon}_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s] \mathbf{1}_{y_i^2 > 2\sigma^2 (X^T X)_{ii}^{-1} \log(1+p/s^2)} \right)^2 \\ &= \sigma^4 \sum_{i \notin S} \sum_{j \notin S} \mathbf{E} \left((X^T X)_{ii}^{-1} (X^T X)_{jj}^{-1} (\tilde{\xi}_i^2 - \alpha_s)(\tilde{\xi}_j^2 - \alpha_s) \mathbf{1}_{|\tilde{\xi}_i| > x} \mathbf{1}_{|\tilde{\xi}_j| > x} \right), \end{aligned} \quad (42)$$

where

$$x = \sqrt{2 \log(1 + p/s^2)}, \quad \tilde{\xi}_i = \frac{\tilde{\epsilon}_i}{\sqrt{\sigma^2 (X^T X)_{ii}^{-1}}}.$$

Note that $\mathbf{E}(\tilde{\xi}_i^2 | X) = \mathbf{E}(\tilde{\xi}_j^2 | X) = 1$ and, conditionally on X , $(\tilde{\xi}_i, \tilde{\xi}_j)$ is a centered Gaussian vector with covariance ρ_{ij} where

$$\rho_{ij} = \frac{(X^T X)_{ij}^{-1}}{\sqrt{(X^T X)_{ii}^{-1}} \sqrt{(X^T X)_{jj}^{-1}}}.$$

Using Lemma 5 we obtain that, for some absolute positive constants C ,

$$\begin{aligned} & \sum_{i \notin S} \sum_{j \notin S} \mathbf{E} \left((X^T X)_{ii}^{-1} (X^T X)_{jj}^{-1} (\tilde{\xi}_i^2 - \alpha_s)(\tilde{\xi}_j^2 - \alpha_s) \mathbf{1}_{|\tilde{\xi}_i| > x} \mathbf{1}_{|\tilde{\xi}_j| > x} \right) \\ &= \sum_{i \notin S} \sum_{j \notin S} \mathbf{E} \left((X^T X)_{ii}^{-1} (X^T X)_{jj}^{-1} \mathbf{E} \left[(\tilde{\xi}_i^2 - \alpha_s)(\tilde{\xi}_j^2 - \alpha_s) \mathbf{1}_{|\tilde{\xi}_i| > x} \mathbf{1}_{|\tilde{\xi}_j| > x} \mid X \right] \right) \\ &\leq C \sum_{i,j=1}^p \mathbf{E} \left[(X^T X)_{ii}^{-1} (X^T X)_{jj}^{-1} \rho_{ij}^2 \right] x^4 \exp(-x^2/2) \\ &= C \mathbf{E} \left[\|(X^T X)^{-1}\|_F^2 \right] x^4 \exp(-x^2/2) \\ &\leq C \sigma^4 \mathbf{E} \left[\|(X^T X)^{-1}\|_F^2 \right] \frac{s^2}{p} \log^2(1 + p/s^2) \\ &\leq C \sigma^4 \mathbf{E} \left[\lambda_{\min}^{-2}(X^T X) \right] s^2 \log^2(1 + p/s^2), \end{aligned}$$

where $\|(X^T X)^{-1}\|_F$ is the Frobenius norm of matrix $(X^T X)^{-1}$. Finally, Lemma 3, (42) and the last display imply that, for some constant C depending only on γ ,

$$\mathbf{E}_\theta \left(\sum_{i \notin S} \left[\hat{\epsilon}_i^2 - \sigma^2 (X^T X)_{ii}^{-1} \alpha_s \right] \mathbb{1}_{y_i^2 > 2\sigma^2 (X^T X)_{ii}^{-1} \log(1+p/s^2)} \right)^2 \leq C \frac{\sigma^4 s^2 \log^2(1+p/s^2)}{n^2}. \quad (43)$$

The proof is completed by combining (36), (40), (41) and (43).

6. Preliminary lemmas for the proof of Theorem 4

We first recall some general facts about lower bounds for the risks of tests. Let Θ be a measurable set, not necessarily the set $\Theta(s, \tau)$, and let μ be a probability measure on Θ . Consider any family of probability measures \mathbf{P}_θ indexed by $\theta \in \Theta$. Denote by \mathbb{P}_μ the mixture probability measure

$$\mathbb{P}_\mu = \int_{\Theta} \mathbf{P}_\theta \mu(d\theta).$$

Let

$$\chi^2(P', P) = \int (dP'/dP)^2 dP - 1$$

be the chi-square divergence between two probability measures P' and P if $P' \ll P$, and set $\chi^2(P', P) = +\infty$ otherwise. The following lemma is a version of Le Cam's method in proving lower bounds (see, e.g., (Collier, Comminges and Tsybakov, 2017, Lemma 3)).

Lemma 6 *Let μ be a probability measure on Θ , and let $\{\mathbf{P}_\theta : \theta \in \Theta\}$ be a family of probability measures indexed by $\theta \in \Theta$ on \mathcal{X} . Then, for any probability measure Q on \mathcal{X} ,*

$$\inf_{\Delta} \left\{ Q(\Delta = 1) + \sup_{\theta \in \Theta} \mathbf{P}_\theta(\Delta = 0) \right\} \geq 1 - \sqrt{\chi^2(\mathbb{P}_\mu, Q)}$$

where \inf_{Δ} is the infimum over all $\{0, 1\}$ -valued statistics.

Applying Lemma 6 with $Q = \mathbf{P}_0$, we see that it suffices to choose a suitable measure μ and to bound $\chi^2(\mathbb{P}_\mu, \mathbf{P}_0)$ from above by a small enough value in order to obtain the desired lower bound on $\mathcal{R}_{s, \tau}$. The following lemma is useful to evaluate $\chi^2(\mathbb{P}_\mu, \mathbf{P}_0)$.

Lemma 7 *Let μ be a probability measure on Θ , and let $\{\mathbf{P}_\theta : \theta \in \Theta\}$ be a family of probability measures indexed by $\theta \in \Theta$ on \mathcal{X} . Let Q be a probability measure on \mathcal{X} such that $\mathbf{P}_\theta \ll Q$ for all $\theta \in \Theta$. Then,*

$$\chi^2(\mathbb{P}_\mu, Q) = \mathbb{E}_{(\theta, \theta') \sim \mu^2} \left(\int \frac{d\mathbf{P}_\theta d\mathbf{P}_{\theta'}}{dQ} \right) - 1.$$

Here, $\mathbb{E}_{(\theta, \theta') \sim \mu^2}$ denotes the expectation with respect to the distribution of the pair (θ, θ') where θ and θ' are independent and each of them is distributed according to μ .

Proof. It suffices to note that

$$\chi^2(\mathbb{P}_\mu, Q) = \int \frac{(d\mathbb{P}_\mu)^2}{dQ} - 1$$

whereas

$$\int \frac{(d\mathbb{P}_\mu)^2}{dQ} = \int \int_{\Theta} \frac{d\mathbf{P}_\theta \mu(d\theta) \int_{\Theta} d\mathbf{P}_{\theta'} \mu(d\theta')}{dQ} = \int_{\Theta} \int_{\Theta} \mu(d\theta) \mu(d\theta') \int \frac{d\mathbf{P}_\theta d\mathbf{P}_{\theta'}}{dQ}.$$

■

We now specify the expression for the χ^2 divergence in Lemma 7 when \mathbf{P}_θ is the probability distribution generated by model (1) and $Q = \mathbf{P}_0$.

Lemma 8 Let \mathbf{P}_θ be the distribution of (X, Y) satisfying (1). Then,

$$\chi^2(\mathbb{P}_\mu, \mathbf{P}_0) = \mathbb{E}_{(\theta, \theta') \sim \mu^2} E_X \exp(\langle X\theta, X\theta' \rangle / \sigma^2) - 1.$$

Proof. We apply Lemma 7 and notice that, for any $(\theta, \theta') \in \Theta \times \Theta$,

$$\begin{aligned} \int \frac{d\mathbf{P}_\theta d\mathbf{P}_{\theta'}}{d\mathbf{P}_0} &= \frac{1}{(2\pi\sigma)^{n/2}} E_X \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2\sigma^2}(\|y - X\theta\|_2^2 + \|y - X\theta'\|_2^2 - \|y\|_2^2)\right) dy \\ &= \frac{1}{(2\pi\sigma)^{n/2}} E_X \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2\sigma^2}(\|y\|_2^2 - 2\langle y, X(\theta + \theta') \rangle + \|X(\theta + \theta')\|_2^2 - 2\langle X\theta, X\theta' \rangle)\right) dy \\ &= E_X \left(\frac{\exp(\langle X\theta, X\theta' \rangle / \sigma^2)}{(2\pi\sigma)^{n/2}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2\sigma^2}\|y - X(\theta + \theta')\|_2^2\right) dy \right) \\ &= E_X \exp(\langle X\theta, X\theta' \rangle / \sigma^2). \end{aligned}$$

■

Lemma 9 Let $a \in \mathbb{R}$ be a constant and let W be a random variable. Then,

$$\mathbf{E} \exp(W) \leq \exp(a) \left(1 + \int_0^\infty e^t p(t) dt\right)$$

where $p(t) = \mathbf{P}(|W - a| \geq t)$.

Proof. We have

$$\begin{aligned} \mathbf{E} \exp(W) &\leq \exp(a) \mathbf{E} \exp(|W - a|) \\ &= \exp(a) \int_0^\infty \mathbf{P}(\exp(|W - a|) \geq x) dx \\ &= \exp(a) \left[1 + \int_1^\infty \mathbf{P}(\exp(|W - a|) \geq x) dx\right] \\ &= \exp(a) \left[1 + \int_0^\infty e^t p(t) dt\right]. \end{aligned}$$

■

Lemma 10 Assume that matrix X has an isotropic distribution with independent σ_X -subgaussian rows for some $\sigma_X > 0$. Then, for all $x > 0$ and all $\theta, \theta' \in \mathbb{R}^p$ we have

$$P_X\left(|\langle X\theta, X\theta' \rangle - n\langle \theta, \theta' \rangle| \geq \|\theta\|_2 \|\theta'\|_2 x\right) \leq 6 \exp(-C_1 \min(x, x^2/n))$$

where the constant $C_1 > 0$ depends only on σ_X .

Proof. By homogeneity, it is enough to consider the case $\|\theta\|_2 = \|\theta'\|_2 = 1$, which will be assumed in the rest of the proof. Then we have

$$\langle X\theta, X\theta' \rangle = \frac{\|X\theta\|_2^2 + \|X\theta'\|_2^2 - \|X(\theta - \theta')\|_2^2}{2}, \quad \langle \theta, \theta' \rangle = \frac{2 - \|\theta - \theta'\|_2^2}{2},$$

which implies

$$\begin{aligned} \left| \frac{1}{n} \langle X\theta, X\theta' \rangle - \langle \theta, \theta' \rangle \right| &\leq \frac{1}{2} \left(\left| \frac{1}{n} \|X\theta\|_2^2 - 1 \right| + \left| \frac{1}{n} \|X\theta'\|_2^2 - 1 \right| \right. \\ &\quad \left. + \left| \frac{1}{n} \|X(\theta - \theta')\|_2^2 - \|\theta - \theta'\|_2^2 \right| \right), \end{aligned} \quad (44)$$

By renormalization, the third summand on the right hand side of (44) is reduced to the same form as the first two summands. Thus, to prove the lemma it suffices to show that

$$P_X \left(\left| \frac{1}{n} \|X\theta\|_2^2 - 1 \right| \geq v \right) \leq 2 \exp(-C'_1 \min(v, v^2)n), \quad \forall v > 0, \|\theta\|_2 = 1, \quad (45)$$

where the constant $C'_1 > 0$ depends only on σ_X .

To prove (45), we apply a standard argument, cf., e.g., Vershynin (2012). Denote by \mathbf{x}_i the i th row of matrix X . For $i = 1, \dots, n$, the random variables $Z_i = \mathbf{x}_i^T \theta$ are independent, σ_X -subgaussian and such that $\mathbf{E}(Z_i^2) = 1$. Therefore, $Z_i^2 - 1$, $i = 1, \dots, n$, are independent centered sub-exponential random variables, while

$$\frac{1}{n} \|X\theta\|_2^2 - 1 = \frac{1}{n} \sum_{i=1}^n (Z_i^2 - 1).$$

Then (45) follows from Bernstein's inequality for sub-exponential random variables (cf., e.g., Vershynin (2012), Corollary 5.17). ■

Lemma 11 *Assume that matrix X has an isotropic distribution with independent σ_X -subgaussian rows for some $\sigma_X > 0$. Then, there exists $u_0 > 0$ depending only on σ_X such that, for all θ, θ' with $\|\theta\|_2, \|\theta'\|_2 \leq un^{-1/4}$ and $u \in (0, u_0)$ we have*

$$E_X \exp(\langle X\theta, X\theta' \rangle) \leq \exp(n\langle \theta, \theta' \rangle)(1 + C_0 u^2)$$

where the constant $C_0 > 0$ depends only on σ_X .

Proof. By Lemma 10, for any $x > 0$ with P_X -probability at least $1 - 6e^{-C_1 \min(x, x^2/n)}$ we have

$$\left| \langle X\theta, X\theta' \rangle - n\langle \theta, \theta' \rangle \right| \leq \|\theta\|_2 \|\theta'\|_2 x \leq u^2 n^{-1/2} x.$$

Therefore, for any $t > 0$ with P_X -probability at least $1 - 6e^{-C_1 \min(\sqrt{nt}/u^2, t^2/u^4)}$ we have

$$\left| \langle X\theta, X\theta' \rangle - n\langle \theta, \theta' \rangle \right| \leq t.$$

This and Lemma 9 imply that, for all $u \leq u_0 := (C_1/2)^{1/2}$,

$$\begin{aligned} E_X \exp(\langle X\theta, X\theta' \rangle) &\leq \exp(n\langle \theta, \theta' \rangle) \left(1 + 6 \int_0^\infty e^{t - C_1 \min(\sqrt{nt}/u^2, t^2/u^4)} dt \right) \\ &\leq \exp(n\langle \theta, \theta' \rangle) \left(1 + 6 \int_0^\infty e^{t(1 - C_1 \sqrt{n}/u^2)} dt + 6 \int_0^\infty e^{t - C_1 t^2/u^4} dt \right) \\ &\leq \exp(n\langle \theta, \theta' \rangle) \left(1 + 6 \int_0^\infty e^{-C_1 \sqrt{nt}/(2u^2)} dt + 6 \int_0^\infty e^{-t(C_1 t/u^4 - 1)} dt \right) \quad (\text{as } C_1 \sqrt{n}/u^2 > 2) \\ &\leq \exp(n\langle \theta, \theta' \rangle) \left(1 + \frac{12u^2}{C_1 \sqrt{n}} + \frac{12u^4}{C_1} e^{2u^4/C_1} + 6 \int_{2u^4/C_1}^\infty e^{-t^2 C_1/(2u^4)} dt \right) \\ &\leq \exp(n\langle \theta, \theta' \rangle) (1 + C_0 u^2), \end{aligned} \quad (46)$$

where the constant $C_0 > 0$ depends only on C_1 , and thus only on σ_X . ■

7. Proof of Theorem 4

For an integer s such that $1 \leq s \leq p$ and $\tau > 0$, we denote by μ_τ the uniform distribution on the set of vectors in \mathbb{R}^p having exactly s nonzero coefficients, all equal to τ/\sqrt{s} . Note that the support of measure μ_τ is contained in $\Theta(s, \tau)$.

We now take $\tau = \tau(s, n, p)$ defined by (11) and set $\mu = \mu_\tau$. In view of Lemmas 6 - 8, to prove Theorem 4 it is enough to show that

$$\mathbb{E}_{(\theta, \theta') \sim \mu_\tau^2} E_X \exp(\langle X\theta, X\theta' \rangle / \sigma^2) \leq 1 + o_A(1) \quad (47)$$

where $o_A(1)$ tends to 0 as $A \rightarrow 0$.

Before proving (47) we proceed to some simplifications. First, note that for τ defined by (11) the left hand side of (47) does not depend on σ . Thus, in what follows we set $\sigma = 1$ without loss of generality. Next, notice that it is enough to prove the theorem for the case $s \leq \sqrt{p}$. Indeed, for $s > \sqrt{p}$ we can use the inclusions $\Theta(s, \tau(s, n, p)) \supseteq \Theta(s', \tau(s, n, p)) \supseteq \Theta(s', \tau(s', n, p))$ where s' is the greatest integer smaller than or equal to \sqrt{p} . Since

$$\tau(s', n, p) \asymp \min\left(\frac{p^{1/4}}{\sqrt{n}}, n^{-1/4}\right)$$

and the rate (11) is also of this order for $s > \sqrt{p}$, it suffices to prove the lower bound for $s \leq s'$, and thus for $s \leq \sqrt{p}$. Taking onto account these simplifications, we assume in what follows without loss of generality that $s \leq \sqrt{p}$, $\sigma = 1$, and

$$\tau := A \min\left(\sqrt{\frac{s \log(1 + p/s^2)}{n}}, n^{-1/4}\right). \quad (48)$$

We now prove (47) under these assumptions. By Lemma 11, for any $0 < A < u_0$ we have

$$\mathbb{E}_{(\theta, \theta') \sim \mu_\tau^2} E_X \exp(\langle X\theta, X\theta' \rangle) \leq \mathbb{E}_{(\theta, \theta') \sim \mu_\tau^2} \exp(n\langle \theta, \theta' \rangle) (1 + C_0 A^2). \quad (49)$$

Assume that $A < 1$. Arguing exactly as in the proof of Lemma 1 in Collier, Comminges and Tsybakov (2017), we find

$$\begin{aligned} \mathbb{E}_{(\theta, \theta') \sim \mu_\tau^2} \exp(n\langle \theta, \theta' \rangle) &= \mathbb{E}_{(\theta, \theta') \sim \mu_\tau^2} \exp\left(n\tau^2 s^{-1} \sum_{j=1}^p \mathbf{1}_{\theta_j \neq 0} \mathbf{1}_{\theta'_j \neq 0}\right) \\ &\leq \left(1 - \frac{s}{p} + \frac{s}{p} \exp(n\tau^2 s^{-1})\right)^s \\ &\leq \left(1 - \frac{s}{p} + \frac{s}{p} \left(1 + \frac{p}{s^2}\right)^{A^2}\right)^s \\ &\leq \left(1 + \frac{A^2}{s}\right)^s \leq \exp(A^2) \end{aligned} \quad (50)$$

where we have used the inequality $(1+x)^{A^2} - 1 \leq A^2 x$ valid for $0 < A < 1$ and $x > 0$. Combining (49) and (50) we obtain that, for all $0 < A < \min(1, u_0)$,

$$\mathbb{E}_{(\theta, \theta') \sim \mu_\tau^2} E_X \exp(\langle X\theta, X\theta' \rangle) \leq \exp(A^2) (1 + C_0 A^2)$$

with some $u_0 > 0$ and $C_0 > 0$ depending only on σ_X . This completes the proof of the theorem.

Acknowledgements

The work of Alexandra Carpentier is supported by the Emmy Noether grant MuSyAD CA 1488/1-1, by the GK 2297 MathCoRe on Mathematical Complexity Reduction” 314838170, GRK 2297 MathCoRe, and by the SFB 1294 Data Assimilation on Data Assimilation “The seamless integration of data and models”, Project A03, all funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), and by the Deutsch-Französisches Doktorandenkolleg/Collège doctoral franco-allemand Potsdam-Toulouse CDFA 01-18, funded by the French-German University. Olivier Collier’s research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). The work of A.B.Tsybakov was supported by GENES and by the French National Research Agency (ANR) under the grants IPANEMA (ANR-13-BSH1-0004-02) and Labex Ecodec (ANR-11-LABEX-0047).

References

- ARIAS-CASTRO, E. CANDÈS, E. AND PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556.
- BARAUD, Y. (2002). Non asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606.
- BORDENAVE, C. AND CHAFAÏ, D. (2012). Around the circular law. *Probability Surveys*. **9** 1-89.
- COLLIER, O., COMMINGES, L., AND TSYBAKOV, A.B. (2017). Minimax estimation of linear and quadratic functionals under sparsity constraints. *Ann. Statist.* **45** 923–958.
- DONOHO, D.L. AND JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994.
- INGSTER, Y.I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–49.
- INGSTER, Y.I. AND SUSLINA, I.A. (2003). *Nonparametric Goodness-of-Fit Testing under Gaussian Models*. Springer, New York.
- INGSTER, Y.I., TSYBAKOV, A.B. AND VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526.
- DAVIDSON, K.R. AND SZAREK, S.J. (2001) Local operator theory, random matrices and Banach spaces. *Handbook of the geometry of Banach spaces* **1** 317–366
- TAO, T. AND VU, V. (2012) Random matrices: Universality of ESDs and the circular law. *The Annals of Probability*, 2023–2065.
- VERSHYNIN, R. (2012) Introduction to the non-asymptotic analysis of random matrices. In: *Compressed sensing*, 210–268, Cambridge Univ. Press, Cambridge.
- VERZELEN N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomena. *Electron. J. Stat.* **6** 38–90.