PRICE OF QUALITY: SUFFICIENT CONDITIONS FOR SPARSE RECOVERY USING MIXED-QUALITY DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

We study sparse recovery when observations come from mixed-quality sources: a small collection of high-quality measurements with small noise variance and a larger collection of lower-quality measurements with higher variance. For this heterogeneous-noise setting, we establish sample-size conditions for information-theoretic and algorithmic recovery. On the information-theoretic side, we show that (n_1, n_2) must satisfy a linear trade-off defining the *Price of Quality*: the number of low-quality samples needed to replace one high-quality sample. In the agnostic setting, where the decoder is completely agnostic to the quality of the data, it is uniformly bounded, and in particular one high-quality sample is never worth more than two low-quality samples. In the informed setting, where the decoder is informed of per-sample variances, the price of quality can grow arbitrarily large. On the algorithmic side, we analyze the LASSO in the agnostic setting and show that the recovery threshold matches the homogeneous-noise case and only depends on the average noise level, revealing a striking robustness of computational recovery to data heterogeneity. Together, these results give the first conditions for sparse recovery with mixed-quality data and expose a fundamental difference between how the information-theoretic and algorithmic thresholds adapt to changes in data quality.

1 INTRODUCTION

1.1 OVERVIEW AND PREVIOUS WORK

1.1.1 SPARSE RECOVERY

Sparse recovery is a central problem in high-dimensional statistics and machine learning. Its applications include compressive sensing (Foucart et al., 2013; Candès et al., 2006; Donoho, 2006), signal denoising (Chen et al., 2001), sparse regression (Miller, 2002), data-stream algorithms (Cormode & Hadjieleftheriou, 2009; Indyk, 2007; Muthukrishnan et al., 2005), and combinatorial group testing (Du & Hwang, 1999). Other applications range from medical imaging to communications and compression (Foucart et al., 2013, Chap. 1).

We formulate the problem as follows. A high-dimensional *signal* $\beta^\star \in \mathbb{R}^p$ (also called *model* or *ground truth*), unknown but a-priori s-sparse, is transmitted through a noisy channel that projects it onto a collection of n random vectors $\{x_i\}_{i\in[n]}$ in $\mathbb{R}^p$. This is expressed as:

$$Y \coloneqq X\beta^\star + Z, \tag{1}$$

where $X = (x_1, \dots, x_n)^T$ is called *measurements*, *design* or *features*; Y *observations*, *annotations* or *labels*; and Z *noise*. On the other end of the channel, a decoder who observes (X, Y) is interested in recovering the support of the original signal $\beta^\star$, i.e. the subset $S^\star \coloneqq \{i \in [p] : \beta^\star_i \neq 0\} \subseteq [p]$, known a-priori to be of cardinality s. How many observations n (as a function of p and s) does the decoder need to recover the support of the signal as the dimension of the problem grows to infinity?

Previous works have shown that the sparse recovery problem exhibits two phase transitions at two thresholds, one *information-theoretic* and one *algorithmic*:

$$n_{\text{INF}} = \frac{2s \log(p/s)}{\log s} \quad \text{and} \quad n_{\text{ALG}} = 2s \log(p - s) + s + 1, \tag{2}$$

leading to three regimes:

- $n < n_{\text{INF}}$: Signal support recovery is information-theoretically impossible. (Reeves et al., 2019).
- $n_{\text{INF}} < n < n_{\text{ALG}}$: The maximum likelihood estimator (MLE) recovers $S^\star$. However, it is believed that no algorithm can do it in polynomial time since the problem exhibits an Overlap Gap Property (OGP) (Gamarnik & Zadik, 2022).
- $n > n_{\text{ALG}}$: The ℓ_1-regularized least-squares estimator (also known as the LASSO (Tibshirani, 1996)) recovers $S^\star$ (Wainwright, 2009).

Of particular interest is the signal-to-noise ratio (SNR), known to be an important quantity for characterizing the difficulty of sparse recovery problems (Wang et al., 2010; Reeves et al., 2019; Chaabouni & Gamarnik, 2025). It's defined as follows:

$$\text{SNR} := \frac{\mathbb{E}\|X\beta^\star\|_2^2}{\mathbb{E}\|Z\|_2^2}. \tag{3}$$

1.1.2 MIXED QUALITY DATA

A recent body of work has explored how low-quality data, e.g. labeled by an LLM or weak annotator (Ratner et al., 2017; Frénay & Verleysen, 2013), should be combined with fewer but higher-quality data, e.g. labeled by humans or experts, for prediction and inference tasks (Gligorić et al., 2024; Li et al., 2023; Zhang et al., 2023; Egami et al., 2023).

In this paper, we formalize the mixed-quality data setting for sparse signal recovery: the decoder has access to n_1 noisy projections of the signal $\beta^\star$ with a small noise level $\sigma_1^2 > 0$ that we denote $\{(y_i, x_i)\}_{i=1}^{n_1}$ and call *high-quality* data. In addition, the decoder also observes a larger set of $n_2 > n_1$ noisy projections of the same signal $\beta^\star$, but with a higher noise level $\sigma_2^2 > \sigma_1^2$, that we denote $\{(y_i, x_i)\}_{i=n_1+1}^{n_2}$ and call *low-quality* data. We distinguish two settings:

- **Agnostic setting**: The decoder lacks access to observation-level noise variances and treats all measurements as if drawn from a single homogeneous model. This occurs when heterogeneous data sources lose provenance: for example in web-scale text corpora (Ratner et al., 2017; Frénay & Verleysen, 2013) or citizen-science campaigns lacking sensor calibration (Silvertown, 2009). The decoder simply applies standard sparse-recovery methods without noise estimation or reweighting.
- **Informed setting**: where the decoder has access to the per-sample noise variance of the data. This regime captures situations where provenance information accompanies each observation, so the decoder knows which measurements are high- or low-quality. Examples include multi-site clinical trials or sensor networks that log calibration statistics (Loh & Wainwright, 2011; Delaigle et al., 2008), and medical-imaging datasets with per-rater confidence scores (Rajpurkar et al., 2018).

1.2 OUR WORK

In this paper, we consider the sparse recovery problem described above (1). Specifically, we study the setting where the measurements are drawn i.i.d. from a standard normal Gaussian distribution, and the noise is unbiased and drawn independently from Gaussian distributions of variance σ_1^2 for the high-quality samples and $\sigma_2^2 > \sigma_1^2$ for the low-quality ones:

$$\{X_{ij}\}_{i\in[n], j\in[p]} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \text{ and } Z = \Sigma W; \text{ where } \Sigma = \begin{pmatrix} \sigma_1 I_{n_1} & 0 \\ 0 & \sigma_2 I_{n_2} \end{pmatrix}, W \sim \mathcal{N}(0, I_n). \tag{4}$$

Although much of the literature on sparse recovery in the homogeneous noise setting assumes constant noise level σ^2, we don't assume in this work that σ_1^2 and σ_2^2 are constant. In fact, the reason previous work can assume constant noise variance without loss of generality is that the model (1) could be scaled down by σ when the noise is homogeneous with variance σ^2 to make it constant. However, it is not the case anymore when the noise is heterogeneous.

Since data come from two different sources, we define in addition to (3) two signal-to-noise ratios: SNR_1 for high-quality observations and SNR_2 corresponding to low-quality observations.

We are interested in the two following questions:

- **Sampling complexity of sparse recovery:** How large do the sample sizes (n_1, n_2) need to be for the decoder to be able, information-theoretically, to recover the support of the signal?
- **Algorithmic recovery:** How large do the sample sizes (n_1, n_2) need to be for the decoder to be able to recover the support of the signal using a polynomial-time algorithm?

We summarize below our findings on each of these questions in the agnostic and informed settings.

1.2.1 SAMPLING COMPLEXITY OF SPARSE RECOVERY

In the first part of work (section 3), we focus on the question of sampling complexity. For simplicity, we assume the signal is binary, i.e. $\beta^\star \in \{0,1\}^p$. Note that in this case, recovering the support is equivalent to recovering the signal. This assumption is very common in the literature (Aeron et al., 2010; Reeves et al., 2019; Gamarnik & Zadik, 2022; Chaabouni & Gamarnik, 2025). Intuitively, detecting a component of size 1 is at least as hard as detecting a stronger component, so the resulting thresholds are representative of signals with non-zero entries bounded away from zero. We discuss this assumption in more detail in Remark 3.1.

Our main results, Theorem 1 for the agnostic setting and Theorem 2 for the informed one, each provide a sufficient condition (9, 15) on the sample sizes (n_1, n_2) for support recovery. In both results, the condition requires that a linear combination of has the form $\alpha_1 n_1 + \alpha_2 n_2 > n^\star$, for some coefficients $\alpha_1, \alpha_2 > 0$ depending on σ_1^2, σ_2^2 and s, and having different expressions in the agnostic and informed settings. In particular, we note that if (n_1, n_2) verify this condition (i.e. are together large enough), then so do $(n_1 - 1, n_2 + \alpha_1/\alpha_2)$. In this sense, we say that 1 unit of high-quality data is worth:

$$\gamma\left(s, \sigma_1^2, \sigma_2^2\right) := \frac{\alpha_1}{\alpha_2} \tag{5}$$

units of low-quality data. We label γ the *Price of Quality* and study its behavior in the agnostic and informed settings and for different regimes of SNR_1 and SNR_2. In the agnostic setting, it is uniformly bounded. In particular, one high-quality sample is never worth more than *two* low-quality samples (12, 13). In the informed setting, where the decoder is informed of per-sample variances, the price of quality goes to infinity in the low SNR_2 & high SNR_1 regime (18), and can be arbitrarily large in both low and high SNR regimes (17, 19).

1.2.2 ALGORITHMIC RECOVERY

In the second part of our work (section 4), we focus on the question of algorithmic recovery. Unlike for sampling complexity, we don't assume that the signal is binary, but still require non-zero components to be bounded away from zero, i.e. there exists $\rho > 0$ such that $\min_{i \in S^\star} |\beta_i^\star| \geq \rho$. This is standard in the literature (Aeron et al., 2010; Ndaoud & Tsybakov, 2020; Wang et al., 2010) since we can't hope to detect non-zero signal components if they can have arbitrarily small amplitude.

Specifically, we study the question of *signed support* recovery, that is, recovering not only the indices of the non-zero components of the signal but also their sign ($+$ or $-$). This is usual in the algorithmic sparse recovery literature (Wainwright, 2009; Wang et al., 2010; Omidiran & Wainwright, 2008), as it follows naturally from the standard proof techniques.

Our main result, Theorem 3, provides necessary and sufficient conditions for the ℓ_1-regularized least-squares estimator (known as the LASSO) to recover the signed support of $\beta^\star$ in the agnostic setting. Our result reveals that the problem behaves like the homogeneous-noise setting (Wainwright, 2009) with a homogeneous noise level equal to the average noise level of Z:

$$\sigma_{\text{avg}}^2 := \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n}. \tag{6}$$

In particular, the sample size conditions (22, 23) do not depend on the noise levels σ_1^2, σ_2^2. The condition on the LASSO regularization parameter (24) only depends on σ_1^2 and σ_2^2 through σ_{avg}^2 and is the same as the one for homogeneous noise σ_{avg}^2 (see equation (28) in Wainwright (2009)). We further provide a necessary and sufficient condition on noise scaling (Proposition 4.1).

This shows that, unlike in sampling complexity, high-quality and low-quality data contribute equally to the sample size condition under which the LASSO recovers the support of the signal.

Although we don't address algorithmic recovery in the informed setting, we briefly discuss it in Remark 4.2, where we discuss why the proof of Theorem 3 cannot be easily extended to the informed case.

1.3 Contributions, Outline and Notations

To the best of our knowledge, this paper is the first to:

1. Provide a sufficient condition for sparse recovery in the heterogeneous noise case, and quantify the trade-off between high-quality and low-quality data in the agnostic and informed settings.
2. Extend necessary and sufficient conditions for Lasso sparse recovery to the heterogeneous-noise, agnostic setting and show that high-quality and low-quality data contribute equally to reaching the algorithmic threshold.

We organize the rest of the paper as follows. Section 2 introduces the problem setup. Section 3 studies the sampling complexity of sparse recovery under heterogeneous noise. Section 4 investigates algorithmic recovery using the Lasso. Section 5 concludes and outlines directions for future work.

Throughout this document, we will use the following notations:

- We say that $f(x) \simeq g(x)$ as $x \to a \in \mathbb{R} \cup \{-\infty, +\infty\}$ if and only if $f(x) = g(x)(1 + o(1))$.
- We denote by $h(\cdot)$ the binary entropy: $h(x) = -x\log x - (1-x)\log(1-x)$, $x \in (0,1)$.
- We call ℓ_0-*norm* the number of non-zero coordinates of $x \in \mathbb{R}^d$, that is $\|x\|_0 := \sum_{i=1}^d \mathbb{1}(x_i \neq 0)$.

2 Preliminaries

The problem of sparse signal recovery is defined above (1). The decoder a-priori knows that $\beta^\star$ is s-sparse and belongs to a known set $\mathcal{A} \subseteq \mathbb{R}^p$. The design and noise are random with $(X_{ij})_{i\in[n],j\in[p]} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $Z := \Sigma W$ with Σ and W defined as in (4) and $n_1 + n_2 = n$. The signal-to-noise ratio (3) writes:

$$\mathrm{SNR} := \frac{\mathbb{E}\|X\beta\|_2^2}{\mathbb{E}\|Z\|_2^2} = \frac{ns}{n_1\sigma_1^2 + n_2\sigma_2^2} = \frac{s}{\sigma_{\text{avg}}^2}, \tag{7}$$

where σ_{avg}^2 denotes the average noise level (6). In addition, we define the *high-quality SNR* and the *low-quality SNR* respectively by:

$$\mathrm{SNR}_1 := \frac{\mathbb{E}\left\|\left[y_i - x_i^T\beta^\star\right]_{i=1}^{n_1}\right\|_2^2}{\mathbb{E}\|Z^1\|_2^2} = \frac{s}{\sigma_1^2}\ ,\ \mathrm{SNR}_2 := \frac{\mathbb{E}\left\|\left[y_i - x_i^T\beta^\star\right]_{i=n_1+1}^{n_2}\right\|_2^2}{\mathbb{E}\|Z^2\|_2^2} = \frac{s}{\sigma_2^2}.$$

In particular, we always have $\mathrm{SNR}_2 < \mathrm{SNR}_1$, which reveals three regimes of interest:

- High SNR: when $\mathrm{SNR}_1, \mathrm{SNR}_2 \to +\infty$, or equivalently $\sigma_2^2 = o(s)$.
- Low SNR_2, High SNR_1: $\mathrm{SNR}_2 \to 0$, $\mathrm{SNR}_1 \to +\infty$ or equivalently $\sigma_2^2 = \omega(s)$, $\sigma_1^2 = o(s)$.
- Low SNR: when $\mathrm{SNR}_1, \mathrm{SNR}_2 \to 0$, or equivalently $\sigma_1^2 = \omega(s)$.

3 Sampling complexity of sparse recovery

In this section, we are interested in determining whether it is possible, information-theoretically, to recover the support of the signal, depending on the sample size n. We assume that $\beta^\star$ is binary and a priori s-sparse, that is: $\mathcal{A} := \mathcal{B}_{p,s} = \{\beta \in \{0,1\}^p : \|\beta\|_0 = s\}$.

Remark 3.1 (Binary-signal assumption)**.** Our results for sparse recovery can be viewed as applying to signals whose non-zero components are at least 1 in magnitude, i.e. $\beta^\star \in \mathcal{C}_{p,s}(1) := \left\{\beta \in \mathbb{R}^d : \min_{i\in\mathrm{Supp}(\beta)} |\beta_i| \geq 1\right\}$. Assuming that the non-zero entries are exactly equal to 1 serves only to simplify computations. Intuitively, detecting a component of magnitude 1 is at least as hard as detecting a stronger component, so stronger signals can only make recovery easier. Conversely, detecting a signal in $\mathcal{C}_{p,s}(1)$ is at least as hard as detecting a binary signal, since $\{0,1\}^p \subseteq \mathcal{C}_{p,s}(1)$. More generally, recovering any signal whose non-zero entries are bounded below by some $\rho > 0$ can be reduced to the case of $\mathcal{C}_{p,s}(1)$ by rescaling the model (1) by ρ.

Let $A \triangle B := (A \cup B) \setminus (A \cap B)$ denote the symmetric difference between any two finite sets A and B, and $\mathrm{Supp}(\beta) := \{i \in [p] \colon \beta_i \neq 0\}$ denote the support of any vector $\beta \in \mathbb{R}^p$. Let $\delta \in (0,1)$. We say that $\hat{\beta} \in \mathcal{B}_{p,s}$ *recovers the support of* $\beta^\star$ *up to error* δ if $\left|\mathrm{Supp}\left(\hat{\beta}\right) \triangle \mathrm{Supp}(\beta^\star)\right| < 2\delta s$.

3.1 Agnostic setting

In the agnostic setting where the decoder ignores the quality of each observation, the sample sizes (n_1, n_2) and the noise levels $\left(\sigma_1^2, \sigma_2^2\right)$, we define the estimator:

$$\hat{\beta} := \underset{\beta \in \mathcal{B}_{p,s}}{\arg\min} \|Y - X\beta\|_2^2 . \tag{8}$$

Theorem 1 (Sufficient condition for support recovery in the agnostic setting)**.**

1. *Assume $s = o(p)$ and $s \to +\infty$ as $p \to +\infty$. Then let $n^\star := 2s\log(p/s)$.*
2. *Assume $s = \alpha p$ for some constant $\alpha \in (0,1)$. Then let $n^\star := 2h(\alpha)\, p$.*

In both settings described above, if there exists $\varepsilon > 0$ such that:

$$n_1 \log\left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^4}\right) + n_2 \log\left(1 + \frac{\delta s}{2\sigma_2^2}\right) \geq (1+\varepsilon)\, n^\star, \tag{9}$$

then $\hat{\beta}$ recovers the support of $\beta^\star$ up to error δ w.h.p.:

$$\mathbb{P}\left(\left|\mathit{Supp}(\beta^\star) \triangle \mathit{Supp}\left(\hat{\beta}\right)\right| < 2\delta s\right) \geq 1 - \exp\left\{-\left(\varepsilon + o(1)\right) n^\star/2\right\} \overset{p\to+\infty}{\longrightarrow} 1.$$

Proof Sketch. The proof of Theorem 1 is in appendix A and uses standard techniques. We control the probability that a high-error support attains a lower objective value in (8) than the ground truth and then take a union bound over such supports. For any β, we have:

$$\|Y - X\beta\|_2^2 - \|Y - X\beta^\star\|_2^2 = \sum_{i=1}^{n} \left\{\langle X_i, \beta^\star - \beta\rangle^2 + 2Z_i\langle X_i, \beta^\star - \beta\rangle\right\}. \tag{10}$$

Applying a Chernoff bound to the LHS above and analyzing the MGF of the summands yields an exponent that factorizes across two blocks (see Proposition A.1). We conclude using a union bound over supports S with $|S \triangle S^\star| \geq 2\delta s$ (there are at most $\binom{p}{s}$ of them). □

We interpret Theorem 1 as follows.

- **Price of Quality.** The sufficient condition for recovery (9) is equivalent to a linear combination of the sample size n_1 and n_2 being larger than the threshold $n^\star$. The coefficients of the sample sizes reveal that 1 unit of high-quality data is worth:

$$\gamma := \frac{\log\left(1 + \delta\left(2\sigma_2^2 - \sigma_1^2\right)s/\left(2\sigma_4^2\right)\right)}{\log\left(1 + \delta s/\left(2\sigma_2^2\right)\right)} > 1 \tag{11}$$

  units of low-quality data. We call γ the *Price of Quality*. In fact, one unit of high-quality data can be replaced by γ units of low-quality data: that is, if (n_1, n_2) are sufficient for recovering $\beta^\star$, then so are $(n_1 - 1, n_2 + \gamma)$.

- **High SNR_2 regime.** Assume $s = \omega\left(\sigma_2^2\right)$. The price of quality (11) writes:

$$\gamma \simeq \frac{\log\left(\delta s/\left(2\sigma_2^2\right)\right) + \log\left(2 - \sigma_1^2/\sigma_2^2\right)}{\log\left(\delta s/\left(2\sigma_2^2\right)\right)} \simeq 1, \tag{12}$$

  which means that when $\sigma_1^2, \sigma_2^2 = o(s)$, the high-quality and low-quality data contribute equally to the recovery condition (9).

- **Low SNR$_2$ regime.** Assume $s = o\left(\sigma_2^2\right)$. The price of quality (11) writes:

$$\gamma \simeq \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right) s / \left(2\sigma_4^2\right)}{\delta s / \left(2\sigma_2^2\right)} \simeq 2 - \frac{\sigma_1^2}{\sigma_2^2}. \tag{13}$$

Note that $\gamma < 2$ for any σ_1^2, σ_2^2. We conclude that in the low SNR regime, one unit of high-quality data is worth at most 2 units of low-quality data, regardless of the noise ratio.

Remark 3.2 (Limitations)**.**

- Unlike results on sufficient conditions for sparse recovery in the homogeneous-noise setting (Gamarnik & Zadik, 2022), Theorem 1 might not be tight. In fact, Wang et al. (2010) showed that in the homogeneous-noise setting, the sufficient condition in (Gamarnik & Zadik, 2022) is also necessary. However, we believe it might not be the case for Theorem 1 since, unlike the results in the homogeneous-noise setting, the sufficient condition (9) was not obtained using the *best Chernoff bound* (see Remark A.1) due to the complexity of the minimization of the Chernoff upper bound, which we reduce to finding the root of a third degree polynomial (33).

- Even under the assumption that the decoder is agnostic to the quality of the data, the estimator $\hat{\beta}$ (8), might not constitute the best approach to recover the support of $\beta^\star$. For instance, especially in the low SNR regime, the decoder might re-weight the loss of each observation by the magnitude of its observed label, i.e.:

$$\arg\min_{\beta \in \mathcal{B}_{p,s}} \sum_{i=1}^{n} \frac{1}{Y_i^2} \Big(Y_i - \langle x_i, \beta \rangle\Big)^2,$$

as an attempt to rescale each row of data by its noise level. In fact, in the low SNR regime we have $\mathbb{E}Y_i^2 \simeq \sigma_i^2$ where σ_i^2 denotes the noise level corresponding to the i^{th} observation, which motivates the use of Y_i^2 as a proxy for σ_i^2 when the noise levels are unknown.

3.2 Informed setting

In this section, we assume that the decoder knows the distribution of each noise entry: $\mathcal{N}\left(0, \sigma_1^2\right)$ or $\mathcal{N}\left(0, \sigma_2^2\right)$. Recall the distributions of Z and W from (4). The MLE is (see appendix B for a proof):

$$\hat{\beta}_{\text{MLE}} = \arg\min_{\beta \in \mathcal{B}_{p,s}} \left\|\Sigma^{-1}\left(Y - X\beta\right)\right\|_2^2. \tag{14}$$

Theorem 2 (Sufficient condition for support recovery in the informed setting)**.**

1. Assume $s = o(p)$ and $s \to +\infty$ as $p \to +\infty$. Then let $n^\star := 2s\log(p/s)$.

2. Assume $s = \alpha p$ for some constant $\alpha \in (0,1)$. Then let $n^\star := 2h(\alpha)\,p$.

In both settings described above, if there exists $\varepsilon > 0$ such that:

$$n_1 \log\left(1 + \frac{\delta s}{2\sigma_1^2}\right) + n_2 \log\left(1 + \frac{\delta s}{2\sigma_2^2}\right) \geq (1+\varepsilon)\, n^\star, \tag{15}$$

then $\hat{\beta}_{MLE}$ recovers the support of $\beta^\star$ up to error δ w.h.p.:

$$\mathbb{P}\left(\left|\mathit{Supp}\left(\beta^\star\right) \triangle \mathit{Supp}\left(\hat{\beta}_{MLE}\right)\right| < 2\delta s\right) \geq 1 - \exp\left\{-\left(\varepsilon + o(1)\right) n^\star/2\right\} \overset{p\to+\infty}{\longrightarrow} 1.$$

Proof Sketch. The proof of Theorem 2 is given in appendix C and follows a similar argument as Theorem 1. Here, the rescaled loss in (14) leads to a Chernoff bound that can be optimized in closed-form, yielding a sharp convergence rate. □

We interpret Theorem 2 as follows.

- **Price of Quality.** In the informed setting, the expression of the price of quality is different from the one in the agnostic case (11). It writes:

$$\gamma = \log\left(1 + \frac{\delta s}{2\sigma_1^2}\right) \Big/ \log\left(1 + \frac{\delta s}{2\sigma_2^2}\right). \tag{16}$$

- **Low SNR regime.** Assume $\sigma_1^2 = \omega(s)$. Then:

$$\gamma \simeq \sigma_2^2/\sigma_1^2. \tag{17}$$

- **Low SNR$_2$, High SNR$_1$ regime.** Assume $\sigma_2^2 = \omega(s)$ and $\sigma_1^2 = o(s)$. Then:

$$\gamma = \Theta\left(\frac{\log\left(s/\sigma_1^2\right)}{s/\sigma_2^2}\right) = \Theta\left(\frac{\log \mathrm{SNR}_1}{\mathrm{SNR}_2}\right) \overset{p\to+\infty}{\longrightarrow} +\infty. \tag{18}$$

- **High SNR regime.** Assume $\sigma_2^2 = o(s)$. Then:

$$\gamma \simeq \log\left(s/\sigma_1^2\right)/\log\left(s/\sigma_2^2\right) = \log \mathrm{SNR}_1/\log \mathrm{SNR}_2. \tag{19}$$

Remark 3.3.

- Compared to the agnostic setting (Theorem 1), the appropriate rescaling of the loss in the MLE (14) constitutes a better use of the high-quality data, in the sense that it leads to a higher price of quality γ. In particular, γ is infinite in the low SNR$_2$ & high SNR$_1$ setting (18) and can be arbitrarily large in both low and high SNR regimes (17, 19).
- We believe that the sufficient condition for recovery in the informed setting (15) is also necessary (and therefore tight), as it was obtained using the tightest Chernoff bound (see equations (35) and (38) in the proof of Theorem 2), which has been shown to yield a necessary and sufficient condition in the homogeneous noise setting in different sparse recovery settings (Gamarnik & Zadik, 2022; Wang et al., 2010; Chaabouni & Gamarnik, 2025).

4 ALGORITHMIC RECOVERY

In this section, we are interested in the existence of a tractable algorithm to recover the support of the underlying signal. We assume that the components of the signal $\beta^\star$ take real values and are bounded away from zero: that is $\mathcal{A} \coloneqq \mathcal{C}_{p,s}(\rho) = \left\{\beta \in \mathbb{R}^p \colon \|\beta\|_0 = s,\ \min_{i\in\mathrm{Supp}(\beta)} |\beta_i| \geq \rho\right\}$, for some $\rho \in \mathbb{R}_+$. We say that $\hat\beta \in \mathbb{R}^p$ recovers the signed support of $\beta^\star$ if $\mathrm{sign}(\hat\beta) = \mathrm{sign}(\beta^\star)$, where the $\mathrm{sign}\colon \mathbb{R} \longrightarrow \{-1, 0, 1\}$ function is defined by $\mathrm{sign}(0) = 0$ and $\mathrm{sign}(x) = x/|x|$ for all $x \neq 0$, and is applied coordinate-wise. A common approach to recovering the signed support of the signal is using the solution to the following ℓ_1-constrained quadratic program, also known as the LASSO:

$$\mathcal{B}_{\mathrm{Lasso}} \coloneqq \min_{\beta\in\mathbb{R}^p}\left\{\frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda_p\|\beta\|_1\right\}, \tag{20}$$

where $\lambda_p \geq 0$ denotes a sequence of regularization parameters converging to 0 as $p \to +\infty$. We are interested in characterizing the regime where the LASSO recovers the signed support of the true signal. Specifically, we call "recovery" the event:

$$\mathcal{R}(X, \beta^\star, Z, \lambda_p) \coloneqq \left\{\exists\ \hat\beta \in \mathcal{B}_{\mathrm{Lasso}} \colon\ \mathrm{sign}\left(\hat\beta\right) = \mathrm{sign}(\beta^\star)\right\}. \tag{21}$$

In the homogeneous noise setting, Wainwright (2009) showed that the performance of the LASSO in estimating the signed support of $\beta^\star$ exhibits a phase transition with respect to the sample size. In fact, there exists a threshold n_{ALG} such that:

- If $n > n_{\mathrm{ALG}}$: then the LASSO correctly recovers the signed support of $\beta^\star$.
- If $n < n_{\mathrm{ALG}}$: then the LASSO fails to recover the signed support of $\beta^\star$.

In addition, it is widely believed that no algorithm can recover the support of $\beta^\star$ in polynomial time when $n < n_{\mathrm{ALG}}$. Indeed, Gamarnik & Zadik (2022) showed that the problem exhibits an OGP. This motivates the use of (20) to estimate $\beta^\star$ in the agnostic setting where the decoder treats the data impartially. Our main result of this section, Theorem 3, extends the result mentioned above on the LASSO threshold (by Wainwright (2009)) to the heterogeneous, agnostic noise setting.

Theorem 3 (Lasso recovery phase transition)**.** *Assume that, as $p \to +\infty$; s goes to infinity, $s = o(p)$ and $n_1, n_2 = \omega(s)$. Let $n_{ALG} \coloneqq 2s\log(p-s) + s + 1$.*

i. If there exists $\varepsilon > 0$ *such that:*

$$n < (1-\varepsilon)\, n_{ALG}, \tag{22}$$

then, for any sequence $\lambda_p > 0$ *such that* $\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{\lambda_p^2 n^2}$ *has a limit in* $\mathbb{R}_{\geq 0} \cup \{+\infty\}$*, we have* $\mathbb{P}_{X,Z}\Big(\mathcal{R}\left(X, \beta^\star, Z, \lambda_p\right)\Big) \to 0$.

ii. If there exists $\varepsilon > 0$ *such that:*

$$n > (1+\varepsilon)\, n_{ALG}, \tag{23}$$

and $\left(\lambda_p\right)_{p\geq 1} \to 0$ *is chosen such that:*

$$\frac{n\lambda_p^2}{\sigma_{avg}^2 \log\left(p-s\right)} \to +\infty, \quad \textit{and} \quad \frac{1}{\rho}\left[\lambda_p\sqrt{s} + \sqrt{\frac{\sigma_{avg}^2 \log s}{n}}\right] \to 0, \tag{24}$$

then $\mathbb{P}_{X,Z}\Big(\mathcal{R}\left(X, \beta^\star, Z, \lambda_p\right)\Big) \to 1$.

The full proof of Theorem 3 is given in Appendix D and follows the core LASSO threshold argument of Wainwright (2009). We use the same argument but generalize it to the heterogeneous-noise setting, where the presence of the matrix Σ, no longer a scalar multiple of the identity, causes key steps of the classical proof to fail. We overcome this by applying a Gram–Schmidt (QR) decomposition of X_S (45) and analyzing the resulting orthogonal matrix using properties of the Haar measure on the orthogonal group (e.g. see Lemma D.6). The monograph of Meckes (2019) on Haar-distributed matrices was particularly valuable in understanding this component from random-matrix theory.

Proof Sketch of Theorem 3. We express the recovery property (21) via the first-order optimality conditions of the LASSO (20):

$$\mathcal{R}\left(X, \beta^\star, \Sigma w, \lambda_p\right) \Longleftrightarrow \begin{cases} \left|\left(\frac{1}{n}X_S^T X_S\right)^{-1}\left(\frac{1}{n}X_S^T \Sigma w - \lambda_p \ \mathrm{sign}\left(\beta_S^\star\right)\right)\right| < |\beta_S^\star| \\ \left|X_{S^c}^T X_S \left(X_S^T X_S\right)^{-1}\left(\frac{1}{n}X_S^T \Sigma w - \lambda_p\, \mathrm{sign}\left(\beta_S^\star\right)\right) - \frac{1}{n}X_{S^c}^T \Sigma w\right| \leq \lambda_p \end{cases} \tag{25}$$

where absolute values and inequalities are taken component-wise. This well-known result (Wainwright, 2009; Fuchs, 2004; Meinshausen & Bühlmann, 2006; Tropp, 2006; Zhao & Yu, 2006) is stated in Proposition $D.1$. When (23) and (24) hold, the random variables inside the absolute values on the RHS of (25) concentrate below their respective upper bounds, establishing sufficiency. When (22) holds, the second absolute value in (25) concentrates above λ_p, showing necessity. □

Although Theorem 3 does not explicitly state any condition on the scaling on the noise, the existence of $\lambda_p \to 0$ such that (24) holds requires that the noise does not scale arbitrarily large. The next result explicitly states this condition.

Proposition 4.1 (Necessary and sufficient condition on noise scaling)**.** *If there exists* $\left(\lambda_p\right)_{p\geq 1} \to 0$ *such that (24) holds, then:*

$$\sigma_{avg}^2 = o\left(\frac{n}{\left(1 + s/\rho^2\right)\log\left(p-s\right)}\right). \tag{26}$$

Conversely, if (26) holds, let:

$$\lambda_p := \left(\frac{\sigma_{avg}^2 \log\left(p-s\right)}{\left(1+s/\rho^2\right) n}\right)^{1/4}. \tag{27}$$

Then $\lambda_p \to 0$ *and (24) holds.*

Proof. See appendix E. □

Remark 4.1 (Correlated features)**.** Although we state Theorem 3 only for independent features, i.e. $x_i \sim \mathcal{N}\left(0, I_p\right)$ for all $i \in [n]$, we believe a similar result should hold when the features are correlated under suitable regularity conditions on their covariance matrix. Such a result was proved by Wainwright (2009) in the homogeneous-noise setting. However, the proof of the extension to heterogeneous noise is already technically heavy, and extending it to correlated designs would introduce even more complexity and further reduce the readability of an already extensive proof. Because our focus is on the new heterogeneous-noise aspect rather than on re-deriving the covariance-matrix-dependent bound, we focus our analysis in this paper on the independent-feature case.

Remark 4.2 (Informed setting)**.** Although we only show the phase transition in the scope of the LASSO in the agnostic setting, we believe that a "rescaled" version of the LASSO where the loss is defined by $\left\|\Sigma^{-1}\left(Y-X\beta\right)\right\|_2^2$ instead of $\|Y-X\beta\|_2^2$ in (20) would exhibit a similar behavior in the informed setting. However, using the same argument as Theorem 3 and Wainwright (2009) is tricky in the rescaled setting, as it introduces Σ^{-1} factors alongside the X terms in (25), which breaks down not only the original proof of Wainwright (2009) but also our Gram-Schmidt argument. In fact, rescaling the rows of X invalidates the property $X_S^T X_S \sim \mathcal{W}\left(I_s, n\right)$, which both arguments use to compute the moments of its inverse using properties of inverse Wishart matrices (Anderson et al., 1958; Siskind, 1972). We believe this can be overcome by finding a way to study the moments of $\left(X_S^T \Sigma^{-2} X_S\right)^{-1}$, which we leave for future work.

5 CONCLUSION AND FUTURE WORK

We study the problem of sparse recovery when observations come from mixed-quality sources. We establish sufficient conditions on the sample sizes (n_1, n_2) for both information-theoretic and algorithmic recovery purposes and in two settings, one when the decoder is completely agnostic to the noise and one where they are informed of the per-sample noise variance.

At the level of the information-theoretic threshold, we study the trade-off between high-quality and low-quality samples, and label the number of low-quality samples required to replace one high-quality sample *the Price of Quality*. In the agnostic setting, we reveal that this entity is quite low: in particular, one high-quality sample is never worth more than two low-quality samples. However, in the informed setting, the price of quality can grow arbitrarily large depending on the noise variances and the signal-to-noise regime. This constitutes the most important takeaway from our work: whenever possible, quantify uncertainty in the annotations and rescale the loss accordingly.

At the algorithmic threshold, we show in the agnostic setting that the classical LASSO recovery results from the homogeneous setting remain valid in the heterogeneous case and depend only on the total sample size $n_1 + n_2$. First, the threshold itself is independent of the individual noise levels. Second, the sufficient condition on the penalization coefficient involves the noise only through its average, exactly as if all observations had that average noise. Consequently, high-quality and low-quality samples contribute *equally* to the sample-size requirement for LASSO recovery. This reveals an unexpected difference in the effect of data heterogeneity on the information-theoretic and algorithmic thresholds for recovery.

In a broader discussion on how the information-theoretic and algorithmic thresholds interact across different problem settings, our result further emphasizes that the algorithmic threshold seems to be more "robust" to changes in the traditional problem settings (Gamarnik & Zadik, 2022; Wainwright, 2009). In fact, Wang et al. (2010) and Chaabouni & Gamarnik (2025) observed that when the noise is homogeneous but the design is sparse (i.e. X_{ij} set to 0 uniformly at random) the information-theoretic threshold increases, while Omidiran & Wainwright (2008) showed that the algorithmic threshold remains the same and is unaffected by changes in the sparsity level of the data (although this was shown only for the sufficient condition, with no corresponding result on necessity).

Although we don't study LASSO recovery in the informed setting, we believe this is a promising direction for future work. It would be interesting to study the price of quality there, and compare it to LASSO recovery in the agnostic setting on one hand, and to the price of quality of information-theoretic recovery on the other.

REFERENCES

Shuchin Aeron, Venkatesh Saligrama, and Manqi Zhao. Information theoretic bounds for compressed sensing. *IEEE Transactions on Information Theory*, 56(10):5111–5130, 2010. doi: 10.1109/TIT.2010.2059891.

Theodore Wilbur Anderson, Theodore Wilbur Anderson, Theodore Wilbur Anderson, Theodore Wilbur Anderson, and Etats-Unis Mathématicien. *An introduction to multivariate statistical analysis*, volume 2. Wiley New York, 1958.

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

Youssef Chaabouni and David Gamarnik. The price of sparsity: Sufficient conditions for sparse recovery using sparse and sparsified measurements. *arXiv preprint arXiv:2509.01809*, 2025.

Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

Graham Cormode and Marios Hadjieleftheriou. Finding the frequent items in streams of data. *Communications of the ACM*, 52(10):97–105, 2009.

Laurens De Haan and Ana Ferreira. *Extreme value theory: an introduction*. Springer, 2006.

Aurore Delaigle, Peter Hall, and Alexander Meister. On deconvolution with repeated measurements. 2008.

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

Ding-Zhu Du and Frank Kwang-ming Hwang. *Combinatorial group testing and its applications*, volume 12. World Scientific, 1999.

Naoki Egami, Musashi Hinck, Brandon Stewart, and Hanying Wei. Using imperfect surrogates for downstream inference: Design-based supervised learning for social science applications of large language models. *Advances in Neural Information Processing Systems*, 36:68589–68601, 2023.

Simon Foucart, Holger Rauhut, Simon Foucart, and Holger Rauhut. *An invitation to compressive sensing*. Springer, 2013.

Benoît Frénay and Michel Verleysen. Classification in the presence of label noise: a survey. *IEEE transactions on neural networks and learning systems*, 25(5):845–869, 2013.

J-J Fuchs. Recovery of exact sparse representations in the presence of noise. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pp. ii–533. IEEE, 2004.

David Gamarnik and Ilias Zadik. Sparse high-dimensional linear regression. estimating squared error and a phase transition. *The Annals of Statistics*, 50(2):880–903, 2022.

Kristina Gligorić, Tijana Zrnic, Cinoo Lee, Emmanuel J Candès, and Dan Jurafsky. Can unconfident llm annotations be used for confident conclusions? *arXiv preprint arXiv:2408.15204*, 2024.

Yinzheng Gu. Moments of random matrices and. 2013.

Piotr Indyk. Sketching, streaming and sublinear-space algorithms. *Graduate course notes, available at*, 33:617, 2007.

Michel Ledoux. *The concentration of measure phenomenon*. Number 89. American Mathematical Soc., 2001.

Minzhi Li, Taiwei Shi, Caleb Ziems, Min-Yen Kan, Nancy F Chen, Zhengyuan Liu, and Diyi Yang. Coannotating: Uncertainty-guided work allocation between human and large language models for data annotation. *arXiv preprint arXiv:2310.15638*, 2023.

Po-Ling Loh and Martin J Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Advances in neural information processing systems*, 24, 2011.

Pascal Massart. *Concentration inequalities and model selection: Ecole d'Eté de Probabilités de Saint-Flour XXXIII-2003*. Springer, 2007.

Elizabeth S Meckes. *The random matrix theory of the classical compact groups*, volume 218. Cambridge University Press, 2019.

Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. 2006.

Alan Miller. *Subset selection in regression*. chapman and hall/CRC, 2002.

Shanmugavelayutham Muthukrishnan et al. Data streams: Algorithms and applications. *Foundations and Trends® in Theoretical Computer Science*, 1(2):117–236, 2005.

Mohamed Ndaoud and Alexandre B Tsybakov. Optimal variable selection and adaptive noisy compressed sensing. *IEEE Transactions on Information Theory*, 66(4):2517–2532, 2020.

Dapo Omidiran and Martin J Wainwright. High-dimensional subset recovery in noise: Sparsified measurements without loss of statistical efficiency. *arXiv preprint arXiv:0805.3005*, 2008.

Pranav Rajpurkar, Jeremy Irvin, Robyn L Ball, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis P Langlotz, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. *PLoS medicine*, 15(11):e1002686, 2018.

Alexander Ratner, Stephen H Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. Snorkel: Rapid training data creation with weak supervision. In *Proceedings of the VLDB endowment. International conference on very large data bases*, volume 11, pp. 269, 2017.

Galen Reeves, Jiaming Xu, and Ilias Zadik. The all-or-nothing phenomenon in sparse linear regression. In *Conference on Learning Theory*, pp. 2652–2663. PMLR, 2019.

Alvin C Rencher and G Bruce Schaalje. *Linear models in statistics*. John Wiley & Sons, 2008.

Jonathan Silvertown. A new dawn for citizen science. *Trends in ecology & evolution*, 24(9):467–471, 2009.

Victor Siskind. Second moments of inverse wishart-matrix elements. *Biometrika*, 59(3):690–691, 1972.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

Joel A Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE transactions on information theory*, 52(3):1030–1051, 2006.

Martin J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using $\backslash ell\_\{1\}$-constrained quadratic programming (lasso). *IEEE transactions on information theory*, 55(5):2183–2202, 2009.

Wei Wang, Martin J Wainwright, and Kannan Ramchandran. Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices. *IEEE Transactions on Information Theory*, 56(6):2967–2979, 2010.

Ruoyu Zhang, Yanzeng Li, Yongliang Ma, Ming Zhou, and Lei Zou. Llmaaa: Making large language models as active annotators. *arXiv preprint arXiv:2310.19596*, 2023.

Peng Zhao and Bin Yu. On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563, 2006.

A PROOF OF THEOREM 1

Proof of Theorem 1. We denote by $S^\star \coloneqq \mathrm{Supp}\left(\beta^\star\right)$. Let $\mathcal{S}_{p,s} \coloneqq \{S \subset [p] : \; |S| = s\}$. We define the function:

$$\begin{aligned} L:\; \mathcal{S}_{p,s} &\longrightarrow \mathbb{R}_{\geq 0} \\ S &\longmapsto \|Y - X\mathbb{1}_S\|_2^2 \,, \end{aligned}$$

where $\mathbb{1}_S$ denotes the vector in $\{0,1\}^p$ such that $\left[\mathbb{1}_S\right]_j = \mathbb{1}\left(j \in S\right)$ for all $j \in [p]$. In particular, note (8) that:

$$\hat{\beta} = \mathbb{1}_{\hat{S}}, \quad \text{where} \quad \hat{S} \coloneqq \underset{S \in \mathcal{S}_{p,s}}{\arg\min}\, L\left(S\right).$$

For every $S \in \mathcal{S}_{p,s}$, we define: $M\left(S\right) \coloneqq \left|S \triangle S^\star\right|/2$, and let $U\left(S\right) \coloneqq S^\star \setminus S$, $V\left(S\right) \coloneqq S \setminus S^\star$. Note that, since $|S| = |S^\star| = s$, we have $|U\left(S\right)| = |V\left(S\right)| = M\left(S\right)$. We also define:

$$\begin{aligned} \Delta:\; \mathcal{S}_{p,s} &\longrightarrow \mathbb{R} \\ S &\longmapsto L\left(S\right) - L\left(S^\star\right). \end{aligned}$$

Proposition A.1. *For any $S \in \mathcal{S}_{p,s}$: if $M\left(S\right) \geq \delta s$, then:*

$$\mathbb{P}\left(\Delta\left(S\right) \leq 0\right) \leq \left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^2}\right)^{-n_1/2}\left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2}.$$

Proof. See section A.1. □

Hence we have, for any support $S \in \mathcal{S}_{p,s}$ such that $|S \triangle S^\star| \geq 2\delta s$:

$$\mathbb{P}\Big(\|Y - X\mathbb{1}_S\|_2^2 \leq \|Y - X\mathbb{1}_{S^\star}\|_2^2\Big) \leq \left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^2}\right)^{-n_1/2}\left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2} \tag{28}$$

Using (28) and a union bound over $\{S \in \mathcal{S}_{p,s} : \; |S \triangle S^\star| \geq 2\delta s\}$ we have:

$$\begin{aligned} &\mathbb{P}_{X,Z}\left(\left|\mathrm{Supp}\left(\hat{\beta}\right) \triangle \mathrm{Supp}\left(\beta^\star\right)\right| < 2\delta s\right) \\ &\geq \mathbb{P}_{X,Z}\Big(\|Y - X\mathbb{1}_S\|_2^2 > \|Y - X\mathbb{1}_{S^\star}\|_2^2\,, \forall\, S \in \mathcal{S}_{p,s} : \; |S \triangle S^\star| \geq 2\delta s\Big) \\ &= 1 - \mathbb{P}_{X,Z}\Big(\exists\, S \in \mathcal{S}_{p,s}:\; |S \triangle S^\star| \geq 2\delta s\,,\, \|Y - X\mathbb{1}_S\|_2^2 \leq \|Y - X\mathbb{1}_{S^\star}\|_2^2\Big) \\ &\overset{\text{U.B.}}{\geq} 1 - \sum_{S \in \mathcal{S}_{p,s}\,:\; |S \triangle S^\star| \geq 2\delta s} \mathbb{P}_{X,Z}\Big(\|Y - X\mathbb{1}_S\|_2^2 \leq \|Y - X\mathbb{1}_{S^\star}\|_2^2\Big) \\ &\overset{(28)}{\geq} 1 - \sum_{S \in \mathcal{S}_{p,s}\,:\; |S \triangle S^\star| \geq 2\delta s} \left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^2}\right)^{-n_1/2}\left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2} \\ &\geq 1 - |\mathcal{S}_{p,s}|\left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^2}\right)^{-n_1/2}\left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2} \\ &= 1 - \binom{p}{s}\left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^2}\right)^{-n_1/2}\left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2}. \end{aligned}$$

Case 1: Assume $s = o\left(p\right)$. We use the corollary of Stirling:

$$\binom{p}{s} = \exp\Big(s\log\left(p/s\right)\left(1 + o\left(1\right)\right)\Big),$$

which yields:

$$\mathbb{P}_{X,Z}\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right)$$
$$\geq 1 - \exp\left\{s \log\left(p/s\right)\big(1 + o\left(1\right)\big) - \frac{n_1}{2}\log\left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^2}\right) - \frac{n_2}{2}\log\left(1 + \frac{\delta s}{2\sigma_2^2}\right)\right\}.$$

Now using (9) in above, we have:

$$\begin{aligned}\mathbb{P}_{X,Z}&\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right)\\ &\geq 1 - \exp\left\{s\log\left(p/s\right)\big(1 + o\left(1\right)\big) - \left(1+\varepsilon\right)s\log\left(p/s\right)\right\}\\ &\geq 1 - \exp\left\{-\varepsilon s \log\left(p/s\right) - o\left(s\log\left(p/s\right)\right)\right\}.\end{aligned}$$

Finally we conclude:

$$\mathbb{P}_{X,Z}\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right) \geq 1 - \exp\left\{-\left(\varepsilon + o\left(1\right)\right)n^\star/2\right\} \overset{p\to+\infty}{\longrightarrow} 1,$$

where $n^\star = 2s\log\left(p/s\right)$.

Case 2: Assume $s = \alpha\left(p\right)$ for some constant $\alpha \in (0,1)$. We use the corollary of Stirling:

$$\binom{p}{s} = \exp\Big(h\left(\alpha\right)p\big(1 + o\left(1\right)\big)\Big),$$

which yields:

$$\mathbb{P}_{X,Z}\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right)$$
$$\geq 1 - \exp\left\{h\left(\alpha\right)p\big(1 + o\left(1\right)\big) - \frac{n_1}{2}\log\left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^2}\right) - \frac{n_2}{2}\log\left(1 + \frac{\delta s}{2\sigma_2^2}\right)\right\}.$$

Now using (9) in above, we have:

$$\begin{aligned}\mathbb{P}_{X,Z}&\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right)\\ &\geq 1 - \exp\left\{h\left(\alpha\right)p\big(1 + o\left(1\right)\big) - \left(1+\varepsilon\right)h\left(\alpha\right)p\right\}\\ &\geq 1 - \exp\left\{-\varepsilon h\left(\alpha\right)p - o\left(p\right)\right\}.\end{aligned}$$

Finally we conclude:

$$\mathbb{P}_{X,Z}\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right) \geq 1 - \exp\left\{-\left(\varepsilon + o\left(1\right)\right)n^\star/2\right\} \overset{p\to+\infty}{\longrightarrow} 1,$$

where $n^\star = 2h\left(\alpha\right)p$. ☐

A.1 Proof of Proposition A.1

Proof of Proposition A.1. Fix $S \in \mathcal{S}_{p,s}$ such that $M\left(S\right) \geq \delta s$. We have:

$$\begin{aligned}\Delta\left(S\right) &= L\left(S\right) - L\left(S^\star\right)\\ &= \|Y - X\mathbb{1}_S\|_2^2 - \|Y - X\mathbb{1}_{S^\star}\|_2^2\\ &= \|X\beta^\star + Z - X\mathbb{1}_S\|_2^2 - \|X\beta^\star + Z - X\mathbb{1}_{S^\star}\|_2^2\\ &= \|X\left(\mathbb{1}_{S^\star} - \mathbb{1}_S\right)\|_2^2 + 2\langle Z, X\left(\mathbb{1}_{S^\star} - \mathbb{1}_S\right)\rangle\\ &= \sum_{i=1}^n \langle X_i, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2\sum_{i=1}^n Z_i\langle X_i, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle.\end{aligned}$$

Let $X_1 \in \mathbb{R}^{n_1\times p}$, $X_2 \in \mathbb{R}^{n_2\times p}$ such that:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

Then the above expression of $\Delta(S)$ writes:

$$\Delta(s) = \sum_{i=1}^{n_1}\left\{\langle X_i^1, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2Z_i^1\langle X_i^1, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle\right\} + \sum_{i=1}^{n_2}\left\{\langle X_i^2, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2Z_i^2\langle X_i^2, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle\right\}.$$

We denote by $\left(\Delta_i^1\right)_{i\in[n_1]}$ and $\left(\Delta_i^2\right)_{i\in[n_2]}$ the terms of the sums above, that is:

$$\begin{cases}\Delta_i^1 := \langle X_i^1, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2Z_i^1\langle X_i^1, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle \\ \Delta_i^2 := \langle X_i^2, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2Z_i^2\langle X_i^2, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle\end{cases}.$$

Note that each of the elements of each of $\left\{\Delta_i^1 : i\in[n_1]\right\}$ and $\left\{\Delta_i^2 : i\in[n_2]\right\}$ are i.i.d. and:

$$\Delta(S) = \sum_{i=1}^{n_1}\Delta_i^1 + \sum_{i=1}^{n_2}\Delta_i^2.$$

Using the Chernoff bound, we have:

$$\begin{aligned}\mathbb{P}(\Delta(S)\le 0) &= \mathbb{P}(-\Delta(S)\ge 0) \\ &= \inf_{\theta\ge 0}\mathbb{P}\left(e^{-\theta\Delta(S)}\ge 1\right) \\ &\le \inf_{\theta\ge 0}\mathbb{E}\left[e^{-\theta\Delta(S)}\right] \\ &= \inf_{\theta\ge 0}\mathbb{E}\left[e^{-\sum_{i=1}^{n_1}\theta\Delta_i^1 + \sum_{i=1}^{n_2}\theta\Delta_i^2}\right] \\ &\overset{\text{ind.}}{=} \inf_{\theta\ge 0}\prod_{i=1}^{n_1}\mathbb{E}\left[e^{-\theta\Delta_i^1}\right]\prod_{i=1}^{n_2}\mathbb{E}\left[e^{-\theta\Delta_i^2}\right] \\ &= \inf_{\theta\ge 0}\prod_{i=1}^{n_1}M_{-\Delta_i^1}(\theta)\prod_{i=1}^{n_2}M_{-\Delta_i^2}(\theta) \\ &\overset{\text{i.d.}}{=} \inf_{\theta\ge 0}\left\{M_{-\Delta_i^1}(\theta)\right\}^{n_1}\left\{M_{-\Delta_i^2}(\theta)\right\}^{n_2}.\end{aligned}$$

Therefore:

$$\mathbb{P}(\Delta(S)\le 0) \le \inf_{\theta\ge 0}\left\{M_{-\Delta_i^1}(\theta)\right\}^{n_1}\left\{M_{-\Delta_i^2}(\theta)\right\}^{n_2}. \tag{29}$$

Now we have, for any $\theta\ge 0$:

$$\begin{aligned}M_{-\Delta_i^1}(\theta) &= \mathbb{E}_{X_i^1,Z_i^1}\left[e^{-\theta\left[\langle X_i^1,\beta^\star-\beta(S)\rangle^2 + 2Z_i^1\langle X_i^1,\beta^\star-\beta(S)\rangle\right]}\right] \\ &= \mathbb{E}_{X_i^1}\left[e^{-\theta\langle X_i^1,\beta^\star-\beta(S)\rangle^2}\mathbb{E}_{Z_i^1}\left[e^{-2\theta Z_i^1\langle X_i^1,\beta^\star-\beta(S)\rangle}\middle|X_i^1\right]\right] \\ &= \mathbb{E}_{X_i^1}\left[e^{-\theta\langle X_i^1,\beta^\star-\beta(S)\rangle^2}M_{Z_i^1|X_i^1}\left(-2\theta\langle X_i^1,\beta^\star-\beta(S)\rangle\right)\right] \\ &= \mathbb{E}_{X_i^1}\left[e^{-\theta\langle X_i^1,\beta^\star-\beta(S)\rangle^2}e^{\frac{1}{2}\left(-2\theta\langle X_i^1,\beta^\star-\beta(S)\rangle\right)^2\sigma_1^2}\right] \\ &= \mathbb{E}_{X_i}\left[e^{\left(-\theta+2\theta^2\sigma_1^2\right)\langle X_i^1,\beta^\star-\beta(S)\rangle^2}\right] \\ &= \mathbb{E}_{X_i}\left[e^{\left(-\theta+2\theta^2\sigma_1^2\right)\left(\sum_{j\in U}X_{ij}^1 - \sum_{j\in V}X_{ij}^1\right)^2}\right],\end{aligned}$$

where we write U and V for $U(S)$ and $V(S)$, respectively, for simplicity. We know that:

$$\sum_{j\in U}X_{ij}^1 - \sum_{j\in V}X_{ij}^1 \overset{d}{=} \sum_{j\in U\cup V}X_{ij}^1 \sim \mathcal{N}\left(0, |U\cup V|\right).$$

Hence:

$$M_{-\Delta_i^1}(\theta) = \mathbb{E}_{X_i}\left[e^{\left(-\theta+2\theta^2\sigma_1^2\right)\left(\sum_{j\in U} X_{ij}^1 - \sum_{j\in V} X_{ij}^1\right)^2}\right]$$
$$= \mathbb{E}_{X_i}\left[e^{|U\cup V|\left(-\theta+2\theta^2\sigma_1^2\right)\Gamma}\right],$$

where:

$$\Gamma = \left(\frac{\sum_{j\in U} X_{ij}^1 - \sum_{j\in V} X_{ij}^1}{\sqrt{|U\cup V|}}\right)^2 \sim \chi^2(1).$$

Therefore:

$$M_{-\Delta_i^1}(\theta) = \begin{cases} \frac{1}{\sqrt{1-2|U\cup V|\left(-\theta+2\theta^2\sigma_1^2\right)}} & \text{if } |U\cup V|\left(-\theta+2\theta^2\sigma_1^2\right) < 1/2 \\ +\infty & \text{else.} \end{cases} \tag{30}$$

Similarly to (30), we obtain the following expression for $M_{-\Delta_i^2}(\theta)$:

$$M_{-\Delta_i^2}(\theta) = \begin{cases} \frac{1}{\sqrt{1-2|U\cup V|\left(-\theta+2\theta^2\sigma_2^2\right)}} & \text{if } |U\cup V|\left(-\theta+2\theta^2\sigma_2^2\right) < 1/2 \\ +\infty & \text{else.} \end{cases} \tag{31}$$

Therefore, for any $\theta \geq 0$:

$$M_{-\Delta_i^1}(\theta)^{n_1} M_{-\Delta_i^2}(\theta)^{n_2} = \begin{cases} \left(1-2\left|U\cup V\right|\left(-\theta+2\theta^2\sigma_1^2\right)\right)^{-n_1/2} \\ \quad \times\left(1-2\left|U\cup V\right|\left(-\theta+2\theta^2\sigma_2^2\right)\right)^{-n_2/2} \\ \qquad \text{if } \begin{cases} |U\cup V|\left(-\theta+2\theta^2\sigma_1^2\right) < 1/2 \\ |U\cup V|\left(-\theta+2\theta^2\sigma_2^2\right) < 1/2 \end{cases} \\ +\infty \qquad \text{else.} \end{cases}$$

Remark A.1 (Best Chernoff bound)**.** To find the best Chernoff bound (29), we need to solve the optimization problem in (29), defined by:

$$\inf_{\theta\geq 0} \left\{M_{-\Delta_i^1}(\theta)\right\}^{n_1} \left\{M_{-\Delta_i^2}(\theta)\right\}^{n_2}. \tag{32}$$

Using the First Order Optimality Condition, (32) reduces to finding the roots of $\xi'(\theta) = 0$ is closed form, where $\xi(\cdot)$ is defined by:

$$\xi(\theta) := M_{-\Delta_i^1}(\theta)^{n_1} M_{-\Delta_i^2}(\theta)^{n_2}.$$

Using the closed-form solution of $\xi(\cdot)$ obtained above, we conclude that solving 32 reduces to finding the positive roots of the following third-degree polynomial in θ:

$$n_1\left(4\sigma_1^2\theta-1\right)\left(1-2\left|U\cup V\right|\left(-\theta+2\theta^2\sigma_2^2\right)\right) + n_2\left(4\sigma_2^2\theta-1\right)\left(1-2\left|U\cup V\right|\left(-\theta+2\theta^2\sigma_1^2\right)\right). \tag{33}$$

To the best of our knowledge, this doesn't lead to any "reasonable" closed-form expression for the minimizer $\theta^\star_{\text{min}}$. Instead, we note that:

$$\left\{\theta\colon\ |U\cup V|\left(-\theta+2\theta^2\sigma_2^2\right) < 1/2\right\} \subseteq \left\{\theta\colon\ |U\cup V|\left(-\theta+2\theta^2\sigma_1^2\right) < 1/2\right\},$$

and choose θ to the middle of the LHS interval.

In particular, setting $\theta^\star := \frac{1}{4\sigma_2^2}$, we have:

$$|U\cup V|\left(-\theta^\star+2{\theta^\star}^2\sigma_2^2\right) = -\theta^\star\left|U\cup V\right|\left(-1+2\theta^\star\sigma_2^2\right) = -\theta^\star\left|U\cup V\right|/2 < 0 < 1/2,$$

and:

$$|U\cup V|\left(-\theta^\star+2{\theta^\star}^2\sigma_1^2\right) = -\theta^\star\left|U\cup V\right|\left(-1+\frac{\sigma_1^2}{2\sigma_2^2}\right) < 0 < 1/2 \quad (\text{since } \sigma_1^2 < \sigma_2^2).$$

Therefore:

$$
\begin{aligned}
&M_{-\Delta_i^1}(\theta^\star)^{n_1} M_{-\Delta_i^2}(\theta^\star)^{n_2} \\
&\quad = \left(1 - 2\left|U \cup V\right| \left(-\theta^\star + 2{\theta^\star}^2 \sigma_1^2\right)\right)^{-n_1/2} \left(1 - 2\left|U \cup V\right| \left(-\theta^\star + 2{\theta^\star}^2 \sigma_2^2\right)\right)^{-n_2/2} \\
&\quad = \left(1 - 2\left|U \cup V\right| \left(-\frac{1}{4\sigma_2^2} + 2\left(\frac{1}{4\sigma_2^2}\right)^2 \sigma_1^2\right)\right)^{-n_1/2} \\
&\qquad\quad \times \left(1 - 2\left|U \cup V\right| \left(-\frac{1}{4\sigma_2^2} + 2\left(\frac{1}{4\sigma_2^2}\right)^2 \sigma_2^2\right)\right)^{-n_2/2} \\
&\quad = \left(1 + \left|U \cup V\right| \left(\frac{2\sigma_2^2 - \sigma_1^2}{4\sigma_2^4}\right)\right)^{-n_1/2} \left(1 + \frac{\left|U \cup V\right|}{4\sigma_2^2}\right)^{-n_2/2} \\
&\quad \leq \left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^4}\right)^{-n_1/2} \left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2},
\end{aligned}
$$

where the last inequality holds because $\left|U \cup V\right| = 2M\left(S\right) \geq 2\delta s$. Finally, using this in (29) we conclude:

$$
\mathbb{P}\left(\Delta\left(S\right) \leq 0\right) \leq \left(1 + \frac{\delta\left(2\sigma_2^2 - \sigma_1^2\right)s}{2\sigma_2^4}\right)^{-n_1/2} \left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2}.
$$

□

B PROOF OF (14)

Proof of (14). Let $\beta \in \mathcal{B}_{p,s}$. We know from (1) that:

$$
Y \mid X, \beta \sim \mathcal{N}\left(X\beta, \Sigma^2\right).
$$

Its pdf writes:

$$
f_{Y \mid X, \beta}\left(y\right) = \left(2\pi\right)^{-n/2} \det\left(\Sigma\right)^{-1} \exp\left\{-\left(y - X\beta\right)^T \Sigma^{-2}\left(y - X\beta\right)\right\}.
$$

The MLE is defined as:

$$
\begin{aligned}
\hat{\beta}_{\text{MLE}} &= \operatorname*{arg\,max}_{\beta \in \mathcal{B}_{p,s}} f_{Y \mid X, \beta}\left(Y\right) \\
&= \operatorname*{arg\,min}_{\beta \in \mathcal{B}_{p,s}} \left(Y - X\beta\right)^T \Sigma^{-2}\left(Y - X\beta\right) \\
&= \operatorname*{arg\,min}_{\beta \in \mathcal{B}_{p,s}} \left(Y - X\beta\right)^T \left(\Sigma^{-1}\right)^T \Sigma^{-1}\left(Y - X\beta\right) \\
&= \operatorname*{arg\,min}_{\beta \in \mathcal{B}_{p,s}} \left\|\Sigma^{-1}\left(Y - X\beta\right)\right\|_2^2.
\end{aligned}
$$

□

C PROOF OF THEOREM 2

Proof of Theorem 2. We denote by $S^\star \coloneqq \operatorname{Supp}\left(\beta^\star\right)$. Let $\mathcal{S}_{p,s} \coloneqq \left\{S \subset [p] \colon \left|S\right| = s\right\}$. We define the Σ-rescaled loss:

$$
\begin{aligned}
L_\Sigma \colon \mathcal{S}_{p,s} &\longrightarrow \mathbb{R}_{\geq 0} \\
S &\longmapsto \left\|\Sigma^{-1}\left(Y - X\mathbb{1}_S\right)\right\|_2^2,
\end{aligned}
$$

where $\mathbb{1}_S$ denote the vector in $\{0,1\}^p$ such that $\left[\mathbb{1}_S\right]_j = \mathbb{1}\left(j \in S\right)$ for all $j \in [p]$. In particular, note from (14) that:

$$
\hat{\beta}_{\text{MLE}} = \mathbb{1}_{\hat{S}_{\text{MLE}}}, \quad \text{where} \quad \hat{S}_{\text{MLE}} \coloneqq \operatorname*{arg\,min}_{S \in \mathcal{S}_{p,s}} L_\Sigma\left(S\right).
$$

For every $S \in \mathcal{S}_{p,s}$, we define: $M(S) \coloneqq |S \triangle S^\star|/2$, and let $U(S) \coloneqq S^\star \setminus S$, $V(S) \coloneqq S \setminus S^\star$. Note that, since $|S| = |S^\star| = s$, we have $|U(S)| = |V(S)| = M(S)$. We also define:

$$\begin{aligned} \Delta : \ \mathcal{S}_{p,s} &\longrightarrow \mathbb{R} \\ S &\longmapsto L_\Sigma(S) - L_\Sigma(S^\star). \end{aligned}$$

Proposition C.1. *For any $S \in \mathcal{S}_{p,s}$: if $M(S) \geq \delta s$, then:*

$$\mathbb{P}(\Delta(S) \leq 0) \leq \left(1 + \frac{\delta s}{2\sigma_1^2}\right)^{-n_1/2} \left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2}.$$

Proof. See section C.1. □

Hence we have, for any support $S \in \mathcal{S}_{p,s}$ such that $|S \triangle S^\star| \geq 2\delta s$:

$$\mathbb{P}\Big(\left\|\Sigma^{-1}(Y - X\mathbb{1}_S)\right\|_2^2 \leq \left\|\Sigma^{-1}(Y - X\mathbb{1}_{S^\star})\right\|_2^2\Big) \leq \left(1 + \frac{\delta s}{2\sigma_1^2}\right)^{-n_1/2} \left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2}. \quad (34)$$

Using (34) and a union bound over $\{S \in \mathcal{S}_{p,s} : \ |S \triangle S^\star| \geq 2\delta s\}$ we have:

$$\begin{aligned} &\mathbb{P}_{X,Z}\left(\left|\mathrm{Supp}\left(\hat\beta\right) \triangle \mathrm{Supp}\left(\beta^\star\right)\right| < 2\delta s\right) \\ &\quad\geq \mathbb{P}_{X,Z}\Big(\left\|\Sigma^{-1}(Y - X\mathbb{1}_S)\right\|_2^2 > \left\|\Sigma^{-1}(Y - X\mathbb{1}_{S^\star})\right\|_2^2, \forall\, S \in \mathcal{S}_{p,s} : \ |S \triangle S^\star| \geq 2\delta s\Big) \\ &\quad= 1 - \mathbb{P}_{X,Z}\Big(\exists\, S \in \mathcal{S}_{p,s} : \ |S \triangle S^\star| \geq 2\delta s\,,\ \left\|\Sigma^{-1}(Y - X\mathbb{1}_S)\right\|_2^2 \leq \left\|\Sigma^{-1}(Y - X\mathbb{1}_{S^\star})\right\|_2^2\Big) \\ &\quad\overset{\text{U.B.}}{\geq} 1 - \sum_{S \in \mathcal{S}_{p,s} : \ |S \triangle S^\star| \geq 2\delta s} \mathbb{P}_{X,Z}\Big(\left\|\Sigma^{-1}(Y - X\mathbb{1}_S)\right\|_2^2 \leq \left\|\Sigma^{-1}(Y - X\mathbb{1}_{S^\star})\right\|_2^2\Big) \\ &\quad\overset{(34)}{\geq} 1 - \sum_{S \in \mathcal{S}_{p,s} : \ |S \triangle S^\star| \geq 2\delta s} \left(1 + \frac{\delta s}{2\sigma_1^2}\right)^{-n_1/2} \left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2} \\ &\quad\geq 1 - |\mathcal{S}_{p,s}| \left(1 + \frac{\delta s}{2\sigma_1^2}\right)^{-n_1/2} \left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2} \\ &\quad= 1 - \binom{p}{s} \left(1 + \frac{\delta s}{2\sigma_1^2}\right)^{-n_1/2} \left(1 + \frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2}. \end{aligned}$$

Case 1: Assume $s = o(p)$. We use the corollary of Stirling:

$$\binom{p}{s} = \exp\Big(s \log(p/s)\big(1 + o(1)\big)\Big),$$

which yields:

$$\begin{aligned} &\mathbb{P}_{X,Z}\left(\left|\mathrm{Supp}\left(\hat\beta\right) \triangle \mathrm{Supp}\left(\beta^\star\right)\right| < 2\delta s\right) \\ &\quad\geq 1 - \exp\left\{s \log(p/s)\big(1 + o(1)\big) - \frac{n_1}{2}\log\left(1 + \frac{\delta s}{2\sigma_1^2}\right) - \frac{n_2}{2}\log\left(1 + \frac{\delta s}{2\sigma_2^2}\right)\right\}. \end{aligned}$$

Now using (15) in above, we have:

$$\begin{aligned} &\mathbb{P}_{X,Z}\left(\left|\mathrm{Supp}\left(\hat\beta\right) \triangle \mathrm{Supp}\left(\beta^\star\right)\right| < 2\delta s\right) \\ &\quad\geq 1 - \exp\left\{s \log(p/s)\big(1 + o(1)\big) - (1 + \varepsilon)\, s \log(p/s)\right\} \\ &\quad\geq 1 - \exp\left\{-\varepsilon s \log(p/s) - o\left(s \log(p/s)\right)\right\}. \end{aligned}$$

Finally we conclude:

$$\mathbb{P}_{X,Z}\left(\left|\mathrm{Supp}\left(\hat\beta\right) \triangle \mathrm{Supp}\left(\beta^\star\right)\right| < 2\delta s\right) \geq 1 - \exp\left\{-\left(\varepsilon + o(1)\right) n^\star/2\right\} \overset{p \to +\infty}{\longrightarrow} 1,$$

where $n^\star = 2s\log(p/s)$.

Case 2: Assume $s = h(\alpha)\,p$, for some constant $\alpha \in (0,1)$. We use the corollary of Stirling:

$$\binom{p}{s} = \exp\Big(h(\alpha)\,p\big(1+o(1)\big)\Big),$$

which yields:

$$\begin{aligned}&\mathbb{P}_{X,Z}\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right)\\&\quad\geq 1-\exp\left\{h(\alpha)\,p\big(1+o(1)\big) - \frac{n_1}{2}\log\left(1+\frac{\delta s}{2\sigma_1^2}\right) - \frac{n_2}{2}\log\left(1+\frac{\delta s}{2\sigma_2^2}\right)\right\}.\end{aligned}$$

Now using (15) in above, we have:

$$\begin{aligned}&\mathbb{P}_{X,Z}\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right)\\&\quad\geq 1-\exp\left\{h(\alpha)\,p\big(1+o(1)\big) - (1+\varepsilon)\,h(\alpha)\,p\right\}\\&\quad\geq 1-\exp\left\{-\varepsilon h(\alpha)\,p - o(p)\right\}.\end{aligned}$$

Finally we conclude:

$$\mathbb{P}_{X,Z}\left(\left|\text{Supp}\left(\hat{\beta}\right) \triangle \text{Supp}\left(\beta^\star\right)\right| < 2\delta s\right) \geq 1-\exp\left\{-\left(\varepsilon+o(1)\right)n^\star/2\right\} \overset{p\to+\infty}{\longrightarrow} 1,$$

where $n^\star = 2h(\alpha)\,p$. □

C.1 Proof of Proposition C.1

Proof of Proposition C.1. Fix $S \in \mathcal{S}_{p,s}$ such that $M(S) \geq \delta s$. We have:

$$\begin{aligned}\Delta(S) &= L_\Sigma(S) - L_\Sigma(S^\star)\\&= \left\|\Sigma^{-1}\left(Y - X\mathbb{1}_S\right)\right\|_2^2 - \left\|\Sigma^{-1}\left(Y - X\mathbb{1}_{S^\star}\right)\right\|_2^2\\&= \left\|\Sigma^{-1}\left(X\beta^\star + Z - X\mathbb{1}_S\right)\right\|_2^2 - \left\|\Sigma^{-1}\left(X\beta^\star + Z - X\mathbb{1}_{S^\star}\right)\right\|_2^2\\&= \left\|\Sigma^{-1}X\left(\mathbb{1}_{S^\star} - \mathbb{1}_S\right)\right\|_2^2 + 2\left\langle\Sigma^{-1}Z, \Sigma^{-1}X\left(\mathbb{1}_{S^\star} - \mathbb{1}_S\right)\right\rangle\\&= \sum_{i=1}^{n_1}\frac{1}{\sigma_1^2}\langle X_i, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + \sum_{i=n_1+1}^{n_2}\frac{1}{\sigma_2^2}\langle X_i, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2\\&\quad + 2\sum_{i=1}^{n_1}\frac{1}{\sigma_1^2}Z_i\langle X_i, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle + 2\sum_{i=n_1+1}^{n_2}\frac{1}{\sigma_2^2}Z_i\langle X_i, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle.\end{aligned}$$

Let $X_1 \in \mathbb{R}^{n_1\times p}$, $X_2 \in \mathbb{R}^{n_2\times p}$ such that:

$$X = \begin{pmatrix}X_1\\X_2\end{pmatrix}.$$

Then the above expression of $\Delta(S)$ writes:

$$\begin{aligned}\Delta(s) &= \frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}\left\{\langle X_i^1, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2Z_i^1\langle X_i^1, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle\right\}\\&\quad + \frac{1}{\sigma_2^2}\sum_{i=1}^{n_2}\left\{\langle X_i^2, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2Z_i^2\langle X_i^2, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle\right\}.\end{aligned}$$

We denote by $\left(\Delta_i^1\right)_{i\in[n_1]}$ and $\left(\Delta_i^2\right)_{i\in[n_2]}$ the terms of the sums above, that is:

$$\begin{cases}\Delta_i^1 \coloneqq \langle X_i^1, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2Z_i^1\langle X_i^1, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle\\ \Delta_i^2 \coloneqq \langle X_i^2, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle^2 + 2Z_i^2\langle X_i^2, \mathbb{1}_{S^\star} - \mathbb{1}_S\rangle\end{cases}.$$

Note that each of the elements of each of $\left\{\Delta_i^1 : i \in [n_1]\right\}$ and $\left\{\Delta_i^2 : i \in [n_2]\right\}$ are i.i.d. and:

$$\Delta(S) = \frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}\Delta_i^1 + \frac{1}{\sigma_2^2}\sum_{i=1}^{n_2}\Delta_i^2.$$

Using the Chernoff bound, we have:

$$\begin{aligned}
\mathbb{P}(\Delta(S) \le 0) &= \mathbb{P}(-\Delta(S) \ge 0) \\
&= \inf_{\theta \ge 0} \mathbb{P}\left(e^{-\theta\Delta(S)} \ge 1\right) \\
&\le \inf_{\theta \ge 0} \mathbb{E}\left[e^{-\theta\Delta(S)}\right] \\
&= \inf_{\theta \ge 0} \mathbb{E}\left[e^{-\frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}\theta\Delta_i^1 + \frac{1}{\sigma_2^2}\sum_{i=1}^{n_2}\theta\Delta_i^2}\right] \\
&\overset{\text{ind.}}{=} \inf_{\theta \ge 0} \prod_{i=1}^{n_1}\mathbb{E}\left[e^{-\theta\Delta_i^1/\sigma_1^2}\right]\prod_{i=1}^{n_2}\mathbb{E}\left[e^{-\theta\Delta_i^2/\sigma_2^2}\right] \\
&= \inf_{\theta \ge 0} \prod_{i=1}^{n_1} M_{-\Delta_i^1}\left(\frac{\theta}{\sigma_1^2}\right)\prod_{i=1}^{n_2} M_{-\Delta_i^2}\left(\frac{\theta}{\sigma_2^2}\right) \\
&\overset{\text{i.d.}}{=} \inf_{\theta \ge 0} \left\{M_{-\Delta_i^1}\left(\frac{\theta}{\sigma_1^2}\right)\right\}^{n_1}\left\{M_{-\Delta_i^2}\left(\frac{\theta}{\sigma_1^2}\right)\right\}^{n_2}.
\end{aligned}$$

Therefore:

$$\mathbb{P}(\Delta(S) \le 0) \le \inf_{\theta \ge 0} \left\{M_{-\Delta_i^1}\left(\frac{\theta}{\sigma_1^2}\right)\right\}^{n_1}\left\{M_{-\Delta_i^2}\left(\frac{\theta}{\sigma_2^2}\right)\right\}^{n_2}. \tag{35}$$

Now we have, for any $\theta \ge 0$:

$$\begin{aligned}
M_{-\Delta_i^1}(\theta/\sigma_1^2) &= \mathbb{E}_{X_i^1, Z_i^1}\left[e^{-\theta\left[\langle X_i^1, \beta^\star - \beta(S)\rangle^2 + 2Z_i^1\langle X_i^1, \beta^\star - \beta(S)\rangle\right]/\sigma_1^2}\right] \\
&= \mathbb{E}_{X_i^1}\left[e^{-\theta\langle X_i^1, \beta^\star - \beta(S)\rangle^2/\sigma_1^2}\mathbb{E}_{Z_i^1}\left[e^{-2\theta Z_i^1\langle X_i^1, \beta^\star - \beta(S)\rangle/\sigma_1^2}\middle| X_i^1\right]\right] \\
&= \mathbb{E}_{X_i^1}\left[e^{-\theta\langle X_i^1, \beta^\star - \beta(S)\rangle^2/\sigma_1^2}M_{Z_i^1|X_i^1}\left(-2\theta\langle X_i^1, \beta^\star - \beta(S)\rangle/\sigma_1^2\right)\right] \\
&= \mathbb{E}_{X_i^1}\left[e^{-\theta\langle X_i^1, \beta^\star - \beta(S)\rangle^2/\sigma_1^2}e^{\frac{1}{2}\left(-2\theta\langle X_i^1, \beta^\star - \beta(S)\rangle/\sigma_1^2\right)^2\sigma_1^2}\right] \\
&= \mathbb{E}_{X_i}\left[e^{\left(-\theta + 2\sigma_1^2\right)\langle X_i^1, \beta^\star - \beta(S)\rangle^2/\sigma_1^2}\right] \\
&= \mathbb{E}_{X_i}\left[e^{\left(-\theta + 2\theta^2\right)\left(\sum_{j \in U} X_{ij}^1 - \sum_{j \in V} X_{ij}^1\right)^2/\sigma_1^2}\right],
\end{aligned}$$

where we write U and V for $U(S)$ and $V(S)$, respectively, for simplicity. We know that:

$$\sum_{j \in U} X_{ij}^1 - \sum_{j \in V} X_{ij}^1 \overset{d}{=} \sum_{j \in U \cup V} X_{ij}^1 \sim \mathcal{N}(0, |U \cup V|).$$

Hence:

$$\begin{aligned}
M_{-\Delta_i^1}(\theta) &= \mathbb{E}_{X_i}\left[e^{\left(-\theta + 2\theta^2\right)\left(\sum_{j \in U} X_{ij}^1 - \sum_{j \in V} X_{ij}^1\right)^2/\sigma_1^2}\right] \\
&= \mathbb{E}_{X_i}\left[e^{|U \cup V|\left(-\theta + 2\theta^2\right)\Gamma/\sigma_1^2}\right],
\end{aligned}$$

where:

$$\Gamma = \left(\frac{\sum_{j \in U} X_{ij}^1 - \sum_{j \in V} X_{ij}^1}{\sqrt{|U \cup V|}}\right)^2 \sim \chi^2(1).$$

Therefore:

$$M_{-\Delta_i^1}(\theta/\sigma_1^2) = \begin{cases} \frac{1}{\sqrt{1 - 2|U \cup V|(-\theta + 2\theta^2)/\sigma_1^2}} & \text{if } |U \cup V|\left(-\theta + 2\theta^2\right)/\sigma_1^2 < 1/2 \\ +\infty & \text{else.} \end{cases} \tag{36}$$

Similarly to (36), we obtain the following expression for $M_{-\Delta_i^2}(\theta/\sigma_2^2)$:

$$M_{-\Delta_i^2}(\theta/\sigma_2^2) = \begin{cases} \frac{1}{\sqrt{1-2|U\cup V|(-\theta+2\theta^2)/\sigma_2^2}} & \text{if } |U\cup V|\left(-\theta+2\theta^2\right)/\sigma_2^2 < 1/2 \\ +\infty & \text{else.} \end{cases} \tag{37}$$

From above, we clearly have:

$$\arg\min_{\theta\geq 0} M_{-\Delta_i^1}(\theta/\sigma_1^2) = \arg\min_{\theta\geq 0} M_{-\Delta_i^2}(\theta/\sigma_2^2) = \arg\min_{\theta\geq 0}\left\{-\theta+2\theta^2\right\} = \frac{1}{4},$$

and, taking $\theta^\star \coloneqq 1/4$ we have:

$$M_{-\Delta_i^1}(\theta/\sigma_1^2) = \frac{1}{\sqrt{1+\frac{|U\cup V|}{4\sigma_1^2}}} \quad \text{and} \quad M_{-\Delta_i^2}(\theta/\sigma_2^2) = \frac{1}{\sqrt{1+\frac{|U\cup V|}{4\sigma_2^2}}}.$$

Therefore:

$$\begin{aligned} \inf_{\theta\geq 0}\left\{M_{-\Delta_i^1}\left(\frac{\theta}{\sigma_1^2}\right)\right\}^{n_1}\left\{M_{-\Delta_i^2}\left(\frac{\theta}{\sigma_2^2}\right)\right\}^{n_2} &= \left\{M_{-\Delta_i^1}\left(\frac{\theta^\star}{\sigma_1^2}\right)\right\}^{n_1}\left\{M_{-\Delta_i^2}\left(\frac{\theta^\star}{\sigma_2^2}\right)\right\}^{n_2} \\ &= \left\{\frac{1}{\sqrt{1+\frac{|U\cup V|}{4\sigma_1^2}}}\right\}^{n_1}\left\{\frac{1}{\sqrt{1+\frac{|U\cup V|}{4\sigma_2^2}}}\right\}^{n_2} \end{aligned}$$

Therefore:

$$\inf_{\theta\geq 0}\left\{M_{-\Delta_i^1}\left(\frac{\theta}{\sigma_1^2}\right)\right\}^{n_1}\left\{M_{-\Delta_i^2}\left(\frac{\theta}{\sigma_2^2}\right)\right\}^{n_2} = \left(1+\frac{|U\cup V|}{4\sigma_1^2}\right)^{-n_1/2}\left(1+\frac{|U\cup V|}{4\sigma_2^2}\right)^{-n_2/2}. \tag{38}$$

Since $|U\cup V| = 2M(S) \geq 2\delta s$, the above yields:

$$\inf_{\theta\geq 0}\left\{M_{-\Delta_i^1}\left(\frac{\theta}{\sigma_1^2}\right)\right\}^{n_1}\left\{M_{-\Delta_i^2}\left(\frac{\theta}{\sigma_2^2}\right)\right\}^{n_2} \leq \left(1+\frac{\delta s}{2\sigma_1^2}\right)^{-n_1/2}\left(1+\frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2}.$$

Finally, using this in (35) we conclude:

$$\mathbb{P}\left(\Delta(S)\leq 0\right) \leq \left(1+\frac{\delta s}{2\sigma_1^2}\right)^{-n_1/2}\left(1+\frac{\delta s}{2\sigma_2^2}\right)^{-n_2/2}.$$

□

D Proof of Theorem 3

Proof of Theorem 3. We define $\Sigma \in \mathbb{R}^{n\times n}$ and W random vector in $\mathbb{R}^n$ such that:

$$Z = \Sigma W,$$

where:

$$\Sigma = \begin{pmatrix} \sigma_1 I_{n_1} & 0 \\ 0 & \sigma_2 I_{n_2} \end{pmatrix}, \quad W \sim \mathcal{N}\left(0, I_n\right).$$

Let $S \coloneqq \mathrm{Supp}\left(\beta^\star\right)$ and $S^c \coloneqq [p]\backslash S$. The following proposition characterizes the recovery property in a more tractable way that will help us in the proof.

Proposition D.1. *Assume that the matrix $X_S^T X_S$ is invertible. Then, for any given $\lambda_p > 0$ and noise $\Sigma w \in \mathbb{R}^n$ we have:*

$$\mathcal{R}\left(X,\beta^\star,\Sigma w,\lambda_p\right) \iff \begin{cases} \left|\left(\frac{1}{n}X_S^T X_S\right)^{-1}\left(\frac{1}{n}X_S^T\Sigma w - \lambda_p \ \mathrm{sign}\left(\beta_S^\star\right)\right)\right| < |\beta_S^\star| \\ \left|X_{S^c}^T X_S\left(X_S^T X_S\right)^{-1}\left(\frac{1}{n}X_S^T\Sigma w - \lambda_p\,\mathrm{sign}\left(\beta_S^\star\right)\right) - \frac{1}{n}X_{S^c}^T\Sigma w\right| \leq \lambda_p \end{cases}$$

where the absolute values and inequalities are taken component-wise.

Proof. The equivalence follows from the First Order Optimality Condition of the LASSO (20). It was used in the proof of the LASSO threshold (Wainwright, 2009) and previously by Fuchs (2004); Meinshausen & Bühlmann (2006); Tropp (2006); Zhao & Yu (2006). See appendix D.1 for the complete proof. □

Let $\vec{b} := \text{sign}(\beta^\star)$. We define:

$$U_i := e_i^T \left(\frac{1}{n} X_S^T X_S\right)^{-1} \left[\frac{1}{n} X_S^T \Sigma W - \lambda_p \vec{b}\right], \tag{39}$$

and

$$V_j := X_j^T \left\{ X_S \left(X_S^T X_S\right)^{-1} \lambda_p \vec{b} - \left[X_S \left(X_S^T X_S\right)^{-1} X_S^T - I_n\right] \frac{\Sigma W}{n} \right\}, \tag{40}$$

for all $i \in S$ and $j \in S^c$. Let $\rho := \min_{i \in S} |\beta_i^\star|$. Note that:

$$\max_{i \in S} |U_i| < \rho \implies \left| \left(\frac{1}{n} X_S^T X_S\right)^{-1} \left(\frac{1}{n} X_S^T \Sigma w - \lambda_p \text{ sign}(\beta_S^\star)\right) \right| < |\beta_S^\star|, \tag{41}$$

and

$$\max_{j \in S^c} |V_j| \le \lambda_p \iff \left| X_{S^c}^T X_S \left(X_S^T X_S\right)^{-1} \left(\frac{1}{n} X_S^T \Sigma w - \lambda_p \text{sign}(\beta_S^\star)\right) - \frac{1}{n} X_{S^c}^T \Sigma w \right| \le \lambda_p. \tag{42}$$

In addition, note that when $s < n$, X_S is full-rank a.s., and hence $X_S^T X_S$ is invertible a.s. Therefore, the equivalence in Proposition 3.1 holds a.s. The proof of Theorem 3 relies of the two following propositions:

Proposition D.2.

i. *Under the sample size condition (22) we have:*

$$\mathbb{P}\left(\max_{j \in S^c} |V_j| \le \lambda_p\right) \overset{p \to +\infty}{\longrightarrow} 0.$$

ii. *Under the sample size condition (23) and the regularization condition (24), we have:*

$$\mathbb{P}\left(\max_{j \in S^c} |V_j| \le \lambda_p\right) \overset{p \to +\infty}{\longrightarrow} 1.$$

Proof. See appendix D.2. □

Proposition D.3. *Under the sample size condition (23) and the regularization condition (24), we have:*

$$\mathbb{P}\left(\max_{i \in S} |U_i| < \rho\right) \overset{p \to +\infty}{\longrightarrow} 1.$$

Proof. See appendix D.3. □

Necessity: assume (22) holds. Then we conclude by Proposition D.2 [*i*] and (42).

Sufficiency: assume (23) and (24) hold. Then we conclude by Proposition D.2 [*ii*], Proposition D.3 and (41), (42). □

D.1 Proof of Proposition D.1

Proof of Proposition D.1. Let $\hat{\beta} \in \mathbb{R}^p$. By First Order Optimality Condition of the Lasso, $\hat{\beta}$ is optimal if and only if the exits $z \in \mathbb{R}^p$ such that:

$$\begin{cases} z \in \partial \ell_1\left(\hat{\beta}\right) = \left\{ z \in \mathbb{R}^p \colon z_i = \text{sign}\left(\hat{\beta}_i\right) \text{ for } \hat{\beta}_i \neq 0, \ |z_i| \leq 1 \text{ otherwise} \right\}, \\ \frac{1}{n} X^T \left(X\hat{\beta} - Y \right) + \lambda_p z = 0. \end{cases}$$

The condition above can be equivalently written as:

$$\begin{cases} z_S = \text{sign}\left(\hat{\beta}_S\right), \\ |z_{S^c}| \leq 1, \\ \frac{1}{n} X^T X \hat{\beta} - \frac{1}{n} X^T Y + \lambda_p z = 0. \end{cases}$$

Substituting $Y = X\beta^\star + \Sigma w$, the condition writes:

$$\begin{cases} z_S = \text{sign}\left(\hat{\beta}_S\right), \\ |z_{S^c}| \leq 1, \\ \frac{1}{n} X^T X \left(\hat{\beta} - \beta^\star\right) - \frac{1}{n} X^T \Sigma w + \lambda_p z = 0. \end{cases}$$

Splitting on S and S^c we get:

$$\begin{cases} z_S = \text{sign}\left(\hat{\beta}_S\right), \\ |z_{S^c}| \leq 1, \\ \frac{1}{n} X_S^T X \left(\hat{\beta} - \beta^\star\right) - \frac{1}{n} X_S^T \Sigma w + \lambda_p \text{ sign}\left(\hat{\beta}_S\right) = 0, \\ \frac{1}{n} X_{S^c}^T X \left(\hat{\beta} - \beta^\star\right) - \frac{1}{n} X_{S^c}^T \Sigma w + \lambda_p z_{S^c} = 0. \end{cases}$$

Now the Lasso recovers the support of $\beta^\star$ if and only if there exits $\hat{\beta} \in \mathbb{R}^p$ such that $\text{sign}\left(\hat{\beta}\right) = \text{sign}\left(\beta^\star\right)$ and:

$$\exists \ z \in \mathbb{R}^p \text{ such that } \begin{cases} z_S = \text{sign}\left(\hat{\beta}_S\right), \\ |z_{S^c}| \leq 1, \\ \frac{1}{n} X_S^T X \left(\hat{\beta} - \beta^\star\right) - \frac{1}{n} X_S^T \Sigma w + \lambda_p \text{ sign}\left(\hat{\beta}_S\right) = 0, \\ \frac{1}{n} X_{S^c}^T X \left(\hat{\beta} - \beta^\star\right) - \frac{1}{n} X_{S^c}^T \Sigma w + \lambda_p z_{S^c} = 0. \end{cases}$$

Which is equivalent to:

$$\exists \ z, \hat{\beta} \in \mathbb{R}^p \text{ such that } \begin{cases} \text{sign}\left(\hat{\beta}\right) = \text{sign}\left(\beta^\star\right), \\ z_S = \text{sign}\left(\hat{\beta}_S\right), \\ |z_{S^c}| \leq 1, \\ \frac{1}{n} X_S^T X \left(\hat{\beta} - \beta^\star\right) - \frac{1}{n} X_S^T \Sigma w + \lambda_p \text{ sign}\left(\hat{\beta}_S\right) = 0, \\ \frac{1}{n} X_{S^c}^T X \left(\hat{\beta} - \beta^\star\right) - \frac{1}{n} X_{S^c}^T \Sigma w + \lambda_p z_{S^c} = 0. \end{cases}$$

which is equivalent to

$$\exists \ z, \hat{\beta} \in \mathbb{R}^p \text{ such that } \begin{cases} \text{sign}\left(\hat{\beta}\right) = \text{sign}\left(\beta^\star\right), \\ z_S = \text{sign}\left(\beta_S^\star\right), \\ |z_{S^c}| \leq 1, \\ \frac{1}{n} X_S^T X_S \left(\hat{\beta}_S - \beta_S^\star\right) - \frac{1}{n} X_S^T \Sigma w + \lambda_p \text{ sign}\left(\beta_S^\star\right) = 0, \\ \frac{1}{n} X_{S^c}^T X_S \left(\hat{\beta}_S - \beta_S^\star\right) - \frac{1}{n} X_{S^c}^T \Sigma w + \lambda_p z_{S^c} = 0. \end{cases}$$

which is equivalent to

$$\exists\, z, \hat{\beta} \in \mathbb{R}^p \text{ such that } \begin{cases} \hat{\beta}_{S^c} = 0, \\ z_S = \operatorname{sign}\left(\hat{\beta}_S\right) = \operatorname{sign}\left(\beta_S^\star\right) \neq 0, \\ |z_{S^c}| \leq 1, \\ \hat{\beta}_S = \beta_S^\star + \left(\frac{1}{n} X_S^T X_S\right)^{-1} \left(\frac{1}{n} X_S^T \Sigma w - \lambda_p \operatorname{sign}\left(\beta_S^\star\right)\right), \\ \left|\frac{1}{n} X_{S^c}^T X_S \left(\hat{\beta}_S - \beta_S^\star\right) - \frac{1}{n} X_{S^c}^T \Sigma w\right| \leq \lambda_p. \end{cases}$$

which is equivalent to

$$\begin{cases} \left|\left(\frac{1}{n} X_S^T X_S\right)^{-1} \left(\frac{1}{n} X_S^T \Sigma w - \lambda_p \operatorname{sign}\left(\beta_S^\star\right)\right)\right| < \left|\beta_S^\star\right|, \\ \left|X_{S^c}^T X_S \left(X_S^T X_S\right)^{-1} \left(\frac{1}{n} X_S^T \Sigma w - \lambda_p \operatorname{sign}\left(\beta_S^\star\right)\right) - \frac{1}{n} X_{S^c}^T \Sigma w\right| \leq \lambda_p. \end{cases}$$

□

D.2 Proof of Proposition D.2

D.2.1 Preliminary results

Lemma D.1 (Moments of $(V \mid X_S, W)$)**.** *Conditionally on X_S and W, V is Gaussian vector. In addition, we have:*

$$\mathbb{E}\left[V \mid X_S, W\right] = 0,$$

and:

$$\operatorname{Cov}\left[V \mid X_S, W\right] = M_p I_{|S^c|},$$

where:

$$\begin{aligned} M_p &\coloneqq \left\| X_S \left(X_S^T X_S\right)^{-1} \lambda_p \vec{b} + \left[I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T\right] \frac{\Sigma W}{n} \right\|_2^2, \\ &= \lambda_p^2 \vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b} + \frac{1}{n^2} W^T \Sigma \left[I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T\right] \Sigma W. \end{aligned}$$

Proof. See section D.2.4. □

Lemma D.2 (Bounding the second moment of $(V \mid X_S, W)$)**.** *We have:*

$$\mathbb{E}\left[M_p\right] = \frac{\lambda_p^2}{n - s - 1} \|b\|_2^2 + \frac{(n - s)\left(n_1 \sigma_1^2 + n_2 \sigma_2^2\right)}{n^3}.$$

In addition, for any constant $\delta > 0$, we have:

$$\mathbb{P}\left(\left|M_p - \mathbb{E}\left[M_p\right]\right| \geq \delta \mathbb{E}\left[M_p\right]\right) \to 0,$$

as $p \to +\infty$.

Proof. See Section D.2.5. □

D.2.2 Showing that $\mathbb{P}\left(\max_{j \in S^c} |V_j| \leq \lambda_p\right) \longrightarrow 0$:

Proof of Proposition D.2, part (i.) . Let:

$$T(\delta) \coloneqq \left\{\left|M_p - \mathbb{E}\left[M_p\right]\right| \geq \delta \mathbb{E}\left[M_p\right]\right\}.$$

We have, by total probability:

$$\mathbb{P}\Big(\max_{j \in S^c} |V_j| \leq \lambda_p\Big) \leq \mathbb{P}\Big(\max_{j \in S^c} |V_j| \leq \lambda_p \,\Big|\, T(\delta)^c\Big) + \mathbb{P}\big(T(\delta)\big). \tag{43}$$

Conditioning on X_S and W:

$$\mathbb{P}\Big(\max_{j \in S^c} |V_j| \leq \lambda_p \,\Big|\, T(\delta)^c\Big) = \mathbb{E}\left[\mathbb{P}\Big(\max_{j \in S^c} |V_j| \leq \lambda_p \,\Big|\, X_S, W\Big) \,\Big|\, T(\delta)^c\right].$$

Note that, conditionally on (X_S, W), we have $V_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, M_p)$, for $j \in S^c$. We know the bound on expectation of Gaussian maxima (see Theorem 5.3.1 in (De Haan & Ferreira, 2006)):

$$\mathbb{E}\left[\max_{j\in S^c} V_j \mid X_S, W\right] = \sqrt{2\log(p-s)\, M_p}\,(1+o(1)),$$

Conditionally on $T(\delta)^c$, we have:

$$M_p \geq (1-\delta)\,\mathbb{E}[M_p].$$

Hence, conditionally on $T(\delta)^c$:

$$\begin{aligned}
&\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \mid X_S, W\right] \\
&\geq \frac{1}{\lambda_p}\sqrt{2\log(p-s)(1-\delta)\,\mathbb{E}[M_p]}\,(1+o(1)) \\
&\geq (1+o(1))\,\frac{\sqrt{1-\delta}}{\lambda_p}\sqrt{\frac{2\lambda_p^2 s\log(p-s)}{n-s-1} + \frac{2(n-s)(n_1\sigma_1^2+n_2\sigma_2^2)\log(p-s)}{n^3}} \\
&= (1+o(1))\,\sqrt{1-\delta}\sqrt{\frac{2s\log(p-s)}{n-s-1} + \frac{2(n-s)(n_1\sigma_1^2+n_2\sigma_2^2)\log(p-s)}{\lambda_p^2 n^3}}.
\end{aligned}$$

We consider two cases, depending on the asymptotic behavior of $\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2}$:

- Case 1: $\lim_{p\to+\infty}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2} = 0$.
- Case 2: $\lim_{p\to+\infty}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2} > 0$.

Case 1: Assume $\lim_{p\to+\infty}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2} = 0$. By the above inequality, we have:

$$\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \mid X_S, W\right] \geq (1+o(1))\,\sqrt{1-\delta}\sqrt{\frac{2s\log(p-s)}{n-s-1}}.$$

Using condition (22), the above gives:

$$\begin{aligned}
\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \mid X_S, W\right] &\geq (1+o(1))\,\sqrt{1-\delta}\sqrt{\frac{2s\log(p-s)}{n-s-1}} \\
&\geq (1+o(1))\,\sqrt{1-\delta}\sqrt{\frac{2s\log(p-s)}{(1-\varepsilon)(2s\log(p-s)+s+1)-s-1}} \\
&= (1+o(1))\,\sqrt{1-\delta}\sqrt{\frac{2s\log(p-s)}{2s(1-\varepsilon)\log(p-s)-\varepsilon(s+1)}} \\
&\geq (1+o(1))\,\sqrt{1-\delta}\sqrt{\frac{2s\log(p-s)}{2s(1-\varepsilon)\log(p-s)}} \\
&= (1+o(1))\sqrt{\frac{1-\delta}{1-\varepsilon}}.
\end{aligned}$$

Taking the liminf as $p\to+\infty$ we get, conditionally on $T(\delta)^c$:

$$\liminf_{n\to+\infty}\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \mid X_S, W\right] \geq \sqrt{\frac{1-\delta}{1-\varepsilon}}.$$

Therefore, for n large enough we have:

$$\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \mid X_S, W\right] \geq \frac{1}{2}\left(1+\sqrt{\frac{1-\delta}{1-\varepsilon}}\right) = \frac{1}{2}+\frac{1}{2}\sqrt{\frac{1-\delta}{1-\varepsilon}}.$$

Taking $\delta \coloneqq \varepsilon/2$, we get:

$$\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \;\Big|\; X_S, W\right] \geq \frac{1}{2}+\frac{1}{2}\sqrt{\frac{1-\varepsilon/2}{1-\varepsilon}} =: \kappa > 1.$$

Next, we use the following the result on concentration of Gaussian maxima:

Lemma D.3 (Concentration of Gaussian maxima). *Let $k \in \mathbb{N}$ and $(N_i)_{i=1}^{i=k} \overset{i.i.d.}{\sim} \mathcal{N}\left(0,\tau^2\right)$. Then for any $\eta > 0$, we have:*

$$\begin{cases} \mathbb{P}\left(\max_{i\in[k]} N_i - \mathbb{E}\left[\max_{i\in[k]} N_i\right] > \eta\right) \leq \exp\left(-\frac{\eta^2}{2\tau^2}\right), \\ \mathbb{P}\left(\max_{i\in[k]} N_i - \mathbb{E}\left[\max_{i\in[k]} N_i\right] < -\eta\right) \leq \exp\left(-\frac{\eta^2}{2\tau^2}\right). \end{cases}$$

Proof. See appendix D.2.6. □

Using Lemma D.3 gives, for all $\eta > 0$:

$$\mathbb{P}\left(\max_{j\in S^c}|V_j| < \mathbb{E}\left[\max_{j\in S^c}|V_j|\right] - \eta\right) \leq \exp\left(-\frac{\eta^2}{2\left(1+\delta\right)\mathbb{E}\left[M_p\right]}\right).$$

Setting $\eta \coloneqq (\kappa-1)\,\lambda_p/2$, we get:

$$\begin{aligned} \mathbb{P}\left(\max_{j\in S^c}|V_j| \leq \lambda_p \;\Big|\; T\left(\delta\right)^c\right) &\leq \mathbb{P}\left(\max_{j\in S^c}|V_j| < (\kappa+1)\,\lambda_p/2 \;\Big|\; T\left(\delta\right)^c\right) \\ &\leq \mathbb{P}\left(\max_{j\in S^c}|V_j| < \mathbb{E}\left[\max_{j\in S^c}|V_j|\right] - (\kappa-1)\,\lambda_p/2 \;\Big|\; T\left(\delta\right)^c\right) \\ &\leq \exp\left(-\frac{(\kappa-1)^2\lambda_p^2}{4\left(2+\varepsilon\right)\mathbb{E}\left[M_p\right]}\right) \\ &= \exp\left(-\frac{(\kappa-1)^2\lambda_p^2}{4\left(2+\varepsilon\right)\left(\frac{\lambda_p^2 s}{n-s-1}+\frac{(n-s)\left(n_1\sigma_1^2+n_2\sigma_2^2\right)}{n^3}\right)}\right) \\ &= \exp\left(-\frac{(\kappa-1)^2}{4\left(2+\varepsilon\right)\left(\frac{s}{n-s-1}+\frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2}\right)}\right). \end{aligned}$$

Now note that, because $\lim_{p\to+\infty}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2} = 0$, the above RHS goes to 0 as $p \to +\infty$. Hence, we get:

$$\mathbb{P}\left(\max_{j\in S^c}|V_j| \leq \lambda_p \;\Big|\; T\left(\delta\right)^c\right) \overset{p\to+\infty}{\longrightarrow} 0.$$

Taking the limit in (43) and using the fact that $\mathbb{P}\big(T\left(\delta\right)\big) \overset{p\to+\infty}{\longrightarrow} 0$, we conclude:

$$\lim_{p\to+\infty}\mathbb{P}\left(\max_{j\in S^c}|V_j| \leq \lambda_p\right) = 0.$$

Case 2: Assume $\lim_{p\to+\infty} \frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2} > 0$. Note that this could be $+\infty$. We have:

$$\begin{aligned}
&\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \;\middle|\; X_S, W\right] \\
&\geq (1+o(1))\sqrt{1-\delta}\sqrt{\frac{2s\log(p-s)}{n-s-1} + \frac{2(n-s)(n_1\sigma_1^2+n_2\sigma_2^2)\log(p-s)}{\lambda_p^2 n^3}} \\
&\geq (1+o(1))\sqrt{1-\delta}\sqrt{\frac{2(n-s)(n_1\sigma_1^2+n_2\sigma_2^2)\log(p-s)}{\lambda_p^2 n^3}} \\
&= (1+o(1))\sqrt{1-\delta}\sqrt{2\,\frac{n-s}{n}\,\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2}}\sqrt{\log(p-s)} \\
&\overset{p\to+\infty}{\longrightarrow} +\infty.
\end{aligned}$$

Therefore, for n large enough, we have:

$$\mathbb{E}\left[\max_{j\in S^c}|V_j| \;\middle|\; X_S, W\right] \geq 4\lambda_p.$$

Now using Lemma D.3 on concentration of Gaussian maxima, we have for all $\eta > 0$:

$$\mathbb{P}\left(\max_{j\in S^c}|V_j| < \mathbb{E}\left[\max_{j\in S^c}|V_j|\right] - \eta\right) \leq \exp\left(-\frac{\eta^2}{2(1+\delta)\mathbb{E}[M_p]}\right).$$

Fixing $\eta := \mathbb{E}[\max_{j\in S^c}|V_j|]/2$ and $\delta := \varepsilon/2$ we get, for n large enough:

$$\begin{aligned}
\mathbb{P}\left(\max_{j\in S^c}|V_j| \leq \lambda_p \;\middle|\; T(\delta)^c\right) &\leq \mathbb{P}\left(\max_{j\in S^c}|V_j| < 2\lambda_p \;\middle|\; T(\delta)^c\right) \\
&\leq \mathbb{P}\left(\max_{j\in S^c}|V_j| < \frac{1}{2}\mathbb{E}\left[\max_{j\in S^c}|V_j|\right] \;\middle|\; T(\delta)^c\right) \\
&= \mathbb{P}\left(\max_{j\in S^c}|V_j| < \mathbb{E}\left[\max_{j\in S^c}|V_j|\right] - \frac{1}{2}\mathbb{E}\left[\max_{j\in S^c}|V_j|\right] \;\middle|\; T(\delta)^c\right) \\
&\leq \exp\left(-\frac{\mathbb{E}[\max_{j\in S^c}|V_j|]^2}{4(2+\varepsilon)\mathbb{E}[M_p]}\right) \\
&= \exp\left(-\frac{\mathbb{E}[\max_{j\in S^c}|V_j|]^2/\lambda_p^2}{4(2+\varepsilon)\mathbb{E}[M_p]/\lambda_p^2}\right) \\
&= \exp\left(-\frac{\mathbb{E}[\max_{j\in S^c}|V_j|]^2/\lambda_p^2}{4(2+\varepsilon)\left(\frac{\lambda_p^2 s}{n-s-1} + \frac{(n-s)(n_1\sigma_1^2+n_2\sigma_2^2)}{n^3}\right)/\lambda_p^2}\right) \\
&= \exp\left(-\frac{\left(\mathbb{E}[\max_{j\in S^c}|V_j|]/\lambda_p\right)^2}{4(2+\varepsilon)\left(\frac{s}{n-s-1} + \frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2}\right)}\right).
\end{aligned}$$

Since:

$$\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \;\middle|\; X_S, W\right] \geq (1+o(1))\sqrt{1-\delta}\sqrt{2\,\frac{n-s}{n}\,\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2}}\sqrt{\log(p-s)},$$

the above yields:

$$\begin{aligned}
\mathbb{P}\left(\max_{j\in S^c}|V_j|\le\lambda_p\,\middle|\,T(\delta)^c\right)&\le\exp\left(-\frac{2(1+o(1))(1-\delta)\left(\frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2n^2}\right)\log(p-s)}{4(2+\varepsilon)\left(\frac{s}{n-s-1}+\frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2n^2}\right)}\right)\\
&=\exp\left(-\frac{2(1+o(1))(1-\delta)\left(\frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2n^2}\right)\log(p-s)}{4(2+\varepsilon)(1+o(1))\left(\frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2n^2}\right)}\right)\\
&=\exp\left(-\frac{(1-\delta)\log(p-s)}{2(2+\varepsilon)}(1+o(1))\right)\\
&\overset{p\to+\infty}{\longrightarrow}0.
\end{aligned}$$

Hence, we get:

$$\mathbb{P}\left(\max_{j\in S^c}|V_j|\le\lambda_p\,\middle|\,T(\delta)^c\right)\overset{p\to+\infty}{\longrightarrow}0.$$

Taking the limit in (43) and using the fact that $\mathbb{P}\big(T(\delta)\big)\overset{p\to+\infty}{\longrightarrow}0$, we conclude:

$$\lim_{p\to+\infty}\mathbb{P}\left(\max_{j\in S^c}|V_j|\le\lambda_p\right)=0.$$

□

D.2.3 Showing that $\mathbb{P}\left(\max_{j\in S^c}|V_j|\le\lambda_p\right)\longrightarrow 1$:

Proof of Proposition D.2, part (ii.) . Let:

$$T(\delta):=\big\{|M_p-\mathbb{E}[M_p]|\ge\delta\mathbb{E}[M_p]\big\}.$$

We have, by total probability:

$$\mathbb{P}\left(\max_{j\in S^c}|V_j|>\lambda_p\right)\le\mathbb{P}\left(\max_{j\in S^c}|V_j|>\lambda_p\,\middle|\,T(\delta)^c\right)+\mathbb{P}(T(\delta)).\tag{44}$$

Conditioning on X_S and W:

$$\mathbb{P}\left(\max_{j\in S^c}|V_j|>\lambda_p\,\middle|\,T(\delta)^c\right)=\mathbb{E}\left[\mathbb{P}\left(\max_{j\in S^c}|V_j|>\lambda_p\,\middle|\,X_S,W\right)\,\middle|\,T(\delta)^c\right]$$

Note that, conditionally on (X_S,W), we have $V_j\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,M_p)$, for $j\in S^c$. Using the bound on expectation of Gaussian maxima (see Theorem 5.3.1 in (De Haan & Ferreira, 2006)):

$$\mathbb{E}\left[\max_{j\in S^c}|V_j|\,\middle|\,X_S,W\right]\le\sqrt{2\log\big(2(p-s)\big)M_p}.$$

Conditionally on $T(\delta)^c$, we have:

$$M_p\le(1+\delta)\,\mathbb{E}[M_p].$$

Hence, conditionally on $T(\delta)^c$:

$$\begin{aligned}
&\frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j|\,\middle|\,X_S,W\right]\\
&\le\frac{1}{\lambda_p}\sqrt{2\log\big(2(p-s)\big)(1+\delta)\,\mathbb{E}[M_p]}\\
&=\frac{1}{\lambda_p}\sqrt{2\log\big(2(p-s)\big)(1+\delta)\left(\frac{\lambda_p^2s}{n-s-1}+\frac{(n-s)(n_1\sigma_1^2+n_2\sigma_2^2)}{n^3}\right)}\\
&=\sqrt{2+2\delta}\sqrt{\frac{\log\big(2(p-s)\big)s}{n-s-1}+\frac{\log\big(2(p-s)\big)(n-s)(n_1\sigma_1^2+n_2\sigma_2^2)}{\lambda_p^2n^3}}\\
&=\sqrt{2+2\delta}\\
&\times\sqrt{\frac{s\log 2}{n-s-1}+\frac{s\log(p-s)}{n-s-1}+\frac{n-s}{n}\left(\frac{(n_1\sigma_1^2+n_2\sigma_2^2)\log(p-s)}{\lambda_p^2n^2}+\frac{(n_1\sigma_1^2+n_2\sigma_2^2)\log 2}{\lambda_p^2n^2}\right)}.
\end{aligned}$$

Taking the limsup as $p \to +\infty$ and using conditions (23) and (24) we get, conditionally on $T(\delta)^c$:

$$\begin{aligned}
&\limsup_{p\to+\infty} \frac{1}{\lambda_p}\mathbb{E}\left[\max_{j\in S^c}|V_j| \;\middle|\; X_S, W\right] \\
&\le \limsup_{p\to+\infty} \sqrt{(1+\delta)\left(\frac{2s\log(p-s)}{(1+\varepsilon)(2s\log(p-s)+s+1)-s-1}\right)} \\
&\le \limsup_{p\to+\infty} \sqrt{(1+\delta)\left(\frac{2s\log(p-s)}{(1+\varepsilon)(2s\log(p-s)+s+1)-s-1}\right)} \\
&\le \sqrt{\frac{1+\delta}{1+\varepsilon}}.
\end{aligned}$$

Fix $\delta \coloneqq \varepsilon/4$. By the above, we know that for large enough n, we have:

$$\mathbb{E}\left[\max_{j\in S^c}|V_j| \;\middle|\; X_S, W\right] \le \lambda_p\sqrt{\frac{1+\varepsilon/2}{1+\varepsilon}}.$$

For simplicity of notation, set $\upsilon \coloneqq 1 - \sqrt{\frac{1+\varepsilon/2}{1+\varepsilon}} > 0$. In addition, we by Lemma D.3 on concentration of Gaussian maxima, for all $\eta > 0$:

$$\mathbb{P}\left(\max_{j\in S^c}|V_j| > \eta + \mathbb{E}\left[\max_{j\in S^c}|V_j|\right] \;\middle|\; X_S, W\right) \le \exp\left(-\frac{\eta^2}{2(1+\delta)\,\mathbb{E}[M_p]}\right).$$

Let $\eta \coloneqq \upsilon\lambda_p$. Then we get, for n large enough:

$$\begin{aligned}
\mathbb{P}\left(\max_{j\in S^c}|V_j| > \lambda_p \;\middle|\; X_S, W\right) &\le \mathbb{P}\left(\max_{j\in S^c}|V_j| > \eta + \mathbb{E}\left[\max_{j\in S^c}|V_j|\right] \;\middle|\; X_S, W\right) \\
&\le \exp\left(-\frac{\eta^2}{2(1+\delta)\,\mathbb{E}[M_p]}\right) \\
&= \exp\left(-\frac{\upsilon^2\lambda_p^2}{2(1+\varepsilon/2)\left(\frac{\lambda_p^2 s}{n-s-1} + \frac{(n-s)\left(n_1\sigma_1^2+n_2\sigma_2^2\right)}{n^3}\right)}\right) \\
&= \exp\left(-\frac{\upsilon^2}{2(1+\varepsilon/2)\left(\frac{s}{n-s-1} + \frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2}\right)}\right).
\end{aligned}$$

Substituting in the above, we get:

$$\begin{aligned}
\mathbb{P}\left(\max_{j\in S^c}|V_j| > \lambda_p \;\middle|\; T(\delta)^c\right) &= \mathbb{E}\left[\mathbb{P}\left(\max_{j\in S^c}|V_j| > \lambda_p \;\middle|\; X_S, W\right) \;\middle|\; T(\delta)^c\right] \\
&\le \mathbb{E}\left[\exp\left(-\frac{\upsilon^2}{2(1+\varepsilon/2)\left(\frac{s}{n-s-1} + \frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2}\right)}\right) \;\middle|\; T(\delta)^c\right] \\
&= \exp\left(-\frac{\upsilon^2}{2(1+\varepsilon/2)\left(\frac{s}{n-s-1} + \frac{n-s}{n}\frac{n_1\sigma_1^2+n_2\sigma_2^2}{\lambda_p^2 n^2}\right)}\right) \\
&\overset{p\to+\infty}{\longrightarrow} 0,
\end{aligned}$$

since, by condition (24), we have:

$$0 < \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{\lambda_p^2 n^2} \le \frac{\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)\log(p-s)}{\lambda_p^2 n^2} \overset{p\to+\infty}{\longrightarrow} 0.$$

Taking the limit in (44) and using the fact that $\mathbb{P}(T(\delta)) \to 0$, we conclude:

$$\mathbb{P}\left(\max_{j\in S^c}|V_j| \le \lambda_p\right) \overset{p\to+\infty}{\longrightarrow} 1.$$

□

D.2.4 PROOF OF LEMMA D.1

Proof of Lemma D.1. Recall from (40) that:

$$V_j = \left\{ X_S \left(X_S^T X_S\right)^{-1} \lambda_p \vec{b} - \left[X_S \left(X_S^T X_S\right)^{-1} X_S^T - I_n \right] \frac{\Sigma W}{n} \right\}^T X_j,$$

for all $j \in S^c$. Conditionally on X_S and W, the first term of the RHS above is constant and $\left(X_j\right)_{j \in S^c} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, I_n\right)$. Therefore:

$$V_j \mid X_S, W \sim \mathcal{N}\left(0, A^T A\right),$$

where:

$$A = \left\{ X_S \left(X_S^T X_S\right)^{-1} \lambda_p \vec{b} - \left[X_S \left(X_S^T X_S\right)^{-1} X_S^T - I_n \right] \frac{\Sigma W}{n} \right\} \in \mathbb{R}^n.$$

Expanding the expression of $A^T A$, we get:

$$A^T A = \lambda_p^2 \vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b} + \frac{1}{n^2} W^T \Sigma^T \left[I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T \right]^T \Sigma W.$$

In addition, we have:

$$\mathrm{Cov}\left(X_{j_1}, X_{j_2}\right) = \delta_{j_1 j_2} I_n,$$

therefore:

$$\mathrm{Cov}\left(V_{j_1}, V_{j_2}\right) = \delta_{j_1 j_2} A^T A.$$

We conclude:

$$V \mid X_S, W \sim \mathcal{N}\left(0, M_p I_{|S^c|}\right),$$

where $M_p \coloneqq A^T A = \|A\|_2^2$. ☐

D.2.5 PROOF OF LEMMA D.2

Proof of Lemma D.2. Recall from Lemma D.1 that:

$$M_p = \lambda_p^2 \vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b} + \frac{1}{n^2} W^T \Sigma \left[I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T \right] \Sigma W.$$

By expectation of inverse Wishart matrices, we have:

$$\mathbb{E}\left[\lambda_p^2 \vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b} \right] = \frac{\lambda_p^2}{n - s - 1} \|b\|_2^2.$$

By Gram-Schmidt decomposition of X_S (Meckes, 2019), we write:

$$X_S = QR \in \mathbb{R}^{n \times s}, \tag{45}$$

with $R_{ii} > 0$ and $Q^T Q = I_s$, where $R \in \mathbb{R}^{s \times s}$ is upper triangular (hence invertible) $Q \in \mathbb{R}^{n \times s}$ corresponds to s columns of a $n \times n$ matrix of Haar distribution over the orthogonal group $O(n)$, which we define as follows:

$$U = [P \quad Q] \sim \text{Haar on } O(n).$$

Then we have:

$$\begin{aligned} X_S \left(X_S^T X_S\right)^{-1} X_S^T &= QR \left(R^T Q^T QR\right)^{-1} R^T Q^T \\ &= QR \left(R^T R\right)^{-1} R^T Q^T \\ &= QRR^{-1} \left(R^T\right)^{-1} R^T Q^T \\ &= QQ^T. \end{aligned}$$

Therefore:

$$I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T = I_n - QQ^T = PP^T = [P \quad Q] \begin{bmatrix} I_{n-s} & 0_{(n-s) \times s} \\ 0_{s \times (n-s)} & 0_{s \times s} \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} = UDU^T,$$

where:

$$D = \begin{bmatrix} I_{n-s} & 0_{(n-s)\times s} \\ 0_{s\times(n-s)} & 0_{s\times s} \end{bmatrix}.$$

Note that unlike U, D is deterministic. Since $W \sim \mathcal{N}(0, I_n)$, we have:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n^2} W^T \Sigma \left[I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T\right] \Sigma W \,\middle|\, X_S\right] &= \frac{1}{n^2} \operatorname{tr}\left\{\Sigma \left[I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T\right] \Sigma\right\} \\ &= \frac{1}{n^2} \operatorname{tr}\left(\Sigma^T U D U^T \Sigma\right). \end{aligned}$$

Hence, by total expectation:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n^2} W^T \Sigma \left[I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T\right] \Sigma W\right] &= \mathbb{E}_{X_S}\left[\frac{1}{n^2} \operatorname{tr}\left(\Sigma^T U D U^T \Sigma\right)\right] \\ &= \frac{1}{n^2} \operatorname{tr}\left(\mathbb{E}_{X_S}\left[\Sigma^T U D U^T \Sigma\right]\right) \\ &= \frac{1}{n^2} \operatorname{tr}\left(\Sigma^T \mathbb{E}_U\left[U D U^T\right] \Sigma\right). \end{aligned}$$

By properties of the Haar distribution (see Example 1.8 of (Gu, 2013)), we have:

$$\mathbb{E}_{U \sim \text{Haar}(n)}\left[U^T D U\right] = \frac{\operatorname{tr}(D)}{n} I_n.$$

Substituting in the above, we get:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n^2} W^T \Sigma \left[I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T\right] \Sigma W\right] &= \frac{1}{n^2} \operatorname{tr}\left(\Sigma \mathbb{E}_U\left[U D U^T\right] \Sigma\right) \\ &= \frac{1}{n^2} \operatorname{tr}\left(\frac{\operatorname{tr}(D)}{n} \Sigma\, I_n \Sigma\right) \\ &= \frac{\operatorname{tr}(D)}{n^3} \operatorname{tr}\left(\Sigma^T I_n \Sigma\right) \\ &= \frac{n-s}{n^3} \operatorname{tr}\left(\Sigma^T \Sigma\right) \\ &= \frac{(n-s)\left(n_1 \sigma_1^2 + n_2 \sigma_2^2\right)}{n^3}. \end{aligned}$$

Hence, we conclude:

$$\mathbb{E}\left[M_p\right] = \frac{\lambda_p^2}{n-s-1} \|b\|_2^2 + \frac{(n-s)\left(n_1 \sigma_1^2 + n_2 \sigma_2^2\right)}{n^3}.$$

We now compute $\text{Var}(M_p)$. For simplicity, we write $\Lambda = I_n - X_S \left(X_S^T X_S\right)^{-1} X_S^T = U D U^T$, so that:

$$M_p = \lambda_p^2 \vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b} + \frac{1}{n^2} W^T \Sigma \Lambda \Sigma W.$$

Therefore:

$$M_p^2 = \lambda_p^4 \left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right]^2 + \frac{1}{n^4} \left(W^T \Sigma \Lambda \Sigma W\right)^2 + 2\,\frac{\lambda_p^2}{n^2} \left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right] \left(W^T \Sigma \Lambda \Sigma W\right).$$

Let:

$$T_1 \coloneqq \left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right]^2, \quad T_2 \coloneqq \left(W^T \Sigma \Lambda \Sigma W\right)^2, \quad T_3 \coloneqq \left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right] \left(W^T \Sigma \Lambda \Sigma W\right),$$

so that:

$$M_p^2 = \lambda_p^4 T_1 + \frac{1}{n^4} T_2 + 2\,\frac{\lambda_p^2}{n^2} T_3.$$

We start by computing $\mathbb{E}[T_1]$. Recall that $X_S^T X_S \sim \mathcal{W}_s(I_s, n)$. We use the following result for second moments of inverse Wishart matrices from (Siskind, 1972).

Lemma D.4 (Second moment of inverse Wishart matrices, (Siskind, 1972))**.** *Let* $a, b \in \mathbb{N}$ *with* $a > b + 3$*. Let* $t \in \mathbb{R}^s$ *and* $M \in R^{s \times s}$ *and* $A \sim \mathcal{W}_a(T, b)$*. Then:*

$$(b-a)(b-a-3)\,\mathbb{E}\left[A^{-1}tt^TA^{-1}\right] = T^{-1}tt^TT^{-1} + T^{-1}\left(t^TT^{-1}t\right)/(b-a-1).$$

Setting $b \coloneqq n$; $a \coloneqq s$; $t \coloneqq \vec{b}$; $T \coloneqq I_s$ and $A \coloneqq X_S^TX_S$ in Lemma D.4, we get:

$$(n-s)(n-s-3)\,\mathbb{E}\left[\left(X_S^TX_S\right)^{-1}\vec{b}\vec{b}^T\left(X_S^TX_S\right)^{-1}\right] = \vec{b}\vec{b}^T + \left\|\vec{b}\right\|_2^2 I_s/(n-s-1). \tag{46}$$

Multiplying by the LHS by $\vec{b}^T$ on the left and by $\vec{b}$ on right we get:

$$\mathbb{E}\left[T_1\right] = \frac{1}{(n-s)(n-s-3)}\left(1 + \frac{1}{n-s-1}\right)\left(\vec{b}^T\vec{b}\right)^2.$$

Hence:

$$\mathbb{E}\left[T_1\right] = \frac{1}{(n-s)(n-s-3)}\left(1 + \frac{1}{n-s-1}\right)\left\|\vec{b}\right\|_2^4. \tag{47}$$

Now let us compute $\mathbb{E}\left[T_2\right]$. Since $W \sim \mathcal{N}(0, I_n)$, we have by moments of the multivariate normal distribution (see Theorem 5.2a and Theorem 5.2b in (Rencher & Schaalje, 2008)):

$$\mathbb{E}\left[T_2 \mid X_S\right] = 2\,\mathrm{tr}\left[(\Sigma\Lambda\Sigma)^2\right] + \left[\mathrm{tr}\,(\Sigma\Lambda\Sigma)\right]^2.$$

Lemma D.5. *We have:*

$$\begin{aligned}
&\mathbb{E}\left\{\left[\mathrm{tr}\,(\Sigma\Lambda\Sigma)\right]^2\right\} \\
&\quad = \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2\left\{\frac{(n+1)(n-s)}{(n-1)\,n\,(n+2)}\right\}\left\{n - s - \frac{2}{n+1}\right\} \\
&\qquad + \left(n_1\sigma_1^4 + n_2\sigma_2^4\right)\left\{\frac{(n-s)(n-s+2)}{n\,(n+2)} - \frac{(n+1)(n-s)\left(n-s-\frac{2}{(n+1)}\right)}{(n-1)\,n\,(n+2)}\right\}.
\end{aligned}$$

and:

$$\begin{aligned}
\mathbb{E}\left\{\mathrm{tr}\left[(\Sigma\Lambda\Sigma)^2\right]\right\} &= \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2\left\{\frac{s\,(n-s)}{(n-1)\,n\,(n+2)}\right\} \\
&\quad + \left(n_1\sigma_1^4 + n_2\sigma_2^4\right)\left\{\frac{n-s}{n\,(n+2)}\right\}\left\{n - s + 1 + \frac{n-s-1}{n-1}\right\}.
\end{aligned}$$

Proof. See appendix D.2.7. □

Substituting in the expression of $\mathbb{E}\left[T_2 \mid X_S\right]$ and using total expectation, we get:

$$
\begin{aligned}
&\mathbb{E}\left[T_2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[T_2 \mid X_S\right]\right] \\
&= \mathbb{E}\left[2\operatorname{tr}\left[(\Sigma\Lambda\Sigma)^2\right] + \left[\operatorname{tr}(\Sigma\Lambda\Sigma)\right]^2\right] \\
&= 2\,\mathbb{E}\left[\operatorname{tr}\left[(\Sigma\Lambda\Sigma)^2\right]\right] + \mathbb{E}\left[\left[\operatorname{tr}(\Sigma\Lambda\Sigma)\right]^2\right] \\
&= \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2 \left\{\frac{2s\,(n-s)}{(n-1)\,n\,(n+2)}\right\} \\
&\quad + \left(n_1\sigma_1^4 + n_2\sigma_2^4\right) \left\{\frac{2\,(n-s)}{n\,(n+2)}\right\} \left\{n-s+1+\frac{n-s-1}{n-1}\right\} \\
&\quad + \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2 \left\{\frac{(n+1)\,(n-s)}{(n-1)\,n\,(n+2)}\right\} \left\{n-s-\frac{2}{n+1}\right\} \\
&\quad + \left(n_1\sigma_1^4 + n_2\sigma_2^4\right) \left\{\frac{(n-s)\,(n-s+2)}{n\,(n+2)} - \frac{(n+1)\,(n-s)\left(n-s-\frac{2}{(n+1)}\right)}{(n-1)\,n\,(n+2)}\right\} \\
&= \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2 \left\{\frac{n-s}{(n-1)\,n\,(n+2)}\right\} \left\{2s + (n+1)\left(n-s-\frac{2}{n+1}\right)\right\} \\
&\quad + \left(n_1\sigma_1^4 + n_2\sigma_2^4\right) \left\{\frac{n-s}{n\,(n+2)}\right\} \\
&\qquad \times \left\{2n-2s+2+\frac{2\,(n-s-1)}{n-1} + n-s+2 - \frac{(n+1)\left(n-s-\frac{2}{(n+1)}\right)}{n-1}\right\} \\
&= \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2 \left\{\frac{n-s}{(n-1)\,n\,(n+2)}\right\} \left\{2s + (n+1)\left(n-s-\frac{2}{n+1}\right)\right\} \\
&\quad + \left(n_1\sigma_1^4 + n_2\sigma_2^4\right) \left\{\frac{n-s}{n\,(n+2)}\right\} \\
&\qquad \times \left\{3n-3s+4+\frac{2\,(n-s-1)}{n-1} - \frac{(n+1)\left(n-s-\frac{2}{(n+1)}\right)}{n-1}\right\}.
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
\mathbb{E}\left[T_2\right] = &\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2 \left\{\frac{n-s}{(n-1)\,n\,(n+2)}\right\} \left\{2s + (n+1)\left(n-s-\frac{2}{n+1}\right)\right\} \\
&+ \left(n_1\sigma_1^4 + n_2\sigma_2^4\right) \left\{\frac{n-s}{n\,(n+2)}\right\} \\
&\times \left\{3n-3s+4+\frac{2\,(n-s-1)}{n-1} - \frac{(n+1)\left(n-s-\frac{2}{(n+1)}\right)}{n-1}\right\}.
\end{aligned} \tag{48}
$$

Finally, we compute $\mathbb{E}\left[T_3\right]$. We have:

$$
\begin{aligned}
\mathbb{E}\left[T_3 \mid X_S\right] &= \mathbb{E}\left\{\left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right] \left(W^T \Sigma\Lambda\Sigma W\right) \,\middle|\, X_S\right\} \\
&= \left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right] \mathbb{E}\left\{W^T \Sigma\Lambda\Sigma W \,\middle|\, X_S\right\} \\
&= \left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right] \operatorname{tr}\left(\Sigma\Lambda\Sigma\right).
\end{aligned}
$$

Recall from (45) that:

$$X_S = QR,$$

where $Q \in \mathbb{R}^{n\times s}$ and $R \in \mathbb{R}^{s\times s}$ are independent, R is upper triangular with $R_{ii} > 0$ and $Q^T Q = I_s$. Therefore, by total expectation:

$$
\begin{aligned}
\mathbb{E}\left[T_3\right] &= \mathbb{E}\Big[\mathbb{E}\left[T_3 \mid X_S\right]\Big] \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right] \operatorname{tr}\left(\Sigma\Lambda\Sigma\right)\right\} \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right] \operatorname{tr}\left[\Sigma\left(I_n - X_S\left(X_S^T X_S\right)^{-1} X_S^T\right)\Sigma\right]\right\} \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(R^T Q^T Q R\right)^{-1} \vec{b}\right] \operatorname{tr}\left[\Sigma\left(I_n - QR\left(R^T Q^T Q R\right)^{-1} R^T Q^T\right)\Sigma\right]\right\} \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(R^T R\right)^{-1} \vec{b}\right] \operatorname{tr}\left[\Sigma\left(I_n - QR\left(R^T R\right)^{-1} R^T Q^T\right)\Sigma\right]\right\} \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(R^T R\right)^{-1} \vec{b}\right] \operatorname{tr}\left[\Sigma\left(I_n - QRR^{-1}\left(R^T\right)^{-1} R^T Q^T\right)\Sigma\right]\right\} \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(R^T R\right)^{-1} \vec{b}\right] \operatorname{tr}\left[\Sigma\left(I_n - QQ^T\right)\Sigma\right]\right\} \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(R^T R\right)^{-1} \vec{b}\right]\right\} \mathbb{E}\left\{\operatorname{tr}\left[\Sigma\left(I_n - QQ^T\right)\Sigma\right]\right\} \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(R^T R\right)^{-1} \vec{b}\right]\right\} \mathbb{E}\left\{\operatorname{tr}\left[\Sigma\left(I_n - QQ^T\right)\Sigma\right]\right\} \\
&= \mathbb{E}\left\{\left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right]\right\} \mathbb{E}\left\{\operatorname{tr}\left(\Sigma U D U^T \Sigma\right)\right\}.
\end{aligned}
$$

By expectation of inverse Wishart matrices (Anderson et al., 1958), we have:

$$
\mathbb{E}\left\{\left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right]\right\} = \frac{1}{n-s-1}\left\|\vec{b}\right\|_2^2.
$$

Using this in the above expression of $\mathbb{E}\left[T_3\right]$ yields:

$$
\begin{aligned}
\mathbb{E}\left[T_3\right] &= \mathbb{E}\left\{\left[\vec{b}^T \left(X_S^T X_S\right)^{-1} \vec{b}\right]\right\} \mathbb{E}\left\{\operatorname{tr}\left(\Sigma U D U^T \Sigma\right)\right\} \\
&= \frac{1}{n-s-1}\left\|\vec{b}\right\|_2^2 \operatorname{tr}\left\{\mathbb{E}\Big[\Sigma U D U^T \Sigma\Big]\right\} \\
&= \frac{1}{n-s-1}\left\|\vec{b}\right\|_2^2 \operatorname{tr}\left\{\Sigma \mathbb{E}\big[U D U^T\big]\Sigma\right\} \\
&= \frac{1}{n-s-1}\left\|\vec{b}\right\|_2^2 \operatorname{tr}\left\{\Sigma\left(\frac{\operatorname{tr}(D)}{n} I_n\right)\Sigma\right\} \\
&= \frac{\operatorname{tr}(D)}{n(n-s-1)}\left\|\vec{b}\right\|_2^2 \operatorname{tr}\left(\Sigma^2\right) \\
&= \frac{(n-s)\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)}{n(n-s-1)}\left\|\vec{b}\right\|_2^2.
\end{aligned}
$$

Therefore:

$$
\mathbb{E}\left[T_3\right] = \frac{(n-s)\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)}{n(n-s-1)}\left\|\vec{b}\right\|_2^2. \tag{49}
$$

Now we have:

$$
\begin{aligned}
&\mathbb{E}\left[M_p^2\right] - \left(\mathbb{E}\left[M_p\right]\right)^2 \\
&= \lambda_p^4 \mathbb{E}\left[T_1\right] + \frac{1}{n^4}\mathbb{E}\left[T_2\right] + \frac{2\lambda_p^2}{n^2}\mathbb{E}\left[T_3\right] - \left(\frac{\lambda_p^2 s}{n-s-1} + \frac{(n-s)\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)}{n^3}\right)^2 \\
&= \lambda_p^4\left(\mathbb{E}\left[T_1\right] - \frac{s^2}{(n-s-1)^2}\right) + \left(\frac{\mathbb{E}\left[T_2\right]}{n^4} - \frac{(n-s)^2\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2}{n^6}\right) \\
&\quad + \frac{2\lambda_p^2}{n^3}\left(\mathbb{E}\left[T_3\right] n - \frac{s(n-s)\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)}{n-s-1}\right).
\end{aligned}
$$

Let:

$$H_1 \coloneqq \lambda_p^4 \left(\mathbb{E}\left[T_1\right] - \frac{s^2}{\left(n-s-1\right)^2} \right),$$

$$H_2 \coloneqq \frac{\mathbb{E}\left[T_2\right]}{n^4} - \frac{\left(n-s\right)^2 \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2}{n^6},$$

and:

$$H_3 \coloneqq \frac{2\lambda_p^2}{n^3} \left\{ \mathbb{E}\left[T_3\right] n - \frac{s\left(n-s\right)\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)}{n-s-1} \right\}.$$

Then we have, using (47):

$$\begin{aligned}
\frac{H_1}{\left(\mathbb{E}\left[M_p\right]\right)^2} &= \frac{\lambda_p^4 \left(\mathbb{E}\left[T_1\right] - \frac{s^2}{(n-s-1)^2} \right)}{\left(\frac{\lambda_p^2 s}{n-s-1} + \frac{(n-s)\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)}{n^3} \right)^2} \\
&\leq \frac{\lambda_p^4 s^2 \left(\frac{1}{(n-s)(n-s-3)} \left(1 + \frac{1}{n-s-1}\right) - \frac{1}{(n-s-1)^2} \right)}{\frac{\lambda_p^4 s^2}{(n-s-1)^2}} \\
&= \left(n-s-1\right)^2 \left(\frac{1}{\left(n-s\right)\left(n-s-3\right)} \left(1 + \frac{1}{n-s-1}\right) - \frac{1}{\left(n-s-1\right)^2} \right) \\
&= \frac{\left(n-s-1\right)^2}{\left(n-s\right)\left(n-s-3\right)} \left(1 + \frac{1}{n-s-1}\right) - 1 \\
&= o\left(1\right).
\end{aligned}$$

Similarly, using (48):

$$\begin{aligned}
&H_2 \\
&= \frac{\mathbb{E}\left[T_2\right]}{n^4} - \frac{\left(n-s\right)^2 \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2}{n^6} \\
&= \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2 \left\{ \frac{2s\left(n-s\right)}{\left(n-1\right) n^5 \left(n+2\right)} + \frac{\left(n-s\right)\left(n+1\right)}{\left(n-1\right) n^5 \left(n+2\right)} \left(n - s - \frac{2}{n+1}\right) - \frac{\left(n-s\right)^2}{n^6} \right\} \\
&+ \left(n_1\sigma_1^4 + n_2\sigma_2^4\right) \left\{ \frac{n-s}{n^5\left(n+2\right)} \right\} \left\{ 3n - 3s + 4 + \frac{2\left(n-s-1\right)}{n-1} - \frac{\left(n+1\right)\left(n - s - \frac{2}{(n+1)}\right)}{n-1} \right\}.
\end{aligned}$$

We have:

$$\left(\mathbb{E}\left[M_p\right]\right)^2 \geq \frac{\left(n-s\right)^2 \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2}{n^6}.$$

Hence:

$$
\begin{aligned}
&\frac{H_2}{\left(\mathbb{E}\left[M_p\right]\right)^2} \\
&\leq \left\{\frac{n^6}{(n-s)^2}\right\}\left\{\frac{2s\,(n-s)}{(n-1)\,n^5\,(n+2)} + \frac{(n-s)\,(n+1)}{(n-1)\,n^5\,(n+2)}\left(n-s-\frac{2}{n+1}\right) - \frac{(n-s)^2}{n^6}\right\} \\
&\quad + \frac{n_1\sigma_1^4 + n_2\sigma_2^4}{\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2}\left\{\frac{n^6}{(n-s)^2}\right\}\left\{\frac{(n-s)}{n^5\,(n+2)}\right\} \\
&\qquad \times \left\{3n - 3s + 4 + \frac{2\,(n-s-1)}{n-1} - \frac{(n+1)\left(n-s-\frac{2}{(n+1)}\right)}{n-1}\right\} \\
&= \left\{\frac{2sn}{(n-1)\,(n-s)\,(n+2)} + \frac{n\,(n+1)}{(n-s)\,(n-1)\,(n+2)}\left(n-s-\frac{2}{n+1}\right) - 1\right\} \\
&\quad + \frac{n_1\sigma_1^4 + n_2\sigma_2^4}{\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2}\left\{\frac{n}{(n+2)\,(n-s)}\right\} \\
&\qquad \times \left\{3n - 3s + 4 + \frac{2\,(n-s-1)}{n-1} - \frac{(n+1)\left(n-s-\frac{2}{(n+1)}\right)}{n-1}\right\} \\
&= \left\{\frac{2sn}{(n-1)\,(n-s)\,(n+2)} + \frac{n\,(n+1)}{(n-s)\,(n-1)\,(n+2)}\left(n-s-\frac{2}{n+1}\right) - 1\right\} \\
&\quad + \frac{n_1\sigma_1^4 + n_2\sigma_2^4}{\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2} \\
&\quad \times \left\{\frac{3n}{(n+2)} + \frac{4n}{(n+2)\,(n-s)} + \frac{2\,(n-s-1)\,n}{(n-s)\,(n-1)\,(n+2)} - \frac{n\,(n+1)\left(n-s-\frac{2}{(n+1)}\right)}{(n-s)\,(n-1)\,(n+2)}\right\} \\
&= \left\{\frac{2sn}{(n-1)\,(n-s)\,(n+2)} + \frac{n\,(n+1)}{(n-s)\,(n-1)\,(n+2)}\left(n-s-\frac{2}{n+1}\right) - 1\right\} \\
&\quad + \left\{\frac{n_1\sigma_1^4}{\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2} + \frac{n_2\sigma_2^4}{\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2}\right\} \\
&\qquad \times \left\{\frac{3n}{(n+2)} + \frac{4n}{(n+2)\,(n-s)} + \frac{2\,(n-s-1)\,n}{(n-s)\,(n-1)\,(n+2)} - \frac{n\,(n+1)\left(n-s-\frac{2}{(n+1)}\right)}{(n-s)\,(n-1)\,(n+2)}\right\} \\
&\leq \left\{\frac{2sn}{(n-1)\,(n-s)\,(n+2)} + \frac{n\,(n+1)}{(n-s)\,(n-1)\,(n+2)}\left(n-s-\frac{2}{n+1}\right) - 1\right\} \\
&\quad + \left\{\frac{1}{n_1} + \frac{1}{n_2}\right\} \\
&\qquad \times \left\{\frac{3n}{(n+2)} + \frac{4n}{(n+2)\,(n-s)} + \frac{2\,(n-s-1)\,n}{(n-s)\,(n-1)\,(n+2)} - \frac{n\,(n+1)\left(n-s-\frac{2}{(n+1)}\right)}{(n-s)\,(n-1)\,(n+2)}\right\} \\
&= o\left(1\right).
\end{aligned}
$$

For H_3, using (49):

$$\begin{aligned} H_3 &\coloneqq \frac{2\lambda_p^2}{n^3}\left\{\mathbb{E}\left[T_3\right]n - \frac{s\left(n-s\right)\left(n_1\sigma_1^2+n_2\sigma_2^2\right)}{n-s-1}\right\} \\ &= \frac{2\lambda_p^2}{n^3}\left\{\frac{n\left(n-s\right)s\left(n_1\sigma_1^2+n_2\sigma_2^2\right)}{n\left(n-s-1\right)} - \frac{s\left(n-s\right)\left(n_1\sigma_1^2+n_2\sigma_2^2\right)}{n-s-1}\right\} \\ &= 0. \end{aligned}$$

Therefore, we conclude:

$$\frac{\mathbb{E}\left[M_p^2\right]-\left(\mathbb{E}\left[M_p\right]\right)^2}{\left(\mathbb{E}\left[M_p\right]\right)^2} = \frac{H_1}{\left(\mathbb{E}\left[M_p\right]\right)^2} + \frac{H_2}{\left(\mathbb{E}\left[M_p\right]\right)^2} + \frac{H_3}{\left(\mathbb{E}\left[M_p\right]\right)^2} = o\left(1\right).$$

□

D.2.6 Proof of Lemma D.3

Proof of Lemma D.3. Define:

$$\begin{aligned} f\colon \mathbb{R}^k &\longrightarrow \mathbb{R} \\ w &\longmapsto \max_{i\in[k]} w_i. \end{aligned}$$

Note that for any $u, v \in \mathbb{R}^k$:

$$\begin{aligned} \left|f\left(u\right)-f\left(v\right)\right| &= \left|\max_{i\in[k]} u_i - \max_{i\in[k]} v_i\right| \\ &\leq \max_{i\in[k]} \left|u_i - v_i\right| \\ &\leq \sqrt{\sum_{i\in[k]}\left(u_i-v_i\right)^2} \\ &= \left\|u-v\right\|_2. \end{aligned}$$

Therefore f is 1-Lipschitz. Assume $\tau^2 = 1$. By Gaussian concentration of measure for Lipschitz functions (Ledoux, 2001; Massart, 2007), we have for all $t \geq 0$:

$$\begin{cases} \mathbb{P}\left(\max_{i\in[k]} N_i - \mathbb{E}\left[\max_{i\in[k]} N_i\right] > t\right) \leq \exp\left(-t^2/2\right), \\ \mathbb{P}\left(\max_{i\in[k]} N_i - \mathbb{E}\left[\max_{i\in[k]} N_i\right] < -t\right) \leq \exp\left(-t^2/2\right). \end{cases}$$

The result for general τ^2 follows by substituting $t \coloneqq \eta/\tau$. □

D.2.7 PROOF OF LEMMA D.5

Proof of Lemma D.5. We have, by Einstein notation:

$$\begin{aligned}
&\left[\operatorname{tr}(\Sigma\Lambda\Sigma)\right]^2 \\
&= \left(\sum_{a=1}^{n}\left[\Sigma U D U^T \Sigma\right]_{aa}\right)^2 \\
&= \left(\sum_{a=1}^{n}\sum_{b=1}^{n}\sum_{c=1}^{n}\sum_{d=1}^{n}\sum_{e=1}^{n}\Sigma_{ab}U_{bc}D_{cd}\left[U^T\right]_{de}\Sigma_{ea}\right)^2 \\
&= \left(\sum_{a=1}^{n}\sum_{c=1}^{n}\Sigma_{aa}U_{ac}D_{cc}U_{ac}\Sigma_{aa}\right)^2 \\
&= \left(\sum_{a=1}^{n}\sum_{c=1}^{n}D_{aa}U_{ac}^2\Sigma_{cc}^2\right)^2 \\
&= \sum_{a=1}^{n}\sum_{c=1}^{n}D_{cc}^2U_{ac}^4\Sigma_{aa}^4 + \sum_{a=1}^{n}\sum_{c\neq d\in[n]}D_{cc}D_{dd}U_{ac}^2U_{ad}^2\Sigma_{aa}^4 \\
&\quad + \sum_{a\neq b\in[n]}\sum_{c=1}^{n}D_{cc}^2U_{ac}^2U_{bc}^2\Sigma_{aa}^2\Sigma_{bb}^2 + \sum_{a\neq b\in[n]}\sum_{c\neq d\in[n]}D_{cc}D_{dd}U_{ac}^2U_{bd}^2\Sigma_{aa}^2\Sigma_{bb}^2.
\end{aligned}$$

Therefore:

$$\begin{aligned}
\mathbb{E}\left\{\left[\operatorname{tr}(\Sigma\Lambda\Sigma)\right]^2\right\} &= \sum_{a=1}^{n}\sum_{c=1}^{n}D_{cc}^2\Sigma_{aa}^4\mathbb{E}\left[U_{ac}^4\right] \\
&\quad + \sum_{a=1}^{n}\sum_{c\neq d\in[n]}D_{cc}D_{dd}\Sigma_{aa}^4\mathbb{E}\left[U_{ac}^2U_{ad}^2\right] \\
&\quad + \sum_{a\neq b\in[n]}\sum_{c=1}^{n}D_{cc}^2\Sigma_{aa}^2\Sigma_{bb}^2\mathbb{E}\left[U_{ac}^2U_{bc}^2\right] \\
&\quad + \sum_{a\neq b\in[n]}\sum_{c\neq d\in[n]}D_{cc}D_{dd}\Sigma_{aa}^2\Sigma_{bb}^2\mathbb{E}\left[U_{ac}^2U_{bd}^2\right].
\end{aligned}$$

We use the following result from (Meckes, 2019) on fourth-moments of Haar(n) matrices:

Lemma D.6 (Fourth-moments of Haar(n) matrices, Lemma 2.22 in (Meckes, 2019))**.** *Let $U \sim$ Haar(n). Then for all $i, j, r, s, \alpha, \beta, \lambda, \mu \in [n]$ we have:*

$$\begin{aligned}
&\mathbb{E}\left[U_{ij}U_{rs}U_{\alpha\beta}U_{\lambda\mu}\right] \\
&= -\frac{1}{(n-1)\,n\,(n+2)}\Big[\delta_{ir}\delta_{\alpha\lambda}\delta_{j\beta}\delta_{s\mu} + \delta_{ir}\delta_{\alpha\lambda}\delta_{j\mu}\delta_{s\beta} + \delta_{i\alpha}\delta_{r\lambda}\delta_{js}\delta_{\beta\mu} \\
&\qquad + \delta_{i\alpha}\delta_{r\lambda}\delta_{j\mu}\delta_{\beta s} + \delta_{i\lambda}\delta_{r\alpha}\delta_{js}\delta_{\beta\mu} + \delta_{i\lambda}\delta_{r\alpha}\delta_{j\beta}\delta_{s\mu}\Big] \\
&\quad + \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{ir}\delta_{\alpha\lambda}\delta_{js}\delta_{\beta\mu} + \delta_{i\alpha}\delta_{r\lambda}\delta_{j\beta}\delta_{s\mu} + \delta_{i\lambda}\delta_{r\alpha}\delta_{j\mu}\delta_{s\beta}\Big].
\end{aligned}$$

Substituting: $i, r, \alpha, \lambda \coloneqq a$; and $j, s, \beta, \mu \coloneqq c$, we get:

$$\begin{aligned}
&\mathbb{E}\left[U_{ac}^4\right] \\
&= -\frac{1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc} + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc} + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc} \\
&\qquad\qquad + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc} + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc} + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc}\Big] \\
&\quad + \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc} + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc} + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{cc}\Big] \\
&= -\frac{6}{(n-1)\,n\,(n+2)} + \frac{3\,(n+1)}{(n-1)\,n\,(n+2)} \\
&= \frac{3}{n\,(n+2)}.
\end{aligned}$$

Thus:

$$\mathbb{E}\left[U_{ac}^4\right] = \frac{3}{n\,(n+2)}. \tag{50}$$

Now substituting $i, r, \alpha, \lambda \coloneqq a$; $j, s \coloneqq c$; and $\beta, \mu \coloneqq d$, we get:

$$\begin{aligned}
&\mathbb{E}\left[U_{ac}^2 U_{ad}^2\right] \\
&= -\frac{1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{aa}\delta_{cd}\delta_{cd} + \delta_{aa}\delta_{aa}\delta_{cd}\delta_{cd} + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{dd} \\
&\qquad\qquad + \delta_{aa}\delta_{aa}\delta_{cd}\delta_{dc} + \delta_{aa}\delta_{aa}\delta_{cc}\delta_{dd} + \delta_{aa}\delta_{aa}\delta_{cd}\delta_{cd}\Big] \\
&\quad + \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{aa}\delta_{cc}\delta_{dd} + \delta_{aa}\delta_{aa}\delta_{cd}\delta_{cd} + \delta_{aa}\delta_{aa}\delta_{cd}\delta_{cd}\Big] \\
&= -\frac{2}{(n-1)\,n\,(n+2)} + \frac{n+1}{(n-1)\,n\,(n+2)} \\
&= \frac{1}{n\,(n+2)}.
\end{aligned}$$

Thus:

$$\mathbb{E}\left[U_{ac}^2 U_{ad}^2\right] = \frac{1}{n\,(n+2)}. \tag{51}$$

Now substituting $i, r \coloneqq a$; $j, s, \beta, \mu \coloneqq c$; and $\alpha, \lambda \coloneqq b$, we get:

$$\begin{aligned}
&\mathbb{E}\left[U_{ac}^2 U_{bc}^2\right] \\
&= -\frac{1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{bb}\delta_{cc}\delta_{cc} + \delta_{aa}\delta_{bb}\delta_{cc}\delta_{cc} + \delta_{ab}\delta_{ab}\delta_{cc}\delta_{cc} \\
&\qquad\qquad + \delta_{ab}\delta_{ab}\delta_{cc}\delta_{cc} + \delta_{ab}\delta_{ab}\delta_{cc}\delta_{cc} + \delta_{ab}\delta_{ab}\delta_{cc}\delta_{cc}\Big] \\
&\quad + \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{bb}\delta_{cc}\delta_{cc} + \delta_{ab}\delta_{ab}\delta_{cc}\delta_{cc} + \delta_{ab}\delta_{ab}\delta_{cc}\delta_{cc}\Big] \\
&\quad + \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{aa}\delta_{cc}\delta_{dd} + \delta_{aa}\delta_{aa}\delta_{cd}\delta_{cd} + \delta_{aa}\delta_{aa}\delta_{cd}\delta_{cd}\Big] \\
&= -\frac{2}{(n-1)\,n\,(n+2)} + \frac{n+1}{(n-1)\,n\,(n+2)} \\
&= \frac{1}{n\,(n+2)}.
\end{aligned}$$

Thus:

$$\mathbb{E}\left[U_{ac}^2 U_{bc}^2\right] = \frac{1}{n\,(n+2)}. \tag{52}$$

Now substituting $i, r \coloneqq a$; $j, s \coloneqq c$; $\alpha, \lambda \coloneqq b$; and $\beta, \mu \coloneqq d$, we get:

$$\begin{aligned}
&\mathbb{E}\left[U_{ac}^2 U_{bd}^2\right] \\
&\quad = -\frac{1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{bb}\delta_{cd}\delta_{cd} + \delta_{aa}\delta_{bb}\delta_{cd}\delta_{cd} + \delta_{ab}\delta_{ab}\delta_{cc}\delta_{dd} \\
&\qquad\qquad\qquad\qquad + \delta_{ab}\delta_{ab}\delta_{cd}\delta_{dc} + \delta_{ab}\delta_{ab}\delta_{cc}\delta_{dd} + \delta_{ab}\delta_{ab}\delta_{cd}\delta_{cd}\Big] \\
&\qquad + \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{aa}\delta_{bb}\delta_{cc}\delta_{dd} + \delta_{ab}\delta_{ab}\delta_{cd}\delta_{cd} + \delta_{ab}\delta_{ab}\delta_{cd}\delta_{cd}\Big] \\
&\quad = \frac{n+1}{(n-1)\,n\,(n+2)}.
\end{aligned}$$

Thus:

$$\mathbb{E}\left[U_{ac}^2 U_{bd}^2\right] = \frac{n+1}{(n-1)\,n\,(n+2)}. \tag{53}$$

Substituting (50), (51), (52), and (53) in the expression of $\mathbb{E}\left\{\left[\operatorname{tr}(\Sigma\Lambda\Sigma)\right]^2\right\}$ above, we get:

$$
\begin{aligned}
&\mathbb{E}\left\{\left[\operatorname{tr}(\Sigma\Lambda\Sigma)\right]^2\right\} \\
&\quad= \sum_{a=1}^{n}\sum_{c=1}^{n} D_{cc}^2 \Sigma_{aa}^4 \mathbb{E}\left[U_{ac}^4\right] + \sum_{a=1}^{n}\sum_{c\neq d\in[n]} D_{cc}D_{dd}\Sigma_{aa}^4 \mathbb{E}\left[U_{ac}^2 U_{ad}^2\right] \\
&\qquad+ \sum_{a\neq b\in[n]}\sum_{c=1}^{n} D_{cc}^2 \Sigma_{aa}^2\Sigma_{bb}^2 \mathbb{E}\left[U_{ac}^2 U_{bc}^2\right] + \sum_{a\neq b\in[n]}\sum_{c\neq d\in[n]} D_{cc}D_{dd}\Sigma_{aa}^2\Sigma_{bb}^2 \mathbb{E}\left[U_{ac}^2 U_{bd}^2\right] \\
&\quad= \frac{3}{n(n+2)}\left(\sum_{a=1}^{n}\Sigma_{aa}^4\right)\left(\sum_{c=1}^{n}D_{cc}^2\right) + \frac{1}{n(n+2)}\left(\sum_{a=1}^{n}\Sigma_{aa}^4\right)\left(\sum_{c\neq d\in[n]}D_{cc}D_{dd}\right) \\
&\qquad+ \frac{1}{n(n+2)}\left(\sum_{a\neq b\in[n]}\Sigma_{aa}^2\Sigma_{bb}^2\right)\left(\sum_{c=1}^{n}D_{cc}^2\right) \\
&\qquad+ \frac{n+1}{(n-1)n(n+2)}\left(\sum_{a\neq b\in[n]}\Sigma_{aa}^2\Sigma_{bb}^2\right)\left(\sum_{c\neq d\in[n]}D_{cc}D_{dd}\right) \\
&\quad= \frac{1}{n(n+2)}\left(\sum_{a=1}^{n}\Sigma_{aa}^4\right)\left\{3\left(\sum_{c=1}^{n}D_{cc}^2\right) + \left(\sum_{c\neq d\in[n]}D_{cc}D_{dd}\right)\right\} \\
&\qquad+ \frac{n+1}{(n-1)n(n+2)}\left(\sum_{a\neq b\in[n]}\Sigma_{aa}^2\Sigma_{bb}^2\right)\left\{\frac{n-1}{n+1}\left(\sum_{c=1}^{n}D_{cc}^2\right) + \left(\sum_{c\neq d\in[n]}D_{cc}D_{dd}\right)\right\} \\
&\quad= \frac{1}{n(n+2)}\left(\sum_{a=1}^{n}\Sigma_{aa}^4\right)\left\{2\left(\sum_{c=1}^{n}D_{cc}^2\right) + \left(\sum_{c=1}^{n}D_{cc}\right)^2\right\} \\
&\qquad+ \frac{n+1}{(n-1)n(n+2)}\left\{\left(\sum_{a=1}^{n}\Sigma_{aa}^2\right)^2 - \sum_{a=1}^{n}\Sigma_{aa}^4\right\}\left\{-\frac{2}{n+1}\left(\sum_{c=1}^{n}D_{cc}^2\right) + \left(\sum_{c=1}^{n}D_{cc}\right)^2\right\} \\
&\quad= \frac{\operatorname{tr}\left(\Sigma^4\right)}{n(n+2)}\left\{2\operatorname{tr}\left(D^2\right) + \operatorname{tr}(D)^2\right\} \\
&\qquad+ \frac{n+1}{(n-1)n(n+2)}\left\{\operatorname{tr}\left(\Sigma^2\right)^2 - \operatorname{tr}\left(\Sigma^4\right)\right\}\left\{-\frac{2\operatorname{tr}\left(D^2\right)}{n+1} + \operatorname{tr}(D)^2\right\} \\
&\quad= \frac{n_1\sigma_1^4 + n_2\sigma_2^4}{n(n+2)}\left\{2(n-s) + (n-s)^2\right\} \\
&\qquad+ \frac{n+1}{(n-1)n(n+2)}\left\{\left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2 - \left(n_1\sigma_1^4 + n_2\sigma_2^4\right)\right\}\left\{-\frac{2(n-s)}{n+1} + (n-s)^2\right\} \\
&\quad= \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2\left\{\frac{(n+1)(n-s)}{(n-1)n(n+2)}\right\}\left\{n-s-\frac{2}{n+1}\right\} \\
&\qquad+ \left(n_1\sigma_1^4 + n_2\sigma_2^4\right)\left\{\frac{(n-s)(n-s+2)}{n(n+2)} - \frac{(n+1)(n-s)\left(n-s-\frac{2}{(n+1)}\right)}{(n-1)n(n+2)}\right\}.
\end{aligned}
$$

Now we have, by Einstein notation:

$$\begin{aligned}
&\operatorname{tr}\left[(\Sigma\Lambda\Sigma)^2\right] \\
&\quad= \operatorname{tr}\left[\Sigma U D U^T \Sigma^2 U D U^T \Sigma\right] \\
&\quad= \sum_{a=1}^{n} \left[\Sigma U D U^T \Sigma^2 U D U^T \Sigma\right]_{aa} \\
&\quad= \sum_{a=1}^{n}\sum_{b=1}^{n}\sum_{c=1}^{n}\sum_{d=1}^{n}\sum_{e=1}^{n}\sum_{f=1}^{n}\sum_{g=1}^{n}\sum_{h=1}^{n}\sum_{j=1}^{n} \Sigma_{ab} U_{bc} D_{cd} \left[U^T\right]_{de} \left[\Sigma^2\right]_{ef} U_{fg} D_{gh} \left[U^T\right]_{hj} \Sigma_{ja} \\
&\quad= \sum_{a=1}^{n}\sum_{c=1}^{n}\sum_{e=1}^{n}\sum_{h=1}^{n} \Sigma_{aa} U_{ac} D_{cc} U_{ec} \Sigma_{ee}^2 U_{eh} D_{hh} U_{ah} \Sigma_{aa} \\
&\quad= \sum_{a=1}^{n}\sum_{c=1}^{n}\sum_{e=1}^{n}\sum_{h=1}^{n} D_{cc} D_{hh} \Sigma_{aa}^2 \Sigma_{ee}^2 U_{ac} U_{ec} U_{eh} U_{ah} \\
&\quad= \sum_{a=1}^{n}\sum_{c=1}^{n} D_{cc}^2 \Sigma_{aa}^4 U_{ac}^4 + \sum_{a\neq e\in[n]}\sum_{c=1}^{n} D_{cc}^2 \Sigma_{aa}^2 \Sigma_{ee}^2 U_{ac}^2 U_{ec}^2 + \sum_{a=1}^{n}\sum_{c\neq h\in[n]} D_{cc} D_{hh} \Sigma_{aa}^4 U_{ac}^2 U_{ah}^2 \\
&\qquad+ \sum_{a\neq e\in[n]}\sum_{c\neq h\in[n]} D_{cc} D_{hh} \Sigma_{aa}^2 \Sigma_{ee}^2 U_{ac} U_{ec} U_{eh} U_{ah}.
\end{aligned}$$

Therefore:

$$\begin{aligned}
\mathbb{E}\left\{\operatorname{tr}\left[(\Sigma\Lambda\Sigma)^2\right]\right\} &= \sum_{a=1}^{n}\sum_{c=1}^{n} D_{cc}^2 \Sigma_{aa}^4 \mathbb{E}\left[U_{ac}^4\right] \\
&\quad+ \sum_{a\neq e\in[n]}\sum_{c=1}^{n} D_{cc}^2 \Sigma_{aa}^2 \Sigma_{ee}^2 \mathbb{E}\left[U_{ac}^2 U_{ec}^2\right] \\
&\quad+ \sum_{a=1}^{n}\sum_{c\neq h\in[n]} D_{cc} D_{hh} \Sigma_{aa}^4 \mathbb{E}\left[U_{ac}^2 U_{ah}^2\right] \\
&\quad+ \sum_{a\neq e\in[n]}\sum_{c\neq h\in[n]} D_{cc} D_{hh} \Sigma_{aa}^2 \Sigma_{ee}^2 \mathbb{E}\left[U_{ac} U_{ec} U_{eh} U_{ah}\right].
\end{aligned}$$

Recall that:

$$\mathbb{E}\left[U_{ac}^4\right] = \frac{3}{n\,(n+2)},$$

and:

$$\mathbb{E}\left[U_{ac}^2 U_{ec}^2\right] = \mathbb{E}\left[U_{ac}^2 U_{ah}^2\right] = \frac{1}{n\,(n+2)}.$$

Now substituting $i, \lambda \coloneqq a$; $j, s \coloneqq c$; $r, \alpha \coloneqq e$; and $\beta, \mu \coloneqq h$ in Lemma D.6, we get:

$$\begin{aligned}
&\mathbb{E}\left[U_{ac} U_{ec} U_{eh} U_{ah}\right] \\
&\quad= -\frac{1}{(n-1)\,n\,(n+2)}\Big[\delta_{ae}\delta_{ea}\delta_{ch}\delta_{ch} + \delta_{ae}\delta_{ea}\delta_{ch}\delta_{ch} + \delta_{ae}\delta_{ea}\delta_{cc}\delta_{hh} \\
&\qquad\qquad+ \delta_{ae}\delta_{ea}\delta_{ch}\delta_{hc} + \delta_{aa}\delta_{ee}\delta_{cc}\delta_{hh} + \delta_{aa}\delta_{ee}\delta_{ch}\delta_{ch}\Big] \\
&\qquad+ \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{ae}\delta_{ea}\delta_{cc}\delta_{hh} + \delta_{ae}\delta_{ea}\delta_{ch}\delta_{ch} + \delta_{aa}\delta_{ee}\delta_{ch}\delta_{ch}\Big] \\
&\quad= -\frac{1}{(n-1)\,n\,(n+2)}.
\end{aligned}$$

Therefore:

$$\mathbb{E}\left[U_{ac} U_{ec} U_{eh} U_{ah}\right] = -\frac{1}{(n-1)\,n\,(n+2)}. \tag{54}$$

Substituting (50), (52), (51), and (54) in the expression of $\mathbb{E}\left\{\mathrm{tr}\left[(\Sigma\Lambda\Sigma)^2\right]\right\}$ above, we get:

$$
\begin{aligned}
&\mathbb{E}\left\{\mathrm{tr}\left[(\Sigma\Lambda\Sigma)^2\right]\right\} \\
&\quad= \frac{3}{n(n+2)}\sum_{a=1}^{n}\sum_{c=1}^{n} D_{cc}^2\Sigma_{aa}^4 + \frac{1}{n(n+2)}\sum_{a\neq e\in[n]}\sum_{c=1}^{n} D_{cc}^2\Sigma_{aa}^2\Sigma_{ee}^2 \\
&\qquad+ \frac{1}{n(n+2)}\sum_{a=1}^{n}\sum_{c\neq h\in[n]} D_{cc}D_{hh}\Sigma_{aa}^4 \\
&\qquad- \frac{1}{(n-1)n(n+2)}\sum_{a\neq e\in[n]}\sum_{c\neq h\in[n]} D_{cc}D_{hh}\Sigma_{aa}^2\Sigma_{ee}^2 \\
&\quad= \frac{3}{n(n+2)}\left(\sum_{a=1}^{n}\Sigma_{aa}^4\right)\left(\sum_{c=1}^{n}D_{cc}^2\right) + \frac{1}{n(n+2)}\left(\sum_{a\neq e\in[n]}\Sigma_{aa}^2\Sigma_{ee}^2\right)\left(\sum_{c=1}^{n}D_{cc}^2\right) \\
&\qquad+ \frac{1}{n(n+2)}\left(\sum_{a=1}^{n}\Sigma_{aa}^4\right)\left(\sum_{c\neq h\in[n]}D_{cc}D_{hh}\right) \\
&\qquad- \frac{1}{(n-1)n(n+2)}\left(\sum_{a\neq e\in[n]}\Sigma_{aa}^2\Sigma_{ee}^2\right)\left(\sum_{c\neq h\in[n]}D_{cc}D_{hh}\right) \\
&\quad= \frac{1}{n(n+2)}\left(\sum_{c=1}^{n}D_{cc}^2\right)\left\{3\sum_{a=1}^{n}\Sigma_{aa}^4 + \sum_{a\neq e\in[n]}\Sigma_{aa}^2\Sigma_{ee}^2\right\} \\
&\qquad+ \frac{1}{n(n+2)}\left(\sum_{a=1}^{n}\Sigma_{aa}^4\right)\left\{\left(\sum_{c=1}^{n}D_{cc}\right)^2 - \sum_{c=1}^{n}D_{cc}^2\right\} \\
&\qquad- \frac{1}{(n-1)n(n+2)}\left\{\left(\sum_{a=1}^{n}\Sigma_{aa}^2\right)^2 - \sum_{a=1}^{n}\Sigma_{aa}^4\right\}\left\{\left(\sum_{c=1}^{n}D_{cc}\right)^2 - \sum_{c=1}^{n}D_{cc}^2\right\} \\
&\quad= \frac{1}{n(n+2)}\left(\sum_{c=1}^{n}D_{cc}^2\right)\left\{2\sum_{a=1}^{n}\Sigma_{aa}^4 + \left(\sum_{a=1}^{n}\Sigma_{aa}^2\right)^2\right\} \\
&\qquad+ \frac{1}{n(n+2)}\left(\sum_{a=1}^{n}\Sigma_{aa}^4\right)\left\{\left(\sum_{c=1}^{n}D_{cc}\right)^2 - \sum_{c=1}^{n}D_{cc}^2\right\} \\
&\qquad- \frac{1}{(n-1)n(n+2)}\left\{\left(\sum_{a=1}^{n}\Sigma_{aa}^2\right)^2 - \sum_{a=1}^{n}\Sigma_{aa}^4\right\}\left\{\left(\sum_{c=1}^{n}D_{cc}\right)^2 - \sum_{c=1}^{n}D_{cc}^2\right\}.
\end{aligned}
$$

Thus:

$$
\begin{aligned}
&\mathbb{E}\left\{\operatorname{tr}\left[(\Sigma\Lambda\Sigma)^2\right]\right\} \\
&\quad= \frac{\operatorname{tr}\left(D^2\right)}{n\,(n+2)}\left\{2\operatorname{tr}\left(\Sigma^4\right)+\left(\operatorname{tr}\Sigma^2\right)^2\right\}+\frac{\operatorname{tr}\left(\Sigma^4\right)}{n\,(n+2)}\left\{\operatorname{tr}\left(D\right)^2-\operatorname{tr}\left(D^2\right)\right\} \\
&\quad\quad-\frac{1}{(n-1)\,n\,(n+2)}\left\{\operatorname{tr}\left(\Sigma^2\right)^2-\operatorname{tr}\left(\Sigma^4\right)\right\}\left\{\operatorname{tr}\left(D\right)^2-\operatorname{tr}\left(D^2\right)\right\} \\
&\quad= \frac{n-s}{n\,(n+2)}\left\{2\left(n_1\sigma_1^4+n_2\sigma_2^4\right)+\left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2\right\}+\frac{n_1\sigma_1^4+n_2\sigma_2^4}{n\,(n+2)}\left\{(n-s)^2-(n-s)\right\} \\
&\quad\quad-\frac{1}{(n-1)\,n\,(n+2)}\left\{\left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2-\left(n_1\sigma_1^4+n_2\sigma_2^4\right)\right\}\left\{(n-s)^2-(n-s)\right\} \\
&\quad= \left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2\left\{\frac{n-s}{n\,(n+2)}\right\}\left\{1-\frac{n-s-1}{n-1}\right\} \\
&\quad\quad+\left(n_1\sigma_1^4+n_2\sigma_2^4\right)\left\{\frac{2\,(n-s)}{n\,(n+2)}+\frac{(n-s)\,(n-s-1)}{n\,(n+2)}\frac{(n-s)\,(n-s-1)}{(n-1)\,n\,(n+2)}\right\} \\
&\quad= \left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2\left\{\frac{s\,(n-s)}{(n-1)\,n\,(n+2)}\right\} \\
&\quad\quad+\left(n_1\sigma_1^4+n_2\sigma_2^4\right)\left\{\frac{n-s}{n\,(n+2)}\right\}\left\{n-s+1+\frac{n-s-1}{n-1}\right\}.
\end{aligned}
$$

□

D.3 Proof of Proposition D.3

Proof of Proposition D.3. Recall that:

$$U_i := e_i^T\left(\frac{1}{n}X_S^TX_S\right)^{-1}\left[\frac{1}{n}X_S^T\Sigma W-\lambda_p\vec{b}\right].$$

Note that conditionally on X_S, U_i is Gaussian for each $i\in S$ and:

$$
\begin{aligned}
Y_i &:= \mathbb{E}\left[U_i\,|\,X_S\right] = -\lambda_p e_i^T\left(\frac{1}{n}X_S^TX_S\right)^{-1}\vec{b}, \\
Y_i' &:= \operatorname{Var}\left[U_i\,|\,X_S\right] = \frac{1}{n^2}e_i^T\left(\frac{1}{n}X_S^TX_S\right)^{-1}X_S^T\Sigma^2X_S\left(\frac{1}{n}X_S^TX_S\right)^{-1}e_i.
\end{aligned}
$$

Lemma D.7.

(a) The random variables Y_i and Y_i' have means:

$$\mathbb{E}\left[Y_i\right] = \frac{-\lambda_p n}{n-s-1}e_i^T\vec{b}, \quad \textit{and} \quad \mathbb{E}\left[Y_i'\right] = \frac{n_1\sigma_1^2+n_2\sigma_2^2}{n\,(n-s-1)}.$$

(b) Moreover, each pair (Y_i, Y_i') is concentrated such that:

$$\mathbb{P}\left(|Y_i|\geq\frac{n\lambda_p\sqrt{s}}{n-s-1},\ \textit{or}\ |Y_i'|\geq 2\mathbb{E}\left[Y_i'\right]\right)\leq K\left(\frac{1}{n_1}+\frac{1}{n_2}\right),$$

where K is a universal constant.

Proof. See appendix D.3.1. □

Now define the event:

$$T := \bigcup_{i=1}^{s}\left\{|Y_i|\geq\frac{n\lambda_p\sqrt{s}}{n-s-1}\ \text{ or }\ |Y_i'|\geq 2\mathbb{E}\left[Y_i'\right]\right\}.$$

By union bound and statement (b) of Lemma D.7, we have:

$$\begin{aligned}\mathbb{P}\left(T\right) &\leq \sum_{i\in S}\mathbb{P}\left(|Y_i| \geq \frac{n\lambda_p\sqrt{s}}{n-s-1}, \text{ or } |Y_i'| \geq 2\mathbb{E}\left[Y_i'\right]\right)\\ &\leq sK\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\\ &= K\left(\frac{s}{n_1}+\frac{s}{n_2}\right),\end{aligned}$$

which converges to 0 as $p \to +\infty$. Conditionning on T, we have by total probability:

$$\begin{aligned}\mathbb{P}\left(\max_{i\in S} U_i \geq \rho\right) &\leq \mathbb{P}\left(\max_{i\in S} U_i \geq \rho \mid T^c\right) + \mathbb{P}\left(T\right)\\ &\leq \mathbb{P}\left(\max_{i\in S} U_i \geq \rho \mid T^c\right) + K\left(\frac{s}{n_1}+\frac{s}{n_2}\right).\end{aligned}$$

In addition:

$$\begin{aligned}\mathbb{P}\left(\max_{i\in S} U_i \geq \rho \mid T^c\right) &= \mathbb{E}\left[\mathbb{1}\left\{\max_{i\in S} U_i \geq \rho\right\} \mid T^c\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\left\{\max_{i\in S} U_i \geq \rho\right\} \mid T^c, X_S\right]\right]\\ &= \mathbb{E}\left[\mathbb{P}\left(\max_{i\in S} U_i \geq \rho \mid T^c, X_S\right)\right].\end{aligned}$$

Now we have:

$$\begin{aligned}\mathbb{P}\left(\max_{i\in S} U_i \geq \rho \mid T^c, X_S\right) &\leq \mathbb{P}_{N_i \overset{\text{ind.}}{\sim} \mathcal{N}\left(Y_i, Y_i'\right)}\left(\max_{i\in S} N_i \geq \rho \mid T^c\right)\\ &\leq \mathbb{P}_{N_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\frac{n\lambda_p\sqrt{s}}{n-s-1}, \frac{2\left(n_1\sigma_1^2+n_2\sigma_2^2\right)}{n(n-s-1)}\right)}\left(\max_{i\in S} N_i \geq \rho\right),\end{aligned}$$

where:

- The first inequality holds because the maximum of independent Gaussians has a heavier positive tail than the maximum of correlated ones (under the same distributions).
- The second inequality holds because a Gaussian with a larger mean and variance has a heavier positive tail than one with a smaller mean and variance (therefore each of the N_is would have a heavier tail if its mean and variance were equal to their respective upper bounds).

For simplicity, we drop the long subscript and assume $N_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\frac{n\lambda_p\sqrt{s}}{n-s-1}, \frac{2\left(n_1\sigma_1^2+n_2\sigma_2^2\right)}{n(n-s-1)}\right)$. Using Markov's inequality in the above, we have:

$$\mathbb{P}\left(\max_{i\in S} U_i \geq \rho \mid T^c, X_S\right) \leq \mathbb{P}\left(\max_{i\in S} N_i \geq \rho\right) \leq \frac{\mathbb{E}\left[\max_{i\in S} N_i\right]}{\rho}. \tag{55}$$

Using the formula for expectation of Gaussian (see Theorem 5.3.1 in (De Haan & Ferreira, 2006)) maxima, we have:

$$\begin{aligned}\mathbb{E}\left[\max_{i\in S} N_i\right] &\leq \frac{n\lambda_p\sqrt{s}}{n-s-1} + \sqrt{2\log\left(s\right)\frac{2\left(n_1\sigma_1^2+n_2\sigma_2^2\right)}{n\left(n-s-1\right)}}\\ &= \frac{n\lambda_p\sqrt{s}}{n-s-1} + 2\sqrt{\frac{\left(n_1\sigma_1^2+n_2\sigma_2^2\right)\log\left(s\right)}{n\left(n-s-1\right)}}.\end{aligned}$$

Therefore, we have:

$$\begin{aligned}\mathbb{P}\left(\max_{i\in S} U_i \geq \rho\right) &\leq \mathbb{P}\left(\max_{i\in S} U_i \geq \rho \mid T^c\right) + K\left(\frac{s}{n_1} + \frac{s}{n_2}\right) \\ &\leq \mathbb{E}\left[\mathbb{P}\left(\max_{i\in S} U_i \geq \rho \mid T^c, X_S\right)\right] + K\left(\frac{s}{n_1} + \frac{s}{n_2}\right) \\ &\overset{(55)}{\leq} \frac{1}{\rho}\left(\frac{n\lambda_p\sqrt{s}}{n-s-1} + 2\sqrt{\frac{(n_1\sigma_1^2 + n_2\sigma_2^2)\log(s)}{n(n-s-1)}}\right) + K\left(\frac{s}{n_1} + \frac{s}{n_2}\right).\end{aligned}$$

Hence we have:

$$\mathbb{P}\left(\max_{i\in S} U_i \geq \rho\right) \leq \frac{1}{\rho}\left(\lambda_p\sqrt{s} + \sqrt{\frac{(n_1\sigma_1^2 + n_2\sigma_2^2)\log(s)}{n^2}}\right)(1 + o_p(1)) + o_p(1), \tag{56}$$

which converges to 0 as $p \to +\infty$ under condition (24). Using a similar argument, we establish the same bound for $\{-U_i\}_{i\in S}$, that:

$$\mathbb{P}\left(\max_{i\in S}\{-U_i\} \geq \rho\right) \leq \frac{1}{\rho}\left(\lambda_p\sqrt{s} + \sqrt{\frac{(n_1\sigma_1^2 + n_2\sigma_2^2)\log(s)}{n^2}}\right)(1 + o_p(1)) + o_p(1). \tag{57}$$

Bringing together (56) and (57) and using a union bound, we conclude:

$$\mathbb{P}\left(\max_{i\in S} |U_i| < \rho\right) \overset{p\to+\infty}{\longrightarrow} 1.$$

□

D.3.1 PROOF OF LEMMA D.7

Proof of Lemma D.7.

Mean of Y_i. Recall that:

$$Y_i = -\lambda_p e_i^T \left(\frac{1}{n} X_S^T X_S\right)^{-1} \vec{b} = -\lambda_p n e_i^T \left(X_S^T X_S\right)^{-1} \vec{b}$$

Note that $X_S^T X_S \sim \mathcal{W}_s(I_s, n)$. Using properties of the Wishart distribution (see Lemma 7.7.1 of (Anderson et al., 1958)), we have:

$$\mathbb{E}\left[\left(X_S^T X_S\right)^{-1}\right] = \left(\frac{1}{n-s-1}\right) I_s.$$

Therefore, we get:

$$\mathbb{E}[Y_i] = \frac{-\lambda_p n}{n-s-1} e_i^T \vec{b}. \tag{58}$$

Mean of Y_i'. Recall that:

$$\begin{aligned}Y_i' &= \frac{1}{n^2} e_i^T \left(\frac{1}{n} X_S^T X_S\right)^{-1} X_S^T \Sigma^2 X_S \left(\frac{1}{n} X_S^T X_S\right)^{-1} e_i \\ &= e_i^T \left(X_S^T X_S\right)^{-1} X_S^T \Sigma^2 X_S \left(X_S^T X_S\right)^{-1} e_i.\end{aligned}$$

Now recall from (45) in the proof of Lemma D.2 the matrices $Q \in \mathbb{R}^{n\times s}, R \in \mathbb{R}^{s\times s}, U \in \mathbb{R}^{n\times n}$ such that:

$$X_S = QR, \quad U = [P \quad Q],$$

where $Q^T Q = I_s$, R is upper triangular and $U \sim \text{Haar}(n)$. We have:

$$\begin{aligned}Y_i' &= e_i^T \left(X_S^T X_S\right)^{-1} X_S^T \Sigma^2 X_S \left(X_S^T X_S\right)^{-1} e_i \\ &= e_i^T \left(R^T R\right)^{-1} R^T Q^T \Sigma^2 Q R \left(R^T R\right)^{-1} e_i \\ &= e_i^T R^{-1} Q^T \Sigma^2 Q \left(R^T\right)^{-1} e_i.\end{aligned}$$

Note that:

$$U^T\Sigma^2U = \begin{bmatrix} P^T \\ Q^T \end{bmatrix} \Sigma^2 \begin{bmatrix} P & Q \end{bmatrix} = \begin{bmatrix} P^T\Sigma^2 \\ Q^T\Sigma^2 \end{bmatrix} \begin{bmatrix} P & Q \end{bmatrix} = \begin{bmatrix} P^T\Sigma^2P & P^T\Sigma^2Q \\ Q^T\Sigma^2P & Q^T\Sigma^2Q \end{bmatrix}.$$

Therefore:

$$\mathbb{E}\left[U^T\Sigma^2U\right] = \begin{bmatrix} \mathbb{E}\left[P^T\Sigma^2P\right] & \mathbb{E}\left[P^T\Sigma^2Q\right] \\ \mathbb{E}\left[Q^T\Sigma^2P\right] & \mathbb{E}\left[Q^T\Sigma^2Q\right] \end{bmatrix}.$$

On the other hand, we know by the properties of the Haar distribution (see Example 1.8 of (Gu, 2013)) that:

$$\mathbb{E}\left[U^T\Sigma^2U\right] = \frac{\operatorname{tr}\left(\Sigma^2\right)}{n} I_n = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n} I_n.$$

Hence:

$$\mathbb{E}\left[Q^T\Sigma^2Q\right] = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n} I_s.$$

Therefore:

$$\begin{aligned}
\mathbb{E}\left[Y_i'\right] &= \mathbb{E}_{Q,R}\left[e_i^T R^{-1} Q^T\Sigma^2 Q \left(R^T\right)^{-1} e_i\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e_i^T R^{-1} Q^T\Sigma^2 Q \left(R^T\right)^{-1} e_i \mid R\right]\right] \\
&= \mathbb{E}\left[e_i^T R^{-1} \mathbb{E}\left[Q^T\Sigma^2 Q \mid R\right] \left(R^T\right)^{-1} e_i\right] \\
&= \mathbb{E}\left[e_i^T R^{-1} \mathbb{E}\left[Q^T\Sigma^2 Q\right] \left(R^T\right)^{-1} e_i\right] \\
&= \mathbb{E}\left[e_i^T R^{-1} \left(\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n} I_s\right) \left(R^T\right)^{-1} e_i\right] \\
&= \left(\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n}\right) \mathbb{E}\left[e_i^T \left(R^T R\right)^{-1} e_i\right] \\
&= \left(\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n}\right) \mathbb{E}\left[e_i^T \left(R^T R\right)^{-1} e_i\right] \\
&= \left(\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n}\right) \mathbb{E}\left[e_i^T \left(X_S^T X_S\right)^{-1} e_i\right] \\
&= \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n\left(n-s-1\right)} e_i^T e_i \\
&= \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n\left(n-s-1\right)}.
\end{aligned}$$

Concentration of Y_i. Recall that:

$$Y_i = -\lambda_p n e_i^T \left(X_S^T X_S\right)^{-1} \vec{b}.$$

Thus:

$$Y_i^2 = Y_i Y_i^T = \lambda_p^2 n^2 e_i^T \left(X_S^T X_S\right)^{-1} \vec{b}\vec{b}^T \left(X_S^T X_S\right)^{-1} e_i.$$

Taking the expectation and recalling (46), we have:

$$\begin{aligned}
\mathbb{E}\left[Y_i^2\right] &= \lambda_p^2 n^2 e_i^T \mathbb{E}\left[\left(X_S^T X_S\right)^{-1} \vec{b}\vec{b}^T \left(X_S^T X_S\right)^{-1}\right] e_i \\
&\overset{(46)}{=} \frac{\lambda_p^2 n^2}{\left(n-s-3\right)\left(n-s\right)} e_i^T \left[\vec{b}\vec{b}^T + \left\|\vec{b}\right\|_2^2 I_s / \left(n-s-1\right)\right] e_i \\
&= \frac{\lambda_p^2 n^2}{\left(n-s-3\right)\left(n-s\right)} \left[\vec{b}_i^2 + \left\|\vec{b}\right\|_2^2 / \left(n-s-1\right)\right] \\
&= \frac{\lambda_p^2 n^2}{\left(n-s-3\right)\left(n-s\right)} \left[1 + \frac{s}{n-s-1}\right],
\end{aligned}$$

where the last equality above holds because $\vec{b} = \text{sign}\left(\beta^\star\right)$ and $i \in S = \text{Supp}\left(\beta^\star\right)$. Thus:

$$\mathbb{E}\left[Y_i^2\right] = \frac{\lambda_p^2 n^2 \left(n-1\right)}{\left(n-s-3\right)\left(n-s-1\right)\left(n-s\right)}.$$

Recalling (58), we get:

$$\begin{aligned}
\text{Var}\left[Y_i\right] &= \frac{\lambda_p^2 n^2 \left(n-1\right)}{\left(n-s-3\right)\left(n-s-1\right)\left(n-s\right)} - \left(\frac{-\lambda_p n}{n-s-1} e_i^T \vec{b}\right)^2 \\
&= \frac{\lambda_p^2 n^2 \left(n-1\right)}{\left(n-s-3\right)\left(n-s-1\right)\left(n-s\right)} - \frac{\lambda_p^2 n^2 \vec{b}_i^2}{\left(n-s-1\right)^2} \\
&= \frac{\lambda_p^2 n^2}{\left(n-s-1\right)} \left[\frac{\left(n-1\right)}{\left(n-s-3\right)\left(n-s\right)} - \frac{1}{n-s-1}\right].
\end{aligned}$$

Therefore:

$$\begin{aligned}
\frac{\text{Var}\left[Y_i\right]}{s\left(\mathbb{E}\left[Y_i\right]\right)^2} &= \frac{\left(n-s-1\right)^2}{s\lambda_p^2 n^2} \times \frac{\lambda_p^2 n^2}{\left(n-s-1\right)} \left[\frac{\left(n-1\right)}{\left(n-s-3\right)\left(n-s\right)} - \frac{1}{n-s-1}\right] \\
&= \frac{1}{s}\left(\frac{\left(n-1\right)\left(n-s-1\right)}{\left(n-s-3\right)\left(n-s\right)} - 1\right) \\
&= \frac{n^2 - n - ns + s - n + 1 - n^2 + ns + ns - s^2 + 3n - 3s}{s\left(n-s-3\right)\left(n-s\right)} \\
&= \frac{1 + ns - s^2 + n - 2s}{s\left(n-s-3\right)\left(n-s\right)} \\
&= \Theta\left(\frac{1}{n}\right).
\end{aligned}$$

Now using inclusion and Chebyshev's inequality, we have:

$$\begin{aligned}
\mathbb{P}\left(\left|Y_i\right| \geq \frac{n\lambda_p\sqrt{s}}{n-s-1}\right) &\overset{(58)}{\leq} \mathbb{P}\left(\left|Y_i - \mathbb{E}\left[Y_i\right]\right| \geq \left|\mathbb{E}\left[Y_i\right]\right|\left(\sqrt{s}-1\right)\right) \\
&\leq \frac{\text{Var}\left[Y_i\right]}{\left(\sqrt{s}-1\right)^2\left(\mathbb{E}\left[Y_i\right]\right)^2} \\
&= \Theta\left(\frac{\text{Var}\left[Y_i\right]}{s\left(\mathbb{E}\left[Y_i\right]\right)^2}\right) \\
&= \Theta\left(\frac{1}{n}\right).
\end{aligned}$$

In particular, the above implies:

$$\mathbb{P}\left(\left|Y_i\right| \geq \frac{n\lambda_p\sqrt{s}}{n-s-1}\right) = \mathcal{O}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

Therefore, the exists a universal constant $K_1 > 0$ such that:

$$\mathbb{P}\left(\left|Y_i\right| \geq \frac{n\lambda_p\sqrt{s}}{n-s-1}\right) \leq K_1\left(\frac{1}{n_1} + \frac{1}{n_2}\right). \tag{59}$$

Concentration of Y_i'. We have:

$$\begin{aligned}
\mathbb{E}\left[\left(Y_i'\right)^2 \mid R\right] &= \mathbb{E}\left[\left(e_i^T R^{-1} Q^T \Sigma^2 Q \left(R^T\right)^{-1} e_i\right)^2 \mid R\right] \\
&= \mathbb{E}\left[\left(\left[R^{-1} Q^T \Sigma^2 Q \left(R^T\right)^{-1}\right]_{ii}\right)^2 \;\Big|\; R\right]
\end{aligned}$$

We have, by Einstein notation:

$$\begin{aligned}\left[R^{-1}Q^T\Sigma^2Q\left(R^T\right)^{-1}\right]_{ii} &= \sum_{a=1}^{s}\sum_{b=1}^{n}\sum_{c=1}^{n}\sum_{e=1}^{s}\left[R^{-1}\right]_{ia}\left[Q^T\right]_{ab}\left[\Sigma^2\right]_{bc}Q_{ce}\left[\left(R^T\right)^{-1}\right]_{ei} \\ &= \sum_{a=1}^{s}\sum_{b=1}^{n}\sum_{c=1}^{n}\sum_{e=1}^{s}\left[R^{-1}\right]_{ia}\left[Q^T\right]_{ab}\left[\Sigma^2\right]_{bc}Q_{ce}\left[\left(R^{-1}\right)^T\right]_{ei} \\ &= \sum_{a=1}^{s}\sum_{b=1}^{n}\sum_{c=1}^{n}\sum_{e=1}^{s}\left[R^{-1}\right]_{ia}\left[Q^T\right]_{ab}\left[\Sigma^2\right]_{bc}Q_{ce}\left[R^{-1}\right]_{ie} \\ &= \sum_{a=1}^{s}\sum_{c=1}^{n}\sum_{e=1}^{s}\left[R^{-1}\right]_{ia}Q_{ca}\Sigma^2_{cc}Q_{ce}\left[R^{-1}\right]_{ie} \\ &= \sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e=1}^{s}\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}.\end{aligned}$$

Therefore:

$$\begin{aligned}&\left(\left[R^{-1}Q^T\Sigma^2Q\left(R^T\right)^{-1}\right]_{ii}\right)^2 \\ &= \left(\sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e=1}^{s}\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)^2 \\ &= \sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e=1}^{s}\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)^2 \\ &\quad + \sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e\neq f\in[s]}\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{if}Q_{ca}Q_{cf}\right) \\ &\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e=1}^{s}\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}Q_{cb}Q_{ce}\right) \\ &\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]}\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{if}Q_{cb}Q_{cf}\right) \\ &\quad + \sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e=1}^{s}\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)\left(\Sigma^2_{dd}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{da}Q_{de}\right) \\ &\quad + \sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e\neq f\in[s]}\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)\left(\Sigma^2_{dd}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{if}Q_{da}Q_{df}\right) \\ &\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e=1}^{s}\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)\left(\Sigma^2_{dd}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}Q_{db}Q_{de}\right) \\ &\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]}\left(\Sigma^2_{cc}\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ie}Q_{ca}Q_{ce}\right)\left(\Sigma^2_{dd}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{if}Q_{db}Q_{df}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e=1}^{s} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie}^2 Q_{ca}^2 Q_{ce}^2 \\
&\quad + \sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e\neq f\in[s]} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie} \left[R^{-1}\right]_{if} Q_{ca}^2 Q_{ce} Q_{cf} \\
&\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e=1}^{s} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia} \left[R^{-1}\right]_{ib} \left[R^{-1}\right]_{ie}^2 Q_{ca} Q_{cb} Q_{ce}^2 \\
&\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia} \left[R^{-1}\right]_{ib} \left[R^{-1}\right]_{ie} \left[R^{-1}\right]_{if} Q_{ca} Q_{cb} Q_{ce} Q_{cf} \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e=1}^{s} \Sigma_{cc}^2 \Sigma_{dd}^2 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie}^2 Q_{ca} Q_{ce} Q_{da} Q_{de} \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e\neq f\in[s]} \Sigma_{cc}^2 \Sigma_{dd}^2 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie} \left[R^{-1}\right]_{if} Q_{ca} Q_{ce} Q_{da} Q_{df} \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e=1}^{s} \Sigma_{cc}^2 \Sigma_{dd}^2 \left[R^{-1}\right]_{ia} \left[R^{-1}\right]_{ib} \left[R^{-1}\right]_{ie}^2 Q_{ca} Q_{ce} Q_{db} Q_{de} \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^2 \Sigma_{dd}^2 \left[R^{-1}\right]_{ia} \left[R^{-1}\right]_{ib} \left[R^{-1}\right]_{ie} \left[R^{-1}\right]_{if} Q_{ca} Q_{ce} Q_{db} Q_{df}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e=1}^{s} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie}^2 Q_{ca}^2 Q_{ce}^2 \\
&\quad + 2\sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e\neq f\in[s]} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie} \left[R^{-1}\right]_{if} Q_{ca}^2 Q_{ce} Q_{cf} \\
&\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia} \left[R^{-1}\right]_{ib} \left[R^{-1}\right]_{ie} \left[R^{-1}\right]_{if} Q_{ca} Q_{cb} Q_{ce} Q_{cf} \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e=1}^{s} \Sigma_{cc}^2 \Sigma_{dd}^2 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie}^2 Q_{ca} Q_{ce} Q_{da} Q_{de} \\
&\quad + 2\sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e\neq f\in[s]} \Sigma_{cc}^2 \Sigma_{dd}^2 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie} \left[R^{-1}\right]_{if} Q_{ca} Q_{ce} Q_{da} Q_{df} \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^2 \Sigma_{dd}^2 \left[R^{-1}\right]_{ia} \left[R^{-1}\right]_{ib} \left[R^{-1}\right]_{ie} \left[R^{-1}\right]_{if} Q_{ca} Q_{ce} Q_{db} Q_{df}.
\end{aligned}
$$

Taking the expectation of the above conditionally on R and by independence of Q and R, we get:

$$
\begin{aligned}
&\mathbb{E}\left[\left(Y_i'\right)^2 \mid R\right] \\
&= \mathbb{E}\left[\left(\left[R^{-1}Q^T\Sigma^2 Q\left(R^T\right)^{-1}\right]_{ii}\right)^2 \,\middle|\, R\right] \\
&= \sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e=1}^{s}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ie}^2\mathbb{E}\left[Q_{ca}^2Q_{ce}^2\right] \\
&\quad + 2\sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e\neq f\in[s]}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if}\mathbb{E}\left[Q_{ca}^2Q_{ce}Q_{cf}\right] \\
&\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if}\mathbb{E}\left[Q_{ca}Q_{cb}Q_{ce}Q_{cf}\right] \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e=1}^{s}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ie}^2\mathbb{E}\left[Q_{ca}Q_{ce}Q_{da}Q_{de}\right] \\
&\quad + 2\sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e\neq f\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if}\mathbb{E}\left[Q_{ca}Q_{ce}Q_{da}Q_{df}\right] \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if}\mathbb{E}\left[Q_{ca}Q_{ce}Q_{db}Q_{df}\right],
\end{aligned}
$$

Now recall from Lemma D.6 above (by Meckes (2019)) that for all $i, j, r, s, \alpha, \beta, \lambda, \mu \in [n]$ we have:

$$
\begin{aligned}
&\mathbb{E}\left[U_{ij}U_{rs}U_{\alpha\beta}U_{\lambda\mu}\right] \\
&= -\frac{1}{(n-1)\,n\,(n+2)}\Big[\delta_{ir}\delta_{\alpha\lambda}\delta_{j\beta}\delta_{s\mu} + \delta_{ir}\delta_{\alpha\lambda}\delta_{j\mu}\delta_{s\beta} + \delta_{i\alpha}\delta_{r\lambda}\delta_{js}\delta_{\beta\mu} \\
&\qquad\qquad + \delta_{i\alpha}\delta_{r\lambda}\delta_{j\mu}\delta_{\beta s} + \delta_{i\lambda}\delta_{r\alpha}\delta_{js}\delta_{\beta\mu} + \delta_{i\lambda}\delta_{r\alpha}\delta_{j\beta}\delta_{s\mu}\Big] \\
&\quad + \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{ir}\delta_{\alpha\lambda}\delta_{js}\delta_{\beta\mu} + \delta_{i\alpha}\delta_{r\lambda}\delta_{j\beta}\delta_{s\mu} + \delta_{i\lambda}\delta_{r\alpha}\delta_{j\mu}\delta_{s\beta}\Big].
\end{aligned}
$$

Also, recall from (45) that:

$$
U = \begin{bmatrix} P & Q \end{bmatrix},
$$

hence for any $\alpha \in [n], \beta \in [s]$:

$$
Q_{\alpha\beta} = U_{\alpha(n-s+\beta)}. \tag{60}
$$

Since the above fourth-order formula only depends on indices through Kronecker deltas, it holds that for any $i, r, \alpha, \lambda \in [n]$, $j, s, \alpha, \beta, \mu \in [s]$:

$$
\begin{aligned}
&\mathbb{E}\left[Q_{ij}Q_{rs}Q_{\alpha\beta}Q_{\lambda\mu}\right] \\
&= -\frac{1}{(n-1)\,n\,(n+2)}\Big[\delta_{ir}\delta_{\alpha\lambda}\delta_{j\beta}\delta_{s\mu} + \delta_{ir}\delta_{\alpha\lambda}\delta_{j\mu}\delta_{s\beta} + \delta_{i\alpha}\delta_{r\lambda}\delta_{js}\delta_{\beta\mu} \\
&\qquad\qquad + \delta_{i\alpha}\delta_{r\lambda}\delta_{j\mu}\delta_{\beta s} + \delta_{i\lambda}\delta_{r\alpha}\delta_{js}\delta_{\beta\mu} + \delta_{i\lambda}\delta_{r\alpha}\delta_{j\beta}\delta_{s\mu}\Big] \\
&\quad + \frac{n+1}{(n-1)\,n\,(n+2)}\Big[\delta_{ir}\delta_{\alpha\lambda}\delta_{js}\delta_{\beta\mu} + \delta_{i\alpha}\delta_{r\lambda}\delta_{j\beta}\delta_{s\mu} + \delta_{i\lambda}\delta_{r\alpha}\delta_{j\mu}\delta_{s\beta}.\Big]
\end{aligned}
$$

In addition, note that the expression above is equal to zero when one of $\{j, s, \alpha, \mu\}$ is different than all the others. This observation simplifies the expression of $\mathbb{E}\left[(Y_i')^2 \mid R\right]$ above as follows:

$$
\begin{aligned}
&\mathbb{E}\left[(Y_i')^2 \mid R\right] \\
&= \sum_{c=1}^{n}\sum_{a=1}^{s}\sum_{e=1}^{s}\Sigma_{cc}^4 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie}^2 \mathbb{E}\left[Q_{ca}^2 Q_{ce}^2\right] \\
&\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if} \mathbb{E}\left[Q_{ca}Q_{cb}Q_{ce}Q_{cf}\right] \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a=1}^{s}\sum_{e=1}^{s}\Sigma_{cc}^2\Sigma_{dd}^2 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie}^2 \mathbb{E}\left[Q_{ca}Q_{ce}Q_{da}Q_{de}\right] \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^2\Sigma_{dd}^2 \left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if} \mathbb{E}\left[Q_{ca}Q_{ce}Q_{db}Q_{df}\right].
\end{aligned}
$$

Thus:

$$
\begin{aligned}
&\mathbb{E}\left[(Y_i')^2 \mid R\right] \\
&= \sum_{c=1}^{n}\sum_{a=1}^{s}\Sigma_{cc}^4 \left[R^{-1}\right]_{ia}^4 \mathbb{E}\left[Q_{ca}^4\right] \\
&\quad + \sum_{c=1}^{n}\sum_{a\neq e\in[s]} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie}^2 \mathbb{E}\left[Q_{ca}^2 Q_{ce}^2\right] \\
&\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if} \\
&\qquad\qquad \times \mathbb{E}\left[Q_{ca}Q_{cb}Q_{ce}Q_{cf}\right] \mathbb{1}\left\{(a,b)=(e,f)\right\} \\
&\quad + \sum_{c=1}^{n}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^4 \left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if} \\
&\qquad\qquad \times \mathbb{E}\left[Q_{ca}Q_{cb}Q_{ce}Q_{cf}\right] \mathbb{1}\left\{(a,b)=(f,e)\right\} \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a=1}^{s}\Sigma_{cc}^2\Sigma_{dd}^2 \left[R^{-1}\right]_{ia}^4 \mathbb{E}\left[Q_{ca}^2 Q_{da}^2\right] \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a\neq e\in[s]} \Sigma_{cc}^2\Sigma_{dd}^2 \left[R^{-1}\right]_{ia}^2 \left[R^{-1}\right]_{ie}^2 \mathbb{E}\left[Q_{ca}Q_{ce}Q_{da}Q_{de}\right] \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^2\Sigma_{dd}^2 \left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if} \\
&\qquad\qquad \times \mathbb{E}\left[Q_{ca}Q_{ce}Q_{db}Q_{df}\right] \mathbb{1}\left\{(a,b)=(e,f)\right\} \\
&\quad + \sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\sum_{e\neq f\in[s]} \Sigma_{cc}^2\Sigma_{dd}^2 \left[R^{-1}\right]_{ia}\left[R^{-1}\right]_{ib}\left[R^{-1}\right]_{ie}\left[R^{-1}\right]_{if} \\
&\qquad\qquad \times \mathbb{E}\left[Q_{ca}Q_{ce}Q_{db}Q_{df}\right] \mathbb{1}\left\{(a,b)=(f,e)\right\}.
\end{aligned}
$$

Thus:

$$
\begin{aligned}
\mathbb{E}\left[\left(Y_i'\right)^2 \mid R\right] = &\sum_{c=1}^{n}\sum_{a=1}^{s}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^4\mathbb{E}\left[Q_{ca}^4\right] \\
&+\sum_{c=1}^{n}\sum_{a\neq e\in[s]}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ie}^2\mathbb{E}\left[Q_{ca}^2Q_{ce}^2\right] \\
&+\sum_{c=1}^{n}\sum_{a\neq b\in[s]}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}^2Q_{cb}^2\right] \\
&+\sum_{c=1}^{n}\sum_{a\neq b\in[s]}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}^2Q_{cb}^2\right] \\
&+\sum_{c\neq d\in[n]}\sum_{a=1}^{s}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^4\mathbb{E}\left[Q_{ca}^2Q_{da}^2\right] \\
&+\sum_{c\neq d\in[n]}\sum_{a\neq e\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ie}^2\mathbb{E}\left[Q_{ca}Q_{ce}Q_{da}Q_{de}\right] \\
&+\sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}^2Q_{db}^2\right] \\
&+\sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}Q_{cb}Q_{db}Q_{da}\right].
\end{aligned}
$$

Hence:

$$
\begin{aligned}
\mathbb{E}\left[\left(Y_i'\right)^2 \mid R\right] = &\sum_{c=1}^{n}\sum_{a=1}^{s}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^4\mathbb{E}\left[Q_{ca}^4\right] \\
&+3\sum_{c=1}^{n}\sum_{a\neq b\in[s]}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}^2Q_{cb}^2\right] \\
&+\sum_{c\neq d\in[n]}\sum_{a=1}^{s}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^4\mathbb{E}\left[Q_{ca}^2Q_{da}^2\right] \\
&+\sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}^2Q_{db}^2\right] \\
&+2\sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}Q_{cb}Q_{db}Q_{da}\right].
\end{aligned}
$$

Now recalling (50), (51), (52), (53), (54), and using (60) we have:

$$
\begin{aligned}
&\mathbb{E}\left[Q_{ca}^4\right] = \frac{3}{n\left(n+2\right)} \\
&\mathbb{E}\left[Q_{ca}^2Q_{cb}^2\right] = \mathbb{E}\left[Q_{ca}^2Q_{da}^2\right] = \frac{1}{n\left(n+2\right)} \\
&\mathbb{E}\left[Q_{ca}^2Q_{db}^2\right] = \frac{n+1}{\left(n-1\right)n\left(n+2\right)} \\
&\mathbb{E}\left[Q_{ca}Q_{cb}Q_{db}Q_{da}\right] = -\frac{1}{\left(n-1\right)n\left(n+2\right)}.
\end{aligned}
$$

Therefore:

$$\begin{aligned}
\mathbb{E}\left[(Y_i')^2 \mid R\right] = &\sum_{c=1}^{n}\sum_{a=1}^{s}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^4\mathbb{E}\left[Q_{ca}^4\right] \\
&+3\sum_{c=1}^{n}\sum_{a\neq b\in[s]}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}^2Q_{cb}^2\right] \\
&+\sum_{c\neq d\in[n]}\sum_{a=1}^{s}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^4\mathbb{E}\left[Q_{ca}^2Q_{da}^2\right] \\
&+\sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}^2Q_{db}^2\right] \\
&+2\sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\mathbb{E}\left[Q_{ca}Q_{cb}Q_{db}Q_{da}\right]
\end{aligned}$$

Thus:

$$\begin{aligned}
\mathbb{E}\left[(Y_i')^2 \mid R\right] = &\frac{3}{n(n+2)}\sum_{c=1}^{n}\sum_{a=1}^{s}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^4 \\
&+\frac{3}{n(n+2)}\sum_{c=1}^{n}\sum_{a\neq b\in[s]}\Sigma_{cc}^4\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2 \\
&+\frac{1}{n(n+2)}\sum_{c\neq d\in[n]}\sum_{a=1}^{s}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^4 \\
&+\frac{n+1}{(n-1)\,n\,(n+2)}\sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2 \\
&-\frac{2}{(n-1)\,n\,(n+2)}\sum_{c\neq d\in[n]}\sum_{a\neq b\in[s]}\Sigma_{cc}^2\Sigma_{dd}^2\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2
\end{aligned}$$

Thus:

$$\begin{aligned}
&\mathbb{E}\left[(Y_i')^2 \mid R\right] \\
&=\frac{3}{n(n+2)}\left(\sum_{c=1}^{n}\Sigma_{cc}^4\right)\left\{\sum_{a=1}^{s}\left[R^{-1}\right]_{ia}^4+\sum_{a\neq b\in[s]}\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\right\} \\
&\quad+\frac{1}{n(n+2)}\left(\sum_{c\neq d\in[n]}\Sigma_{cc}^2\Sigma_{dd}^2\right)\left\{\sum_{a=1}^{s}\left[R^{-1}\right]_{ia}^4+\sum_{a\neq b\in[s]}\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\right\} \\
&=\frac{1}{n(n+2)}\left\{3\sum_{c=1}^{n}\Sigma_{cc}^4+\sum_{c\neq d\in[n]}\Sigma_{cc}^2\Sigma_{dd}^2\right\}\left\{\sum_{a=1}^{s}\left[R^{-1}\right]_{ia}^4+\sum_{a\neq b\in[s]}\left[R^{-1}\right]_{ia}^2\left[R^{-1}\right]_{ib}^2\right\} \\
&=\frac{1}{n(n+2)}\left\{3\sum_{c=1}^{n}\Sigma_{cc}^4+\left[\left(\sum_{c=1}^{n}\Sigma_{cc}^2\right)^2-\sum_{c=1}^{n}\Sigma_{cc}^4\right]\right\}\left(\sum_{a=1}^{s}\left[R^{-1}\right]_{ia}^2\right)^2 \\
&=\frac{1}{n(n+2)}\left\{2\operatorname{tr}\left(\Sigma^4\right)+\operatorname{tr}\left(\Sigma^2\right)^2\right\}\left(\sum_{a=1}^{s}\left[R^{-1}\right]_{ia}^2\right)^2.
\end{aligned}$$

Now note that:

$$\begin{aligned} e_i^T \left(X_S^T X_S\right)^{-1} e_i &= \left[\left(R^T R\right)^{-1}\right]_{ii} \\ &= \left[R^{-1}\left(R^{-1}\right)^T\right]_{ii} \\ &= \sum_{a=1}^{s} \left[R^{-1}\right]_{ia} \left[\left(R^{-1}\right)^T\right]_{ai} \\ &= \sum_{a=1}^{s} \left[R^{-1}\right]_{ia} \left[R^{-1}\right]_{ia} \\ &= \sum_{a=1}^{s} \left[R^{-1}\right]_{ia}^2 . \end{aligned}$$

On the other hand, recall that $X_S^T X_S \sim \mathcal{W}_s\left(I_s, n\right)$. Setting $b := n$; $a := s$; $t := e_i$; $T := I_s$ and $A := X_S^T X_S$ in Lemma D.4, we get:

$$\begin{aligned} \mathbb{E}\left[\left(X_S^T X_S\right)^{-1} e_i e_i^T \left(X_S^T X_S\right)^{-1}\right] &= \frac{1}{(n-s)(n-s-3)} \left(e_i e_i^T + \left(e_i^T e_i\right) I_s / (n-s-1)\right) \\ &= \frac{1}{(n-s)(n-s-3)} \left(e_i e_i^T + \frac{1}{n-s-1} I_s\right). \end{aligned}$$

Therefore:

$$\begin{aligned} \mathbb{E}\left[\left(e_i^T \left(X_S^T X_S\right)^{-1} e_i\right)^2\right] &= e_i^T \mathbb{E}\left[\left(X_S^T X_S\right)^{-1} e_i e_i^T \left(X_S^T X_S\right)^{-1}\right] e_i \\ &= \left[\mathbb{E}\left[\left(X_S^T X_S\right)^{-1} e_i e_i^T \left(X_S^T X_S\right)^{-1}\right]\right]_{ii} \\ &= \left[\frac{1}{(n-s)(n-s-3)} \left(e_i e_i^T + \frac{1}{n-s-1} I_s\right)\right]_{ii} . \end{aligned}$$

Hence:

$$\mathbb{E}\left[\left(\sum_{a=1}^{s} \left[R^{-1}\right]_{ia}^2\right)^2\right] = \mathbb{E}\left[\left(e_i^T \left(X_S^T X_S\right)^{-1} e_i\right)^2\right] = \frac{1}{(n-s)(n-s-3)} \left(1 + \frac{1}{n-s-1}\right).$$

Hence, we get:

$$\begin{aligned} &\mathbb{E}\left[\left(Y_i'\right)^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(Y_i'\right)^2 \mid R\right]\right] \\ &= \mathbb{E}\left[\frac{1}{n(n+2)} \left\{2\operatorname{tr}\left(\Sigma^4\right) + \operatorname{tr}\left(\Sigma^2\right)^2\right\} \left(\sum_{a=1}^{s} \left[R^{-1}\right]_{ia}^2\right)^2\right] \\ &= \frac{1}{n(n+2)} \left\{2\operatorname{tr}\left(\Sigma^4\right) + \operatorname{tr}\left(\Sigma^2\right)^2\right\} \mathbb{E}\left[\left(\sum_{a=1}^{s} \left[R^{-1}\right]_{ia}^2\right)^2\right] \\ &= \frac{1}{(n-s)(n-s-3)\, n(n+2)} \left(1 + \frac{1}{n-s-1}\right) \left\{2\operatorname{tr}\left(\Sigma^4\right) + \operatorname{tr}\left(\Sigma^2\right)^2\right\} \\ &= \frac{2\left(n_1\sigma_1^4 + n_2\sigma_2^4\right) + \left(n_1\sigma_1^2 + n_2\sigma_2^2\right)^2}{(n-s)(n-s-3)\, n(n+2)} \left(1 + \frac{1}{n-s-1}\right). \end{aligned}$$

Therefore:

$$\begin{aligned}
\frac{\operatorname{Var}\left(Y_i'\right)}{\left(\mathbb{E}\left[Y_i'\right]\right)^2} &= \frac{\mathbb{E}\left[\left(Y_i'\right)^2\right]-\left(\mathbb{E}\left[Y_i'\right]\right)^2}{\left(\mathbb{E}\left[Y_i'\right]\right)^2} \\
&= \frac{\frac{2\left(n_1\sigma_1^4+n_2\sigma_2^4\right)+\left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2}{(n-s)(n-s-3)n(n+2)}\left(1+\frac{1}{n-s-1}\right)-\left(\frac{n_1\sigma_1^2+n_2\sigma_2^2}{n(n-s-1)}\right)^2}{\left(\frac{n_1\sigma_1^2+n_2\sigma_2^2}{n(n-s-1)}\right)^2} \\
&= \frac{\frac{2\left(n_1\sigma_1^4+n_2\sigma_2^4\right)+\left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2}{(n-s)(n-s-3)n(n+2)}\left(1+\frac{1}{n-s-1}\right)-\left(\frac{n_1\sigma_1^2+n_2\sigma_2^2}{n(n-s-1)}\right)^2}{\left(\frac{n_1\sigma_1^2+n_2\sigma_2^2}{n(n-s-1)}\right)^2} \\
&= \frac{n^2\left(n-s-1\right)^2}{\left(n-s\right)\left(n-s-3\right)n\left(n+2\right)}\left(\frac{2\left(n_1\sigma_1^4+n_2\sigma_2^4\right)+\left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2}{\left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2}\right) \\
&\quad\times\left(1+\frac{1}{n-s-1}\right)-1 \\
&= \frac{n\left(n-s-1\right)}{\left(n-s-3\right)\left(n+2\right)}\left(\frac{2\left(n_1\sigma_1^4+n_2\sigma_2^4\right)}{\left(n_1\sigma_1^2+n_2\sigma_2^2\right)^2}+1\right)-1 \\
&= \frac{n\left(n-s-1\right)}{\left(n-s-3\right)\left(n+2\right)}\left(\frac{2n_1\sigma_1^4+2n_2\sigma_2^4}{n_1^2\sigma_1^4+n_2^2\sigma_2^4+2n_1n_2\sigma_1^2\sigma_2^2}+1\right)-1 \\
&\leq \frac{n\left(n-s-1\right)}{\left(n-s-3\right)\left(n+2\right)}\left(\frac{2}{n_1}+\frac{2}{n_2}+1\right)-1 \\
&= \frac{n\left(n-s-1\right)}{\left(n-s-3\right)\left(n+2\right)}\left(\frac{2}{n_1}+\frac{2}{n_2}\right)+\left\{\frac{n\left(n-s-1\right)}{\left(n-s-3\right)\left(n+2\right)}-1\right\} \\
&= \frac{n\left(n-s-1\right)}{\left(n-s-3\right)\left(n+2\right)}\left(\frac{2}{n_1}+\frac{2}{n_2}\right) \\
&\quad+\frac{n^2-ns-n-n^2+ns+3n-2n+2s+6}{\left(n-s-3\right)\left(n+2\right)} \\
&= \frac{n\left(n-s-1\right)}{\left(n-s-3\right)\left(n+2\right)}\left(\frac{2}{n_1}+\frac{2}{n_2}\right)+\frac{2\left(s+3\right)}{\left(n-s-3\right)\left(n+2\right)}.
\end{aligned}$$

Hence there exists a universal constant K_2 such that:

$$\frac{\operatorname{Var}\left(Y_i'\right)}{\left(\mathbb{E}\left[Y_i'\right]\right)^2} \leq K_2\left(\frac{1}{n_1}+\frac{1}{n_2}\right).$$

By Chebyshev's inequality, we conclude:

$$\mathbb{P}\Big(Y_i' \geq 2\mathbb{E}\left[Y_i'\right]\Big) \leq K_2\left(\frac{1}{n_1}+\frac{1}{n_2}\right). \tag{61}$$

Bringing together (59) and (61) and using a union bound, we conclude that there exists a universal constant $K>0$ such that:

$$\mathbb{P}\left(|Y_i| \geq \frac{n\lambda_p\sqrt{s}}{n-s-1},\ \text{ or }\ |Y_i'| \geq 2\mathbb{E}\left[Y_i'\right]\right) \leq K\left(\frac{1}{n_1}+\frac{1}{n_2}\right). \tag{62}$$

□

E Proof of Proposition 4.1

Proof of Proposition 4.1. First, assume there exists $\left(\lambda_p\right)_{p\geq 1} \to 0$ such that (24) holds. By the first part of (24), we have:

$$\frac{\sigma_{\text{avg}}^2 \log\left(p-s\right)}{n} = o\left(\lambda_p^2\right). \tag{63}$$

Using (63) and $(\lambda_p)_{p\geq 1} \to 0$, we get:

$$\frac{\sigma_{\text{avg}}^2 \log(p-s)}{n} \longrightarrow 0. \tag{64}$$

In addition, by the second part of (24), we have:

$$\lambda_p^2 = o\left(\frac{\rho^2}{s}\right). \tag{65}$$

Using (63) and (65), we get:

$$\frac{\sigma_{\text{avg}}^2 \log(p-s)\, s}{n\rho^2} \longrightarrow 0. \tag{66}$$

Taking the sum of (64) and (66), we obtain:

$$\frac{\sigma_{\text{avg}}^2 \log(p-s)\left(1+s/\rho^2\right)}{n} \longrightarrow 0.$$

Hence, we conclude:

$$\sigma_{\text{avg}}^2 = o\left(\frac{n}{(1+s/\rho^2)\log(p-s)}\right).$$

Second, assume (26) holds and let:

$$\lambda_p \coloneqq \left(\frac{\sigma_{\text{avg}}^2 \log(p-s)}{n\,(1+s/\rho^2)}\right)^{1/4}.$$

Then we have:

$$\lambda_p^2 = \sqrt{\frac{\sigma_{\text{avg}}^2 \log(p-s)}{(1+s/\rho^2)\, n}}.$$

By (26), we have:

$$\begin{aligned}\lambda_p^2 &= o\left(\sqrt{\frac{n\log(p-s)}{(1+s/\rho^2)^2 \log(p-s)\, n}}\right)\\ &= o\left(\frac{1}{1+s/\rho^2}\right),\end{aligned}$$

thus:

$$\lambda_p^2\left(1+s/\rho^2\right) \longrightarrow 0.$$

Therefore

$$\lambda_p \longrightarrow 0 \quad \text{and} \quad \frac{\lambda_p\sqrt{s}}{\rho} \longrightarrow 0. \tag{67}$$

In addition, we have by definition of λ_p:

$$\begin{aligned}\frac{n\lambda_p^2}{\sigma_{\text{avg}}^2 \log(p-s)} &= \sqrt{\frac{\sigma_{\text{avg}}^2 \log(p-s)\, n^2}{\sigma_{\text{avg}}^4 \log(p-s)^2 (1+s/\rho^2)\, n}}\\ &= \sqrt{\frac{n}{\sigma_{\text{avg}}^2 (1+s/\rho^2)\log(p-s)}}.\end{aligned}$$

Therefore we have, by (26):

$$\frac{n\lambda_p^2}{\sigma_{\text{avg}}^2 \log(p-s)} \longrightarrow +\infty. \tag{68}$$

Finally, by (26) it holds that:

$$\begin{aligned}\frac{\sigma_{\text{avg}}^2 \log s}{n\rho^2} &= o\left(\frac{n\log s}{n\rho^2 (1+s/\rho^2)\log(p-s)}\right)\\ &= o\left(\frac{1}{\rho^2+s}\right) = o\left(\frac{1}{s}\right).\end{aligned}$$

Therefore:

$$\frac{1}{\rho}\sqrt{\frac{\sigma_{\text{avg}}^2 \log s}{n}} \longrightarrow 0. \tag{69}$$

Bring together (67), (68) and (69), we conclude that $\lambda_p \longrightarrow 0$ and (24) holds. □