



Estimation of the ℓ_2 -norm and testing in sparse linear regression with unknown variance

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

► To cite this version:

Alexandra Carpentier, Olivier Collier, Laëtitia Comminges, Alexandre B. Tsybakov, Yuhao Wang. Estimation of the ℓ_2 -norm and testing in sparse linear regression with unknown variance. *Bernoulli*, 2022, 28 (4). hal-03879507

HAL Id: hal-03879507

<https://hal.science/hal-03879507>

Submitted on 30 Nov 2022

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.

Estimation of the ℓ_2 -norm and testing in sparse linear regression with unknown variance

A. Carpentier, O. Collier, L. Comminges, A.B. Tsybakov and Y. Wang*

October 27, 2020

Abstract

We consider the related problems of estimating the ℓ_2 -norm and the squared ℓ_2 -norm in sparse linear regression with unknown variance, as well as the problem of testing the hypothesis that the regression parameter is null under sparse alternatives with ℓ_2 separation. We establish the minimax optimal rates of estimation (respectively, testing) in these three problems.

1 Introduction

Assume that we have observations (\mathbb{X}, \mathbf{Y}) satisfying the model

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\theta} + \sigma\boldsymbol{\xi}, \quad (1)$$

where \mathbb{X} is a $N \times p$ random matrix, $\boldsymbol{\theta} \in \mathbb{R}^p$ is an unknown parameter of interest, $\boldsymbol{\xi} \in \mathbb{R}^N$ is the random noise, and $\sigma > 0$ is the unknown noise level. We assume that the entries $X_{ij}, i = 1, \dots, N, j = 1, \dots, p$, and $\xi_i, i = 1, \dots, N$, of matrix \mathbb{X} and of vector $\boldsymbol{\xi}$ constitute a collection of $N(p+1)$ jointly independent zero mean random variables with variance 1.

We use the notation $\mathbf{P}_{\boldsymbol{\theta}, \sigma}$ (or \mathbf{P} when there is no ambiguity) for the probability distribution of (\mathbb{X}, \mathbf{Y}) satisfying model (1), and $\mathbf{E}_{\boldsymbol{\theta}, \sigma}$ (or \mathbf{E} when there is no ambiguity) for the expectation with respect to this distribution.

For a vector $\mathbf{u} = (u_i)_{1 \leq i \leq p} \in \mathbb{R}^p$, we define the " ℓ_0 -norm" $\|\cdot\|_0$, the ℓ_1 -norm $\|\cdot\|_1$ and the ℓ_2 -norm $\|\cdot\|_2$ by the relations

$$\|\mathbf{u}\|_0 = \sum_{i=1}^p \mathbb{1}\{u_i \neq 0\}, \quad \|\mathbf{u}\|_1 = \sum_{i=1}^p |u_i|, \quad \|\mathbf{u}\|_2^2 = \sum_{i=1}^p u_i^2,$$

where $\mathbb{1}\{\cdot\}$ is the indicator function.

We assume that $\boldsymbol{\theta}$ belongs to a set of s -sparse vectors defined as follows:

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

*A. Carpentier is with the University of Magdeburg, O. Collier is with Modal'X, Université Paris-Nanterre and CREST. L. Comminges is with CEREMADE, Université Paris-Dauphine and CREST. A.B. Tsybakov is with CREST, ENSAE, Institut Polytechnique de Paris. Y. Wang is with Tsinghua University and Shanghai Qi Zhi Institute.

MSC 2010 subject classifications: 62J05, 62G10. **Keywords and phrases:** linear regression, sparsity, signal detection.

where s is an integer such that $1 \leq s \leq p$.

In this paper, we consider the related problems of estimating the ℓ_2 -norm and the squared ℓ_2 -norm of $\boldsymbol{\theta}$, as well as testing the null hypothesis that $\boldsymbol{\theta} = 0$ under sparse alternatives with ℓ_2 separation (signal detection in ℓ_2 -norm).

Estimation of $\|\boldsymbol{\theta}\|_2$ and $\|\boldsymbol{\theta}\|_2^2$. We consider the problem of minimax optimal estimation of the norm $\|\boldsymbol{\theta}\|_2$ and of the squared norm $Q(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_2^2$ on the class of s -sparse vectors $B_0(s)$ when the noise level σ is unknown. Namely, for a fixed $\delta \in (0, 1)$, we find estimators $\hat{\Lambda}$ of $\|\boldsymbol{\theta}\|_2$ such that

$$\sup_{\sigma > 0} \sup_{\boldsymbol{\theta} \in B_0(s)} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\Lambda} - \|\boldsymbol{\theta}\|_2| > \phi_*) \leq \delta, \quad (2)$$

where

$$\phi_* = \phi_*(p, N, s, \sigma, \delta) > 0,$$

is the non-asymptotic minimax optimal rate. That is, there exists a constant $c_\delta > 0$ that depends only on δ such that

$$\inf_{\hat{T}} \sup_{\sigma > 0} \sup_{\boldsymbol{\theta} \in B_0(s)} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{T} - \|\boldsymbol{\theta}\|_2| > c_\delta \phi_*) \geq \delta.$$

Here and below, $\inf_{\hat{T}}$ denotes the infimum over all estimators. Next, for a fixed $\delta \in (0, 1)$, we find estimators \hat{Q} of $Q(\boldsymbol{\theta})$ such that

$$\sup_{\sigma > 0} \sup_{\boldsymbol{\theta} \in B_0(s)} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{Q} - Q(\boldsymbol{\theta})| > q_\theta) \leq \delta, \quad (3)$$

where the scaling

$$q_\theta = q_\theta(p, N, \sigma, \delta) > 0,$$

depends on $\boldsymbol{\theta}$ only through the ℓ_2 and ℓ_0 -norms $\|\boldsymbol{\theta}\|_2$ and $\|\boldsymbol{\theta}\|_0$. We find the minimax optimal scaling $q_\theta^* = q_\theta^*(p, N, \sigma, \delta) > 0$ in the sense that there exists $c_\delta > 0$ that depends only on δ such that

$$\inf_{\hat{T}} \sup_{\sigma > 0} \sup_{\boldsymbol{\theta} \in B_0(s)} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{T} - Q(\boldsymbol{\theta})| > c_\delta q_\theta^*) \geq \delta,$$

and

$$\inf_{\hat{T}} \sup_{\sigma > 0} \sup_{\boldsymbol{\theta} \in B_0(s)} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{T} - Q(\boldsymbol{\theta})| > q_\theta^*) \leq \delta.$$

Furthermore, as q_θ^* is an increasing function of the ℓ_2 -norm $\|\boldsymbol{\theta}\|_2$, it is natural to consider the minimax setting on the intersection $B_0(s) \cap \{\|\boldsymbol{\theta}\|_2 \leq \kappa\}$ with $\kappa > 0$, for which we prove that the rate

$$q_{s, \kappa}^* = q_{s, \kappa}^*(p, N, \sigma, \delta) := \max\{q_\theta^*(p, N, \sigma, \delta) : \|\boldsymbol{\theta}\|_0 \leq s, \|\boldsymbol{\theta}\|_2 \leq \kappa\},$$

is minimax optimal. All the rates in this paper are obtained either in non-asymptotic form or for $N \geq N_0$, where N_0 does not depend on p, s, σ .

Signal detection under ℓ_2 separation. When σ is unknown, we consider the problem of testing the hypothesis $H_0 : \boldsymbol{\theta} = 0$ against the sparse alternative $H_1 : \boldsymbol{\theta} \in B_0(s), \|\boldsymbol{\theta}\|_2 \geq \sigma\rho$, where $\rho > 0$. For a test $\Delta = \Delta(\mathbb{X}, \mathbf{Y})$, i.e., a measurable function of (\mathbb{X}, \mathbf{Y}) with values in $\{0, 1\}$, we introduce the risk $R(\Delta, p, N, s, \rho)$ as the sum of the type I and type II error probabilities:

$$R(\Delta, p, N, s, \rho) = \sup_{\sigma > 0} \mathbf{P}_{\mathbf{0}, \sigma}(\Delta = 1) + \sup_{\sigma > 0} \sup_{\boldsymbol{\theta} \in B_0(s), \|\boldsymbol{\theta}\|_2 \geq \sigma\rho} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\Delta = 0). \quad (4)$$

For a fixed $\delta \in (0, 1)$, we define the minimax separation rate $\rho_* = \rho_*(p, N, s, \delta) > 0$ as

$$\rho_* := \rho_*(p, N, s, \delta) := \inf\{\rho > 0 : \inf_{\bar{\Delta}} R(\bar{\Delta}, p, N, s, \rho) \leq \delta\}.$$

Here and below, $\inf_{\bar{\Delta}}$ denotes the infimum over all tests $\bar{\Delta}$. We find a minimax rate optimal test Δ , i.e., a test such that $R(\Delta, p, N, s, \rho) \leq \delta$ for all $\rho \geq c_\delta \rho_*$ where $c_\delta > 0$ is a constant that depends only on δ . We obtain the non-asymptotic expression for the separation rate ρ_* .

2 Related work

The problems of estimating the ℓ_2 -norm and of signal detection under ℓ_2 separation are closely related to each other. We provide here an overview of the previous results for both problems.

Signal detection. In the Gaussian mean model, which corresponds to an orthogonal non-random design \mathbb{X} , the problem of signal detection has been extensively studied in the last fifteen years (see, e.g., [16, 17, 3, 13, 8, 9] and the references therein). More recently, this problem has also been investigated in the random design linear regression model, which is most related to the present paper [18, 2, 25, 6]. Among these, [2] and [6] deal with the case of known σ , while [18, 25] consider both known and unknown σ . The setting in [2] is different from ours since it does not consider the alternative defined by separation in the ℓ_2 -norm. It does not allow one to compare [2] directly with [18, 25, 6] and with the present work. Note also that [2] explores an asymptotic setting as p, N, s tend to ∞ , while in this paper we are interested in non-asymptotic results. The setting is also asymptotic in [18] where it is assumed that p, N, s tend to ∞ in such a way that $s = p^a$ for some $0 < a < 1$, and $s \log(p) = o(N)$. It can be deduced from the asymptotic argument in [18] that in this regime, and if the matrix \mathbb{X} has i.i.d. standard normal entries and $\boldsymbol{\xi}$ is standard normal, the minimax separation rate satisfies, in the case of *known* σ , the relations:

$$\begin{aligned} \rho_* &\asymp \min\left(\sqrt{\frac{s \log(p)}{N}}, N^{-1/4}\right) && \text{if } 0 < a < 1/2, \\ \rho_* &\asymp \min\left(\frac{p^{1/4}}{\sqrt{N}}, N^{-1/4}\right) && \text{if } 1/2 < a < 1. \end{aligned} \tag{5}$$

A refined analysis of the minimax risk asymptotics for testing in this model with known σ is given in [20].

Next, in the setting with *unknown* σ , [18] obtained that, in the same asymptotic regime, the asymptotic minimax separation rate is of the order of $\sqrt{\frac{s \log(p)}{N}}$ when $0 < a < 1/2$.

More insight into the behavior of tests under *unknown* σ is provided in [25], which is the first paper where non-asymptotic separation rates were explored in the linear regression context. Namely, considering the case of Gaussian \mathbb{X} and the configuration of parameters where $p > N$ and $s < p^a$ with $0 < a < 1/2$, [25] proves that the minimax separation rate ρ_* is of the order of $\sqrt{\frac{s \log(p)}{N}}$ whenever $p \lesssim N^{1/a}$, where the sign \lesssim means inequality up to a logarithmic factor. On the other hand, [25] shows that if $p \gtrsim N^{1/a}$ and $s = p^a$ with $0 < a < 1/2$, the rate ρ_* explodes becoming at least greater than a constant independent of p, N, s . This is proved for the case of \mathbb{X} with i.i.d. standard normal entries and $\boldsymbol{\xi}$ standard normal. A conclusion is that there is no much interest in considering signal detection in the

regime $p > N^b$ with $b > 2$ when σ is unknown. Regarding the case of *known* σ , [25] proves that, for \mathbb{X} with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is standard normal, the non-asymptotic rate is

$$\rho_* \asymp \min \left(\sqrt{\frac{s \log(p)}{N}}, N^{-1/4} \right), \quad \text{if } p \geq N^2, s \leq N,$$

which is a non-asymptotic version of the corresponding result in [18] providing a refinement of (5) in the zone $p \geq N^2$ in terms of conditions on p, N, s .

Finally, [6] considers the setting with *known* σ and establishes the following non-asymptotic lower bound valid for all configurations of p, N, s :

$$\rho_* \geq c \min \left(\sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}, \frac{p^{1/4}}{\sqrt{N}}, N^{-1/4} \right) \quad (6)$$

where $c > 0$ is a constant. Furthermore, [6] proves that this bound is attained in the low-dimensional regime $p \leq \gamma N$ if \mathbb{X} is a matrix with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is standard normal. Here $\gamma \in (0, 1)$ is an absolute constant specified in [6]. Noticing that for $p \leq N$ the expression in (9) reduces to its first term, the results of [6] lead to the conclusion that, for *known* σ , the non-asymptotic minimax separation rate satisfies

$$\rho_* \asymp \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}} \quad \text{if } p \leq \gamma N$$

when \mathbb{X} is a matrix with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is standard normal. In what follows, the regime $p \leq \gamma N$ where $0 < \gamma < 1$ will be called the *low-dimensional regime*.

Estimation of $\|\boldsymbol{\theta}\|_2$ and $\|\boldsymbol{\theta}\|_2^2$. The problem of estimating $\|\boldsymbol{\theta}\|_2$ is closely related to signal detection under ℓ_2 separation. Indeed, any estimator of $\|\boldsymbol{\theta}\|_2$ can be used as a test statistic and thus it is easy to check that the minimax separation rate for tests is not of greater order than the minimax rate of estimating $\|\boldsymbol{\theta}\|_2$. In other words, the problem of estimating $\|\boldsymbol{\theta}\|_2$ is more difficult than of signal detection. Intuitively, it is clear since in signal detection the objective is only to decide whether $\|\boldsymbol{\theta}\|_2$ is large or zero.

Estimation of $\|\boldsymbol{\theta}\|_2$ and $\|\boldsymbol{\theta}\|_2^2$ is well understood in various settings of the Gaussian mean model - see, for example, [14, 5, 11, 8, 10] and the references therein. In particular, non-asymptotic minimax rates of estimating $\|\boldsymbol{\theta}\|_2$ and $\|\boldsymbol{\theta}\|_2^2$ in the Gaussian mean model on the sparsity classes $B_0(s)$ are established in [8] for the case of known σ , while [10] derives such rates for estimating $\|\boldsymbol{\theta}\|_2$ in the case of unknown σ . It turns out that the rates are quite different in the two cases. Moreover, [10] obtains the minimax rates of estimating $\|\boldsymbol{\theta}\|_2$ when the distribution of $\boldsymbol{\xi}$ belongs to a class of laws with polynomially or exponentially decreasing tails, which are also quite different from those for the Gaussian noise.

In the linear regression model, much less is known and only partial results are available. One may immediately notice the following. As there exist estimators $\hat{\boldsymbol{\theta}}$ achieving the rate $\sigma \sqrt{s \log(ep/s)/N}$ in the ℓ_2 -norm whenever $s \log(ep/s)/N \leq \kappa_0$ where $\kappa_0 > 0$ is a small enough constant (see, e.g., [22, 4, 12]), a direct plug-in estimator $\|\hat{\boldsymbol{\theta}}\|_2$ of $\|\boldsymbol{\theta}\|_2$ achieves the same rate. This gives the following upper bound on the minimax rate ϕ_* of estimation of $\|\boldsymbol{\theta}\|_2$:

$$\phi_* \leq c\sigma \sqrt{s \log(ep/s)/N}$$

for a constant $c > 0$. This bound is valid with known or unknown σ provided \mathbb{X} has independent subGaussian rows. However, this bound is not tight. Indeed, it is shown in [6] that, for the case of *known* σ , the non-asymptotic minimax rate satisfies

$$\phi_* \asymp \sigma \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}} \quad \text{if } p \leq \gamma N \text{ (low-dimensional regime),} \quad (7)$$

when \mathbb{X} is a matrix with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is standard normal. Note that the result (7) slightly improves upon the plug-in rate $\sqrt{s \log(ep/s)/N}$ in the *sparse zone* $s \leq \sqrt{p}$ and substantially improves upon this rate in the *dense zone* $s > \sqrt{p}$. Indeed, it is not hard to check that

$$\sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}} \asymp \frac{p^{1/4}}{\sqrt{N}} \quad \text{if } s > \sqrt{p}. \quad (8)$$

Next, under the same conditions (low-dimensional regime $p \leq \gamma N$ with known σ), [6] establishes the following non-asymptotic upper bound on the rate of estimation of $\|\boldsymbol{\theta}\|_2^2$:

$$q_{\boldsymbol{\theta}} \leq c \left(\frac{\sigma^2 s \log(1 + \sqrt{p}/s)}{N} + \frac{\sigma \|\boldsymbol{\theta}\|_2}{N} \right) \quad (9)$$

for a constant $c > 0$ depending only δ . However, [6] does not explore whether this bound is optimal.

The recent paper [15] considers estimation of $\|\boldsymbol{\theta}\|_2^2$ in the *highly sparse zone*: $s < p^a$ for some $0 < a < 1/2$ when σ is unknown and when the additional condition $\sqrt{s \log(p)/N} = O(\|\boldsymbol{\theta}\|_2)$ holds. Under the assumptions that matrix \mathbb{X} has independent Gaussian rows with zero mean and covariance matrix Σ and that $s \log(p)/N \leq c_0$ where $c_0 > 0$ is a small enough constant, [15] proves the following non-asymptotic upper bound

$$q_{\boldsymbol{\theta}} \leq c(\sigma) \left(\max(\|\boldsymbol{\theta}\|_2, 1) \frac{s \log(p)}{N} + \frac{\|\boldsymbol{\theta}\|_2}{N} \right) \quad \text{if } s < p^a, \ 0 < a < 1/2, \quad (10)$$

for a constant $c(\sigma) > 0$ depending only δ and σ . Note that, disregarding the dependence on σ and on the constants, the bound (9) is better than (10).

In the present paper, we derive non-asymptotic minimax rates for both functionals, $\|\boldsymbol{\theta}\|_2$ and $\|\boldsymbol{\theta}\|_2^2$. This is done both in the low-dimensional and high-dimensional ($p > \gamma N$) regimes and both in the dense and sparse zones. We focus on the case of unknown σ .

One of our conclusions is that the bound (10) is not rate optimal *when Σ is the identity matrix* and in this case the minimax rate is driven by (9). On the other hand, [15] establishes that (10) is rate optimal when the minimax setting is considered over a sufficiently large class of matrices Σ , and the additional condition $\sqrt{s \log(p)/N} = O(\|\boldsymbol{\theta}\|_2)$ holds. In other words, the worsening of the rate observed in (10) should be explained by the complexity of matrix Σ and not by the complexity of $\boldsymbol{\theta}$ or the fact that σ is unknown.

A related work [26] deals with estimation of yet another quantity $\|\Sigma^{1/2} \boldsymbol{\theta}\|_2^2 / \sigma^2$ (called the signal-to-noise ratio) and establishes the minimax rates for this problem on the sparsity class $B_0(s)$ that are quite different from the minimax rates for estimation of $\|\boldsymbol{\theta}\|_2^2$. In particular, unlike (9) or (10) those rates show no dependence on $\|\boldsymbol{\theta}\|_2^2$.

3 Summary of the results

The main results of the present paper can be summarized, in a simplified form, as follows.

- **Signal detection.** We propose a test Δ such that under suitable conditions on \mathbb{X} and $\boldsymbol{\xi}$ the following holds. Let $\delta \in (0, 1)$. If $s \log(ep/s)/N \leq \kappa_0$ where $\kappa_0 > 0$ is a small enough constant depending on δ , then $R(\Delta, s, \rho) \leq \delta$ whenever

$$\rho \geq c_\delta \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}$$

where $c_\delta > 0$ is a constant depending only on δ . Moreover, if \mathbb{X} is a matrix with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is standard normal, we show that, under the same conditions, the non-asymptotic minimax separation rate in the setting with *unknown* σ has the form

$$\rho_* \asymp \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}. \quad (11)$$

Remark that for $s = p$, the condition $s \log(p/s)/N \leq \kappa_0$ reduces to $p \leq \kappa_0 N$, so that only the low-dimensional case is covered. Moreover, since the function $s \mapsto s \log(ep/s)/N$ is increasing for $1 \leq s \leq p$ the condition $s \log(p/s)/N \leq \kappa_0$ in the dense zone $s > \sqrt{p}$ implies that $p \leq cN^2/(\log p)^2$ with some constant $c > 0$ that does not depend on p, s, N . Let us emphasize that this restriction $p \leq cN^2/(\log p)^2$ is rather natural since for $p \geq N^2$ we get into the divergence regime described in [25]. Indeed, due to (8), for $p \geq N^2$ the value ρ_* in (11) becomes greater than a constant independent of p, s, N in the dense zone $s > \sqrt{p}$. As follows from Theorem 6 below, the minimax risk under unknown σ is also greater than a constant for $p \geq N^2$ and $s > \sqrt{p}$. Therefore, considering $p \geq N^2$ is not of much statistical interest since in this regime increasing the sample size does not improve the quality of testing. The same remark applies to the problems of estimation of $\|\boldsymbol{\theta}\|_2$ and of $\|\boldsymbol{\theta}\|_2^2$.

- **Estimation of $\|\boldsymbol{\theta}\|_2$.** We propose an estimator $\hat{\Lambda}$ such that, under the same conditions as in the previous item, the bound (2) is satisfied with

$$\phi_* \geq c_\delta \sigma \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}$$

where $c_\delta > 0$ is a constant depending only on δ . Moreover, if \mathbb{X} is a matrix with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is standard normal, we show that, under the same conditions, the non-asymptotic minimax estimation rate on the class $B_0(s)$ in the setting with *unknown* σ has the form

$$\phi_* \asymp \sigma \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}.$$

- **Estimation of $\|\boldsymbol{\theta}\|_2^2$.** We propose an estimator \hat{Q} such that, under the same conditions as in the previous two items, the bound (3) is satisfied with q_θ as in (9) where $c > 0$ is a constant depending only on δ . Furthermore, if \mathbb{X} is a matrix with i.i.d. standard

normal entries and $\boldsymbol{\xi}$ is standard normal, we show that, under the same conditions, the non-asymptotic minimax estimation rate on the class $B_0(s) \cap \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_2 \leq \kappa\}$ in the setting with *unknown* σ has the form

$$q_{s,\kappa}^* \asymp \min \left(\frac{\sigma^2 s \log(1 + \sqrt{p}/s)}{N} + \frac{\sigma \kappa}{\sqrt{N}}, \kappa^2 \right).$$

4 Assumptions and notation

If \mathbf{u}, \mathbf{v} are two vectors, we denote by $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v}$ their inner product. If \mathbb{A} is a matrix, \mathbb{A}^T denotes its transpose, $\|\mathbb{A}\|_\infty$ its spectral norm and $\|\mathbb{A}\|_F$ its Frobenius norm, \mathbf{A}_j denotes the j th column of \mathbb{A} , and \mathbf{A}^j its j th row. Throughout the paper, we use blackboard bold symbols (e.g., \mathbb{A}) to denote matrices, bold symbols (e.g., \mathbf{A}) to denote vectors, and light symbols (e.g., A) to denote scalars. In particular, A_j 's are the entries of vector \mathbf{A} and A_{ij} 's are the entries of matrix \mathbb{A} . Given a symmetric matrix \mathbb{A} , $\lambda_{\min}(\mathbb{A})$ and $\lambda_{\max}(\mathbb{A})$ stand for its largest and smallest eigenvalue, respectively. We denote by \mathbb{I}_p the identity matrix in dimension p . We denote by C, C' positive constants whose values can differ on different occurrences.

We will assume depending on the considered method that we have either $N = 2n$, or $N = 3n$ for an integer n , and the sample (\mathbb{X}, \mathbf{Y}) is divided into either two or three subsamples of size n that we denote $(\mathbb{X}_1, \mathbf{Y}_1)$, $(\mathbb{X}_2, \mathbf{Y}_2)$, and $(\mathbb{X}_3, \mathbf{Y}_3)$. In each part of the paper, we specify how large N should be (*i.e.*, $2n$ or $3n$). Since we are interested only in the rates of convergence, taking N in this form is without loss of generality apart from imposing the conditions $N \geq 2$ or $N \geq 3$, respectively.

Assumptions. We assume throughout the paper that $p \geq 2$ and the entries $X_{ij}, i = 1, \dots, N, j = 1, \dots, p$, and $\xi_i, i = 1, \dots, N$, of matrix \mathbb{X} and of vector $\boldsymbol{\xi}$ constitute a collection of $N(p+1)$ jointly independent zero mean random variables with variance 1. In addition, in different parts of the paper we will refer to some conditions from the following list.

Condition on p, s, N .

(P) It holds that $s \log(ep/s)/N \leq \kappa_0$ for some small enough positive constant κ_0 .

Conditions on $\boldsymbol{\xi}$.

($\boldsymbol{\xi}^*1$) $\mathbf{E}(\xi_i^8) \leq c_8$ with $c_8 > 0$ for $i = 1, \dots, N$.

($\boldsymbol{\xi}^*2$) Random variables $\xi_i, i = 1, \dots, N$, are L -subGaussian for some constant $L > 0$, that is $\mathbf{E}(\exp(t\xi_i)) \leq \exp(L^2 t^2/2)$ for all $t \in \mathbb{R}$.

Conditions on \mathbb{X} .

(\mathbb{X}^*1) Random variables X_{ij} have a density with respect to the Lebesgue measure bounded by a constant $K > 0$ for $i = 1, \dots, N, j = 1, \dots, p$.

(\mathbb{X}^*2) $\mathbf{E}(X_{ij}^4) \leq d_4$ with $d_4 > 0$ for $i = 1, \dots, N, j = 1, \dots, p$.

(\mathbb{X}^*3) Random variables $X_{ij}, i = 1, \dots, N, j = 1, \dots, p$, are M -subGaussian for some constant $M > 0$.

5 Generic estimators

In this section, we present generic estimators that will be used throughout the paper. First, we consider some preliminary estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$ based on the subsample $(\mathbb{X}_1, \mathbf{Y}_1)$. Then, using the second subsample $(\mathbb{X}_2, \mathbf{Y}_2)$ we define the following estimators of the values θ_j^2 :

$$a_j(\hat{\boldsymbol{\theta}}) := \hat{\theta}_j^2 + \frac{2\hat{\theta}_j}{n} \mathbf{X}_{2,j}^T (\mathbf{Y}_2 - \mathbb{X}_2 \hat{\boldsymbol{\theta}}) + \frac{1}{n(n-1)} \sum_{k \neq l} X_{2,kj} X_{2,lj} (Y_{2,k} - \mathbf{X}_{2,k}^T \hat{\boldsymbol{\theta}}) (Y_{2,l} - \mathbf{X}_{2,l}^T \hat{\boldsymbol{\theta}}), \quad 1 \leq j \leq p.$$

Here, $X_{2,j}$ and $X_{2,kj}$ are the j th column and the (k, j) th entry of matrix \mathbb{X}_2 , respectively, and $Y_{2,k}$ is the k th entry of vector \mathbf{Y}_2 . The estimator $a_j(\hat{\boldsymbol{\theta}})$ can be viewed as a centered version of $\hat{\theta}_j^2$ where $\hat{\theta}_j$ is the j th component of the conditionally unbiased estimator

$$\begin{aligned} \tilde{\boldsymbol{\theta}} &:= \hat{\boldsymbol{\theta}} + \frac{1}{n} \mathbb{X}_2^T (\mathbf{Y}_2 - \mathbb{X}_2 \hat{\boldsymbol{\theta}}) \\ &= \boldsymbol{\theta} + \left[\frac{1}{n} \mathbb{X}_2^T \mathbb{X}_2 - \mathbb{I}_p \right] (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \frac{\sigma}{n} \mathbb{X}_2^T \boldsymbol{\xi}_2. \end{aligned}$$

Note that $\tilde{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$ if we consider $\hat{\boldsymbol{\theta}}$ as fixed.

5.1 Generic estimator for the dense case ($s > \sqrt{p}$)

Given an estimator $\hat{\boldsymbol{\theta}}$ based on the first subsample $(\mathbb{X}_1, \mathbf{Y}_1)$, we define an estimator of $Q(\boldsymbol{\theta})$ as

$$\hat{Q}_D(\hat{\boldsymbol{\theta}}) := \sum_{j=1}^p a_j(\hat{\boldsymbol{\theta}}). \quad (12)$$

The following general result allows one to link the risk of $\hat{Q}_D(\hat{\boldsymbol{\theta}})$ to the risk of the preliminary estimator $\hat{\boldsymbol{\theta}}$.

Theorem 1. *Assume that Condition (\mathbb{X}^*2) holds. There exists a constant $C > 0$ depending only on d_4 such that, for all $\boldsymbol{\theta} \in \mathbb{R}^p$,*

$$\mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[(\hat{Q}_D(\hat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}))^2 \mid \hat{\boldsymbol{\theta}} \right] \leq C \left[\frac{\|\boldsymbol{\theta}\|_2^2}{n} (\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2 + \sigma^2) + \frac{p}{n^2} (\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^4 + \sigma^4) \right].$$

In what follows, we will use $\hat{Q}_D(\hat{\boldsymbol{\theta}})$ with two different preliminary estimators $\hat{\boldsymbol{\theta}}$ depending on whether we are in the low-dimensional or high-dimensional regime.

5.2 Generic estimator for the sparse case ($s \leq \sqrt{p}$)

In the sparse case, along with the first preliminary estimator $\hat{\boldsymbol{\theta}}$ we consider a second one that we denote by $\bar{\boldsymbol{\theta}}$. Given a matrix \mathbb{M} of dimension $p \times p$, two preliminary estimators $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}$, an estimator $\hat{\sigma}$ of σ , and a constant α , we define an estimator of $Q(\boldsymbol{\theta})$ as

$$\hat{Q}_S(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}, \hat{\sigma}, \mathbb{M}, \alpha) := \sum_{j=1}^p a_j(\hat{\boldsymbol{\theta}}) \mathbb{1} \left\{ |\bar{\theta}_j| > \alpha \hat{\sigma} \sqrt{M_{jj} \log(1 + p/s^2)} \right\}. \quad (13)$$

The risk of this estimator depends on the distribution of $\bar{\boldsymbol{\theta}}$. Thus, it is difficult to obtain a result similar to Theorem 1, that is to bound the risk of \hat{Q}_S by a function of $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2$ for any estimator $\hat{\boldsymbol{\theta}}$. Nevertheless, the risk of \hat{Q}_S for specific choices of estimators can be readily controlled. It turns out that if we find estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$ such that $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2$ and $|\hat{\sigma} - \sigma|$ are of the order σ with high probability, and an estimator $\bar{\boldsymbol{\theta}}$ satisfying $\bar{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2 \mathbb{I}_p/n)$, then with high probability

$$(\hat{Q}_S(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}, \hat{\sigma}, \mathbb{I}_p, \alpha) - Q(\boldsymbol{\theta}))^2 \leq C \left[\sigma^2 \frac{\|\boldsymbol{\theta}\|_2^2}{n} + \sigma^4 \frac{s^2 \log^2(1 + \sqrt{p}/s)}{n^2} \right],$$

which is the desired bound. In the following, we construct estimators $\hat{\boldsymbol{\theta}}, \hat{\sigma}$ with such a property and an estimator $\bar{\boldsymbol{\theta}}$ distributed approximately as $\mathcal{N}(\boldsymbol{\theta}, \sigma^2 \mathbb{I}_p/n)$.

6 Upper bounds for the low-dimensional regime: $p \leq \gamma n$ with $0 < \gamma < 1$

In this section, we assume that $N = 2n$ and we divide the sample in two subsamples of size n , namely $(\mathbb{X}_1, \mathbf{Y}_1)$ and $(\mathbb{X}_2, \mathbf{Y}_2)$. We also assume that there exists a constant $\gamma \in (0, 1)$ such that $p \leq \gamma n$ (the low-dimensional regime).

In the low-dimensional regime, we take as a preliminary estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{OLS}$ where $\hat{\boldsymbol{\theta}}_{OLS}$ is the least squares estimator based on the first subsample $(\mathbb{X}_1, \mathbf{Y}_1)$. If Condition (\mathbb{X}^*1) holds, then $\mathbb{X}_1^T \mathbb{X}_1$ is almost surely invertible, so that the least squares estimator can be almost surely written in the form

$$\hat{\boldsymbol{\theta}}_{OLS} = (\mathbb{X}_1^T \mathbb{X}_1)^{-1} \mathbb{X}_1^T \mathbf{Y}_1. \quad (14)$$

We also consider the following standard estimator of σ based on $\hat{\boldsymbol{\theta}}_{OLS}$:

$$\hat{\sigma}_{OLS} = \frac{\|\mathbf{Y}_1 - \mathbb{X}_1 \hat{\boldsymbol{\theta}}_{OLS}\|_2}{\sqrt{n - p}}. \quad (15)$$

6.1 Estimation of $Q(\boldsymbol{\theta})$ and $\|\boldsymbol{\theta}\|_2$ in the dense case: $s > \sqrt{p}$

In the dense case $s > \sqrt{p}$ of the low-dimensional regime, we use the estimators

$$\hat{Q}_D^{LD} := \hat{Q}_D(\hat{\boldsymbol{\theta}}_{OLS}), \quad \hat{\Lambda}_D^{LD} := \left| \hat{Q}_D^{LD} \right|^{1/2},$$

where Q_D and $\hat{\boldsymbol{\theta}}_{OLS}$ are defined in (12) and (14) respectively. Using Theorem 1 we obtain the following result.

Theorem 2. *Let $1 \leq s \leq p$ and let Conditions (ξ^*1) , (\mathbb{X}^*1) , (\mathbb{X}^*2) hold. Then there exist constants $C > 0$, $\gamma \in (0, 1)$ depending only on K, c_8, d_4 such that, for $p \leq \min(\gamma n, n - 14)$,*

$$\forall \boldsymbol{\theta} \in \mathbb{R}^p, \forall \sigma > 0, \quad \mathbf{E}_{\boldsymbol{\theta}, \sigma} [(\hat{Q}_D^{LD} - Q(\boldsymbol{\theta}))^2] \leq C \left(\sigma^4 \frac{p}{n^2} + \sigma^2 \frac{\|\boldsymbol{\theta}\|_2^2}{n} \right),$$

and

$$\forall \boldsymbol{\theta} \in \mathbb{R}^p, \forall \sigma > 0, \quad \mathbf{E}_{\boldsymbol{\theta}, \sigma} [(\hat{\Lambda}_D^{LD} - \|\boldsymbol{\theta}\|_2)^2] \leq C \sigma^2 \frac{\sqrt{p}}{n}.$$

It follows from the Markov inequality that, under the assumptions of Theorem 2, for any $\delta \in (0, 1)$ the estimators $\hat{Q} = \hat{Q}_D^{LD}$ and $\hat{\Lambda} = \hat{\Lambda}_D^{LD}$ satisfy

$$\begin{aligned} \forall \boldsymbol{\theta} \in \mathbb{R}^p, \forall \sigma > 0, \quad \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left[(\hat{\Lambda}_D^{LD} - \|\boldsymbol{\theta}\|_2)^2 \geq \frac{C\sigma^2 \sqrt{p}}{\delta n} \right] &\leq \delta \\ \forall \boldsymbol{\theta} \in \mathbb{R}^p, \forall \sigma > 0, \quad \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left[(\hat{Q}_D^{LD} - Q(\boldsymbol{\theta}))^2 \geq \frac{C}{\delta} \left(\sigma^4 \frac{p}{n^2} + \sigma^2 \frac{\|\boldsymbol{\theta}\|_2^2}{n} \right) \right] &\leq \delta. \end{aligned}$$

Note also that the zero estimator $\hat{Q} \equiv 0$ satisfies (3) with $q_{\boldsymbol{\theta}}(p, N, \sigma, \delta) \leq \|\boldsymbol{\theta}\|_2^2$. Thus, we have the following corollary.

Corollary 1. *Let the assumptions of Theorem 2 hold. Then, for any $\delta \in (0, 1)$ the estimators $\hat{\Lambda} = \hat{\Lambda}_D^{LD}$ and $\hat{Q} = \hat{Q}_D^{LD}$ satisfy (2) and (3), respectively, with*

$$\begin{aligned} \phi_*(p, N, s, \sigma, \delta) &\leq \frac{C}{\sqrt{\delta}} \sigma \frac{p^{1/4}}{\sqrt{N}}, \\ q_{\boldsymbol{\theta}}(p, N, \sigma, \delta) &\leq \frac{C}{\sqrt{\delta}} \left(\sigma^2 \frac{\sqrt{p}}{N} + \sigma \frac{\|\boldsymbol{\theta}\|_2}{\sqrt{N}} \right). \end{aligned}$$

Furthermore, for all $1 \leq s \leq p$, $\boldsymbol{\theta} \in B_0(s)$, and $\kappa > 0$,

$$\begin{aligned} q_{\boldsymbol{\theta}}^*(p, N, \sigma, \delta) &\leq \min \left(\frac{C}{\sqrt{\delta}} \left(\sigma^2 \frac{\sqrt{p}}{N} + \sigma \frac{\|\boldsymbol{\theta}\|_2}{\sqrt{N}} \right), \|\boldsymbol{\theta}\|_2^2 \right), \\ q_{s, \kappa}^*(p, N, \sigma, \delta) &\leq \min \left(\frac{C}{\sqrt{\delta}} \left(\sigma^2 \frac{\sqrt{p}}{N} + \sigma \frac{\kappa}{\sqrt{N}} \right), \kappa^2 \right). \end{aligned}$$

Note that Corollary 1 and Theorem 2 are non-asymptotic in all parameters and valid for all $1 \leq s \leq p$. However, their upper bounds are minimax optimal only in the dense case $s > \sqrt{p}$. The matching lower bounds are given in Theorems 6 and 7, respectively. Taking them into account we obtain that, under the assumptions of Theorem 2, the estimators $\hat{\Lambda}_D^{LD}$ and \hat{Q}_D^{LD} are minimax optimal for $s > \sqrt{p}$.

6.2 Estimation of $Q(\boldsymbol{\theta})$ and $\|\boldsymbol{\theta}\|_2$ in the sparse case: $s \leq \sqrt{p}$

In the sparse case $s \leq \sqrt{p}$ of the low-dimensional regime, we use the estimators

$$\hat{Q}_S^{LD} := \hat{Q}_S(\hat{\boldsymbol{\theta}}_{OLS}, \hat{\boldsymbol{\theta}}_{OLS}, \hat{\sigma}_{OLS}, (\mathbb{X}_1^T \mathbb{X}_1)^{-1}, \alpha), \quad \hat{\Lambda}_S^{LD} := \left| \hat{Q}_S^{LD} \right|^{1/2}, \quad (16)$$

where \hat{Q}_S , $\hat{\boldsymbol{\theta}}_{OLS}$ and $\hat{\sigma}_{OLS}$ are defined in (13), (14) and (15), and $\alpha > 0$ is a large enough constant. The following theorem holds.

Theorem 3. *Let $s \leq \sqrt{p}$ and let Conditions $(\boldsymbol{\xi}^*2)$, (\mathbb{X}^*1) and (\mathbb{X}^*2) be satisfied. Then there exists a tuning parameter $\alpha > 0$ depending only on K, L and constants $C > 0$, $\gamma \in (0, 1)$ depending only on K, L such that, for $p \leq \min(\gamma n, n - 14)$,*

$$\forall \boldsymbol{\theta} \in B_0(s), \forall \sigma > 0, \quad \mathbf{E}_{\boldsymbol{\theta}, \sigma} [(\hat{Q}_S^{LD} - Q(\boldsymbol{\theta}))^2] \leq C \left(\sigma^4 \frac{s^2 \log^2(1 + \sqrt{p}/s)}{n^2} + \sigma^2 \frac{\|\boldsymbol{\theta}\|_2^2}{n} \right) \quad (17)$$

and

$$\forall \boldsymbol{\theta} \in B_0(s), \forall \sigma > 0, \quad \mathbf{E}_{\boldsymbol{\theta}, \sigma} [(\hat{\Lambda}_S^{LD} - \|\boldsymbol{\theta}\|_2)^2] \leq C \sigma^2 \frac{s \log(1 + \sqrt{p}/s)}{n}. \quad (18)$$

Theorem 3 immediately implies the following corollary analogous to Corollary 1.

Corollary 2. *Let the assumptions of Theorem 3 hold. Then, for any $\delta \in (0, 1)$ the estimators $\hat{\Lambda} = \hat{\Lambda}_S^{LD}$ and $\hat{Q} = \hat{Q}_S^{LD}$ satisfy (2) and (3), respectively, with*

$$\begin{aligned}\phi_*(p, N, s, \sigma, \delta) &\leq \frac{C}{\sqrt{\delta}} \sigma \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}, \\ q_{\theta}(p, N, \sigma, \delta) &\leq \frac{C}{\sqrt{\delta}} \left(\sigma^2 \frac{s \log(1 + \sqrt{p}/s)}{N} + \sigma \frac{\|\theta\|_2}{\sqrt{N}} \right).\end{aligned}$$

Furthermore, for all $s \leq \sqrt{p}$, $\theta \in B_0(s)$, and $\kappa > 0$,

$$\begin{aligned}q_{\theta}^*(p, N, \sigma, \delta) &\leq \min \left(\frac{C}{\sqrt{\delta}} \left(\sigma^2 \frac{s \log(1 + \sqrt{p}/s)}{N} + \sigma \frac{\|\theta\|_2}{\sqrt{N}} \right), \|\theta\|_2^2 \right), \\ q_{s, \kappa}^*(p, N, \sigma, \delta) &\leq \min \left(\frac{C}{\sqrt{\delta}} \left(\sigma^2 \frac{s \log(1 + \sqrt{p}/s)}{N} + \sigma \frac{\kappa}{\sqrt{N}} \right), \kappa^2 \right).\end{aligned}$$

Corollary 2 combined with Theorems 6 and 7 below implies that, under the assumptions of Theorem 3, the estimators \hat{Q}_S^{LD} and $\hat{\Lambda}_S^{LD}$ are minimax optimal when $s \leq \sqrt{p}$ in the sense defined in the Introduction.

6.3 Signal detection

Define the testing procedure

$$\Delta^{LD} = \mathbb{1} \left\{ \hat{\Lambda}^{LD} \geq \beta \hat{\sigma}_{\text{srs}} \sigma \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}} \right\},$$

where $\beta > 0$ is a constant, $\hat{\Lambda}^{LD} = \hat{\Lambda}_S^{LD}$ if $s \leq \sqrt{p}$, and $\hat{\Lambda}^{LD} = \hat{\Lambda}_D^{LD}$ if $s > \sqrt{p}$. Theorems 2, 3 and Lemma 2 in the appendix imply the following corollary.

Corollary 3. *Let Conditions (ξ^*2) , (\mathbb{X}^*1) and (\mathbb{X}^*2) hold and $\delta \in (0, 1)$. Then there exist positive constants $\alpha, \beta, \gamma \in (0, 1)$ depending only on K, d_4, L and a constant $C_{\delta} > 0$ depending only on δ, K, d_4, L such that, for any $p \leq \min(\gamma n, n - 14)$, and any*

$$\rho \geq C_{\delta} \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{n}},$$

we have

$$R(\Delta^{LD}, p, N, s, \rho) \leq \delta,$$

where $R(\cdot, p, N, s, \rho)$ is defined in (4).

The proof of this corollary is straightforward (see, for example, the argument leading to Theorem 3 in [6]). It follows that, for $C'_{\delta} > 0$ depending only on δ, K, d_4, L ,

$$\rho_*(p, N, s, \delta) \leq C'_{\delta} \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}.$$

Moreover, Corollary 3 and Theorem 6 below imply that, under the assumptions of Corollary 3, the test Δ^{LD} is minimax optimal in the sense defined in the Introduction.

7 Upper bounds for the high-dimensional regime

As the results of Section 6 are based on the preliminary least squares estimator, it is problematic to extend them to the zone $p > n$ where this estimator is not unique and cannot be written in the form (14). In this section, we will use another preliminary estimator and provide analogs of Theorems 2 and 3 in the high-dimensional regime that we define by the following condition: $p \geq \gamma n$, where $\gamma \in (0, 1)$ is the maximal constant, for which Theorems 2 and 3 hold. Noteworthy, the theorems of this section are valid without the condition $p \geq \gamma n$. However, to obtain their corollaries about the minimax optimal rates for all n large enough that we state in this section we need a lower bound on p as a function n . To make a connection to the results of Section 6, we will use the lower bound of the form $p \geq \gamma n$ for some $\gamma > 0$. Nevertheless, these corollaries also hold in a more general situation: $p \geq a_n$ for a given sequence a_n that tends to ∞ with n .

7.1 Preliminary estimators

Instead of the least squares estimator, we now choose the preliminary estimator $\hat{\boldsymbol{\theta}}$ as the Square-Root Slope estimator based on the first subsample $(\mathbb{X}_1, \mathbf{Y}_1)$. This estimator is defined by the relation

$$\hat{\boldsymbol{\theta}}_{\text{srs}} \in \arg \min_{\mathbf{t} \in \mathbb{R}^p} \left\{ \|\mathbf{Y}_1 - \mathbb{X}_1 \mathbf{t}\|_2 + \|\mathbf{t}\|_* \right\}. \quad (19)$$

Here, $\|\cdot\|_*$ denotes the sorted ℓ_1 -norm, that is,

$$\|\mathbf{t}\|_* := \sum_{i=1}^d \lambda_i |t|_{(d-i+1)},$$

where $|t|_{(1)} \leq \dots \leq |t|_{(d)}$ are the order statistics of $|t_1|, \dots, |t_d|$, and

$$\lambda_j = c_{\text{srs},1} \sqrt{\frac{\log(2p/j)}{n}}, \quad j = 1, \dots, p,$$

for some constant $c_{\text{srs},1} > 0$. The next proposition follows by combining Corollary 6.2 in [12] and Theorem 8.3 in [4].

Proposition 1. *Let Conditions $(\boldsymbol{\xi}^*2)$, (\mathbb{X}^*3) and (P) hold. Then there exist positive constants $c_{\text{srs},1}, c_{\text{srs},2}, c_{\text{srs},3}$ depending only on L, M, κ_0 such that*

$$\begin{aligned} \inf_{\boldsymbol{\theta} \in B_0(s)} \inf_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma} & \left(\|\hat{\boldsymbol{\theta}}_{\text{srs}} - \boldsymbol{\theta}\|_2^2 \leq c_{\text{srs},2} \sigma^2 \frac{s}{n} \log(ep/s), \|\mathbb{X}(\hat{\boldsymbol{\theta}}_{\text{srs}} - \boldsymbol{\theta})\|_2^2 \leq c_{\text{srs},2} \sigma^2 s \log(ep/s), \right. \\ & \left. \|\hat{\boldsymbol{\theta}}_{\text{srs}} - \boldsymbol{\theta}\|_* \leq c_{\text{srs},2} \sigma \frac{s}{n} \log(ep/s) \right) \\ & \geq 1 - c_{\text{srs},3} \exp(-(n \wedge s \log(ep/s))/c_{\text{srs},3}). \end{aligned}$$

In the high-dimensional regime $p \geq \gamma n$ that we study in this section, we have $s \log(ep/s) \geq \log(ep) \geq \log(e\gamma n)$, so that

$$\exp(-(n \wedge s \log(ep/s))/c_{\text{srs},3}) \leq C(\gamma n)^{-1/C}$$

for some constant $C > 0$ depending only on L, M, κ_0 .

Next, we will use the following natural estimator of σ based on the estimator $\hat{\boldsymbol{\theta}}_{\text{srs}}$:

$$\hat{\sigma}_{\text{srs}} = \frac{\|\mathbf{Y}_1 - \mathbb{X}_1 \hat{\boldsymbol{\theta}}_{\text{srs}}\|_2}{\sqrt{n}}. \quad (20)$$

The next proposition will be useful.

Proposition 2. *Let Conditions $(\boldsymbol{\xi}^*2)$, (\mathbb{X}^*3) and (P) hold. Then there exist positive constants $c_{\text{srs},1}, C, C'$ depending only on L, M, κ_0 such that, for any $t > 0$,*

$$\begin{aligned} \inf_{\boldsymbol{\theta} \in B_0(s)} \inf_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\left| \frac{\hat{\sigma}_{\text{srs}}^2}{\sigma^2} - 1 \right| \leq C \left(\sqrt{\frac{t}{n}} + \frac{s}{n} \log(ep/s) \right) \right) \\ \geq 1 - C' \exp(-(n \wedge t \wedge s \log(ep/s))/C'). \end{aligned}$$

Setting here $t = \kappa_0^2 n$ where κ_0 is the constant from Condition (P) and using the remark after Proposition 1, we deduce from Proposition 2 that if Conditions $(\boldsymbol{\xi}^*2)$, (\mathbb{X}^*3) and (P) hold and we are in the high-dimensional regime $p \geq \gamma n$ then there exist constants $C, C' > 0$ depending only on L, M, κ_0 such that

$$\inf_{\boldsymbol{\theta} \in B_0(s)} \inf_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(|\hat{\sigma}_{\text{srs}}^2 - \sigma^2| \leq \frac{\sigma^2}{2} \right) \geq 1 - C e^{-\frac{n \wedge s \log(ep/s)}{C}} \geq 1 - C' (\gamma n)^{-1/C'}.$$

7.2 Estimation of $Q(\boldsymbol{\theta})$ and $\|\boldsymbol{\theta}\|_2$ in the dense case: $s > \sqrt{p}$

Here, we assume that $N = 2n$ and we divide the sample (\mathbb{X}, \mathbf{Y}) in two sub-samples $(\mathbb{X}_1, \mathbf{Y}_1)$ and $(\mathbb{X}_2, \mathbf{Y}_2)$, each of size n . In the dense case $s > \sqrt{p}$ of the high-dimensional regime, we use the estimators

$$\hat{Q}_D^{HD} := \hat{Q}_D(\hat{\boldsymbol{\theta}}_{\text{srs}}), \quad \hat{\Lambda}_D^{HD} := |\hat{Q}_D^{HD}|^{1/2},$$

where \hat{Q}_D and $\hat{\boldsymbol{\theta}}_{\text{srs}}$ are defined in (12) and (19). The risks of these estimators admit the following upper bounds.

Theorem 4. *Let Conditions $(\boldsymbol{\xi}^*2)$, (\mathbb{X}^*3) and (P) hold. Then there exists a tuning constant $c_{\text{srs},1}$ in the definition of $\hat{\boldsymbol{\theta}}_{\text{srs}}$ depending only on L, M, κ_0 , and two constants $C, C' > 0$ depending only on L, M, κ_0 such that for any $0 < v \leq n^{1/3}$ we have*

$$\sup_{\boldsymbol{\theta} \in B_0(s)} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\left[\hat{Q}_D^{HD} - Q(\boldsymbol{\theta}) \right]^2 > C v \left(\sigma^4 \frac{p}{n^2} + \sigma^2 \frac{\|\boldsymbol{\theta}\|_2^2}{n} \right) \right) \leq C' \left[e^{-v} + e^{-\frac{n \wedge s \log(ep/s)}{C'}} \right]$$

and

$$\sup_{\boldsymbol{\theta} \in B_0(s)} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\left| \hat{\Lambda}_D^{HD} - \|\boldsymbol{\theta}\|_2 \right| > C \sigma \frac{(pv)^{1/4}}{\sqrt{n}} \right) \leq C' \left[e^{-v} + e^{-\frac{n \wedge s \log(ep/s)}{C'}} \right].$$

The next corollary follows immediately from Theorem 4 and the remark after Proposition 1.

Corollary 4. *Let the assumptions of Theorem 4 hold. Let $p \geq \gamma n$ for some $\gamma > 0$ and $\delta \in (0, 1)$. Then, there exist $C_\delta > 0$ and an integer N_0 , both depending only on $L, M, \kappa_0, \delta, \gamma$,*

such that for $N \geq N_0$ the estimators $\hat{\Lambda} = \hat{\Lambda}_D^{HD}$ and $\hat{Q} = \hat{Q}_D^{HD}$ satisfy (2) and (3), respectively, with

$$\begin{aligned}\phi_*(p, N, s, \sigma, \delta) &\leq C_\delta \sigma \frac{p^{1/4}}{\sqrt{N}}, \\ q_\theta(p, N, \sigma, \delta) &\leq C_\delta \left(\sigma^2 \frac{\sqrt{p}}{N} + \sigma \frac{\|\theta\|_2}{\sqrt{N}} \right).\end{aligned}$$

Furthermore, for all $1 \leq s \leq p$, $\theta \in B_0(s)$, $\kappa > 0$ and $N \geq N_0$,

$$\begin{aligned}q_\theta^*(p, N, \sigma, \delta) &\leq \min \left(C_\delta \left(\sigma^2 \frac{\sqrt{p}}{N} + \sigma \frac{\|\theta\|_2}{\sqrt{N}} \right), \|\theta\|_2^2 \right), \\ q_{s,\kappa}^*(p, N, \sigma, \delta) &\leq \min \left(C_\delta \left(\sigma^2 \frac{\sqrt{p}}{N} + \sigma \frac{\kappa}{\sqrt{N}} \right), \kappa^2 \right).\end{aligned}$$

It follows from Corollary 4 and Theorems 6 and 7 below that, under the assumptions of Theorem 4, the above bounds are minimax optimal when $s > \sqrt{p}$ in the sense defined in the Introduction.

Remark 1. If Condition (P) is not satisfied, that is we do not have $s \log(ep/s) < \kappa_0 N$ for $\kappa_0 > 0$ small enough, there still exists an estimator of $\|\theta\|_2^2$ that is consistent whenever $p = o(N^2)$. Indeed, consider the estimator $\hat{Q}_D(\mathbf{0})$, i.e., the generic estimator of $\|\theta\|_2^2$ defined in (12) applied to the null vector. It follows immediately from Theorem 1 that for any $\delta \in (0, 1)$ the estimator $\hat{Q}_D(\mathbf{0})$ satisfies (3) with

$$q_\theta(p, N, \sigma, \delta) \leq C_\delta \left[\frac{\|\theta\|_2}{\sqrt{N}} (\|\theta\|_2 + \sigma) + \frac{\sqrt{p}}{N} (\sigma^2 + \|\theta\|_2^2) \right],$$

where $C_\delta > 0$ depends only on δ and d_4 .

7.3 Estimation of $Q(\theta)$ and $\|\theta\|_2$ in the sparse case: $s \leq \sqrt{p}$

Here, we assume that $N = 3n$ and we divide the sample (\mathbb{X}, \mathbf{Y}) in three sub-samples $(\mathbb{X}_1, \mathbf{Y}_1)$, $(\mathbb{X}_2, \mathbf{Y}_2)$ and $(\mathbb{X}_3, \mathbf{Y}_3)$, each of size n . In the sparse case $s \leq \sqrt{p}$ of the high-dimensional regime, we use the estimators

$$\hat{Q}_S^{HD} := \hat{Q}_S(\hat{\theta}_{\text{srs}}, \tilde{\theta}_{\text{srs}}, \hat{\sigma}_{\text{srs}}, 2\mathbb{I}_p, \alpha), \quad \hat{\Lambda}_S^{HD} := \left| \hat{Q}_S^{HD} \right|^{1/2},$$

where the estimators $\hat{\theta}_{\text{srs}}$ and $\hat{\sigma}_{\text{srs}}$ are defined in (19) and (20), $\alpha > 0$ is a constant large enough and $\tilde{\theta}_{\text{srs}}$ is a debiased estimator derived from $\hat{\theta}_{\text{srs}}$:

$$\tilde{\theta}_{\text{srs}} := \hat{\theta}_{\text{srs}} + \frac{1}{n} \mathbb{X}_3^T (\mathbf{Y}_3 - \mathbb{X}_3 \hat{\theta}_{\text{srs}}).$$

Note that the triplet $(\hat{\theta}_{\text{srs}}, \tilde{\theta}_{\text{srs}}, \hat{\sigma}_{\text{srs}})$ is independent of $(\mathbb{X}_2, \mathbf{Y}_2)$. The following theorem gives bounds on the rates of convergence of estimators \hat{Q}_S^{HD} and $\hat{\Lambda}_S^{HD}$.

Theorem 5. *Let $s \leq \sqrt{p}$ and let Conditions (ξ^*2) , (\mathbb{X}^*3) and (P) hold. Then there exist a tuning constant $c_{\text{srs},1}$ in the definition of $\hat{\theta}_{\text{srs}}$ depending only on L, M, κ_0 , and two positive constants C, C' depending only on L, M, κ_0 such that*

$$\sup_{\theta \in B_0(s)} \sup_{\sigma > 0} \mathbf{P}_{\theta, \sigma} \left[(\hat{Q}_S^{HD} - Q(\theta))^2 \leq C \left(\sigma^4 \frac{s^2 \log^2(1 + \sqrt{p}/s)}{n^2} + \sigma^2 \frac{\|\theta\|_2^2}{n} \right) \right] \geq 1 - C' e^{-\frac{n \wedge s \log(ep/s)}{C'}},$$

and

$$\sup_{\boldsymbol{\theta} \in B_0(s)} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left[(\hat{\Lambda}_S^{HD} - \|\boldsymbol{\theta}\|_2)^2 \leq C\sigma^2 \frac{s \log(1 + \sqrt{p}/s)}{n} \right] \geq 1 - C' e^{-\frac{n \wedge s \log(ep/s)}{C'}}.$$

The next corollary follows immediately from Theorem 5 and the remark after Proposition 1.

Corollary 5. *Let the assumptions of Theorem 5 hold. Let $p \geq \gamma n$ for some $\gamma > 0$ and $\delta \in (0, 1)$. Then, there exist $C_\delta > 0$ and an integer N_0 , both depending only on $L, M, \kappa_0, \delta, \gamma$, such that for $N \geq N_0$ the estimators $\hat{\Lambda} = \hat{\Lambda}_S^{HD}$ and $\hat{Q} = \hat{Q}_S^{HD}$ satisfy (2) and (3), respectively, with*

$$\begin{aligned} \phi_*(p, N, s, \sigma, \delta) &\leq C_\delta \sigma \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}, \\ q_{\boldsymbol{\theta}}(p, N, \sigma, \delta) &\leq C_\delta \left(\sigma^2 \frac{s \log(1 + \sqrt{p}/s)}{N} + \sigma \frac{\|\boldsymbol{\theta}\|_2}{\sqrt{N}} \right). \end{aligned}$$

Furthermore, for all $s \leq \sqrt{p}$, $\boldsymbol{\theta} \in B_0(s)$, $\kappa > 0$ and $N \geq N_0$,

$$\begin{aligned} q_{\boldsymbol{\theta}}^*(p, N, \sigma, \delta) &\leq \min \left(C_\delta \left(\sigma^2 \frac{s \log(1 + \sqrt{p}/s)}{N} + \sigma \frac{\|\boldsymbol{\theta}\|_2}{\sqrt{N}} \right), \|\boldsymbol{\theta}\|_2^2 \right), \\ q_{s, \kappa}^*(p, N, \sigma, \delta) &\leq \min \left(C_\delta \left(\sigma^2 \frac{s \log(1 + \sqrt{p}/s)}{N} + \sigma \frac{\kappa}{\sqrt{N}} \right), \kappa^2 \right). \end{aligned}$$

It follows from Corollary 5 and Theorems 6 and 7 below that, under the assumptions of Theorem 5, the above bounds are minimax optimal when $s \leq \sqrt{p}$ in the sense defined in the Introduction.

7.4 Signal detection

Define the testing procedure

$$\Delta^{HD} = \mathbf{1} \left\{ \hat{\Lambda}^{HD} \geq \beta \hat{\sigma}_{\text{srs}} \sigma \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}} \right\},$$

where $\beta > 0$ is a constant, $\hat{\Lambda}^{HD} = \hat{\Lambda}_S^{HD}$ if $s \leq \sqrt{p}$, and $\hat{\Lambda}^{HD} = \hat{\Lambda}_D^{HD}$ if $s > \sqrt{p}$. Theorems 4, 5 and Proposition 2 imply the following corollary.

Corollary 6. *Let Conditions (ξ^*2) , (\mathbb{X}^*3) and (P) hold and $\delta \in (0, 1)$. Let $p \geq \gamma n$ for some $\gamma > 0$. Then there exist positive constants $\alpha, \beta, C_\delta, N_0$ depending only on $\delta, L, M, \kappa_0, \gamma$ such that, for any $N \geq N_0$ and any*

$$\rho \geq C_\delta \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}},$$

we have

$$R(\Delta^{HD}, p, N, s, \rho) \leq \delta,$$

where $R(\cdot, p, N, s, \rho)$ is defined in (4).

The proof of this corollary is straightforward (see, for example, the argument leading to Theorem 3 in [6]). It follows that, for $C'_\delta > 0, N_0 > 0$ depending only on $\delta, L, M, \kappa_0, \gamma$ and any $N \geq N_0$,

$$\rho_*(p, N, s, \delta) \leq C'_\delta \sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}.$$

Moreover, Corollary 6 and Theorem 6 below imply that, under the assumptions of Corollary 6, the test Δ^{HD} is minimax optimal in the sense defined in the Introduction.

8 Lower bounds

Theorem 6. *Assume that \mathbb{X} is a matrix with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is an i.i.d. standard normal noise. For any $\delta \in (0, 1)$, there exists $c_\delta > 0$ depending only on δ such that*

$$\inf_{\Delta} R(\Delta, p, N, s, \rho) \geq \delta, \quad (21)$$

for

$$\rho \leq \rho(p, N, s, \delta) := c_\delta \min \left(\sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}, 1 \right), \quad (22)$$

where $R(\bar{\Delta}, p, N, s, \rho)$ is defined in (4), and \inf_{Δ} denotes the infimum over all tests Δ . As a consequence, for all $\delta \in (0, 1)$ we have

$$\rho^*(p, N, s, \delta) \geq c_\delta \min \left(\sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}, 1 \right), \quad (23)$$

and

$$\phi^*(p, N, s, \sigma, \delta) \geq \frac{c_\delta}{2} \sigma \min \left(\sqrt{\frac{s \log(1 + \sqrt{p}/s)}{N}}, 1 \right). \quad (24)$$

Note that Theorem 6 can be obtained by combining [7, Lemma 17] and the proof of [7, Proposition 6] if we take into account that [7, Proposition 6] is stated with the maximum over the covariance matrices of the design while its proof uses only the identity matrix. For completeness, we provide a direct proof of Theorem 6 in Section 12.1 below.

Theorem 7. *Assume that \mathbb{X} is a matrix with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is an i.i.d. standard normal noise. Then for any $\delta \in (0, 1)$, there exists $c_\delta > 0$ that depends only on δ such that*

$$\inf_{\hat{T}} \sup_{\boldsymbol{\theta} \in B_0(s): \|\boldsymbol{\theta}\|_2 \leq \kappa} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(|\hat{T} - Q(\boldsymbol{\theta})| \geq c_\delta \bar{q}_{s, \kappa}(p, N, \sigma) \right) \geq \delta,$$

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

$$\bar{q}_{s, \kappa}(p, N, \sigma) = \min \left(\sigma^2 \min \left(\frac{s \log(1 + \sqrt{p}/s)}{N}, 1 \right) + \frac{\sigma \kappa}{\sqrt{N}}, \kappa^2 \right).$$

This theorem immediately implies the following corollary.

Corollary 7. *Under the assumptions of Theorem 7 we have*

$$q_{\boldsymbol{\theta}}^*(p, N, \sigma, \delta) \geq c_{\delta} \min \left(\sigma^2 \min \left(\frac{\|\boldsymbol{\theta}\|_0 \log(1 + \sqrt{p}/\|\boldsymbol{\theta}\|_0)}{N}, 1 \right) + \frac{\sigma \|\boldsymbol{\theta}\|_2}{\sqrt{N}}, \|\boldsymbol{\theta}\|_2^2 \right),$$

and

$$q_{s, \kappa}^*(p, N, \sigma, \delta) \geq c_{\delta} \min \left(\sigma^2 \min \left(\frac{s \log(1 + \sqrt{p}/s)}{N}, 1 \right) + \frac{\sigma \kappa}{\sqrt{N}}, \kappa^2 \right).$$

9 Proof of Theorem 1

In this proof, we write for brevity (\mathbb{X}, \mathbf{Y}) instead of $(\mathbb{X}_2, \mathbf{Y}_2)$. Using the notation $\mathbf{u} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$, and $\mathbf{w}_j = \sum_{k \neq j} u_k \mathbf{X}_k$ where u_k 's are the components of \mathbf{u} we have

$$\begin{aligned} \hat{Q}_D &= Q(\boldsymbol{\theta}) + \sum_{j=1}^p \left[2\theta_j u_j \left(1 - \frac{\|\mathbf{X}_j\|_2^2}{n} \right) + 2\theta_j \frac{\mathbf{X}_j^T}{n} (\sigma \boldsymbol{\xi} - \mathbf{w}_j) \right] \\ &+ \sum_{j=1}^p \left[u_j^2 + \frac{2}{n} u_j \mathbf{X}_j^T (\sigma \boldsymbol{\xi} - \mathbb{X} \mathbf{u}) + \frac{1}{n(n-1)} \sum_{k \neq l} X_{kj} X_{lj} (\sigma \xi_k - \mathbf{X}^k \mathbf{u}) (\sigma \xi_l - \mathbf{X}^l \mathbf{u}) \right]. \end{aligned} \quad (25)$$

Note that $\mathbf{E}_{\boldsymbol{\theta}, \sigma}(\hat{Q}_D | \hat{\boldsymbol{\theta}}) = Q(\boldsymbol{\theta})$. Therefore, we need to evaluate the conditional variance of \hat{Q}_D given $\hat{\boldsymbol{\theta}}$. We have

$$\begin{aligned} \mathbf{Var}_{\boldsymbol{\theta}, \sigma} \left(\sum_{j=1}^p \left[2\theta_j u_j \left(1 - \frac{\|\mathbf{X}_j\|_2^2}{n} \right) + 2\sigma \theta_j \frac{\mathbf{X}_j^T}{n} \boldsymbol{\xi} \right] \mid \hat{\boldsymbol{\theta}} \right) &\leq \frac{C}{n} \sum_{j=1}^p \left[\theta_j^2 u_j^2 + \sigma^2 \theta_j^2 \right] \\ &\leq \frac{C}{n} \|\boldsymbol{\theta}\|_2^2 (\|\mathbf{u}\|_2^2 + \sigma^2), \end{aligned} \quad (26)$$

and, since $\mathbf{E}[(\mathbf{X}_j^T \mathbf{X}_l)^2] = \mathbf{E} \sum_{m=1}^n (X_{jm} X_{lm})^2 = n$ for $j \neq l$,

$$\begin{aligned} \mathbf{Var}_{\boldsymbol{\theta}, \sigma} \left(\sum_{j=1}^p \theta_j \frac{\mathbf{X}_j^T}{n} \mathbf{w}_j \mid \hat{\boldsymbol{\theta}} \right) &= \sum_{j=1}^p \frac{\theta_j^2}{n^2} \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[\left(\sum_{k \neq j} u_k \mathbf{X}_k^T \mathbf{X}_j \right)^2 \mid \hat{\boldsymbol{\theta}} \right] + \sum_{j \neq l} \frac{\theta_j \theta_l}{n^2} \mathbf{E}_{\boldsymbol{\theta}, \sigma} (\mathbf{X}_j^T \mathbf{w}_j \mathbf{X}_l^T \mathbf{w}_l \mid \hat{\boldsymbol{\theta}}) \\ &= \sum_{j \neq k} \frac{\theta_j^2}{n} u_k^2 + \sum_{j \neq l} \frac{\theta_j \theta_l}{n} u_j u_l \\ &\leq \frac{2}{n} \|\boldsymbol{\theta}\|_2^2 \|\mathbf{u}\|_2^2. \end{aligned} \quad (27)$$

Finally, the double sum in (25) can be written as $U = \frac{2}{n(n-1)} \sum_{i < j} h((\mathbf{X}^i, \xi_i), (\mathbf{X}^j, \xi_j))$ where

$$h((\mathbf{X}^i, \xi_i), (\mathbf{X}^j, \xi_j)) = \sum_{k=1}^p [X_{ik}(\sigma \xi_i - \mathbf{X}^i \mathbf{u}) + u_k] [X_{jk}(\sigma \xi_j - \mathbf{X}^j \mathbf{u}) + u_k].$$

Since, conditionally on $\hat{\boldsymbol{\theta}}$, the terms in U are zero-mean and uncorrelated, the variance of U conditionally on $\hat{\boldsymbol{\theta}}$ equals

$$\mathbf{Var}_{\boldsymbol{\theta}, \sigma}(U \mid \hat{\boldsymbol{\theta}}) = \frac{2}{n(n-1)} \mathbf{Var}_{\boldsymbol{\theta}, \sigma} \left[h((\mathbf{X}^1, \xi_1), (\mathbf{X}^2, \xi_2)) \mid \hat{\boldsymbol{\theta}} \right].$$

Now, for all i, j ,

$$\mathbf{E}_{\theta, \sigma} \left[(X_{ij}(\sigma \xi_i - \mathbf{X}^i \mathbf{u}) + u_j)^2 \mid \hat{\boldsymbol{\theta}} \right] \leq C(\|\mathbf{u}\|_2^2 + \sigma^2) \quad (28)$$

and, for $j \neq k$,

$$\mathbf{E}_{\theta, \sigma} \left[(X_{ij}(\sigma \xi_i - \mathbf{X}^i \mathbf{u}) + u_j)(X_{ik}(\sigma \xi_i - \mathbf{X}^i \mathbf{u}) + u_k) \mid \hat{\boldsymbol{\theta}} \right] = u_j u_k. \quad (29)$$

Hence, the conditional variance of U given $\hat{\boldsymbol{\theta}}$ satisfies

$$\mathbf{Var}_{\theta, \sigma}(U \mid \hat{\boldsymbol{\theta}}) \leq C \frac{p}{n^2} (\|\mathbf{u}\|_2^2 + \sigma^2)^2. \quad (30)$$

The result follows by combining (26), (27) and (30).

10 Proofs for Section 6

10.1 Preliminary lemmas

The following lemma will be useful.

Lemma 1. *Let $p \leq \min(\gamma n, n - 14)$ for some $\gamma \in (0, 1)$ small enough. If Condition (\mathbb{X}^*1) holds then there exists a constant $C > 0$ depending only on K such that*

$$\mathbf{E}(\lambda_{\min}^{-4}(\mathbb{X}_1^T \mathbb{X}_1)) \leq C n^{-4}.$$

*If Conditions $(\boldsymbol{\xi}^*1)$ and (\mathbb{X}^*1) hold then there exists a constant $C > 0$ depending only on K, c_8 such that*

$$\mathbf{E}_{\theta, \sigma} \|\hat{\boldsymbol{\theta}}_{OLS} - \boldsymbol{\theta}\|_2^8 \leq C \sigma^8.$$

Proof. In this proof, we write for brevity (\mathbb{X}, \mathbf{Y}) instead of $(\mathbb{X}_1, \mathbf{Y}_1)$. First, we note that, almost surely,

$$\|\hat{\boldsymbol{\theta}}_{OLS} - \boldsymbol{\theta}\|_2^8 \leq \sigma^8 \lambda_{\max}^4(\mathbb{X}(\mathbb{X}^T \mathbb{X})^{-2} \mathbb{X}^T) \|\boldsymbol{\xi}\|_2^8 = \sigma^8 \lambda_{\min}^{-4}(\mathbb{X}^T \mathbb{X}) \|\boldsymbol{\xi}\|_2^8. \quad (31)$$

Thus, the proof of the lemma will be complete if we show that $\mathbf{E}[\lambda_{\min}^{-4}(\mathbb{X}^T \mathbb{X})] \leq C n^{-4}$, where $C > 0$ is a constant depending only on K . To this end, we use the following proposition.

Proposition 3. *If Condition (\mathbb{X}^*1) holds and $p \leq \gamma n$ for some $\gamma \in (0, 1)$ small enough then, with probability at least $1 - \exp(-Cn)$,*

$$\lambda_{\min}(\mathbb{X}^T \mathbb{X}) > C' n,$$

where C, C' are positive constants depending only on K .

This proposition follows immediately from combining item (1) of Corollary 2.5 in [19] with the remark on page 884 in [19] that gives the explicit form of the small ball condition for a random vector whose coordinates are independent random variables with bounded Lebesgue density.

Using Proposition 3 we get

$$\mathbf{E}[\lambda_{\min}^{-4}(\mathbb{X}^T \mathbb{X})] \leq (C' n)^{-4} + \sqrt{\mathbf{E}[\lambda_{\min}^{-8}(\mathbb{X}^T \mathbb{X})]} e^{-Cn/2}, \quad (32)$$

where C, C' are the constants from Proposition 3.

To complete the proof, we now show that $\mathbf{E}[\lambda_{\min}^{-8}(\mathbb{X}^T \mathbb{X})] \leq Cn^8$, where $C > 0$ is a constant depending only on K . Denote by \mathbb{X}_{-1} the matrix obtained by removing the first column from \mathbb{X} and by $\text{Span}(\mathbb{X}_{-1})$ the span of its columns. Arguing quite analogously to the proof of Lemma 3 in [6] we obtain

$$\mathbf{E}[\lambda_{\min}^{-8}(\mathbb{X}^T \mathbb{X})] \leq n^8 \mathbf{E}[\text{dist}(\mathbf{X}_1, \text{Span}(\mathbb{X}_{-1}))^{-16}],$$

where $\text{dist}(\mathbf{X}_1, \text{Span}(\mathbb{X}_{-1}))$ is the Euclidean distance from \mathbf{X}_1 to $\text{Span}(\mathbb{X}_{-1})$. Note that, under Condition (\mathbb{X}^*1) , the vector space $\text{Span}(\mathbb{X}_{-1})$ has almost surely dimension $p-1$. Recall also that \mathbf{X}_1 is independent from \mathbb{X}_{-1} . Thus,

$$\mathbf{E}[\text{dist}(\mathbf{X}_1, \text{Span}(\mathbb{X}_{-1}))^{-16}] \leq \max_S \mathbf{E}[\text{dist}(\mathbf{X}_1, S)^{-16}] = \max_S \mathbf{E}[\|P_S^\perp \mathbf{X}_1\|_2^{-16}],$$

where the maximum is taken over all $(p-1)$ -dimensional vector subspaces S of \mathbb{R}^n and we denote by P_S^\perp the orthogonal projector on S^\perp . Furthermore, by Theorem 1.1 in [21], under Condition (\mathbb{X}^*1) the density of $P_S^\perp \mathbf{X}_1$ is bounded almost everywhere by C_0^{n-p+1} , where $C_0 = CK$ for an absolute constant $C > 0$. We have

$$\mathbf{E}[\|P_S^\perp \mathbf{X}_1\|_2^{-16}] \leq C_0^{16} + \mathbf{E}[\|P_S^\perp \mathbf{X}_1\|_2^{-16} \mathbf{1}_{\|P_S^\perp \mathbf{X}_1\|_2 < C_0^{-1}}].$$

As $\max_{k \geq 1} 2\pi^{(k+1)/2} / \Gamma((k+1)/2)$ is less than an absolute constant, it holds that, for $n \geq p+14$,

$$\begin{aligned} \mathbf{E}[\|P_S^\perp \mathbf{X}_1\|_2^{-16} \mathbf{1}_{\|P_S^\perp \mathbf{X}_1\|_2 < C_0^{-1}}] &\leq C_0^{n-p+1} \int_{\|z\|_2 \leq C_0^{-1}} \|z\|^{-16} dz \\ &= C_0^{n-p+1} \frac{2\pi^{(n-p+2)/2}}{\Gamma((n-p+2)/2)} \int_0^{C_0^{-1}} r^{n-p-15} dr \leq C \end{aligned}$$

where $C > 0$ is a constant depending only on K . Hence, $\mathbf{E}[\lambda_{\min}^{-8}(\mathbb{X}^T \mathbb{X})] \leq Cn^8$ where $C > 0$ depends only on K . Together with (31) and (32) this proves the lemma. \square

The following lemma summarizes some properties of the estimator

$$\hat{\sigma}_{OLS} := \frac{\|\mathbf{Y}_1 - \mathbb{X}_1 \hat{\boldsymbol{\theta}}_{OLS}\|_2}{\sqrt{n-p}}.$$

Lemma 2. *If Condition (ξ^*2) holds and $p \leq \gamma n$ for some $\gamma \in (0, 1)$, then there exists a constant $C > 0$ depending only on L such that*

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\left| \frac{\hat{\sigma}_{OLS}^2}{\sigma^2} - 1 \right| \leq \frac{1}{2} \right) \geq 1 - 2e^{-(1-\gamma)n/C} \quad \text{and} \quad \mathbf{E}_{\boldsymbol{\theta}, \sigma}(\hat{\sigma}_{OLS}^4) \leq C\sigma^4.$$

Proof. We have

$$\frac{\hat{\sigma}_{OLS}^2}{\sigma^2} = \frac{\|\mathbb{P}\boldsymbol{\xi}\|_2^2}{n-p},$$

where \mathbb{P} is the orthogonal projector on the orthogonal complement of the range of \mathbb{X}_1 , which has rank $n-p$ almost surely and satisfies $\|\mathbb{P}\|_\infty = 1$, $\|\mathbb{P}\|_F^2 = n-p$. Applying the Hanson-Wright inequality for subGaussian variables (cf. Theorem 6.2.1 in [24]) conditionally on \mathbb{X}_1 we obtain that there is an absolute constant $C > 0$ such that, for all $t \geq 0$,

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\left| \frac{\hat{\sigma}_{OLS}^2}{\sigma^2} - 1 \right| > t \right) \leq 2 \exp \left(-C \min \left\{ \frac{(n-p)t^2}{L^4}, \frac{(n-p)t}{L^2} \right\} \right). \quad (33)$$

Hence for $0 \leq u \leq n - p$,

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\left| \frac{\hat{\sigma}_{OLS}^2}{\sigma^2} - 1 \right| > L^2 \sqrt{\frac{u}{n-p}} \right) \leq 2 \exp(-Cu).$$

Taking $u = (2 \max(L^2, 1))^{-1}(n-p) \geq C'(1-\gamma)n$ yields the first inequality of the lemma. The bound on the fourth moment of $\hat{\sigma}_{OLS}$ follows by integrating (33). \square

Assuming that Condition (\mathbb{X}^*1) holds we set

$$\hat{\tau}_j = 2\alpha \hat{\sigma}_{OLS} \sqrt{[(\mathbb{X}_1^T \mathbb{X}_1)^{-1}]_{jj} \log(1 + p/s^2)},$$

where $[(\mathbb{X}_1^T \mathbb{X}_1)^{-1}]_{jj}$ denotes the j th diagonal entry of matrix $(\mathbb{X}_1^T \mathbb{X}_1)^{-1}$ ¹.

Lemma 3. *Let $p \leq \gamma n$ for some $\gamma \in (0, 1)$ and let Conditions (\mathbb{X}^*1) and $(\boldsymbol{\xi}^*2)$ hold. Then there exist $\alpha > 0$ is large enough and a constant $C > 0$ depending only on γ, L such that*

$$\sup_{\boldsymbol{\theta} \in B_0(s)} \sup_{j: \theta_j=0} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\theta}_{OLS,j}| > \hat{\tau}_j) \leq C \frac{s^4}{p^2}$$

and, for all $j \in \{1, \dots, p\}$,

$$\sup_{\boldsymbol{\theta} \in B_0(s)} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\theta}_{OLS,j}| \leq \hat{\tau}_j, |\theta_j| \geq 2\hat{\tau}_j | \mathbb{X}_1) \leq 2 \exp \left(- \frac{\theta_j^2}{C\sigma^2[(\mathbb{X}_1^T \mathbb{X}_1)^{-1}]_{jj}} \right).$$

Proof. For brevity, in this proof we write (\mathbb{X}, \mathbf{Y}) , $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$ instead of $(\mathbb{X}_1, \mathbf{Y}_1)$, $\hat{\boldsymbol{\theta}}_{OLS}$ and $\hat{\sigma}_{OLS}$, respectively. Set

$$\tau_j = \alpha \sigma \sqrt{[(\mathbb{X}^T \mathbb{X})^{-1}]_{jj} \log(1 + p/s^2)}.$$

Using Lemma 2, we have

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\theta}_j| > \hat{\tau}_j) \leq \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\theta}_j| > \tau_j) + 2e^{-C'(1-\gamma)n}.$$

Next, if $\theta_j = 0$, then the fact that $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \sigma(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \boldsymbol{\xi}$ yields

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\theta}_j| > \tau_j | \mathbb{X}) = \mathbf{P}\{ |[(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \boldsymbol{\xi}]_j| > \frac{\tau_j}{\sigma} | \mathbb{X} \}.$$

Note that, conditionally on \mathbb{X} , the random variable $[(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \boldsymbol{\xi}]_j$ is $L\sqrt{[(\mathbb{X}^T \mathbb{X})^{-1}]_{jj}}$ -subGaussian. Thus, if $\alpha > 2L$,

$$\mathbf{P}\{ |[(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \boldsymbol{\xi}]_j| > \frac{\tau_j}{\sigma} | \mathbb{X} \} \leq 2 \exp \left(- \frac{\tau_j^2}{2\sigma^2 L^2 [(\mathbb{X}^T \mathbb{X})^{-1}]_{jj}} \right) \leq C \frac{s^4}{p^2}.$$

Combining the above bounds and taking into account the condition $p \leq \gamma n$ we obtain the first inequality of the lemma. To prove the second inequality, we note that

$$\begin{aligned} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\theta}_j| \leq \hat{\tau}_j, |\theta_j| \geq 2\hat{\tau}_j | \mathbb{X}) &\leq \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\theta}_j - \theta_j| \geq |\theta_j|/2 | \mathbb{X}) \\ &= \mathbf{P}\{ |[(\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \boldsymbol{\xi}]_j| > |\theta_j|/(2\sigma) | \mathbb{X} \} \\ &\leq 2 \exp \left(- \frac{\theta_j^2}{8L^2 \sigma^2 [(\mathbb{X}^T \mathbb{X})^{-1}]_{jj}} \right). \end{aligned}$$

\square

¹Recall that $\mathbb{X}_1^T \mathbb{X}_1$ is almost surely invertible under Condition (\mathbb{X}^*1) , so that operating with $(\mathbb{X}_1^T \mathbb{X}_1)^{-1}$ is formally legitimate everywhere except an event of zero probability, on which we set, for example, $\hat{\tau}_j = 0$. We do not further invoke this detail since it does not influence the argument.

10.2 Proof of Theorem 2

The first inequality of the theorem is straightforward in view of Theorem 1 and Lemma 1. The second inequality (that is, the upper bound on the squared risk of $\hat{\Lambda}_D^{LD}$) is deduced from the first inequality exactly in the same way as it is done in the proof of Theorem 8 in [8].

10.3 Proof of Theorem 3

For brevity, in this subsection we write (\mathbb{X}, \mathbf{Y}) , $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}$ instead of $(\mathbb{X}_2, \mathbf{Y}_2)$, $\hat{\boldsymbol{\theta}}_{OLS}$ and $\hat{\sigma}_{OLS}$, respectively. We also set $\hat{Q} = \hat{Q}_S^{LD}$, where \hat{Q}_S^{LD} is defined in (16). Let $\hat{\tau}_j$ and τ_j be the same as in Lemma 3.

Denoting by \mathcal{S} the support of $\boldsymbol{\theta}$ and by \mathcal{S}^c its complement, we have

$$\mathbf{E}_{\boldsymbol{\theta}, \sigma}(\hat{Q} - Q(\boldsymbol{\theta}))^2 \leq 2\mathbf{E}_{\boldsymbol{\theta}, \sigma} \left(\sum_{j \in \mathcal{S}^c} a_j(\hat{\boldsymbol{\theta}}) \mathbb{1}_{|\hat{\theta}_j| > \hat{\tau}_j} \right)^2 + 2\mathbf{E}_{\boldsymbol{\theta}, \sigma} \left(\sum_{j \in \mathcal{S}} \{a_j(\hat{\boldsymbol{\theta}}) \mathbb{1}_{|\hat{\theta}_j| > \hat{\tau}_j} - \theta_j^2\} \right)^2. \quad (34)$$

Consider the first sum on the right-hand side of (34). Acting as in the proof of Theorem 1 we get

$$\sum_{j \in \mathcal{S}^c} a_j(\hat{\boldsymbol{\theta}}) \mathbb{1}_{|\hat{\theta}_j| > \hat{\tau}_j} = \frac{2}{n(n-1)} \sum_{i < j} \tilde{h}((\mathbf{X}^i, \xi_i), (\mathbf{X}^j, \xi_j)), \quad (35)$$

where

$$\tilde{h}((\mathbf{X}^i, \xi_i), (\mathbf{X}^j, \xi_j)) = \sum_{k \in \mathcal{S}^c} [X_{ik}(\sigma \xi_i - \mathbf{X}^i \mathbf{u}) + u_k] [X_{jk}(\sigma \xi_j - \mathbf{X}^j \mathbf{u}) + u_k] \mathbb{1}_{|\hat{\theta}_k| > \hat{\tau}_k}.$$

Conditionally on $(\mathbb{X}_1, \mathbf{Y}_1)$, the terms in (35) are zero-mean and distinct terms are uncorrelated. Thus,

$$\mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[\left(\sum_{j \in \mathcal{S}^c} a_j(\hat{\boldsymbol{\theta}}) \mathbb{1}_{|\hat{\theta}_j| > \hat{\tau}_j} \right)^2 \mid \mathbb{X}_1, \mathbf{Y}_1 \right] = \frac{2}{n(n-1)} \mathbf{E}_{\boldsymbol{\theta}, \sigma} [\tilde{h}^2((\mathbf{X}^1, \xi_1), (\mathbf{X}^2, \xi_2)) \mid \mathbb{X}_1, \mathbf{Y}_1] \quad (36)$$

and, by an argument analogous to (28) and (29), we find

$$\begin{aligned} & \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[\left(\sum_{j \in \mathcal{S}^c} a_j(\hat{\boldsymbol{\theta}}) \mathbb{1}_{|\hat{\theta}_j| > \hat{\tau}_j} \right)^2 \right] \\ & \leq \frac{C}{n^2} \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[(\|\mathbf{u}\|_2^2 + \sigma^2)^2 \sum_{k \in \mathcal{S}^c} \mathbb{1}_{|\hat{\theta}_k| > \hat{\tau}_k} + \sum_{k, k' \in \mathcal{S}^c: k' \neq k} u_{k'}^2 u_k^2 \mathbb{1}_{|\hat{\theta}_{k'}| > \hat{\tau}_{k'}} \mathbb{1}_{|\hat{\theta}_k| > \hat{\tau}_k} \right]. \end{aligned} \quad (37)$$

Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned} & \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[\left(\sum_{j \in \mathcal{S}^c} a_j(\hat{\boldsymbol{\theta}}) \mathbb{1}_{|\hat{\theta}_j| > \hat{\tau}_j} \right)^2 \right] \\ & \leq \frac{C}{n^2} \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[(\sigma^4 + \|\mathbf{u}\|_2^4) \sum_{k \in \mathcal{S}^c} \mathbb{1}_{|\hat{\theta}_k| > \hat{\tau}_k} \right] + \frac{C}{n^2} \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left(\sum_{\substack{k', k \in \mathcal{S}^c \\ k' \neq k}} u_{k'}^2 u_k^2 \mathbb{1}_{|\hat{\theta}_{k'}| > \hat{\tau}_{k'}} \mathbb{1}_{|\hat{\theta}_k| > \hat{\tau}_k} \right) \\ & \leq \frac{C}{n^2} \left[\sigma^4 + \sqrt{\mathbf{E}_{\boldsymbol{\theta}, \sigma} \|\mathbf{u}\|_2^8} \right] \sum_{k \in \mathcal{S}^c} \sqrt{\mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{\theta}_k| > \hat{\tau}_k)}. \end{aligned} \quad (38)$$

Using here Lemmas 1 and 3 we conclude that

$$\mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\left(\sum_{j \in \mathcal{S}^c} a_j(\hat{\boldsymbol{\theta}}) \mathbf{1}_{|\hat{\theta}_j| > \hat{\tau}_j} \right)^2 \right] \leq C \sigma^4 \frac{s^2}{n^2}. \quad (39)$$

Next, we consider the second sum on the right-hand side of (34). Similar to the proof of Theorem 1 we obtain

$$\mathbf{E}_{\boldsymbol{\theta},\sigma} \left(\sum_{j \in \mathcal{S}} a_j(\hat{\boldsymbol{\theta}}) \mathbf{1}_{|\hat{\theta}_j| > \hat{\tau}_j} - \theta_j^2 \right)^2 \leq C(S_1 + S_2 + S_3), \quad (40)$$

where

$$\begin{aligned} S_1 &= \mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\sum_{j \in \mathcal{S}} \theta_j u_j \mathbf{1}_{|\hat{\theta}_j| > \hat{\tau}_j} \left(1 - \frac{\|\mathbf{X}_j\|_2^2}{n} \right) + \theta_j \frac{\mathbf{X}_j^T}{n} (\sigma \boldsymbol{\xi} - \mathbf{w}_j) \mathbf{1}_{|\hat{\theta}_j| > \hat{\tau}_j} \right]^2 \\ S_2 &= \mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\frac{2}{n(n-1)} \sum_{i < j} \bar{h}((\mathbf{X}^i, \xi_i), (\mathbf{X}^j, \xi_j)) \right]^2, \quad S_3 = \mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\sum_{j \in \mathcal{S}} \theta_j^2 \mathbf{1}_{|\hat{\theta}_j| \leq \hat{\tau}_j} \right]^2. \end{aligned}$$

and \bar{h} is the equivalent of \tilde{h} for indices in \mathcal{S} instead of \mathcal{S}^c . Thus, S_2 does not exceed the expression in (38) where \mathcal{S}^c is replaced by \mathcal{S} :

$$S_2 \leq \frac{C}{n^2} \left[\sigma^4 + \sqrt{\mathbf{E}_{\boldsymbol{\theta},\sigma} \|\mathbf{u}\|_2^8} \right] \sum_{k \in \mathcal{S}} \sqrt{\mathbf{P}_{\boldsymbol{\theta},\sigma}(|\hat{\theta}_k| > \hat{\tau}_k)}.$$

Bounding in this expression $\mathbf{P}_{\boldsymbol{\theta},\sigma}(|\hat{\theta}_k| > \hat{\tau}_k)$ by 1 we get

$$S_2 \leq C \sigma^4 \frac{s}{n^2}. \quad (41)$$

Next, arguing as in (26) and (27) and using Lemma 1 we obtain

$$S_1 \leq \frac{C}{n} \|\boldsymbol{\theta}\|_2^2 \mathbf{E}_{\boldsymbol{\theta},\sigma} (\|\mathbf{u}\|_2^2 + \sigma^2) \leq \frac{C \sigma^2}{n} \|\boldsymbol{\theta}\|_2^2. \quad (42)$$

Finally, introducing the notation $\mathbb{B} = (\mathbb{X}_1^T \mathbb{X}_1)^{-1}$ we find that the term S_3 satisfies

$$\begin{aligned} S_3 &= \mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\left(\sum_{j \in \mathcal{S}} \theta_j^2 \mathbf{1}_{\{|\hat{\theta}_j| \leq \hat{\tau}_j, |\theta_j| \leq 2\hat{\tau}_j\}} + \sum_{j \in \mathcal{S}} \theta_j^2 \mathbf{1}_{\{|\hat{\theta}_j| \leq \hat{\tau}_j, |\theta_j| > 2\hat{\tau}_j\}} \right)^2 \right] \\ &\leq C \mathbf{E}_{\boldsymbol{\theta},\sigma} \sum_{j,k \in \mathcal{S}} \{ \hat{\tau}_j^2 \hat{\tau}_k^2 + \theta_j^2 \theta_k^2 \mathbf{1}_{\{|\hat{\theta}_j| \leq \hat{\tau}_j, |\theta_j| > 2\hat{\tau}_j\}} \mathbf{1}_{\{|\hat{\theta}_k| \leq \hat{\tau}_k, |\theta_k| > 2\hat{\tau}_k\}} \} \\ &\leq C \sum_{j,k \in \mathcal{S}} \mathbf{E}_{\boldsymbol{\theta},\sigma} \left(\hat{\sigma}^4 B_{jj} B_{kk} \right) \log^2 \left(1 + \frac{p}{s^2} \right) + C \mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\left(\sum_{j \in \mathcal{S}} \theta_j^2 \mathbf{P}_{\boldsymbol{\theta},\sigma}^{1/2}(|\hat{\theta}_j| \leq \hat{\tau}_j, |\theta_j| > 2\hat{\tau}_j | \mathbb{X}_1) \right)^2 \right]. \end{aligned} \quad (43)$$

Here,

$$\sum_{j,k \in \mathcal{S}} \mathbf{E}_{\boldsymbol{\theta},\sigma} \left(\hat{\sigma}^4 B_{jj} B_{kk} \right) \leq \mathbf{E}_{\boldsymbol{\theta},\sigma} (\hat{\sigma}^4 s^2 \lambda_{\max}^2(\mathbb{B})) \leq s^2 (\mathbf{E}_{\boldsymbol{\theta},\sigma}(\hat{\sigma}^8))^{1/2} (\mathbf{E}_{\boldsymbol{\theta},\sigma}(\lambda_{\max}^4(\mathbb{B})))^{1/2} \leq C \sigma^4 \frac{s^2}{n^2},$$

where the last inequality follows from Lemmas 1 and 2. Furthermore, using Lemmas 1 and 3 we get

$$\begin{aligned} \mathbf{E}_{\theta, \sigma} \left[\left(\sum_{j \in \mathcal{S}} \theta_j^2 \mathbf{P}_{\theta, \sigma}^{1/2}(|\hat{\theta}_j| \leq \hat{\tau}_j, |\theta_j| > 2\hat{\tau}_j | \mathbb{X}_1) \right)^2 \right] &\leq \mathbf{E} \left[\left(\sigma^2 \sum_{j \in \mathcal{S}} B_{jj} \max_{x>0} \{x^2 \exp(-Cx^2)\} \right)^2 \right] \\ &\leq C\sigma^4 s^2 \mathbf{E}^2(\lambda_{\max}^2(\mathbb{B})) \leq C\sigma^4 \frac{s^2}{n^2}. \end{aligned}$$

It follows that

$$S_3 \leq C\sigma^4 \frac{s^2}{n^2} \log^2 \left(1 + \frac{p}{s^2} \right). \quad (44)$$

Combining (34), (39) – (44) yields the bound (17) of the theorem for \hat{Q}_S^{LD} . The bound (18) for $\hat{\Lambda}_S^{LD}$ is deduced from (17) exactly in the same way as it is done in the proof of Theorem 8 in [8].

11 Proofs for Section 7

11.1 Proof of Proposition 1

We use a combination of arguments from [4] and [12]. By Theorem 6.1 in [12], if $\boldsymbol{\xi}$ is standard Gaussian and \mathbb{X} satisfies weighted restricted eigenvalue condition then the result of Proposition 1 holds. Now, note that, by Theorem 8.3 in [4], if κ_0 in Condition (P) is small enough and Condition (\mathbb{X}^*3) holds then \mathbb{X} satisfies weighted restricted eigenvalue condition with probability at least $1 - 3\exp(-Cn)$ for some constant $C > 0$ depending only on M . Thus, Proposition 1 holds when $\boldsymbol{\xi}$ is standard Gaussian. In order to extend it to subGaussian $\boldsymbol{\xi}$, we note that, in the proof of Theorem 6.1 in [12], the only ingredients involving $\boldsymbol{\xi}$ are Lemmas 7.6 and 7.7 and those lemmas remain valid for subGaussian noise (with possibly different constants depending on L). Indeed, Lemma 7.6 in [12] states that $\|\boldsymbol{\xi}\|_2^2/n$ is between two absolute constants with probability at least $1 - C'\exp(-n/C')$, which remains true for subGaussian $\boldsymbol{\xi}$. As concerns Lemma 7.7, it can be replaced by Theorem 9.1 from [4] that we state here in the following form.

Proposition 4 (Adapted from Theorem 9.1 in [4]). *Let $t > 0$ and let \mathbb{X} be a design matrix such that $\max_{1 \leq i \leq p} \|\mathbf{X}_i\|_2^2 \leq 2n$. Assume that Condition ($\boldsymbol{\xi}^*2$) holds. Then, there is a constant $C > 0$ such that, for all $\mathbf{u} \in \mathbb{R}^p$,*

$$\frac{1}{n} |\boldsymbol{\xi}^T \mathbb{X} \mathbf{u}| \leq C(G(\mathbf{u}) \vee \|\mathbf{u}\|_*)$$

with probability at least $1 - e^{-t}$, where $G(\mathbf{u}) = \frac{\sqrt{t+1}}{n} \|\mathbb{X} \mathbf{u}\|_2$.

It is not hard to check that the assumption $\max_{1 \leq i \leq p} \|\mathbf{X}_i\|_2^2 \leq 2n$ of Proposition 4 is satisfied with probability at least $1 - C'\exp(-n/C')$ provided \mathbf{X}_i are M -subGaussian, with components having variance 1, and Condition (P) holds. It follows that there is an event with probability at least $1 - C'\exp(-(n \wedge s \log(ep/s))/C')$ such that the conclusions of Theorem 6.1 in [12] (to within the constants that can depend only on L , M and κ_0) hold under the assumptions of Proposition 1.

11.2 Proof of Proposition 2

For brevity, in this proof we write (\mathbb{X}, \mathbf{Y}) instead of $(\mathbb{X}_1, \mathbf{Y}_1)$. Using the definition of $\hat{\sigma}_{\text{srs}}^2$ in (20) we get

$$|\hat{\sigma}_{\text{srs}}^2 - \sigma^2| \leq \frac{1}{n} \|\mathbb{X}(\hat{\boldsymbol{\theta}}_{\text{srs}} - \boldsymbol{\theta})\|_2^2 + \frac{\sigma^2}{n} \left| \|\boldsymbol{\xi}\|_2^2 - n \right| + \frac{2\sigma}{n} \left| \boldsymbol{\xi}^T \mathbb{X}(\hat{\boldsymbol{\theta}}_{\text{srs}} - \boldsymbol{\theta}) \right|. \quad (45)$$

To control the second term on the right-hand side of (45), we apply Bernstein's inequality (cf. e.g. Theorem 2.8.2 in [24]): if Condition $(\boldsymbol{\xi}^*2)$ holds, then

$$\mathbf{P}\left(\left| \|\boldsymbol{\xi}\|_2^2 - n \right| > u\right) \leq 2e^{-c_0\left(\frac{u^2}{n} \wedge u\right)}, \quad \forall u > 0,$$

where $c_0 > 0$ depends only on L . This yields, for any $t > 0$ such that $t \leq c_0 n$,

$$\left| \|\boldsymbol{\xi}\|_2^2 - n \right| \leq \sqrt{nt/c_0}$$

with probability at least $1 - 2e^{-t}$. Next, to bound the first and the third terms on the right-hand side of (45) we place ourselves on the event of probability at least $1 - C' \exp(-(n \wedge s \log(ep/s))/C')$ where the result of Proposition 1 holds. We denote this event by \mathcal{C} . By Proposition 1, on \mathcal{C} we have $\frac{1}{n} \|\mathbb{X}(\hat{\boldsymbol{\theta}}_{\text{srs}} - \boldsymbol{\theta})\|_2^2 \leq C' \sigma^2 \frac{s \log(ep/s)}{n}$.

Finally, to bound the third term on the right-hand side of (45) we use Proposition 4 where we set $\mathbf{u} = \hat{\boldsymbol{\theta}}_{\text{srs}} - \boldsymbol{\theta}$. For $t > 0$, denote by \mathcal{C}'_t the intersection of \mathcal{C} with the event of probability at least $1 - e^{-t}$, on which Proposition 4 holds. Then $\mathbf{P}(\mathcal{C}'_t) \geq 1 - C' \exp(-(n \wedge t \wedge s \log(ep/s))/C')$. Proposition 1 yields that, for any $t > 0$ satisfying $t \leq c_0 n$ we have, on the event \mathcal{C} ,

$$\begin{aligned} \|\mathbf{u}\|_* &\leq C\sigma \frac{s \log(ep/s)}{n}, \\ G(\mathbf{u}) &\leq \frac{\sigma t}{n} + \frac{\|\mathbb{X}\mathbf{u}\|_2^2}{\sigma n} + \frac{\|\mathbb{X}\mathbf{u}\|_2}{n} \leq C\sigma \left(\frac{t}{n} + \frac{s \log(ep/s)}{n} + \frac{(s \log(ep/s))^{1/2}}{n} \right) \\ &\leq C'\sigma \left(\sqrt{\frac{t}{n}} + \frac{s \log(ep/s)}{n} \right), \end{aligned}$$

where we have used the fact that $s \log(ep/s) \geq 1$. These remarks and Proposition 4 imply that, for any $t > 0$ satisfying $t \leq c_0 n$ we have, on the event \mathcal{C}'_t ,

$$\frac{1}{n} |\boldsymbol{\xi}^T \mathbb{X} \mathbf{u}| \leq C\sigma \left(\frac{t}{n} + C' \frac{s \log(ep/s)}{n} \right).$$

The result of the proposition follows.

11.3 Proof of Theorem 4

For brevity, in this subsection we write (\mathbb{X}, \mathbf{Y}) , \hat{Q} and $\hat{\Lambda}$ instead of $(\mathbb{X}_2, \mathbf{Y}_2)$, \hat{Q}_D^{HD} and $\hat{\Lambda}_D^{HD}$, respectively. We also set $\mathbf{u} = \hat{\boldsymbol{\theta}}_{\text{srs}} - \boldsymbol{\theta}$ and $\mathbf{w}_j = \sum_{k \neq j} u_k \mathbf{X}_k$.

Recalling the expressions obtained in the proof of Theorem 1, we have $\hat{Q} = f_1(\mathbf{Z}) + f_2(\mathbf{Z})$, where $\mathbf{Z} = \{(X_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}, (\xi_i)_{1 \leq i \leq n}\}$ and

$$\begin{aligned} f_1(\mathbf{Z}) &= Q(\boldsymbol{\theta}) + \sum_{j=1}^p \left[2\theta_j u_j \left(1 - \frac{\|\mathbf{X}_j\|_2^2}{n} \right) + 2\theta_j \frac{\mathbf{X}_j^T}{n} (\sigma \boldsymbol{\xi} - \mathbf{w}_j) \right] + \frac{2}{n} \sum_{j=1}^p u_j \sum_{i=1}^n X_{ij} (\sigma \xi_i - \mathbf{X}^i \mathbf{u}), \\ f_2(\mathbf{Z}) &= \sum_{i=1}^p u_i^2 + \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k=1}^p X_{ik} X_{jk} (\sigma \xi_i - \mathbf{X}^i \mathbf{u}) (\sigma \xi_j - \mathbf{X}^j \mathbf{u}). \end{aligned}$$

We will use a result from [1]. To this end, we introduce the following notation. Denote by \mathcal{P}_d the set of all partitions of $\{1, \dots, d\}$ into nonempty, pairwise disjoint sets. For a partition $\mathcal{I} = \{\mathcal{I}_1, \dots, \mathcal{I}_K\} \in \mathcal{P}_d$ and a multi-indexed array $\mathfrak{M} = (M_{i_1, \dots, i_d})_{i_1, \dots, i_d=1, \dots, q}$, we define

$$\|\mathfrak{M}\|_{\mathcal{I}} := \max \left\{ \sum_{i_1, \dots, i_d=1}^q M_{i_1, \dots, i_d} \prod_{k=1}^K x_{\mathbf{i}_k} : \|x_{\mathbf{i}_k}\|_2 \leq 1, k = 1, \dots, K \right\},$$

where $x_{\mathbf{i}_k}$ is a $\mathbf{Card}(\mathcal{I}_k)$ -dimensional array indexed by $\mathbf{i}_k = (i_j, j \in \mathcal{I}_k)$. Here, $\mathbf{Card}(\mathcal{I}_k)$ is the cardinality of \mathcal{I}_k and $\|x_{\mathbf{i}_k}\|_2^2 = \sum_{|\mathbf{i}_k| \leq q} x_{\mathbf{i}_k}^2$ with $|\mathbf{i}_k| = \max(i_j, j \in \mathcal{I}_k)$. Examples of this notion can be found in [1]. Finally, denote by $\mathbb{D}^d f$ the d -th derivative of a function $f : \mathbb{R}^q \rightarrow \mathbb{R}$, which we identify with a multi-indexed array \mathfrak{M} , where only entries with indices satisfying $i_1 + \dots + i_d = d$ (that are the corresponding partial derivatives) can be non-zero, and all other entries vanish.

Proposition 5 (Theorem 1.4 in [1]). *Let $f : \mathbb{R}^q \rightarrow \mathbb{R}$ be a polynomial of q variables of degree D . Let $\mathbf{Z} = (Z_1, \dots, Z_q)$ be a vector with independent L -subGaussian components. Then, for any $t > 0$,*

$$\mathbf{P}(|f(\mathbf{Z}) - \mathbf{E}f(\mathbf{Z})| \geq t) \leq 2 \exp \left[-c_D \min_{1 \leq d \leq D} \min_{\mathcal{I} \in \mathcal{P}_d} \left(\frac{t}{L^d \|\mathbf{E}\mathbb{D}^d f(\mathbf{Z})\|_{\mathcal{I}}} \right)^{\frac{2}{\mathbf{Card}(\mathcal{I})}} \right],$$

where c_D is a positive constant depending only on D .

We now apply this proposition with $f = \hat{Q}$, which is a polynomial of degree $D = 4$ (assuming \mathbf{u} fixed) of $q = np + n$ variables. Note that $\mathbf{E}_{\boldsymbol{\theta}, \sigma}(\mathbb{D}^3 \hat{Q} | \mathbf{u}) = \mathbf{E}_{\boldsymbol{\theta}, \sigma}(\mathbb{D}^3 \hat{Q} | \mathbf{u}) = 0$, and $\mathbf{E}_{\boldsymbol{\theta}, \sigma}(\mathbb{D}^4 f_1(\mathbf{Z}) | \mathbf{u}) = 0$. Hence, in order to apply Proposition 5, we only need to evaluate $\|\mathbf{E}_{\boldsymbol{\theta}, \sigma}(\mathbb{D}^2 f_i(\mathbf{Z}) | \mathbf{u})\|_{\mathcal{I}}$, $i = 1, 2$, for every $\mathcal{I} \in \mathcal{P}_2$ and $\|\mathbf{E}_{\boldsymbol{\theta}, \sigma}(\mathbb{D}^4 f_2(\mathbf{Z}) | \mathbf{u})\|_{\mathcal{I}}$ for every $\mathcal{I} \in \mathcal{P}_4$.

We have

$$\begin{aligned} \frac{\partial^2 f_1}{\partial \xi_i \partial \xi_j}(\mathbf{Z}) &= 0, \quad \frac{\partial^2 f_1}{\partial \xi_i \partial X_{jk}}(\mathbf{Z}) = \frac{2\sigma}{n}(u_k + \theta_k) \mathbb{1}_{i=j}, \\ \frac{\partial^2 f_1}{\partial X_{ij} \partial X_{kl}}(\mathbf{Z}) &= -\frac{2}{n}[u_l(u_j + \theta_j) + u_j(u_l + \theta_l)] \mathbb{1}_{i=k}. \end{aligned} \tag{46}$$

Since $\|\cdot\|_{\{1,2\}}$ is simply the Euclidean norm of the array (cf. [1]) we find

$$\|\mathbf{E}_{\boldsymbol{\theta}, \sigma}(\mathbb{D}^2 f_1(\mathbf{Z}) | \mathbf{u})\|_{\{1,2\}}^2 \leq C \frac{(\|\mathbf{u}\|_2^2 + \sigma^2)(\|\mathbf{u}\|_2^2 + \|\boldsymbol{\theta}\|_2^2)}{n}.$$

Furthermore,

$$\begin{aligned} &\|\mathbf{E}_{\boldsymbol{\theta}, \sigma}(\mathbb{D}^2 f_1(\mathbf{Z}) | \mathbf{u})\|_{\{\{1\}, \{2\}\}} \\ &= \max \left\{ \frac{2\sigma}{n} \sum_{i=1}^n \sum_{k=1}^p (u_k + \theta_k) x_{i,k} \bar{y}_i - \frac{2}{n} \sum_{i=1}^n \sum_{j,l=1}^p [u_l(u_j + \theta_j) + u_j(u_l + \theta_l)] x_{i,j} y_{i,l} \right. \\ &\quad \left. + \frac{2\sigma}{n} \sum_{i=1}^n \sum_{k=1}^p (u_k + \theta_k) y_{i,k} \bar{x}_i - \frac{2}{n} \sum_{i=1}^n \sum_{j,l=1}^p [u_l(u_j + \theta_j) + u_j(u_l + \theta_l)] y_{i,j} x_{i,l} \right\}, \end{aligned}$$

where the maximum is taken over $(x_{i,j}, \bar{x}_k, y_{i,j}, \bar{y}_k)$ such that $\sum_{i=1}^n \sum_{j=1}^p x_{i,j}^2 + \sum_{k=1}^n \bar{x}_k^2 \leq 1$ and $\sum_{i=1}^n \sum_{j=1}^p y_{i,j}^2 + \sum_{k=1}^n \bar{y}_k^2 \leq 1$. By the Cauchy-Schwarz inequality,

$$\left| \frac{2\sigma}{n} \sum_{i=1}^n \sum_{k=1}^p (u_k + \theta_k) x_{i,k} \bar{y}_i \right| \leq \frac{2\sigma}{n} (\|\mathbf{u}\|_2 + \|\boldsymbol{\theta}\|_2),$$

and on the other hand,

$$\begin{aligned} \left| \frac{2}{n} \sum_{i=1}^n \sum_{j,l=1}^p [u_l(u_j + \theta_j) + u_j(u_l + \theta_l)] x_{i,j} y_{i,l} \right| &\leq \frac{4}{n} \sqrt{\sum_{i,j,l} u_l^2 x_{i,j}^2} \sqrt{\sum_{i,j,l} (u_j + \theta_j)^2 y_{i,l}^2} \\ &\leq \frac{C}{n} \|\mathbf{u}\|_2 (\|\mathbf{u}\|_2 + \|\boldsymbol{\theta}\|_2). \end{aligned}$$

Hence $\|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^2 f_1(\mathbf{Z}) | \mathbf{u})\|_{\{\{1\}, \{2\}\}} \leq Cn^{-1}(\|\mathbf{u}\|_2 + \sigma)(\|\mathbf{u}\|_2 + \|\boldsymbol{\theta}\|_2)$.

We now turn to evaluation of $\|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^2 f_2(\mathbf{Z}) | \mathbf{u})\|_{\mathcal{I}}$ for $\mathcal{I} \in \mathcal{P}_2$ and $\|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^4 f_2(\mathbf{Z}) | \mathbf{u})\|_{\mathcal{I}}$ for $\mathcal{I} \in \mathcal{P}_4$. We have

$$f_2(\mathbf{Z}) - \sum_{i=1}^p u_i^2 = S_1(\mathbf{Z}) + S_2(\mathbf{Z}) + S_3(\mathbf{Z}) \quad (47)$$

where

$$\begin{aligned} S_1(\mathbf{Z}) &= \frac{\sigma^2}{n(n-1)} \sum_{i,j} \sum_{k_1, l_1} \sum_{k_2, l_2} [\mathbb{1}_{i \neq j, k_1=i, k_2=j, l_1=l_2}] \xi_i \xi_j X_{k_1 l_1} X_{k_2 l_2} \\ S_2(\mathbf{Z}) &= -\frac{2\sigma}{n(n-1)} \sum_i \sum_{k_1, l_1} \sum_{k_2, l_2} \sum_{k_3, l_3} [u_{l_3} \mathbb{1}_{k_1=i, k_2=k_3, k_2 \neq i, l_1=l_2}] \xi_i X_{k_1 l_1} X_{k_2 l_2} X_{k_3 l_3} \\ S_3(\mathbf{Z}) &= \frac{1}{n(n-1)} \sum_{k_1, l_1} \sum_{k_2, l_2} \sum_{k_3, l_3} \sum_{k_4, l_4} [u_{l_3} u_{l_4} \mathbb{1}_{k_1=k_3, k_2=k_4, k_1 \neq k_2, l_1=l_2}] X_{k_1 l_1} X_{k_2 l_2} X_{k_3 l_3} X_{k_4 l_4}. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\frac{\partial^2 f_2}{\partial \xi_i \partial \xi_j}(\mathbf{Z}) | \mathbf{u} \right] &= 0, \quad \mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\frac{\partial^2 f_2}{\partial \xi_i \partial X_{k,l}}(\mathbf{Z}) | \mathbf{u} \right] = -\frac{2\sigma u_l}{n} \mathbb{1}_{i=k}, \\ \mathbf{E}_{\boldsymbol{\theta},\sigma} \left[\frac{\partial^2 f_2}{\partial X_{k_1, l_1} \partial X_{k_2, l_2}}(\mathbf{Z}) | \mathbf{u} \right] &= \frac{2u_{l_1} u_{l_2}}{n} \mathbb{1}_{k_1=k_2}. \end{aligned}$$

By replacing the $u_j + \theta_j$ in (46) with u_k , similar argument as for f_1 yields $\|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^2 f_2(\mathbf{Z}) | \mathbf{u})\|_{\{1,2\}}^2 \leq Cn^{-1}(\|\mathbf{u}\|_2^2 + \sigma^2)\|\mathbf{u}\|_2^2$ and $\|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^2 f_2(\mathbf{Z}) | \mathbf{u})\|_{\{\{1\}, \{2\}\}} \leq Cn^{-1}(\|\mathbf{u}\|_2 + \sigma)\|\mathbf{u}\|_2$.

Furthermore, retrieving the fourth order derivatives of f_2 from (47) and recalling that $\|\cdot\|_{\{1,2,3,4\}}$ is the Euclidean norm we find

$$\|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^4 \hat{Q} | \mathbf{u})\|_{\{1,2,3,4\}}^2 = \|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^4 f_2(\mathbf{Z}) | \mathbf{u})\|_{\{1,2,3,4\}}^2 \leq (\|\mathbf{u}\|_2^4 + \sigma^4) \frac{p}{n^2}.$$

Next, for any partition $\mathcal{I} \in \mathcal{P}_4$,

$$\|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^4 f_2(\mathbf{Z}) | \mathbf{u})\|_{\mathcal{I}} \leq \sum_{i=1}^3 \|\mathbf{E}_{\boldsymbol{\theta},\sigma}(\mathbb{D}^4 S_i(\mathbf{Z}) | \mathbf{u})\|_{\mathcal{I}}. \quad (48)$$

We now evaluate separately the three terms on the right hand side of this inequality. We will only do it for the partition $\mathcal{I} = \{\{1\}, \{2, 3, 4\}\}$ since other partitions are treated analogously. First, consider the term with $S_1(\mathbf{Z})$ in (48). Observe that

$$\|\mathbf{E}_{\theta, \sigma}(\mathbb{D}^4 S_1(\mathbf{Z}) | \mathbf{u})\|_{\{\{1\}, \{2, 3, 4\}\}} \leq \frac{C\sigma^2}{n(n-1)} \max \left\{ \sum_{i \neq j} \sum_{l=1}^p x_i y_{i,j,l} + \sum_{i \neq j} \sum_{l=1}^p x_{i,l} y_{i,j,l} \right\}, \quad (49)$$

where the maximum is taken under the constraints imposed by the definition of the norm $\|\cdot\|_{\mathcal{I}}$. Thus, for the first double sum on the right hand side of (49), the maximum is taken under the constraint $\sum_{i=1}^n x_i^2 \leq 1, \sum_{i \neq j} \sum_{l=1}^p y_{i,j,l}^2 \leq 1$. Using the implication

$$\sum_{i=1}^n x_i^2 \leq 1, \sum_{i \neq j} \sum_{l=1}^p y_{i,j,l}^2 \leq 1 \quad \Rightarrow \quad \sum_{i \neq j} \sum_{l=1}^p x_i y_{i,j,l} \leq \sqrt{np}$$

to control the first double sum and treating quite analogously the second double sum in (49) we find that $\|\mathbf{E}_{\theta, \sigma}(\mathbb{D}^4 S_1(\mathbf{Z}) | \mathbf{u})\|_{\{\{1\}, \{2, 3, 4\}\}} \leq C\sigma^2 \sqrt{pn}^{-3/2}$.

Next, we consider the term with $S_2(\mathbf{Z})$ in (48). Similarly to (49) we have

$$\begin{aligned} \|\mathbf{E}_{\theta, \sigma}(\mathbb{D}^4 S_2(\mathbf{Z}) | \mathbf{u})\|_{\{\{1\}, \{2, 3, 4\}\}} &\leq \frac{C\sigma}{n(n-1)} \max \left\{ \sum_{i \neq j} \sum_{l, l'=1}^p x_i y_{i,j,l,l'} u_{l'} \right. \\ &\quad \left. + \sum_{i \neq j} \sum_{l, l'=1}^p x_{i,l} y_{i,j,l,l'} u_{l'} + \sum_{i \neq j} \sum_{l, l'=1}^p x_{i,l'} y_{i,j,l,l'} u_{l'} \right\}, \end{aligned} \quad (50)$$

where the maximum is taken under the constraints imposed by the definition of the norm $\|\cdot\|_{\mathcal{I}}$. For the constraint corresponding to the first triple sum on the right hand side of (50) we have the implication

$$\sum_{i=1}^n x_i^2 \leq 1, \sum_{i \neq j} \sum_{l, l'=1}^p y_{i,j,l,l'}^2 \leq 1 \quad \Rightarrow \quad \sum_{i \neq j} \sum_{l=1}^p x_i y_{i,j,l,l'} u_{l'} \leq \sqrt{\sum_{i,j,l,l'} u_{l'}^2 x_i^2 \sum_{i,j,l,l'} y_{i,j,l,l'}^2} \leq \|\mathbf{u}\|_2 \sqrt{np}.$$

The other terms on the right hand side of (50) are treated analogously. It follows that

$$\|\mathbf{E}_{\theta, \sigma}(\mathbb{D}^4 S_2(\mathbf{Z}) | \mathbf{u})\|_{\{\{1\}, \{2, 3, 4\}\}} \leq C\sigma \|\mathbf{u}\|_2 \sqrt{pn}^{-3/2}.$$

Similarly, such a bound holds for every partition in \mathcal{P}_4 . Finally, the control of the term with $S_3(\mathbf{Z})$ in (48) follows the same lines by using the implication

$$\begin{aligned} \sum_{i,a} x_{i,a}^2 \leq 1, \sum_{i \neq j} \sum_{a,b,c} y_{i,j,a,b,c}^2 \leq 1 \quad \Rightarrow \quad \sum_{i,j,a,b,c} x_{i,a} y_{i,j,a,b,c} u_b u_c &\leq \sqrt{\sum_{i,j,a,b,c} u_b^2 x_{i,a}^2 \sum_{i,j,a,b,c} u_c^2 y_{i,j,a,b,c}^2} \\ &\leq \|\mathbf{u}\|_2^2 \sqrt{np}, \end{aligned}$$

and analogous implications to obtain the bound $\|\mathbf{E}_{\theta, \sigma}(\mathbb{D}^4 S_3(\mathbf{Z}) | \mathbf{u})\|_{\{\{1\}, \{2, 3, 4\}\}} \leq C\|\mathbf{u}\|_2^2 \sqrt{pn}^{-3/2}$. Again, such a bound holds for every partition in \mathcal{P}_4 (we skip the argument, which is quite analogous).

Using the bounds obtained above for the three terms in (48) we get

$$\max_{\mathcal{I} \in \mathcal{P}_4} \|\mathbf{E}_{\theta, \sigma}(\mathbb{D}^4 \hat{Q} | \mathbf{u})\|_{\mathcal{I}} \leq C(\|\mathbf{u}\|_2^2 + \sigma^2) \sqrt{pn}^{-3/2}.$$

Using Proposition 5 and taking $t = \frac{\sigma^2 \sqrt{pv}}{n} + \sigma \sqrt{\frac{v}{n}} \|\boldsymbol{\theta}\|_2$ with $0 < v \leq n^{1/3}$ we have on the event $\{\|\mathbf{u}\|_2^2 \leq c_{\text{srs},2} \sigma^2 s \log(ep/s)/n\}$ and under Condition (P) that

$$\begin{aligned} \min_{1 \leq d \leq D} \min_{\mathcal{I} \in \mathcal{P}_d} \left(\frac{t}{L^d \|\mathbf{E} \mathbb{D}^d f(\mathbf{Z})\|_{\mathcal{I}}} \right)^{\frac{2}{\text{Card}(\mathcal{I})}} &\geq C \left[\frac{nt^2}{(\|\mathbf{u}\|_2^2 + \sigma^2)(\|\mathbf{u}\|_2^2 + \|\boldsymbol{\theta}\|_2^2)} \wedge \frac{nt}{(\|\mathbf{u}\|_2 + \|\boldsymbol{\theta}\|_2)(\sigma + \|\mathbf{u}\|_2)} \right. \\ &\quad \left. \wedge \frac{n^2 t^2}{(\|\mathbf{u}\|_2^4 + \sigma^4)p} \wedge \min_{k=2,3,4} \left\{ \frac{n^{3/2} t}{(\|\mathbf{u}\|_2^2 + \sigma^2) \sqrt{p}} \right\}^{2/k} \right] \\ &\geq Cv. \end{aligned}$$

Thus, the first assertion of the theorem follows from Propositions 1 and 5, and the fact that the estimator \hat{Q} is conditionally unbiased: $\mathbf{E}_{\boldsymbol{\theta},\sigma}(\hat{Q} | \mathbf{u}) = Q(\boldsymbol{\theta})$.

We now prove the second assertion in the theorem. Set $\tau^2 = \sigma^2 \frac{\sqrt{pv}}{n}$. We bound $|\hat{\Lambda} - \|\boldsymbol{\theta}\|_2|$ by considering the cases $\|\boldsymbol{\theta}\|_2 \leq \tau$ and $\|\boldsymbol{\theta}\|_2 > \tau$ separately.

Case I: $\|\boldsymbol{\theta}\|_2 \leq \tau$. Then we have

$$|\hat{\Lambda} - \|\boldsymbol{\theta}\|_2|^2 \leq 2(\hat{\Lambda}^2 + \|\boldsymbol{\theta}\|_2^2) = 2(\hat{Q} - \|\boldsymbol{\theta}\|_2^2) + 4\|\boldsymbol{\theta}\|_2^2.$$

By combining this inequality with the first assertion of the theorem we get that, with probability at least $1 - C'[e^{-v} + e^{-\frac{n \wedge s \log(ep/s)}{C'}}]$,

$$\begin{aligned} |\hat{\Lambda} - \|\boldsymbol{\theta}\|_2|^2 &\leq C \left(\sqrt{\sigma^4 \frac{pv}{n^2} + \sigma^2 \frac{\tau^2 v}{n}} + \tau^2 \right) \\ &\leq C \left(\sigma^2 \frac{\sqrt{pv}}{n} + \sigma \frac{\tau \sqrt{v}}{\sqrt{n}} + \tau^2 \right) \leq C \left(\sigma^2 \frac{\sqrt{pv}}{n} + \sigma^2 \frac{p^{1/4} v^{3/4}}{n} \right). \end{aligned} \quad (51)$$

Case II: $\|\boldsymbol{\theta}\|_2 > \tau$. Applying the inequality $\forall a > 0, b \geq 0, (a - b)^2 \leq (a^2 - b^2)^2 / a^2$ we find

$$|\hat{\Lambda} - \|\boldsymbol{\theta}\|_2|^2 \leq \frac{(\hat{Q} - \|\boldsymbol{\theta}\|_2^2)^2}{\|\boldsymbol{\theta}\|_2^2}.$$

Combining this inequality with the first assertion of the theorem we obtain that, with probability at least $1 - C'[e^{-v} + e^{-\frac{n \wedge s \log(ep/s)}{C'}}]$,

$$|\hat{\Lambda} - \|\boldsymbol{\theta}\|_2|^2 \leq C \frac{\sigma^4 \frac{pv}{n^2} + \sigma^2 \frac{\|\boldsymbol{\theta}\|_2^2 v}{n}}{\|\boldsymbol{\theta}\|_2^2} \leq C \left(\sigma^4 \frac{pv}{\tau^2 n^2} + \sigma^2 \frac{v}{n} \right) \leq C \left(\sigma^2 \frac{\sqrt{pv}}{n} + \sigma^2 \frac{v}{n} \right). \quad (52)$$

Putting (51) and (52) together we find

$$\mathbf{P}_{\boldsymbol{\theta},\sigma} \left(|\hat{\Lambda} - \|\boldsymbol{\theta}\|_2| > C \frac{\sigma(pv)^{1/4}}{\sqrt{n}} + C \sigma \sqrt{\frac{v}{n}} + C \sigma \frac{v^{3/8} p^{1/8}}{\sqrt{n}} \right) \leq C'(e^{-v/C'} + (\frac{s}{ep})^{s/C'}).$$

Finally, since we assume that $v \leq n^{1/3}$ and $p \geq \gamma n$ for some $0 < \gamma < 1$, the term $\frac{\sigma(pv)^{1/4}}{\sqrt{n}}$ in the above inequality dominates the other two terms. The second assertion of the theorem follows.

11.4 Proof of Theorem 5

For brevity, in this subsection we write $\hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}, \hat{\sigma}$ and \hat{Q} instead of $\hat{\boldsymbol{\theta}}_{\text{srs}}, \tilde{\boldsymbol{\theta}}_{\text{srs}}, \sqrt{2}\hat{\sigma}_{\text{srs}}$ and \hat{Q}_S^{HD} , respectively. We also set $\mathbf{u} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ and

$$\boldsymbol{\nu} = \left(\mathbb{I}_p - \frac{\mathbb{X}_3^T \mathbb{X}_3}{n} \right) \mathbf{u}, \quad \boldsymbol{\epsilon} = \frac{\sigma}{n} \mathbb{X}_3^T \boldsymbol{\xi}_3, \quad \hat{\tau} = \alpha \hat{\sigma} \sqrt{\frac{\log(1+p/s^2)}{n}}, \quad \tau = \alpha \sigma \sqrt{\frac{\log(1+p/s^2)}{n}}. \quad (53)$$

With this notation, we have $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + \boldsymbol{\nu} + \boldsymbol{\epsilon}$. We define the random event

$$\mathcal{A} = \left\{ \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2 \leq \sigma^2 \right\} \cap \left\{ \sigma^2 \leq \hat{\sigma}^2 \leq 3\sigma^2 \right\}. \quad (54)$$

11.4.1 Preliminary lemmas for the proof of Theorem 5

Lemma 4. *Let Conditions $(\boldsymbol{\xi}^*2)$, (\mathbb{X}^*3) and (P) hold. Then there exist positive constants C_1, C_2 depending only on L, M such that for $\alpha > 0$ large enough and any $j \in \{1, \dots, p\}$,*

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\{|\epsilon_j| > \hat{\tau}/2\} \cap \mathcal{A} \right) \leq C_1 \frac{s^2}{p}$$

and

$$\theta_j^4 \mathbf{P}_{\boldsymbol{\theta}, \sigma} (|\epsilon_j| > |\theta_j|/4) \leq C_2 \frac{\sigma^4}{n^2},$$

where $\boldsymbol{\epsilon}, \hat{\tau}$ and \mathcal{A} are defined in (53) and (54) respectively.

Proof. For brevity, in this proof we denote by \mathbf{X}_j the j th column of matrix \mathbb{X}_3 . Using the definition of \mathcal{A} we get

$$\begin{aligned} \mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\{|\epsilon_j| > \hat{\tau}/2\} \cap \mathcal{A} \right) &\leq \mathbf{P}_{\boldsymbol{\theta}, \sigma} (|\epsilon_j| > \tau/2) \\ &\leq \mathbf{P}_{\boldsymbol{\theta}, \sigma} (\|\mathbf{X}_j\|_2^2 > 2n) + \mathbf{P}_{\boldsymbol{\theta}, \sigma} (\|\mathbf{X}_j\|_2^2 \leq 2n, |\epsilon_j| > \tau/2). \end{aligned}$$

Since ξ_1, \dots, ξ_n are independent L -subGaussian variables we have that ϵ_j is a $\sigma L \|\mathbf{X}_j\|_2 / n$ -subGaussian variable for fixed \mathbf{X}_j . It follows that

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma} (|\epsilon_j| > \tau/2 \mid \mathbf{X}_j) \leq 2 \exp \left(- \frac{n^2 \tau^2}{8L^2 \sigma^2 \|\mathbf{X}_j\|_2^2} \right),$$

and thus

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma} (\|\mathbf{X}_j\|_2^2 \leq 2n, |\epsilon_j| > \tau/2) \leq 2 \exp \left(- \frac{n \tau^2}{16L^2 \sigma^2} \right) \leq 2s^2/p$$

if $\alpha > 4L$. Moreover, as X_{1j}, \dots, X_{nj} are independent M -subGaussian variables their squares are subExponential. Since also $\mathbf{E}(X_{ij}^2) = 1$ Bernstein's inequality (cf. Theorem 2.8.2 in [24]) yields that there is a constant $C > 0$ such that $\mathbf{P}_{\boldsymbol{\theta}, \sigma} (\|\mathbf{X}_j\|_2^2 - n > n) \leq \exp(-Cn)$. In view of Condition (P) we have $\exp(-Cn) \leq C' \frac{s^2}{p}$ for $\kappa_0 > 0$ small enough. The first result of the lemma follows.

To prove the second inequality of the lemma, note that the variables $(X_{ij}\xi_i)_{1 \leq i \leq n}$ are independent, and they are subExponential as products of two subGaussians, cf. Lemma 2.7.7 in [24]. Thus, Theorem 2.8.2 in [24] yields

$$\theta_j^4 \mathbf{P}_{\boldsymbol{\theta}, \sigma} (|\epsilon_j| > |\theta_j|/4) \leq 2\theta_j^4 e^{-C \min \left\{ \frac{n\theta_j^2}{\sigma^2}, \frac{n|\theta_j|}{\sigma} \right\}} \leq \frac{2\sigma^4}{n^2} \sup_{x \geq 0} [x^4 e^{-C \min \{x^2, x\}}]. \quad (55)$$

□

Lemma 5. *Let Conditions (ξ^*2) , (\mathbb{X}^*3) and (P) hold. Then there exist positive constants C_1, C_2 depending only on L, M such that for $\alpha > 0$ large enough and any $j \in \{1, \dots, p\}$,*

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{|\nu_j| > \hat{\tau}/2\} \cap \mathcal{A}) \leq C_1 \frac{s^2}{p}$$

and

$$\theta_j^4 \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{|\nu_j| > |\theta_j|/4\} \cap \mathcal{A}) \leq C_2 \frac{\sigma^4}{n^2},$$

where $\boldsymbol{\nu}, \hat{\tau}$ and \mathcal{A} are defined in (53) and (54) respectively.

Proof. For brevity, in this proof we denote by \mathbf{X}_j the j th column of matrix \mathbb{X}_3 , by \mathbb{X}_{-j} the matrix obtained by removing the j th column (that is, \mathbf{X}_j) from \mathbb{X}_3 , and by \mathbf{u}_{-j} the vector obtained by removing the j th entry from \mathbf{u} .

For any $j \in \{1, \dots, p\}$, we have the upper bound

$$|\nu_j| \leq \left| \left(1 - \frac{\|\mathbf{X}_j\|_2^2}{n}\right) u_j \right| + \left| \frac{1}{n} \mathbf{X}_j^T \mathbb{X}_{-j} \mathbf{u}_{-j} \right|. \quad (56)$$

Using the definition of \mathcal{A} and Bernstein's inequality (Theorem 2.8.2 in [24]) we find

$$\begin{aligned} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{ |(1 - n^{-1} \|\mathbf{X}_j\|_2^2) u_j| > \hat{\tau}/4 \} \cap \mathcal{A}) &\leq \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|1 - n^{-1} \|\mathbf{X}_j\|_2^2| > \tau/(4\sigma)) \\ &\leq 2e^{-Cn \min\{\tau^2/\sigma^2, \tau/\sigma\}} \leq C' s^2/p, \end{aligned} \quad (57)$$

where the last inequality holds for α large enough due to the fact that $\tau/\sigma < 1$ for κ_0 small enough, cf. Condition (P) .

Next, notice that $\mathbf{X}_j^T \mathbb{X}_{-j} \mathbf{u}_{-j} = \sum_{i=1}^n \eta_i$ where the random variables $\eta_i := X_{ij} \sum_{k \neq j} X_{ik} u_k$, $i = 1, \dots, n$, are independent and zero-mean conditionally on \mathbf{u} . Furthermore, conditionally on \mathbf{u} , each variable $\sum_{k \neq j} X_{ik} u_k$ is $\|\mathbf{u}\|_2 M$ -subGaussian. Thus, the subGaussian norms (cf. [24]) of X_{ij} and of $\sum_{k \neq j} X_{ik} u_k$, conditionally on \mathbf{u} , do not exceed $C_0 M$ and $C_0 M \|\mathbf{u}\|_2$, respectively, where $C_0 > 0$ is an absolute constant. By Lemma 2.7.7 in [24], the subExponential norm of η_i (for any fixed \mathbf{u}) does not exceed $C_0^2 M^2 \|\mathbf{u}\|_2$. Therefore, conditionally on \mathbf{u} , we can apply Theorem 2.8.1 in [24]. This yields

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma}(\left| \frac{1}{n} \mathbf{X}_j^T \mathbb{X}_{-j} \mathbf{u}_{-j} \right| > \tau/4 \mid \mathbf{u}) \leq 2e^{-Cn \min\{\tau^2/\|\mathbf{u}\|_2^2, \tau/\|\mathbf{u}\|_2\}}. \quad (58)$$

It follows that

$$\begin{aligned} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{ \left| \frac{1}{n} \mathbf{X}_j^T \mathbb{X}_{-j} \mathbf{u}_{-j} \right| > \hat{\tau}/4 \} \cap \mathcal{A}) &\leq \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{ \left| \frac{1}{n} \mathbf{X}_j^T \mathbb{X}_{-j} \mathbf{u}_{-j} \right| > \tau/4 \} \cap \mathcal{A}) \\ &\leq \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[\mathbf{P}_{\boldsymbol{\theta}, \sigma} \left(\left| \frac{1}{n} \mathbf{X}_j^T \mathbb{X}_{-j} \mathbf{u}_{-j} \right| > \tau/4 \mid \mathbf{u} \right) \mathbb{1}_{\|\mathbf{u}\|_2 \leq \sigma} \right] \\ &\leq 2e^{-Cn \min\{\tau^2/\sigma^2, \tau/\sigma\}} \leq C' s^2/p. \end{aligned} \quad (59)$$

The first assertion of the lemma follows.

To prove the second assertion, we use again the bound (56). Considering the first term on the right hand side of (56) and reasoning analogously to (57) we get

$$\theta_j^4 \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{ |(1 - n^{-1} \|\mathbf{X}_j\|_2^2) u_j| > |\theta_j|/8 \} \cap \mathcal{A}) \leq 2\theta_j^4 e^{-C \min\left\{ \frac{n\theta_j^2}{\sigma^2}, \frac{n|\theta_j|}{\sigma} \right\}} \leq C' \frac{\sigma^4}{n^2}, \quad (60)$$

where the last inequality is obtained as in (55). To handle the second term on the right hand side of (56), we note that, analogously to (58) and (59),

$$\theta_j^4 \mathbf{P}_{\theta, \sigma} \left(\left| \frac{1}{n} \mathbf{X}_j^T \mathbb{X}_{-j} \mathbf{u}_{-j} \right| > |\theta_j|/8 \mid \mathbf{u} \right) \leq 2\theta_j^4 e^{-Cn \min\{\theta_j^2/\|\mathbf{u}\|_2^2, |\theta_j|/\|\mathbf{u}\|_2\}},$$

and

$$\theta_j^4 \mathbf{P}_{\theta, \sigma} \left(\left\{ \left| \frac{1}{n} \mathbf{X}_j^T \mathbb{X}_{-j} \mathbf{u}_{-j} \right| > |\theta_j|/8 \right\} \cap \mathcal{A} \right) \leq 2\theta_j^4 e^{-C \min\left\{ \frac{n\theta_j^2}{\sigma^2}, \frac{n|\theta_j|}{\sigma} \right\}} \leq C' \frac{\sigma^4}{n^2}. \quad (61)$$

The second assertion of the lemma follows from (60) and (61). \square

11.4.2 Proof of Theorem 5

Recall that, for brevity, $\hat{Q}_S^{HD} = \hat{Q}$. We follow an argument close to the proof of Theorem 3. Similarly to (34), we first write

$$\begin{aligned} \mathbf{E}_{\theta, \sigma} \left[(\hat{Q} - Q(\theta))^2 \mathbf{1}_{\mathcal{A}} \right] &\leq 2\mathbf{E}_{\theta, \sigma} \left(\mathbf{1}_{\mathcal{A}} \sum_{j \in \mathcal{S}^c} a_j(\hat{\theta}) \mathbf{1}_{|\tilde{\theta}_j| > \hat{\tau}} \right)^2 \\ &\quad + 2\mathbf{E}_{\theta, \sigma} \left(\mathbf{1}_{\mathcal{A}} \sum_{j \in \mathcal{S}} \{a_j(\hat{\theta}) \mathbf{1}_{|\tilde{\theta}_j| > \hat{\tau}} - \theta_j^2\} \right)^2, \end{aligned} \quad (62)$$

where $\hat{\tau}$ is defined in (53), \mathcal{A} is defined in (54), and \mathcal{S} is the support of $\theta \in B_0(s)$. Recall that \mathcal{A} and $\tilde{\theta}$ are measurable with respect to $(\mathbb{X}_1, \mathbf{Y}_1, \mathbb{X}_3, \mathbf{Y}_3)$, $a_j(\hat{\theta})$ is measurable with respect to $(\mathbb{X}_1, \mathbf{Y}_1, \mathbb{X}_2, \mathbf{Y}_2)$, and $\hat{\tau}$ is measurable with respect to $(\mathbb{X}_1, \mathbf{Y}_1)$.

Following the same lines as in deriving (37) (now, to obtain an analog of (36) we condition on $(\mathbb{X}_1, \mathbf{Y}_1, \mathbb{X}_3, \mathbf{Y}_3)$ rather than on $(\mathbb{X}_1, \mathbf{Y}_1)$) we find

$$\begin{aligned} \mathbf{E}_{\theta, \sigma} \left[\left(\mathbf{1}_{\mathcal{A}} \sum_{j \in \mathcal{S}^c} a_j(\hat{\theta}) \mathbf{1}_{|\tilde{\theta}_j| > \hat{\tau}} \right)^2 \right] \\ \leq \frac{C}{n^2} \mathbf{E}_{\theta, \sigma} \left[\mathbf{1}_{\mathcal{A}} \left((\|\mathbf{u}\|_2^2 + \sigma^2)^2 \sum_{k \in \mathcal{S}^c} \mathbf{1}_{|\tilde{\theta}_k| > \hat{\tau}} + \sum_{k, k' \in \mathcal{S}^c: k' \neq k} u_{k'}^2 u_k^2 \mathbf{1}_{|\tilde{\theta}_{k'}| > \hat{\tau}} \mathbf{1}_{|\tilde{\theta}_k| > \hat{\tau}} \right) \right]. \end{aligned} \quad (63)$$

Using (63), the definition of \mathcal{A} and the fact that

$$\sum_{k, k' \in \mathcal{S}^c: k' \neq k} u_{k'}^2 u_k^2 \mathbf{1}_{|\tilde{\theta}_{k'}| > \hat{\tau}} \mathbf{1}_{|\tilde{\theta}_k| > \hat{\tau}} \mathbf{1}_{\mathcal{A}} \leq \|\mathbf{u}\|_2^4 \sum_{j \in \mathcal{S}^c} \mathbf{1}_{|\tilde{\theta}_j| > \hat{\tau}} \mathbf{1}_{\mathcal{A}} \leq \sigma^4 \sum_{j \in \mathcal{S}^c} \mathbf{1}_{|\tilde{\theta}_j| > \hat{\tau}} \mathbf{1}_{\mathcal{A}},$$

we get the upper bound

$$\mathbf{E}_{\theta, \sigma} \left[\mathbf{1}_{\mathcal{A}} \left(\sum_{j \in \mathcal{S}^c} a_j(\hat{\theta}) \mathbf{1}_{|\tilde{\theta}_j| > \hat{\tau}} \right)^2 \right] \leq C \frac{\sigma^4}{n^2} \mathbf{E}_{\theta, \sigma} \left(\mathbf{1}_{\mathcal{A}} \sum_{j \in \mathcal{S}^c} \mathbf{1}_{|\tilde{\theta}_j| > \hat{\tau}} \right).$$

Noticing that for $j \in \mathcal{S}^c$ we have $\tilde{\theta}_j = \nu_j + \epsilon_j$ and applying Lemmas 4 and 5 we obtain

$$\mathbf{E}_{\theta, \sigma} \left[\mathbf{1}_{\mathcal{A}} \left(\sum_{j \in \mathcal{S}^c} a_j(\hat{\theta}) \mathbf{1}_{|\tilde{\theta}_j| > \hat{\tau}} \right)^2 \right] \leq C \sigma^4 \frac{s^2}{n^2}. \quad (64)$$

To analyze the second term on the right hand side of (62), we use a decomposition analogous to (40) with the difference that now we insert $\mathbb{1}_{\mathcal{A}}$ under the expectation. It is easy to check that the analogs of the terms S_1 and S_2 in (40) are bounded quite similarly to (41) and (42) (we skip the details here). This yields

$$\mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[\mathbb{1}_{\mathcal{A}} \left(\sum_{j \in \mathcal{S}} \{a_j(\hat{\boldsymbol{\theta}}) \mathbb{1}_{|\tilde{\theta}_j| > \hat{\tau}} - \theta_j^2\} \right)^2 \right] \leq C \left(\frac{\sigma^2}{n} \|\boldsymbol{\theta}\|_2^2 + \sigma^4 \frac{s}{n^2} + \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[\mathbb{1}_{\mathcal{A}} \left(\sum_{j \in \mathcal{S}} \theta_j^2 \mathbb{1}_{\tilde{\theta}_j \leq \hat{\tau}} \right)^2 \right] \right). \quad (65)$$

Now, notice that, in view of Lemmas 4 and 5,

$$\begin{aligned} \theta_j^4 \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{|\tilde{\theta}_j| \leq \hat{\tau}, |\theta_j| > 2\hat{\tau}\} \cap \mathcal{A}) &\leq \theta_j^4 \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{|\tilde{\theta}_j - \theta_j| \geq |\theta_j|/2\} \cap \mathcal{A}) \\ &\leq \theta_j^4 \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{|\epsilon_j| \geq |\theta_j|/4\} \cap \mathcal{A}) + \theta_j^4 \mathbf{P}_{\boldsymbol{\theta}, \sigma}(\{|\nu_j| \geq |\theta_j|/4\} \cap \mathcal{A}) \\ &\leq C \frac{\sigma^4}{n^2}. \end{aligned}$$

Using this inequality and acting as in (43) we get

$$\begin{aligned} \mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[\mathbb{1}_{\mathcal{A}} \left(\sum_{j \in \mathcal{S}} \theta_j^2 \mathbb{1}_{\tilde{\theta}_j \leq \hat{\tau}} \right)^2 \right] &\leq C s^2 \mathbf{E}_{\boldsymbol{\theta}, \sigma} [\hat{\tau}^4 \mathbb{1}_{\mathcal{A}}] + C \left[\sum_{j \in \mathcal{S}} \theta_j^2 \mathbf{P}_{\boldsymbol{\theta}, \sigma}^{1/2}(\{|\tilde{\theta}_j| \leq \hat{\tau}, |\theta_j| > 2\hat{\tau}\} \cap \mathcal{A}) \right]^2 \\ &\leq C' \sigma^4 \frac{s^2}{n^2} \log^2 \left(1 + \frac{p}{s^2} \right). \end{aligned} \quad (66)$$

Combining (62), (64), (65), and (66) we obtain

$$\mathbf{E}_{\boldsymbol{\theta}, \sigma} \left[(\hat{Q} - Q(\boldsymbol{\theta}))^2 \mathbb{1}_{\mathcal{A}} \right] \leq C \left(\sigma^4 \frac{s^2 \log^2(1 + \sqrt{p}/s)}{n^2} + \sigma^2 \frac{\|\boldsymbol{\theta}\|_2^2}{n} \right). \quad (67)$$

To conclude the proof of the first inequality of the theorem (that is, the bound on the error of \hat{Q}_S^{HD}), it suffices to use (67), Markov's inequality and the fact that, due to Propositions 1 and 2,

$$\mathbf{P}_{\boldsymbol{\theta}, \sigma}(\mathcal{A}) \geq 1 - C \exp(-(n \wedge s \log(ep/s))/C).$$

The second inequality of the theorem (that is, the bound on the error of $\hat{\Lambda}_S^{HD}$) is deduced from the first one exactly in the same way as it is done in the proof of Theorem 8 in [8].

12 Proofs of the lower bounds

12.1 Proof of Theorem 6

It is straightforward to check that (23) and (24) follow from (21) - (22). Thus, we only prove (21) - (22).

For an integer s such that $1 \leq s \leq p$ and $u > 0$, we define $\Theta(s, u) = \{\boldsymbol{\theta} \in B_0(s) : \|\boldsymbol{\theta}\|_2 \geq u\}$. Note first that it suffices to prove (21) for $s \leq \sqrt{p}$ and $\rho \leq r$, where for $A > 0$

$$r = A \min \left(\sqrt{\frac{s \log(1 + p/s^2)}{N}}, 1 \right). \quad (68)$$

Indeed, for $\sqrt{p} < s \leq p$ we have the inclusions

$$\Theta(s, \sigma\rho(s, N, p)) \supseteq \Theta(s', \sigma\rho(s, N, p)) \supseteq \Theta(s', \sigma\rho(s', N, p))$$

where $s' = \lfloor \sqrt{p} \rfloor$ is the greatest integer smaller than or equal to \sqrt{p} . It is easy to check that $\rho(s, N, p)$ is of the same order of magnitude for $s = s'$ as for all $\sqrt{p} < s \leq p$:

$$\rho(s, N, p) \asymp \min\left(\frac{p^{1/4}}{\sqrt{n}}, 1\right), \quad \forall s' \leq s \leq p.$$

It follows that it is enough to prove (21) for $s \leq s'$, and thus for $s \leq \sqrt{p}$. Furthermore, if $s \leq \sqrt{p}$, we have $\log(1 + p/s^2) \geq \log(1 + \sqrt{p}/s)$, so that it is sufficient to prove (21) for $\rho \leq r$ where r is defined in (68).

Thus, in the rest of this proof we assume that $s \leq \sqrt{p}$ and we show that (21) holds for $\bar{\rho} \leq r$, where r is defined in (68).

Let $\tau > 0$ be defined by the formula

$$\tau^2 = \frac{\bar{\rho}^2}{1 + \bar{\rho}^2}. \quad (69)$$

Denote by μ_τ the probability measure corresponding to the uniform distribution on the set $\Theta' = \Theta'(\tau)$ of all vectors in \mathbb{R}^p that have exactly s nonzero components, all equal to τ/\sqrt{s} . Note that the support of measure μ_τ is contained in $\Theta(s, \tau)$. In the rest of this proof, we set

$$\sigma^2 = 1 - \tau^2 = \frac{1}{1 + \bar{\rho}^2}.$$

For any test $\bar{\Delta}$ and $\sigma = \sqrt{1 - \tau^2}$, we have

$$\begin{aligned} R(\bar{\Delta}, s, \rho) &\geq \mathbf{P}_{\mathbf{0},1}(\bar{\Delta} = 1) + \sup_{\boldsymbol{\theta} \in B_0(s): \|\boldsymbol{\theta}\|_2 \geq \sigma\bar{\rho}} \mathbf{P}_{\boldsymbol{\theta},\sigma}(\bar{\Delta} = 0) \\ &\geq \mathbf{P}_{\mathbf{0},1}(\bar{\Delta} = 1) + \sup_{\boldsymbol{\theta} \in \Theta': \|\boldsymbol{\theta}\|_2 = \tau} \mathbf{P}_{\boldsymbol{\theta},\sqrt{1-\tau^2}}(\bar{\Delta} = 0) \\ &\geq \mathbf{P}_{\mathbf{0},1}(\bar{\Delta} = 1) + \int \mathbf{P}_{\boldsymbol{\theta},\sqrt{1-\tau^2}}(\bar{\Delta} = 0) \mu_\tau(d\boldsymbol{\theta}) = \mathbf{P}_{\mathbf{0},1}(\bar{\Delta} = 1) + \mathbb{P}_{\mu_\tau}(\bar{\Delta} = 0), \end{aligned}$$

where we have used the fact that $\tau = \sigma\bar{\rho}$ and the notation $\mathbb{P}_{\mu_\tau}(\cdot) = \int \mathbf{P}_{\boldsymbol{\theta},\sqrt{1-\tau^2}}(\cdot) \mu_\tau(d\boldsymbol{\theta})$. The theorem now follows from the last display and Lemma 8 given below.

To prove Lemma 8, we need the following two lemmas that may be of independent interest.

Lemma 6. *Let $\tau \in (0, 1)$ and $\sigma = \sqrt{1 - \tau^2}$. Let $\boldsymbol{\xi}$ be standard normal and let \mathbb{X} be a random matrix independent of $\boldsymbol{\xi}$. Then, for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^p$ we have*

$$\int \frac{d\mathbf{P}_{\boldsymbol{\theta},\sigma} d\mathbf{P}_{\boldsymbol{\theta}',\sigma}}{d\mathbf{P}_{\mathbf{0},1}} = \lambda^{N/2} \mathbf{E}_{\mathbb{X}} \exp \left(\lambda \left[\langle \mathbb{X}\boldsymbol{\theta}, \mathbb{X}\boldsymbol{\theta}' \rangle - \frac{\tau^2}{2} (\|\mathbb{X}\boldsymbol{\theta}\|_2^2 + \|\mathbb{X}\boldsymbol{\theta}'\|_2^2) \right] \right),$$

where $\lambda = (1 - \tau^4)^{-1}$ and $\mathbf{E}_{\mathbb{X}}$ denotes the expectation with respect to the distribution of \mathbb{X} .

Proof. Set $a = (2/\sigma^2 - 1)^{-1} > 0$. We have

$$\begin{aligned} \int \frac{d\mathbf{P}_{\boldsymbol{\theta},\sigma} d\mathbf{P}_{\boldsymbol{\theta}',\sigma}}{d\mathbf{P}_{\mathbf{0},1}} &= \frac{1}{(2\pi)^{N/2} \sigma^{2N}} \mathbf{E}_{\mathbb{X}} \int_{\mathbb{R}^N} \exp \left(-\frac{1}{2\sigma^2} (\|y - \mathbb{X}\boldsymbol{\theta}\|_2^2 + \|y - \mathbb{X}\boldsymbol{\theta}'\|_2^2) + \frac{\|y\|_2^2}{2} \right) dy \\ &= \frac{1}{(2\pi)^{N/2} \sigma^{2N}} \mathbf{E}_{\mathbb{X}} \int_{\mathbb{R}^N} \exp \left(-\frac{\|y\|_2^2}{2a} - \frac{1}{2\sigma^2} \left[-2\langle y, \mathbb{X}(\boldsymbol{\theta} + \boldsymbol{\theta}') \rangle + \|\mathbb{X}\boldsymbol{\theta}\|_2^2 + \|\mathbb{X}\boldsymbol{\theta}'\|_2^2 \right] \right) dy \\ &= \frac{1}{(2\pi)^{N/2} \sigma^{2N}} \mathbf{E}_{\mathbb{X}} \left[\exp \left(-\frac{1}{2\sigma^2} (\|\mathbb{X}\boldsymbol{\theta}\|_2^2 + \|\mathbb{X}\boldsymbol{\theta}'\|_2^2) \right) \int_{\mathbb{R}^N} \exp \left(-\frac{\|y\|_2^2}{2a} + \frac{\langle y, \mathbb{X}(\boldsymbol{\theta} + \boldsymbol{\theta}') \rangle}{\sigma^2} \right) dy \right]. \end{aligned}$$

Here,

$$\int_{\mathbb{R}^N} \exp \left(-\frac{\|y\|_2^2}{2a} + \frac{\langle y, \mathbb{X}(\boldsymbol{\theta} + \boldsymbol{\theta}') \rangle}{\sigma^2} \right) dy = (2\pi a)^{N/2} \exp \left(\frac{\|\mathbb{X}(\boldsymbol{\theta} + \boldsymbol{\theta}')\|_2^2}{2\sigma^2(2 - \sigma^2)} \right),$$

and the lemma follows. \square

Lemma 7. *Under the assumptions of Lemma 6, let \mathbb{X} be a matrix with i.i.d. standard normal entries and $\|\boldsymbol{\theta}\|_2 = \|\boldsymbol{\theta}'\|_2 = \tau$. Then, for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^p$ we have*

$$\int \frac{d\mathbf{P}_{\boldsymbol{\theta},\sigma} d\mathbf{P}_{\boldsymbol{\theta}',\sigma}}{d\mathbf{P}_{\mathbf{0},1}} = \frac{1}{(1 - \langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle)^N}.$$

Proof. Lemma 6 and the fact that \mathbb{X} is a matrix with i.i.d. standard normal entries imply

$$\begin{aligned} \int \frac{d\mathbf{P}_{\boldsymbol{\theta},\sigma} d\mathbf{P}_{\boldsymbol{\theta}',\sigma}}{d\mathbf{P}_{\mathbf{0},1}} &= \lambda^{N/2} \left(\mathbf{E}_{\mathbb{X}} \exp \left[\lambda ((\mathbf{X}^1 \boldsymbol{\theta})(\mathbf{X}^1 \boldsymbol{\theta}') - \frac{\tau^2}{2} ((\mathbf{X}^1 \boldsymbol{\theta})^2 + (\mathbf{X}^1 \boldsymbol{\theta}')^2)) \right] \right)^N \quad (70) \\ &= \lambda^{N/2} \left(\mathbf{E} \exp(\lambda Z^T \mathbb{D} Z) \right)^N, \end{aligned}$$

where $Z \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_p)$ and

$$\mathbb{D} = \frac{1}{2} \left[(\boldsymbol{\theta} \boldsymbol{\theta}'^T + \boldsymbol{\theta}' \boldsymbol{\theta}^T) - \tau^2 (\boldsymbol{\theta} \boldsymbol{\theta}^T + \boldsymbol{\theta}' \boldsymbol{\theta}'^T) \right].$$

The rank of the matrix \mathbb{D} is at most two and, for all $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ such that $\|\boldsymbol{\theta}\|_2 = \|\boldsymbol{\theta}'\|_2 = \tau$, this matrix has the eigenvalues

$$\lambda_1 = \frac{\langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle + \tau^2}{2} (1 - \tau^2) > 0,$$

$$\lambda_2 = \frac{\langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle - \tau^2}{2} (1 + \tau^2) \leq 0.$$

In particular, since $\tau < 1$ we have $\lambda_1 \lambda \leq \tau^2 / (1 + \tau^2) < 1/2$. It follows that

$$\mathbf{E} \exp(\lambda Z^T \mathbb{D} Z) = \frac{1}{\sqrt{(1 - 2\lambda_1 \lambda)(1 - 2\lambda_2 \lambda)}}. \quad (71)$$

using the fact that $\mathbb{E}_{G \sim \mathcal{N}(0,1)} \exp(tG^2) = (1 - 2t)^{-1/2}$ for any $t < 1/2$.

Combining (70) with (71) we get after some algebra that, for all $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ such that $\|\boldsymbol{\theta}\|_2 = \|\boldsymbol{\theta}'\|_2 = \tau$,

$$\int \frac{d\mathbf{P}_{\boldsymbol{\theta},\sigma} d\mathbf{P}_{\boldsymbol{\theta}',\sigma}}{d\mathbf{P}_{\mathbf{0},1}} = \left(\frac{\lambda}{(1 - 2\lambda_1 \lambda)(1 - 2\lambda_2 \lambda)} \right)^{N/2} = \frac{1}{(1 - \langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle)^N}.$$

\square

Lemma 8. Assume that \mathbb{X} is a matrix with i.i.d. standard normal entries and $\boldsymbol{\xi}$ is standard normal. Let $\delta > 0$. Let $\tau^2 = \frac{\bar{\rho}^2}{1+\bar{\rho}^2}$ as in (69) with $\bar{\rho} \leq r$, where r is defined in (68) with $A = \sqrt{\frac{1}{2} \log((1-\delta)^2 + 1)}$, and let μ_τ be the prior measure defined above in this subsection. Then for any test $\bar{\Delta}$ we have

$$\mathbf{P}_{\mathbf{0},1}(\bar{\Delta} = 1) + \mathbb{P}_{\mu_\tau}(\bar{\Delta} = 0) \geq \delta,$$

where $\mathbb{P}_{\mu_\tau}(\cdot) = \int \mathbf{P}_{\boldsymbol{\theta}, \sqrt{1-\tau^2}}(\cdot) \mu_\tau(d\boldsymbol{\theta})$.

Proof. Set $\sigma = \sqrt{1-\tau^2}$. It follows from Lemmas 6 and 7 in [6] that, for any test $\bar{\Delta}$,

$$\mathbf{P}_{\mathbf{0},1}(\bar{\Delta} = 1) + \mathbb{P}_{\mu_\tau}(\bar{\Delta} = 0) \geq 1 - \left(\mathbb{E}_{(\boldsymbol{\theta}, \boldsymbol{\theta}') \sim \mu_\tau^2} \left(\int \frac{d\mathbf{P}_{\boldsymbol{\theta},\sigma} d\mathbf{P}_{\boldsymbol{\theta}',\sigma}}{d\mathbf{P}_{\mathbf{0},1}} \right) - 1 \right)^{1/2}, \quad (72)$$

where $\mathbb{E}_{(\boldsymbol{\theta}, \boldsymbol{\theta}') \sim \mu_\tau^2}$ denotes the expectation with respect to the distribution of the pair $(\boldsymbol{\theta}, \boldsymbol{\theta}')$ such that $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ are independent and each of them is distributed according to μ_τ . As μ_τ is supported on $\Theta(s, \tau)$ we have $\|\boldsymbol{\theta}\|_2 = \tau$ for all $\boldsymbol{\theta}$ in the support of μ_τ . Since $\rho \leq r \leq A$ we have $|\langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle| \leq \tau^2 \leq A^2 \leq 1/2$ for all $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ in the support of μ_τ . Thus, using Lemma 7 and the fact that $1 - x \geq e^{-2x}$ for $0 < x < 1/2$ we find

$$\begin{aligned} \mathbb{E}_{(\boldsymbol{\theta}, \boldsymbol{\theta}') \sim \mu_\tau^2} \left(\int \frac{d\mathbf{P}_{\boldsymbol{\theta},\sigma} d\mathbf{P}_{\boldsymbol{\theta}',\sigma}}{d\mathbf{P}_{\mathbf{0},1}} \right) &\leq \mathbb{E}_{(\boldsymbol{\theta}, \boldsymbol{\theta}') \sim \mu_\tau^2} \exp(2N \langle \boldsymbol{\theta}, \boldsymbol{\theta}' \rangle) \\ &= \mathbb{E}_{(\boldsymbol{\theta}, \boldsymbol{\theta}') \sim \mu_\tau^2} \exp \left(2N \tau^2 s^{-1} \sum_{j=1}^p \mathbb{1}_{\boldsymbol{\theta}_j \neq 0} \mathbb{1}_{\boldsymbol{\theta}'_j \neq 0} \right). \end{aligned}$$

The last expectation involves a hypergeometric (p, s, s) random variable and it is bounded in the same way as in the proof of Theorem 4 in [6] (see also Lemma 1 in [8]), which yields

$$\mathbb{E}_{(\boldsymbol{\theta}, \boldsymbol{\theta}') \sim \mu_\tau^2} \left(\int \frac{d\mathbf{P}_{\boldsymbol{\theta},\sigma} d\mathbf{P}_{\boldsymbol{\theta}',\sigma}}{d\mathbf{P}_{\mathbf{0},1}} \right) \leq \exp(2A^2). \quad (73)$$

Finally, the bounds (72), (73) and the definition of A imply that

$$\mathbf{P}_{\mathbf{0},1}(\bar{\Delta} = 1) + \mathbb{P}_{\mu_\tau}(\bar{\Delta} = 0) \geq 1 - (\exp(2A^2) - 1)^{1/2} = \delta'.$$

□

12.2 Proof of Theorem 7

We use the following lemma. Its proof is simple and is therefore omitted.

Lemma 9. Let $\kappa > 0, \epsilon > 0$, and $z \geq 1$. Then for

$$q = \min(\max(\epsilon^2 z, \epsilon \kappa), \kappa^2) \quad (74)$$

we have

$$q = \begin{cases} \epsilon \kappa & \text{if } \kappa^2 \geq \epsilon^2 z^2, \\ \epsilon^2 z & \text{if } \epsilon^2 z < \kappa^2 \leq \epsilon^2 z^2, \\ \kappa^2 & \text{if } \kappa^2 < \epsilon^2 z. \end{cases}$$

To prove Theorem 7, we first note that, as in the proof of Theorem 6, it is enough to consider $s \leq \sqrt{p}$ since for $s > \sqrt{p}$ the rate of estimation q^* is of the same order as for $s = \lfloor \sqrt{p} \rfloor$. Next, note that for $s \leq \sqrt{p}$ we have $\log(1 + p/s^2) \geq \log(1 + \sqrt{p}/s)$, so that $q_* \leq \min(\epsilon^2 z + \epsilon \kappa, \kappa^2) \leq 2q$ where q is given by equation (74) with

$$\epsilon = \epsilon(\sigma) := \frac{\sigma}{\sqrt{N}}, \quad z = z(s) := \frac{\min(s \log(1 + p/s^2), N)}{\log 2},$$

and where κ is the bound on the l_2 norm of the parameter θ . We have $z \geq 1$ since $s \log(1 + p/s^2) \geq \log 2$ for $1 \leq s \leq \sqrt{p}$ - we can thus apply Lemma 9. We now prove Theorem 7 separately for the cases $\kappa^2 \geq \epsilon^2 z^2$, $\epsilon^2 z < \kappa^2 \leq \epsilon^2 z^2$ and $\kappa^2 < \epsilon^2 z$.

Case $\kappa^2 \geq \epsilon^2 z^2$. Let $c > 0$. Using Lemma 9 we obtain

$$\begin{aligned} \sup_{\substack{\theta \in B_0(s): \\ \|\theta\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\theta, \sigma}(|\hat{T} - Q(\theta)| \geq cq_*) &\geq \sup_{\substack{\theta \in B_0(s): \\ \|\theta\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\theta, \sigma}(|\hat{T} - Q(\theta)| \geq 2c\epsilon\kappa) \\ &\geq \sup_{\substack{\theta \in B_0(1): \\ \|\theta\|_2 \leq \kappa}} \mathbf{P}_{\theta, 1}(|\hat{T} - Q(\theta)| \geq \frac{2c\kappa}{\sqrt{N}}) \\ &\geq \max_{j=1,2} \mathbf{P}_{\theta_j, 1}(|\hat{T} - Q(\theta_j)| \geq \frac{2c\kappa}{\sqrt{N}}), \end{aligned}$$

where $\theta_1 = (\kappa, 0, \dots, 0)$ and $\theta_2 = (\kappa - \frac{c}{\sqrt{N}}, 0, \dots, 0)$.

Set now $\delta \in (0, 1)$. Using the above display and a standard lower bound for the maximum error of testing two hypotheses (cf., e.g., [23, Chapter 2]) we get that there exists $c_\delta > 0$ depending only on δ such that

$$\sup_{\substack{\theta \in B_0(s): \\ \|\theta\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\theta, \sigma}(|\hat{T} - Q(\theta)| \geq c_\delta q_*) \geq \max_{j=1,2} \mathbf{P}_{\theta_j, 1}(|\hat{T} - Q(\theta_j)| \geq \frac{2c_\delta \kappa}{\sqrt{N}}) \geq \delta,$$

Case $\epsilon^2 z < \kappa^2 \leq \epsilon^2 z^2$. Using again Lemma 9 we obtain $q_* \leq 2q \leq 2\epsilon^2 z$. Let $\delta > 0$. Recalling that $\epsilon = \epsilon(\sigma)$, $z = z(s)$ and setting $\tau^2 = \frac{r^2}{1+r^2}$, $\bar{\sigma} = \sqrt{1 - \tau^2}$, where r is defined in equation (68) with the value $A := A(\delta)$ as in Lemma 8, we have

$$\begin{aligned} \sup_{\substack{\theta \in B_0(s): \\ \|\theta\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\theta, \sigma}(|\hat{T} - Q(\theta)| \geq cq_*) &\geq \sup_{\sigma > 0} \sup_{\substack{\theta \in B_0(s): \\ \|\theta\|_2 \leq \epsilon(\sigma)\sqrt{z(s)}}} \mathbf{P}_{\theta, \sigma}(|\hat{T} - Q(\theta)| \geq 2c\epsilon^2(\sigma)z(s)) \\ &\geq \frac{1}{2} \left[\mathbf{P}_{\mathbf{0}, 1}(|\hat{T}| \geq 2c\epsilon^2(1)z(1)) + \sup_{\substack{\theta \in B_0(s): \\ \|\theta\|_2 \leq \epsilon(\bar{\sigma})\sqrt{z(s)}}} \mathbf{P}_{\theta, \bar{\sigma}}(|\hat{T} - Q(\theta)| \geq 2c\epsilon^2(\bar{\sigma})z(s)) \right] \\ &\geq \frac{1}{2} \left[\mathbf{P}_{\mathbf{0}, 1}(|\hat{T}| \geq r^2/4) + \sup_{\substack{\theta \in B_0(s): \\ \|\theta\|_2 \leq \epsilon(\bar{\sigma})\sqrt{z(s)}}} \mathbf{P}_{\theta, \bar{\sigma}}(|\hat{T} - Q(\theta)| \geq r^2/4) \right] \end{aligned}$$

if we choose $c := c_\delta > 0$ such that $c \leq A^2(\log 2)/8$. In the last line, we have used the relations

$$z(1) \leq z(s), \quad \forall 1 \leq s \leq \sqrt{p}, \quad \epsilon^2(\bar{\sigma}) = \frac{\bar{\sigma}^2}{N} = \frac{1}{(1+r^2)N} \leq \frac{1}{N} = \epsilon^2(1),$$

$$2c\epsilon^2(1)z(s) = \frac{2cz(s)}{N} = \frac{2c}{\log 2} \min\left(\frac{s \log(1 + p/s^2)}{N}, 1\right) = \frac{2cr^2}{A^2 \log 2}.$$

Next, note that $\{\boldsymbol{\theta} \in B_0(s) : \|\boldsymbol{\theta}\|_2 \leq \epsilon(\bar{\sigma})\sqrt{z(s)}\} \supset \{\boldsymbol{\theta} \in B_0(s) : \|\boldsymbol{\theta}\|_2 \leq \tau\}$. Indeed,

$$\tau^2 = \frac{r^2}{1+r^2} \leq \frac{r^2}{(1+r^2)A^2 \log 2} = \epsilon^2(\bar{\sigma})z(s),$$

since $A < 1$ (cf. definition of A in Lemma 8). Thus, recalling that $\Theta'(\tau)$ is the set of all s -sparse vectors with s components equal to τ/\sqrt{s} and all other components equal to 0 we get

$$\begin{aligned} & \sup_{\substack{\boldsymbol{\theta} \in B_0(s): \\ \|\boldsymbol{\theta}\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{T} - Q(\boldsymbol{\theta})| \geq cq_*) \geq \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(|\hat{T}| \geq r^2/4) + \sup_{\substack{\boldsymbol{\theta} \in B_0(s): \\ \|\boldsymbol{\theta}\|_2 \leq \tau}} \mathbf{P}_{\boldsymbol{\theta}, \bar{\sigma}}(|\hat{T} - Q(\boldsymbol{\theta})| \geq r^2/4)] \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(|\hat{T}| \geq r^2/4) + \sup_{\boldsymbol{\theta} \in \Theta'(\tau)} \mathbf{P}_{\boldsymbol{\theta}, \bar{\sigma}}(|\hat{T} - Q(\boldsymbol{\theta})| \geq r^2/4)] \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(\bar{\Delta} = 1) + \sup_{\boldsymbol{\theta} \in \Theta'(\tau)} \mathbf{P}_{\boldsymbol{\theta}, \sqrt{1-\tau^2}}(\bar{\Delta} = 0)] \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(\bar{\Delta} = 1) + \mathbb{P}_{\mu_\tau}(\bar{\Delta} = 0)] \end{aligned}$$

where $\bar{\Delta} = \mathbb{1}\{|\hat{T}| \geq r^2/4\}$ and the penultimate inequality is due to the fact that $Q(\boldsymbol{\theta}) = \frac{r^2}{1+r^2} \geq \frac{r^2}{2}$ for all $\boldsymbol{\theta} \in \Theta'(\tau)$. It follows from Lemma 8 that the minimax risk in the last display is greater than $\delta \in (0, 1)$.

Case $\kappa^2 < \epsilon^2 z$. From Lemma 9 we obtain $q_* \leq 2\kappa^2$. Consider separately the cases $\kappa \geq \tau$ and $\kappa < \tau$. If $\kappa \geq \tau$ we have

$$\begin{aligned} & \sup_{\substack{\boldsymbol{\theta} \in B_0(s): \\ \|\boldsymbol{\theta}\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{T} - Q(\boldsymbol{\theta})| \geq cq_*) \geq \sup_{\substack{\boldsymbol{\theta} \in B_0(s): \\ \|\boldsymbol{\theta}\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{T} - Q(\boldsymbol{\theta})| \geq 2c\epsilon^2(\sigma)z(s)) \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(|\hat{T}| \geq 2c\epsilon^2(1)z(1)) + \sup_{\substack{\boldsymbol{\theta} \in B_0(s): \\ \|\boldsymbol{\theta}\|_2 \leq \tau}} \mathbf{P}_{\boldsymbol{\theta}, \bar{\sigma}}(|\hat{T} - Q(\boldsymbol{\theta})| \geq 2c\epsilon^2(\bar{\sigma})z(s))] \end{aligned}$$

and we use the same argument as in the case $\epsilon^2 z < \kappa^2 \leq \epsilon^2 z^2$ to obtain the result. On the other hand, if $\kappa < \tau$ then for any $c < 1/4$ we get

$$\begin{aligned} & \sup_{\substack{\boldsymbol{\theta} \in B_0(s): \\ \|\boldsymbol{\theta}\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{T} - Q(\boldsymbol{\theta})| \geq cq_*) \geq \sup_{\substack{\boldsymbol{\theta} \in B_0(s): \\ \|\boldsymbol{\theta}\|_2 \leq \kappa}} \sup_{\sigma > 0} \mathbf{P}_{\boldsymbol{\theta}, \sigma}(|\hat{T} - Q(\boldsymbol{\theta})| \geq 2c\kappa^2) \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(|\hat{T}| \geq 2c\kappa^2) + \sup_{\substack{\boldsymbol{\theta} \in B_0(s): \\ \|\boldsymbol{\theta}\|_2 \leq \kappa}} \mathbf{P}_{\boldsymbol{\theta}, \sqrt{1-\kappa^2}}(|\hat{T} - Q(\boldsymbol{\theta})| \geq 2c\kappa^2)] \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(|\hat{T}| \geq 2c\kappa^2) + \sup_{\boldsymbol{\theta} \in \Theta'(\kappa)} \mathbf{P}_{\boldsymbol{\theta}, \sqrt{1-\kappa^2}}(|\hat{T} - Q(\boldsymbol{\theta})| \geq 2c\kappa^2)] \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(\tilde{\Delta} = 1) + \sup_{\boldsymbol{\theta} \in \Theta'(\kappa)} \mathbf{P}_{\boldsymbol{\theta}, \sqrt{1-\kappa^2}}(\tilde{\Delta} = 0)] \\ & \geq \frac{1}{2} [\mathbf{P}_{\mathbf{0}, 1}(\tilde{\Delta} = 1) + \mathbb{P}_{\mu_\kappa}(\tilde{\Delta} = 0)], \end{aligned}$$

where $\tilde{\Delta} = \mathbb{1}\{|\hat{T}| \geq \kappa^2/2\}$ and the penultimate inequality is due to the fact that $Q(\boldsymbol{\theta}) = \kappa^2$ for all $\boldsymbol{\theta} \in \Theta'(\kappa)$. The result now follows by applying Lemma 8 where we replace τ by κ and take $\bar{\rho} \in (0, r)$ such that $\kappa^2 = \frac{\bar{\rho}^2}{1+\bar{\rho}^2}$.

Acknowledgments

The work of O. Collier was supported by the French National Research Agency (ANR) under the grant Labex MME-DII (ANR-11-LBX-0023-01). The work of A.B. Tsybakov was supported by GENES and by ANR under the grant Labex Ecodec (ANR-11-LABEX-0047). The work of A. Carpentier is partially supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether grant MuSyAD (CA 1488/1-1), by the DFG - 314838170, GRK 2297 MathCoRe, by the DFG GRK 2433 DAEDALUS (384950143/GRK2433), by the DFG CRC 1294 'Data Assimilation', Project A03, and by the UFA-DFH through the French-German Doktorandenkolleg CDFA 01-18 and by the UFA-DFH through the French-German Doktorandenkolleg CDFA 01-18 and by the SFI Sachsen-Anhalt for the project RE-BCI.

References

- [1] Radosław Adamczak and Paweł Wolff. Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields*, 162(3-4):531–586, 2015.
- [2] Ery Arias-Castro, Emmanuel Candes, and Yaniv Plan. Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism. *Ann. Statist.*, 39(5):2533–2556, 2011.
- [3] Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.
- [4] Pierre Bellec, Guillaume Lécué, and Alexandre B. Tsybakov. Slope meets Lasso: Improved oracle bounds and optimality. *Ann. Statist.*, 46(6B):3603–3642, 2018.
- [5] Tony T. Cai and Mark G. Low. Nonquadratic estimators of a quadratic functional. *Ann. Statist.*, 33(6):2930–2956, 2005.
- [6] Alexandra Carpentier, Olivier Collier, Laëtitia Comminges, Alexandre B. Tsybakov, and Yuhao Wang. Minimax rate of testing in sparse linear regression. *Automation and Remote Control*, 80(10):1817–1834, 2019.
- [7] Alexandra Carpentier and Nicolas Verzelen. Optimal sparsity testing in linear regression model. *arXiv preprint arXiv:1901.08802 (version 1)*, 2019.
- [8] Olivier Collier, Laëtitia Comminges, and Alexandre B. Tsybakov. Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Statist.*, 45(3):923–958, 2017.
- [9] Olivier Collier, Laëtitia Comminges, Alexandre B. Tsybakov, and Nicolas Verzelen. Optimal adaptive estimation of linear functionals under sparsity. *Ann. Statist.*, 46(6A):3130–3150, 2016.
- [10] Laëtitia Comminges, Olivier Collier, Mohamed Ndaoud, and Alexandre B. Tsybakov. Adaptive robust estimation in sparse vector model. *arXiv preprint arXiv:1802.04230*, 2018.
- [11] Laëtitia Comminges and Arnak S. Dalalyan. Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. *Electronic J. of Statist.*, 7:146–190, 2013.
- [12] Alexis Derumigny. Improved bounds for square-root lasso and square-root slope. *Electronic J. of Statist.*, 12:741–766, 2017.

- [13] David L. Donoho and Jiashun Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994, 2004.
- [14] David L. Donoho and Michael Nussbaum. Minimax quadratic estimation of a quadratic functional. *J. Complexity*, 6(3):290–323, 1990.
- [15] Zijian Guo, Weinjie Wang, Cai Tony T., and Hengzhe Li. Optimal estimation of genetic relatedness in high-dimensional linear models. *J. Amer. Stat. Assoc.*, 114(525):358–369, 2019.
- [16] Yuri I. Ingster. Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Meth. Stat.*, 6:47–69, 1997.
- [17] Yuri I. Ingster and Irina A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lect. Notes Stat.* Springer-Verlag, New York, 2003.
- [18] Yuri I. Ingster, Alexandre B. Tsybakov, and Nicolas Verzelen. Detection boundary in sparse regression. *Electronic J. of Statist.*, 4:1476–1526, 2010.
- [19] Guillaume Lecué and Mendelson Shahar. Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc.*, 19(3):881–904, 2017.
- [20] Rajarshi Mukherjee and Subhabrata Sen. On minimax exponents of sparse testing. *arXiv preprint arXiv:2003.00570*, 2020.
- [21] Mark Rudelson and Roman Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *Inter. Math. Res. Not.*, (19):9594–9617, 2015.
- [22] Tingni Sun and Cun-Hui Zhang. Scaled sparse linear regression. *Biometrika*, 99(4):879–898, 2012.
- [23] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009.
- [24] Roman Vershynin. *High-Dimensional Probability*. Springer Series in Statistics. Springer, New York, 2018.
- [25] Nicolas Verzelen. Minimax risks for sparse regressions: ultra-high dimensional phenomena. *Electronic J. of Statistics*, 6:38–90, 2012.
- [26] Nicolas Verzelen and Elisabeth Gassiat. Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli*, 24(4B):3683–3710, 2018.