Minimax Estimation of Linear Functions of Eigenvectors in the Face of Small Eigen-Gaps

Gen Li* Changxiao Cai† H. Vincent Poor‡ Yuxin Chen*

July 4, 2022

Eigenvector perturbation analysis plays a vital role in various data science applications. A large body of prior works, however, focused on establishing ℓ_2 eigenvector perturbation bounds, which are often highly inadequate in addressing tasks that rely on fine-grained behavior of an eigenvector. This paper makes progress on this by studying the perturbation of linear functions of an unknown eigenvector. Focusing on two fundamental problems — matrix denoising and principal component analysis — in the presence of Gaussian noise, we develop a suite of statistical theory that characterizes the perturbation of arbitrary linear functions of an unknown eigenvector. In order to mitigate a non-negligible bias issue inherent to the natural "plug-in" estimator, we develop de-biased estimators that (1) achieve minimax lower bounds for a family of scenarios (modulo some logarithmic factor), and (2) can be computed in a data-driven manner without sample splitting. Noteworthily, the proposed estimators are nearly minimax optimal even when the associated eigen-gap is *substantially smaller* than what is required in prior statistical theory.

Keywords: linear forms of eigenvectors, matrix denoising, principal component analysis, bias correction, small eigen-gap

Contents

*Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA; email: {ligen,yuxinc}@wharton.upenn.edu.

†Department of Biostatistics, University of Pennsylvania, Philadelphia, PA 19104, USA; email: changxiao.cai@pennmedicine.upenn.edu.

‡Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA; email: poor@princeton.edu.

1 Introduction

A variety of large-scale data science applications involve extracting actionable knowledge from the eigenvectors of a certain low-rank matrix. Representative examples include principal component analysis (PCA) [Johnstone, 2001], phase synchronization [Singer, 2011], clustering in mixture models [Löffler et al., 2019], community recovery [Abbe et al., 2020b, Lei and Rinaldo, 2015], to name just a few. In reality, it is often the case that one only observes a randomly corrupted version of the low-rank matrix of interest, and has to retrieve information from the "empirical" eigenvectors (i.e., the eigenvectors of the observed noisy matrix). This motivates the studies of eigenvector perturbation theory from statistical viewpoints, with particular emphasis on high-dimensional scenarios [Chen et al., 2021b]. In the current paper, we seek to further expand such a statistical theory, focusing on the following two concrete models.

- *Matrix denoising under i.i.d. Gaussian noise.* Let $\boldsymbol{M}^\star \in \mathbb{R}^{n\times n}$ be an unknown rank-r symmetric matrix whose l-th eigenvector (resp. eigenvalue) is $\boldsymbol{u}_l^\star$ (resp. $\lambda_l^\star$). What we have observed is a corrupted version $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$ of $\boldsymbol{M}^\star$, where $\boldsymbol{H} = [H_{i,j}]_{1\le i,j\le n}$ represents a symmetric Gaussian random matrix with

$H_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2), i > j$ and $H_{i,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,2\sigma^2)$. The aim is to estimate $\boldsymbol{u}_l^\star$ based on the l-th eigenvector of the data matrix $\boldsymbol{M}$.

- *Principal component analysis (PCA) and covariance estimation.* Imagine that we have collected n independent p-dimensional sample vectors $\boldsymbol{s}_i \overset{\text{ind.}}{\sim} \mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma})$, $1 \le i \le n$. Suppose that the underlying covariance matrix enjoys a "spiked" structure $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\star + \sigma^2 \boldsymbol{I}_p$, where $\boldsymbol{\Sigma}^\star \succeq \boldsymbol{0}$ is an unknown rank-r matrix whose l-th eigenvector (resp. eigenvalue) is given by $\boldsymbol{u}_l^\star$ (resp. $\lambda_l^\star$). We seek to estimate $\boldsymbol{u}_l^\star$ by examining the l-th eigenvector of the sample covariance matrix $\frac{1}{n}\sum_{i=1}^n \boldsymbol{s}_i\boldsymbol{s}_i^\top$.

While a large body of prior literature has investigated eigenvector perturbation theory for the aforementioned two models, the majority of past works focused on ℓ_2 statistical analysis, namely, quantifying the ℓ_2 estimation error of $\boldsymbol{u}_l$ when it is employed to estimate $\boldsymbol{u}_l^\star$. Such ℓ_2 perturbation theory, however, is often too coarse if the ultimate goal is to retrieve fine-grained information from the eigenvector of interest, say, some linear function of the eigenvector $\boldsymbol{u}_l^\star$ (e.g., the Fourier transform of or any given entry of $\boldsymbol{u}_l^\star$). Motivated by the inadequacy of existing ℓ_2 theory, we seek to investigate how to faithfully estimate linear functionals of the eigenvectors — that is, $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ for some vector $\boldsymbol{a} \in \mathbb{R}^n$ given *a priori*. Towards achieving this goal, two challenges stand out, which merit careful thinking.

- *The need of bias correction.* A natural strategy towards estimating the linear form $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ is to invoke the naive "plug-in" estimator $\boldsymbol{a}^\top\boldsymbol{u}_l$. However, it has already been pointed out in the literature (e.g., Koltchinskii and Xia [2016], Koltchinskii et al. [2016]) that the plug-in estimator might suffer from a non-negligible bias. This calls for careful designs of algorithms that allow for proper bias correction in a data-driven yet efficient manner.
- *How to cope with small eigen-gaps.* When estimating the eigenvector $\boldsymbol{u}_l^\star$, most prior works require the associated eigen-gap $\min_{i\neq l}|\lambda_i^\star - \lambda_l^\star|$ to exceed the spectral norm of the perturbation matrix (i.e., $\boldsymbol{H}$ in the matrix denoising case and $\frac{1}{n}\sum_i \boldsymbol{s}_i\boldsymbol{s}_i^\top - \boldsymbol{\Sigma}$ in the PCA setting) [Chen et al., 2021b, Davis and Kahan, 1970]. However, there is no lower bound in the literature that precludes us from achieving faithful estimation when the eigen-gap violates such requirements. It would thus be of great interest to understand the statistical limits when the eigen-gap of interest is particularly small.

Main contributions. This paper investigates estimating the linear form $\boldsymbol{a}^\top\boldsymbol{u}_i^\star$ for the aforementioned two statistical models under Gaussian noise, with particular emphasis on those scenarios with *small eigen-gaps*. Our main contributions are summarized below.

1. We develop fine-grained perturbation analysis for linear forms of eigenvectors, which is valid even when the eigen-gap $\min_{i\neq l}|\lambda_i^\star - \lambda_l^\star|$ is substantially smaller than the spectral norm of the perturbation matrix. This eigen-gap condition significantly improves upon what is required in prior theory.
2. The natural "plug-in" estimator suffers from a non-negligible bias issue, which is particularly severe when the associated eigen-gap is small. To address this issue, we put forward a de-biased estimator for $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ by multiplying the plug-in estimator by a correction factor, which can be computed in a data-driven manner without the need of sample splitting. The proposed estimator provably achieves enhanced estimation accuracy compared to the plug-in estimator, and is shown to be minimax optimal (up to some logarithmic factor) for a broad class of scenarios.

Organization. The rest of this paper is organized as follows. In Section 2, we formulate the problem precisely and introduce basic definitions. Section 3 presents our main theoretical findings, whereas Section 4 provides a non-exhaustive overview of prior works. The analysis strategy of our main theorems is outlined in Section 5. The detailed proofs and auxiliary lemmas are postponed to the appendix. We conclude this paper with a discussion of future directions in Section 6.

Notation. For any vector $\boldsymbol{v}$, we denote by $\|\boldsymbol{v}\|_2$ and $\|\boldsymbol{v}\|_\infty$ its ℓ_2 norm and ℓ_∞ norm, respectively; for any vectors $\boldsymbol{v}$ and $\boldsymbol{u}$, we use $\langle \boldsymbol{v}, \boldsymbol{u}\rangle$ to represent their inner product. For any matrix $\boldsymbol{M}$, we let $\|\boldsymbol{M}\|$ and $\|\boldsymbol{M}\|_{\mathrm{F}}$ denote the spectral norm and the Frobenius norm of $\boldsymbol{M}$, respectively. For any matrix $\boldsymbol{U}$ whose columns are orthonormal, we use $\boldsymbol{U}^{\perp}$ to denote a matrix whose columns form an orthonormal basis of the orthogonal complement of the column space of $\boldsymbol{U}$, and let $\mathcal{P}_{\boldsymbol{U}}(\boldsymbol{M}) = \boldsymbol{U}\boldsymbol{U}^{\top}\boldsymbol{M}$ be the Euclidean projection of a matrix $\boldsymbol{M}$ onto the column space of $\boldsymbol{U}$. For any two random matrices $\boldsymbol{Z}$ and $\boldsymbol{X}$, the notation $\boldsymbol{Z} \overset{\mathrm{d}}{=} \boldsymbol{X}$ means $\boldsymbol{Z}$ and $\boldsymbol{X}$ are identical in distribution. For notational simplicity, we write $[n]$ for the set $\{1, \cdots, n\}$. For any $a, b \in \mathbb{R}$, we introduce the notation $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, and $\min |a \pm b| = \min\{|a - b|, |a + b|\}$. We denote by $\mathcal{B}_r(\boldsymbol{z}) \coloneqq \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{z}\|_2 \leq r\}$ the ball of radius r centered at $\boldsymbol{z}$. Throughout the paper, we denote by $f(n) \lesssim g(n)$ or $f(n) = O(g(n))$ the condition $|f(n)| \leq Cg(n)$ for some universal constant $C > 0$ when n is sufficiently large; we use $f(n) \gtrsim g(n)$ or $f(n) = \Omega(g(n))$ to indicate that $f(n) \geq C|g(n)|$ for some universal constant $C > 0$ when n is sufficiently large; and we also use $f(n) \asymp g(n)$ or $f(n) = \Theta(g(n))$ to indicate that $f(n) \lesssim g(n)$ and $f(n) \gtrsim g(n)$ hold simultaneously. In addition, the standard notation $\widetilde{O}(g(n))$ (resp. $\widetilde{\Omega}(g(n))$) is similar to $O(g(n))$ (resp. $\Omega(g(n))$) except that it hides the logarithmic dependency. The notation $f(n) = o(g(n))$ means that $\lim_{n\to\infty} f(n)/g(n) = 0$, and $f(n) \gg g(n)$ (resp. $f(n) \ll g(n)$) means that there exists some large (resp. small) constant $c_1 > 0$ (resp. $c_2 > 0$) such that $f(n) \geq c_1 g(n)$ (resp. $f(n) \leq c_2 g(n)$). Finally, for any $1 \leq l \leq r$, we set the expression $\sum_{k \neq l, 1 \leq k \leq r} g(k)$ to be zero for any $g(\cdot)$ if $r = 1$ (that is, the case where no k satisfies the requirement in the summation).

2 Problem formulation

2.1 Matrix denoising

Suppose that we are interested in a symmetric matrix $\boldsymbol{M}^{\star} = [M_{i,j}^{\star}]_{1\leq i,j\leq n} \in \mathbb{R}^{n\times n}$ with eigen-decomposition

$$\boldsymbol{M}^{\star} = \sum_{i=1}^{r} \lambda_i^{\star} \boldsymbol{u}_i^{\star} \boldsymbol{u}_i^{\star\top} =: \boldsymbol{U}^{\star}\boldsymbol{\Lambda}^{\star}\boldsymbol{U}^{\star\top}, \tag{2.1}$$

where the $\boldsymbol{u}_i^{\star}$'s are orthonormal. Here, $\{\lambda_i^{\star}\}$ denotes the set of non-zero eigenvalues of $\boldsymbol{M}^{\star}$, and $\boldsymbol{u}_i^{\star}$ indicates the (normalized) eigenvector associated with $\lambda_i^{\star}$. It is assumed throughout that

$$\lambda_{\min}^{\star} = |\lambda_r^{\star}| \leq \cdots \leq |\lambda_1^{\star}| = \lambda_{\max}^{\star}, \tag{2.2}$$

and the condition number of $\boldsymbol{M}^{\star}$ is defined as

$$\kappa := \frac{\lambda_{\max}^{\star}}{\lambda_{\min}^{\star}}. \tag{2.3}$$

In addition, for any $1 \leq l \leq r$, we introduce an eigen-gap (or eigenvalue separation) metric that quantifies the distance between the eigenvalue $\lambda_l^{\star}$ and the remaining spectrum:

$$\Delta_l^{\star} \coloneqq \begin{cases} \min_{k:\, k\neq l,\, 1\leq k\leq r} \left|\lambda_l^{\star} - \lambda_k^{\star}\right|, & \text{if} \quad r > 1, \\ \lambda_{\max}^{\star}, & \text{if} \quad r = 1, \end{cases} \tag{2.4}$$

which plays a crucial role in our perturbation theory.

What we have observed is a randomly corrupted data matrix $\boldsymbol{M} = [M_{i,j}]_{1\leq i,j\leq n}$ as follows

$$\boldsymbol{M} = \boldsymbol{M}^{\star} + \boldsymbol{H}, \tag{2.5}$$

where $\boldsymbol{H} = [H_{i,j}]_{1\leq i,j\leq n}$ represents a symmetric noise matrix with independent random entries

$$H_{i,j} \overset{\text{ind.}}{\sim} \begin{cases} \mathcal{N}(0, 2\sigma^2), & i = j, \\ \mathcal{N}(0, \sigma^2), & i > j. \end{cases} \tag{2.6}$$

Throughout this paper, we denote by λ_l the l-th largest eigenvalue (in magnitude) of $\boldsymbol{M}$, and let $\boldsymbol{u}_l$ represent the associated eigenvector of $\boldsymbol{M}$. Our goal is to estimate linear functionals of an eigenvector $\boldsymbol{u}_l^{\star}$ — that is, $\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}$ $(1 \leq l \leq r)$ for some fixed vector $\boldsymbol{a} \in \mathbb{R}^n$ — based on the observed noisy data $\boldsymbol{M}$.

2.2 Principal component analysis and covariance estimation

Turning to principal component analysis (PCA) or covariance estimation, we concentrate on the following spiked covariance model. Imagine that we have collected a sequence of n i.i.d. zero-mean Gaussian sample vectors in $\mathbb{R}^p$ as follows

$$\boldsymbol{s}_i \overset{\text{ind.}}{\sim} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right), \qquad 1 \leq i \leq n,$$

where

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\star + \sigma^2 \boldsymbol{I}_p \in \mathbb{R}^{p\times p}$$

denotes the covariance matrix. Here and throughout, we assume that the "spiked component" $\boldsymbol{\Sigma}^\star$ of $\boldsymbol{\Sigma}$ is an unknown rank-r matrix with eigen-decomposition

$$\boldsymbol{\Sigma}^\star = \boldsymbol{U}^\star \boldsymbol{\Lambda}^\star \boldsymbol{U}^{\star\top} = \sum_{i=1}^{r} \lambda_i^\star \boldsymbol{u}_i^\star \boldsymbol{u}_i^{\star\top} \succeq \boldsymbol{0},$$

where $\lambda_i^\star$ denotes the i-th largest eigenvalue of $\boldsymbol{\Sigma}^\star$, with $\boldsymbol{u}_i^\star$ representing the associated eigenvector. Akin to the matrix denoising case, we assume

$$0 < \lambda_{\min}^\star = \lambda_r^\star \leq \cdots \leq \lambda_1^\star = \lambda_{\max}^\star,$$

and introduce the condition number $\kappa := \lambda_{\max}^\star / \lambda_{\min}^\star$ and the eigen-separation metric

$$\Delta_l^\star := \begin{cases} \min_{k:\, k\neq l,\, 1\leq k\leq r} \left|\lambda_l^\star - \lambda_k^\star\right|, & \text{if} \quad r > 1, \\ \lambda_{\max}^\star, & \text{if} \quad r = 1. \end{cases} \tag{2.7}$$

Given a fixed vector $\boldsymbol{a} \in \mathbb{R}^p$, our aim is to develop a reliable estimate of the linear functional $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ of an eigenvector $\boldsymbol{u}_l^\star$ $(1 \leq l \leq r)$, on the basis of the sample vectors $\{\boldsymbol{s}_i\}_{1\leq i\leq n}$ (or the sample covariance matrix $\frac{1}{n}\sum_{i=1}^n \boldsymbol{s}_i \boldsymbol{s}_i^\top$).

3 Main results

With the above description of the problem settings in place, we are ready to present our findings concerning eigenvector perturbation. Given that we cannot distinguish $\boldsymbol{u}_l^\star$ and $-\boldsymbol{u}_l^\star$ based on the observed matrix, the error of an estimator $u_{\boldsymbol{a}}$ for estimating $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ shall be measured via the following metric that accounts for such a global ambiguity issue:

$$\mathsf{dist}\left(u_{\boldsymbol{a}}, \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right) := \min\left\{\left|u_{\boldsymbol{a}} \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\right\} = \min\left\{\left|u_{\boldsymbol{a}} - \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|, \left|u_{\boldsymbol{a}} + \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\right\}. \tag{3.1}$$

3.1 Matrix denoising

We begin with the matrix denoising problem introduced in Section 2.1. Recalling that $\boldsymbol{u}_l$ is the eigenvector of $\boldsymbol{M}$ associated with λ_l $(1 \leq l \leq n)$, we investigate the following two estimators when estimating the linear function $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.

- A plug-in estimator:

$$u_{\boldsymbol{a}}^{\mathsf{plugin}} := \boldsymbol{a}^\top \boldsymbol{u}_l; \tag{3.2a}$$

- A modified estimator that we propose (which we shall refer to as a *de-biased estimator* from now on):

$$u_{\boldsymbol{a}}^{\mathsf{debiased}} := \sqrt{1 + b_l}\, \boldsymbol{a}^\top \boldsymbol{u}_l \qquad \text{with} \quad b_l := \sum_{i:\, r<i\leq n} \frac{\sigma^2}{(\lambda_l - \lambda_i)^2}, \tag{3.2b}$$

  where b_l can be computed directly using the eigenvalues of $\boldsymbol{M}$ without the need of sample splitting. As we shall see shortly, this new estimator is put forward in order to remedy a non-negligible bias issue underlying the naive plug-in estimator.

The following theorem quantifies the estimation errors for both of these estimators.

Theorem 1 (Eigenvector perturbation)**.** *Consider any $1 \le l \le r$, and suppose that*

$$\sigma\sqrt{n} \le c_0 \lambda_{\min}^{\star}, \qquad r \le c_1 n / \log^2 n \qquad \text{and} \qquad \Delta_l^{\star} > C_0 \sigma \sqrt{r} \log n \tag{3.3}$$

for some sufficiently small (resp. large) constants $c_0, c_1 > 0$ (resp. $C_0 > 0$). Let $\boldsymbol{a} \in \mathbb{R}^n$ be any fixed vector with $\|\boldsymbol{a}\|_2 = 1$. With probability at least $1 - O(n^{-10})$, the estimators in (3.2) satisfy

$$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{plugin}}, \boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right) \lesssim E_{\mathsf{md},l} + \frac{\sigma^2 n}{\lambda_l^{\star 2}} \left|\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right|, \tag{3.4a}$$

$$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{debiased}}, \boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right) \lesssim E_{\mathsf{md},l}, \tag{3.4b}$$

where $E_{\mathsf{md},l}$ is defined as

$$E_{\mathsf{md},l} := \frac{\sigma^2 r \log n}{(\Delta_l^{\star})^2} \left|\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right| + \sigma \sqrt{r \log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^{\star}}\right)} \sum_{k:\, k \ne l, 1 \le k \le r} \frac{\left|\boldsymbol{a}^{\top}\boldsymbol{u}_k^{\star}\right|}{\left|\lambda_l^{\star} - \lambda_k^{\star}\right|} + \frac{\sigma\sqrt{\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^{\star}}\right)}}{\left|\lambda_l^{\star}\right|}. \tag{3.5}$$

Remark 1. While the rank r of the true matrix $\boldsymbol{M}^{\star}$ might be unknown *a prior* in practice, it can often be estimated accurately in a data-driven manner. For instance, under the model assumed herein, one might simply choose r by identifying the smallest (in magnitude) eigenvalue λ_l that is larger than λ_{l+1} by an order of $\sigma\sqrt{n}$; see also Han et al. [2019] for a different approach.

Remark 2. While the quantity b_l is provided in a data-driven manner (cf. (3.2b)), we find it helpful to also make note of another expression derived from its asymptotic limit. Specifically, the random matrix theory tells us that the eigenvalues $\{\lambda_i\}_{r<i\le n}$ of $\boldsymbol{M}$ obey the celebrated semi-circular law asymptotically (see, e.g., Knowles and Yin [2013]), and therefore the de-biased term b_l satisfies (as n grows):

$$b_l = \sum_{i:\, r<i\le n} \frac{\sigma^2}{(\lambda_l - \lambda_i)^2} \approx \int_{-2}^{2} \frac{\sqrt{4-\lambda^2}}{2\pi\left(\frac{\lambda_l}{\sigma\sqrt{n}} - \lambda\right)^2} \mathrm{d}\lambda. \tag{3.6}$$

Implications. Theorem 1 develops statistical performance guarantees for the aforementioned two estimators when estimating the linear form $\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}$ for a prescribed vector $\boldsymbol{a} \in \mathbb{R}^n$. We now single out several main implications of our results.

- *Estimation guarantees in the face of a small eigen-gap.* In view of (3.3), the eigen-gap $\Delta_l^{\star}$ is allowed to be substantially smaller than the spectral norm $\|\boldsymbol{H}\|$ of the perturbation matrix. This stands in stark contrast to, and significantly improves upon, the celebrated Davis-Kahan $\sin\boldsymbol{\Theta}$ theorem that requires $\Delta_l^{\star} \gtrsim \|\boldsymbol{H}\|$ [Chen et al., 2021b, Davis and Kahan, 1970]. To be more precise, recalling from standard random matrix theory [Tao, 2012] that $\|\boldsymbol{H}\| \asymp \sigma\sqrt{n}$ with high probability, one can compare our result with classical matrix perturbation theory as follows

$$\text{our eigen-gap requirement:} \quad \Delta_l^{\star} = \widetilde{\Omega}\left(\sigma\sqrt{r}\right);$$
$$\text{eigen-gap requirement in classical theory:} \quad \Delta_l^{\star} = \widetilde{\Omega}\left(\sigma\sqrt{n}\right).$$

  As a comparison, the prior work Bao et al. [2021] studied the distributions of the singular vectors under the matrix denoising setting with $\sigma \asymp n^{-1/2}$, provided that the eigen-gap exceeds $\Omega(1)$; our theory improves their eigen-gap condition by a factor on the order of $\sqrt{n/r}$.

- *Near minimaxity.* In order to assess the effectiveness of our proposed estimator, it is helpful to compare the statistical guarantees in Theorem 1 with minimax lower bounds. Consider, for simplicity, the scenario where $r = O(1)$ and $|\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}| \le (1-\epsilon)\|\boldsymbol{a}\|_2$ for any small non-zero constant $\epsilon > 0$ (so that $\boldsymbol{a}$ is not perfectly aligned with $\boldsymbol{u}_l^{\star}$), and an instance-dependent minimax lower bound has been established in Cheng et al. [2021, Theorem 3] for this scenario. Specifically, if we define the following two sets

$$\mathcal{M}_0(\boldsymbol{M}^{\star}) := \left\{\boldsymbol{A} \mid \mathsf{rank}(\boldsymbol{A}) = r,\, \lambda_i(\boldsymbol{A}) = \lambda_i^{\star}\ (1 \le i \le r),\, \|\boldsymbol{A} - \boldsymbol{M}^{\star}\|_{\mathrm{F}} \le \frac{\sigma}{2}\right\},$$

$$\mathcal{M}_1(\boldsymbol{M}^\star) := \Big\{ \boldsymbol{A} \mid \mathsf{rank}(\boldsymbol{A}) = r,\, \lambda_i(\boldsymbol{A}) = \lambda_i^\star \ (1 \leq i \leq r),\, \|\boldsymbol{u}_l(\boldsymbol{A}) - \boldsymbol{u}_l^\star\|_2 \leq \frac{\sigma}{4\,|\lambda_l^\star|} \Big\},$$

then we necessarily have

$$\inf_{u_{\boldsymbol{a},l}} \sup_{\boldsymbol{A} \in \mathcal{M}_0(\boldsymbol{M}^\star) \cup \mathcal{M}_1(\boldsymbol{M}^\star)} \mathbb{E}\Big[\mathsf{dist}\left(u_{\boldsymbol{a},l},\, \boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{A})\right)\Big] \gtrsim \frac{\sigma^2}{(\Delta_l^\star)^2}\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| + \sigma \max_{k:\, k \neq l} \frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{|\lambda_l^\star - \lambda_k^\star|} + \frac{\sigma}{|\lambda_l^\star|}, \tag{3.8a}$$

where the infimum is over all estimators $u_{\boldsymbol{a},l}$ based on the observed matrix $\boldsymbol{M} = \boldsymbol{A} + \boldsymbol{H}$, and $\boldsymbol{u}_l(\boldsymbol{A})$ denotes the l-th eigenvector of the matrix $\boldsymbol{A}$. In addition, the analysis for Cheng et al. [2021, Theorem 3] directly implies that

$$\inf_{u_{\boldsymbol{a},l}} \sup_{\boldsymbol{A} \in \mathcal{M}_0(\boldsymbol{M}^\star) \cup \mathcal{M}_1(\boldsymbol{M}^\star)} \mathbb{P}\Big\{\mathsf{dist}\left(u_{\boldsymbol{a},l},\, \boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{A})\right) \gtrsim \frac{\sigma^2}{(\Delta_l^\star)^2}\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| + \sigma \max_{k:\, k \neq l} \frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{|\lambda_l^\star - \lambda_k^\star|} + \frac{\sigma}{|\lambda_l^\star|}\Big\} \geq \frac{1}{5}. \tag{3.8b}$$

In comparison, our statistical guarantee (3.4b) for the proposed de-biased estimator obeys

$$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{debiased}}, \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right) \leq \widetilde{O}\bigg(\frac{\sigma^2}{(\Delta_l^\star)^2}\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| + \sigma \sum_{k:\, k \neq l} \frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{|\lambda_l^\star - \lambda_k^\star|} + \frac{\sigma}{|\lambda_l^\star|}\bigg)$$

with high probability in this scenario, thereby matching the minimax lower bound (3.8) (modulo some logarithmic factor). This confirms the near optimality of our de-biased estimator when $r \asymp 1$.

Sub-optimality of the vanilla plug-in estimator. Furthermore, Theorem 1 suggests that the statistical error (3.4a) of the vanilla plug-in estimator $\boldsymbol{a}^\top \boldsymbol{u}_l$ might contain an additional "bias" term

$$E_{\mathsf{md},l}^{\mathsf{bias}} := \frac{\sigma^2 n}{\lambda_l^{\star 2}}\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \tag{3.9}$$

when compared to that of the de-biased estimator (cf. (3.4b)). It is natural to wonder if the theoretical guarantee of the plug-in estimator in (3.4a) is tight or not. To answer the question, we develop the following lower bound on the estimation error of the plug-in estimator $u_{\boldsymbol{a}}^{\mathsf{plugin}}$; the proof is deferred to Appendix F.

Theorem 2. *Instate the assumptions of Theorem 1. Let $\boldsymbol{a} \in \mathbb{R}^n$ be any fixed vector with $\|\boldsymbol{a}\|_2 = 1$. With probability at least $1/3$, the plug-in estimator in (3.2) satisfies*

$$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{plugin}}, \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right) \gtrsim \frac{\sigma^2 n}{\lambda_l^{\star 2}}\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|. \tag{3.10}$$

In short, Theorem 2 demonstrates that it is impossible for the plug-in estimator to get rid of this "bias" term (3.9). The influence of this extra term becomes increasingly large and non-negligible as the correlation of $\boldsymbol{a}$ and $\boldsymbol{u}_l^\star$ increases. To demonstrate the possibly severe impact incurred by this additional term, let us examine a simple case as follows.

- *Example.* Suppose that $r = 2$, $\lambda_1^\star = 2\lambda_2^\star$ (so that $\lambda_1^\star - \lambda_2^\star \asymp \lambda_1^\star$), $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \asymp 1$ and $\sigma\sqrt{n} \asymp |\lambda_1^\star|$. As can be straightforwardly verified, the main term (3.5) and the addition term (3.9) in this example satisfy

$$E_{\mathsf{md},1} = \widetilde{O}\bigg(\frac{\sigma^2}{\lambda_1^{\star 2}}\left|\boldsymbol{a}^\top \boldsymbol{u}_1^\star\right| + \frac{\sigma}{|\lambda_1^\star|}\left|\boldsymbol{a}^\top \boldsymbol{u}_2^\star\right| + \frac{\sigma}{|\lambda_1^\star|}\bigg) = \widetilde{O}\bigg(\frac{\sigma^2}{\lambda_1^{\star 2}} + \frac{\sigma}{|\lambda_1^\star|}\bigg) = \widetilde{O}\bigg(\frac{1}{\sqrt{n}}\bigg);$$

$$E_{\mathsf{md},1}^{\mathsf{bias}} \asymp \frac{\sigma^2 n}{\lambda_1^{\star 2}}\left|\boldsymbol{a}^\top \boldsymbol{u}_1^\star\right| \asymp 1.$$

  In other words, the additional bias term $E_{\mathsf{md},1}^{\mathsf{bias}}$ could be a factor of $\widetilde{O}(\sqrt{n})$ times larger than the main term $E_{\mathsf{md},1}$ in this case, and cannot be neglected.

The above discussion reveals the necessity of proper bias correction in order to mitigate the undesired effect of the bias term $E_{\mathsf{md},l}^{\mathsf{bias}}$. Aimed at addressing this issue, our de-biased estimator $u_{\boldsymbol{a}}^{\mathsf{debiased}}$ compensates for the bias term $E_{\mathsf{md},l}^{\mathsf{bias}}$ by properly rescaling the plug-in estimator by a data-driven correction factor $\sqrt{1 + b_l}$. Note that when the signal-to-noise ratio is sufficiently large such that $|\lambda_l^\star| \gtrsim \sigma n$, then $E_{\mathsf{md},l}$ becomes the dominant term in the error bound; in such a case, there is no need for bias correction.

Comparisons with prior works. While estimation of linear forms of eigenvectors remains largely under-explored in the literature, a small number of prior works have studied this problem or its variants. Among them, perhaps the one that is the closest to the current paper is Koltchinskii and Xia [2016], which considered estimating linear forms of singular vectors under i.i.d. Gaussian noise. In what follows, we briefly compare our result with Koltchinskii and Xia [2016], focusing on the setting where the ground-truth matrix is symmetric (so that the eigenvectors and the singular vectors become identical up to global signs).

- To begin with, the theory in Koltchinskii and Xia [2016, Theorem 1.3] operates under the assumption

  $$\Delta_l^\star = \Omega\big(\mathbb{E}[\|\boldsymbol{H}\|]\big) = \Omega(\sigma\sqrt{n}),$$

  which is $\widetilde{O}\big(\sqrt{n/r}\big)$ times more stringent than the eigen-gap condition imposed in our theory (see (3.7)).

- The estimation bias of the plug-in estimator was already pointed out in Koltchinskii and Xia [2016]. However, the approach proposed in Koltchinskii and Xia [2016] required additional independent copies of $\boldsymbol{M}$ in order to estimate — and hence correct — the bias effect (see Koltchinskii and Xia [2016, Section 1]). By contrast, our de-biased estimator does not require an additional set of data samples and allows one to use all available information fully.

- Next, we compare our theoretical guarantee with the one developed for the de-biased estimator $u_{\boldsymbol{a}}^{\mathsf{debiased,KD}}$ proposed in Koltchinskii and Xia [2016]. When $r \asymp 1$, Koltchinskii and Xia [2016, Theorem 1.3] asserts that

  $$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{debiased,KD}}, \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right) \le \widetilde{O}\left(\frac{\sigma}{\Delta_l^\star}\right) =: E_{\mathsf{md},l}^{\mathsf{KD}},$$

  provided that $\Delta_l^\star \gtrsim \sigma\sqrt{n}$. This result, however, might fall short of attaining minimax optimality. More specifically, comparing our error bound $E_{\mathsf{md},l}$ (cf. (3.5)) with $E_{\mathsf{md},l}^{\mathsf{KD}}$ makes clear that the theoretical gain is on the order of

  $$\frac{E_{\mathsf{md},l}^{\mathsf{KD}}}{E_{\mathsf{md},l}} = \widetilde{O}\left(\frac{\Delta_l^\star}{\sigma\,|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|} \wedge \frac{1}{\sum_{k:k\neq l}|\boldsymbol{a}^\top \boldsymbol{u}_k^\star|} \wedge \frac{|\lambda_l^\star|}{\Delta_l^\star}\right).$$

  For concreteness, consider the case with $r \asymp 1$, $|\boldsymbol{a}^\top \boldsymbol{u}_k^\star| \asymp 1/\sqrt{n}$ for all $k \neq l$, $\Delta_l^\star \asymp |\lambda_l^\star|/\sqrt{n}$, and $\Delta_l^\star \asymp \sigma\sqrt{n}$, thus leading to the gain

  $$\frac{E_{\mathsf{md},l}^{\mathsf{KD}}}{E_{\mathsf{md},l}} = \widetilde{O}\left(\sqrt{n}\right).$$

  In other words, our results might lead to considerable theoretical improvement over Koltchinskii and Xia [2016] in the presence of a small eigen-gap.

  Remark 3. Note that the case $|\boldsymbol{a}^\top \boldsymbol{u}_k^\star| \asymp 1/\sqrt{n}$ is of particular interest if one studies entrywise statistical performance; namely, when $\boldsymbol{a}$ is taken to be the standard basis (i.e. $\boldsymbol{a} = \boldsymbol{e}_i$ for some $i \in [n]$) and when the energy of $\boldsymbol{u}_k^\star$ is more or less spread out across all entries (i.e. $\|\boldsymbol{u}_k^\star\|_\infty \asymp \|\boldsymbol{u}_k^\star\|_2/\sqrt{n}$).

3.2 Principal component analysis

Next, we turn attention to the problem of principal component analysis as formulated in Section 2.2. Denote by

$$\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n] \in \mathbb{R}^{p\times n}$$

the data matrix whose columns consist of i.i.d. samples $\boldsymbol{s}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, and let λ_l represent the l-th largest eigenvalue of $\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top$ with associated eigenvector $\boldsymbol{u}_l$. Our focus is the following two estimators aimed at estimating the linear form $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ $(1 \le l \le r)$.

- A plug-in estimator:

  $$u_{\boldsymbol{a}}^{\mathsf{plugin}} := \boldsymbol{a}^\top \boldsymbol{u}_l; \tag{3.11a}$$

- A "de-biased" estimator:
$$u_{\boldsymbol{a}}^{\mathsf{debiased}} := \sqrt{1+c_l}\,\boldsymbol{a}^{\top}\boldsymbol{u}_l. \tag{3.11b}$$
Here, c_l is a quantity that can be directly computed using the spectrum of $\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^{\top}$ as follows:
$$c_l := \begin{cases} \dfrac{\lambda_l}{n+\sum_{i:\,r<i\le n}\frac{\lambda_i}{\lambda_l-\lambda_i}}\displaystyle\sum_{i:\,r<i\le n}\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}, & \text{if } n\ge p,\\ \dfrac{\frac{\sigma^2 p}{n}}{\lambda_l-\frac{\sigma^2 p}{n}}+\dfrac{\lambda_l}{\lambda_l-\frac{\sigma^2 p}{n}}\dfrac{\lambda_l}{n+\sum_{i:\,r<i\le n}\frac{\lambda_i}{\lambda_l-\lambda_i}}\displaystyle\sum_{i:\,r<i\le n}\frac{\lambda_i-\frac{\sigma^2 p}{n}}{(\lambda_l-\lambda_i)^2}, & \text{if } n<p, \end{cases} \tag{3.12}$$
without any need of using sample splitting.

Akin to the matrix denoising counterpart, the plug-in estimator (3.11a) often incurs some non-negligible estimation bias, which motivates the design of the adjusted estimator (3.11b) to compensate for the bias.

We are now ready to present our statistical guarantees for the two estimators introduced in (3.11).

Theorem 3. *Consider any $1\le l\le r$, and assume that*
$$\lambda_{\max}^{\star}\sqrt{\frac{r}{n}}+\sqrt{\lambda_{\max}^{\star}\sigma^2\frac{p}{n}}+\sigma^2\left(\frac{p}{n}+\sqrt{\frac{p}{n}}\right)\le C_0\frac{\lambda_{\min}^{\star}}{\log^2 n} \tag{3.13a}$$
$$\text{and}\qquad \Delta_l^{\star}>C_1(\lambda_{\max}^{\star}+\sigma^2)\sqrt{\frac{r}{n}}\log n \tag{3.13b}$$
hold for some sufficiently small (resp. large) constant $C_0>0$ (resp. $C_1>0$). Consider any fixed vector $\boldsymbol{a}\in\mathbb{R}^p$ with $\|\boldsymbol{a}\|_2=1$. Then with probability at least $1-O(n^{-10})$, the estimators in (3.11) satisfy
$$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{plugin}},\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right)\lesssim E_{\mathsf{PCA},l}+\frac{(\lambda_l^{\star}+\sigma^2)\sigma^2 p}{\lambda_l^{\star 2}n}\left|\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right|, \tag{3.14a}$$
$$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{debiased}},\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right)\lesssim E_{\mathsf{PCA},l}, \tag{3.14b}$$
where the quantity $E_{\mathsf{PCA},l}$ is defined as
$$\begin{aligned} E_{\mathsf{PCA},l} := {} & \frac{(\lambda_{\max}^{\star}+\sigma^2)(\lambda_l^{\star}+\sigma^2)\,r\log n}{(\Delta_l^{\star})^2 n}\left|\boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right|+\sqrt{\frac{(\lambda_{\max}^{\star}+\sigma^2)\,\sigma^2\kappa^2 r}{\lambda_l^{\star 2}n}}\log^2 n \\ & +\sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^{\top}\boldsymbol{u}_k^{\star}\right|}{\left|\lambda_l^{\star}-\lambda_k^{\star}\right|\sqrt{n}}\sqrt{(\lambda_l^{\star}+\sigma^2)(\lambda_{\max}^{\star}+\sigma^2)(\kappa^2+r)\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^{\star}}\right)}. \end{aligned} \tag{3.15}$$

Remark 4. Akin to the matrix denoising case, while the expression of the de-biasing term c_l (cf. (3.12)) is fully data-driven and preferable in practice, we remark that the Marchenko-Pastur law allows us to approximate the de-biasing term c_l as follows (as n grows)
$$c_l\ \approx\ \begin{cases} \dfrac{\lambda_l}{1+\int\frac{\lambda}{\lambda_l-\lambda}\mu(\mathrm{d}\lambda)}\displaystyle\int\frac{\lambda}{(\lambda_l-\lambda)^2}\mu(\mathrm{d}\lambda), & \text{if } n\ge p,\\ \dfrac{\frac{\sigma^2 p}{n}}{\lambda_l-\frac{\sigma^2 p}{n}}+\dfrac{\lambda_l}{\lambda_l-\frac{\sigma^2 p}{n}}\dfrac{\lambda_l}{1+\int\frac{\lambda}{\lambda_l-\lambda}\mu(\mathrm{d}\lambda)}\displaystyle\int\frac{\lambda-\frac{\sigma^2 p}{n}}{(\lambda_l-\lambda)^2}\mu(\mathrm{d}\lambda), & \text{if } n<p, \end{cases} \tag{3.16}$$
where
$$\mu(\mathrm{d}\lambda)=\frac{n\sqrt{(\lambda_{+}-\lambda)(\lambda-\lambda_{-})}}{2\pi\sigma^2 p\lambda}\mathbb{1}\{\lambda_{-}\le\lambda\le\lambda_{+}\}\mathrm{d}\lambda\quad\text{with}\quad\lambda_{\pm}=\sigma^2(1\pm\sqrt{p/n})^2. \tag{3.17}$$

Implications. In short, Theorem 3 characterizes the statistical accuracy of both the plug-in estimator and the modified de-biased estimator, the latter of which enjoys improved statistical guarantees. In the sequel, we single out a few implications of this result.

- *Estimation guarantees.* Let us first assess the statistical error bound of the de-biased estimator (namely, $E_{\mathsf{PCA},l}$ in (3.15)). For simplicity of presentation, we shall focus on the case with $r, \kappa \asymp 1$, where the error term $E_{\mathsf{PCA},l}$ admits the following simpler expression

$$E_{\mathsf{PCA},l} = \widetilde{\Theta}\bigg(\frac{(\lambda_l^\star + \sigma^2)^2}{(\Delta_l^\star)^2 n}\big|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\big| + (\lambda_l^\star + \sigma^2)\max_{k:k\neq l}\frac{\big|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\big|}{|\lambda_l^\star - \lambda_k^\star|\sqrt{n}} + \frac{\sigma}{\lambda_l^\star\sqrt{n}}\sqrt{\lambda_l^\star + \sigma^2}\bigg). \tag{3.18}$$

  In particular, the first term on the right-hand side of (3.18) quantifies the role of the ground truth $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ on the estimation error, which scales inverse quadratically in the eigen-gap $\Delta_l^\star$; the second term on the right-hand side of (3.18) can be understood as the additional interference resulting from the linear form of other eigenvectors (namely, $\boldsymbol{a}^\top \boldsymbol{u}_k^\star$ for $k \neq l$), which is inversely proportional to the corresponding eigen-gap $|\lambda_l^\star - \lambda_k^\star|$.

- *Relaxed eigen-gap condition.* To simplify discussions, let us again focus on the case with $r, \kappa \asymp 1$ and omit logarithmic factors. Classical matrix perturbation theory (e.g., the Davis-Kahan $\sin\boldsymbol{\Theta}$ theorem [Davis and Kahan, 1970]) requires the eigen-gap to exceed the size of perturbation, namely,

$$\Delta_l^\star \gtrsim \Big\|\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{\Sigma}\Big\|.$$

  As it turns out, the eigen-gap requirement above leads to the following condition (by invoking the high-probability bound to be presented shortly in Lemma 7)

$$\Delta_l^\star \gtrsim \frac{\lambda_l^\star}{\sqrt{n}} + \sqrt{\frac{\lambda_l^\star \sigma^2 p}{n}} + \sigma^2\bigg(\sqrt{\frac{p}{n}} + \frac{p}{n}\bigg) =: \mathsf{gap}_{\mathsf{DK}}.$$

  In comparison, the eigen-gap condition (3.13b) in Theorem 3 reads

$$\Delta_l^\star \gtrsim \frac{\lambda_l^\star + \sigma^2}{\sqrt{n}} =: \mathsf{gap}.$$

  To better understand and compare these two eigen-gap requirements, we shall discuss them for a couple of distinct scenarios.

  - If $\sigma^2\big(\sqrt{\frac{p}{n}} + \frac{p}{n}\big) \lesssim \lambda_l^\star \lesssim \sigma^2$ (the sample size needs to satisfy $n \geq p$ by the assumption (3.13a)), the eigen-gap conditions above simplify to

$$\mathsf{gap} \asymp \frac{\sigma^2}{\sqrt{n}} \qquad \text{and} \qquad \mathsf{gap}_{\mathsf{DK}} \asymp \sigma^2\sqrt{\frac{p}{n}}.$$

$$\Longrightarrow \qquad \frac{\mathsf{gap}_{\mathsf{DK}}}{\mathsf{gap}} \asymp \sqrt{p}.$$

  - If $\sigma^2 \lesssim \lambda_l^\star \lesssim \sigma^2 p$, then one has

$$\mathsf{gap} \asymp \frac{\lambda_l^\star}{\sqrt{n}} \qquad \text{and} \qquad \mathsf{gap}_{\mathsf{DK}} \asymp \sqrt{\frac{\lambda_l^\star\sigma^2 p}{n}} + \frac{\sigma^2 p}{n}.$$

    Comparing these two terms reveals that

$$\frac{\mathsf{gap}_{\mathsf{DK}}}{\mathsf{gap}} \asymp \sqrt{\frac{\sigma^2 p}{\lambda_l^\star}}\bigg(1 + \sqrt{\frac{\sigma^2 p}{\lambda_l^\star n}}\bigg) \overset{\text{(i)}}{\asymp} \sqrt{\frac{\sigma^2 p}{\lambda_l^\star}} \overset{\text{(ii)}}{\gtrsim} 1,$$

    where (i) holds due to the assumption (3.13a) and (ii) follows from the condition $\lambda_l^\star \lesssim \sigma^2 p$.

– If $\lambda_l^\star \gtrsim \sigma^2 p$, then it is straightforward to see that

$$\mathsf{gap} \asymp \mathsf{gap}_{\mathsf{DK}} \asymp \frac{\lambda_l^\star}{\sqrt{n}}.$$

To sum up, our eigen-gap requirement (3.13b) is

$$\Omega\left(\sqrt{p\Big(1 \wedge \frac{\sigma^2}{\lambda_l^\star}\Big)} \vee 1\right)$$

times less stringent than the one demanded in classical matrix perturbation theory, thereby justifying the improvement of our results upon prior art. In addition, we note that [Bao et al., 2022] also considered statistical inference for principal components of spike covariance matrices; when $\sigma = 1$, the eigen-gap therein needs to satisfy $\Delta_l^\star \gtrsim n^{-1/2+\epsilon}$ for an arbitrary small fixed constant $\epsilon > 0$, thereby leading to a more stringent condition than ours.

- *Bias reduction.* Similar to the matrix denoising case, the plug-in estimator $\boldsymbol{a}^\top \boldsymbol{u}_l$ suffers from the following extra "bias" term in comparison to the de-biased estimator (3.14b):

$$E_{\mathsf{pca},l}^{\mathsf{bias}} := \frac{(\lambda_l^\star + \sigma^2)\sigma^2 p}{\lambda_l^{\star 2} n} \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|. \tag{3.19}$$

If $\boldsymbol{a}$ and $\boldsymbol{u}_l^\star$ are fairly correlated, then this additional term becomes non-negligible and might affect the estimation accuracy negatively. To see this, let us consider the following simple case.

– *Example.* Assume that $r = 2$, $\lambda_1^\star = 2\lambda_2^\star > 0$, $|\boldsymbol{a}^\top \boldsymbol{u}_1^\star| \asymp 1$, $\sigma^2 \asymp \lambda_1^\star$ and $p \asymp n$. As can be straightforwardly verified, the error terms (3.15) and (3.19) in this case become

$$\begin{aligned} E_{\mathsf{pca},1} &= \widetilde{O}\bigg(\frac{1}{n}\left|\boldsymbol{a}^\top \boldsymbol{u}_1^\star\right| + \frac{1}{\sqrt{n}}\left|\boldsymbol{a}^\top \boldsymbol{u}_2^\star\right| + \frac{1}{\sqrt{n}}\bigg) = \widetilde{O}\Big(\frac{1}{\sqrt{n}}\Big); \\ E_{\mathsf{pca},1}^{\mathsf{bias}} &\asymp \frac{p}{n}\left|\boldsymbol{a}^\top \boldsymbol{u}_1^\star\right| \asymp 1. \end{aligned}$$

In other words, the bias term $E_{\mathsf{pca},1}^{\mathsf{bias}}$ could be $\sqrt{n}$ times larger than the error term $E_{\mathsf{pca},1}^{\mathsf{bias}}$ (up to some logarithmic factor).

As a takeaway message from the above example, it is crucial to reduce the bias incurred by $E_{\mathsf{pca},1}^{\mathsf{bias}}$. The proposed de-biased estimator $u_{\boldsymbol{a}}^{\mathsf{debiased}}$ achieves bias reduction by enlarging the plug-in estimator by a factor of $\sqrt{1 + c_l}$, where c_l is computable in a data-driven manner. It is worth noting that the factor c_l (cf. (3.12)) takes two different forms, depending on the relative ratio between the sample size n and the dimension p.

Minimax lower bounds and optimality. In order to evaluate the tightness of our statistical guarantees, we develop minimax lower bounds for PCA. Here and below, we denote by $\boldsymbol{u}_l(\boldsymbol{\Sigma}) \in \mathbb{R}^p$ the eigenvector associated with the l-th largest eigenvalue of a matrix $\boldsymbol{\Sigma}$, and we define two sets of covariance matrices as follows:

$$\mathcal{M}_1(\boldsymbol{\Sigma}^\star) := \left\{\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p} : \mathsf{rank}(\boldsymbol{\Sigma}) = r,\ \lambda_i(\boldsymbol{\Sigma}) = \lambda_i^\star\ (1 \le i \le r),\ \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^\star\|_{\mathrm{F}} \le \max_{k:\, k\neq l}\sqrt{\frac{(\lambda_l^\star + \sigma^2)(\lambda_k^\star + \sigma^2)}{n}}\right\}.$$

$$\mathcal{M}_2(\boldsymbol{\Sigma}^\star) := \left\{\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p} : \mathsf{rank}(\boldsymbol{\Sigma}) = r,\ \lambda_i(\boldsymbol{\Sigma}) = \lambda_i^\star\ (1 \le i \le r),\ \|\boldsymbol{u}_l(\boldsymbol{\Sigma}) - \boldsymbol{u}_l^\star\|_2 \le \sqrt{\frac{(\lambda_l^\star + \sigma^2)\sigma^2}{\lambda_l^{\star 2} n}}\right\}.$$

Theorem 4. *Consider any fixed vector $\boldsymbol{a} \in \mathbb{R}^p$. For any given $\boldsymbol{\Sigma}$, let $\{\boldsymbol{s}_i\}_{i=1}^n$ be independent samples satisfying $\boldsymbol{s}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma} + \sigma^2 \boldsymbol{I}_p)$. Assume that the sample size obeys*

$$n \ge \left\{\max_{k:\, k\neq l} \frac{(\lambda_k^\star + \sigma^2)(\lambda_l^\star + \sigma^2)}{|\lambda_l^\star - \lambda_k^\star|^2}\right\} \vee \frac{(\lambda_l^\star + \sigma^2)\sigma^2}{\lambda_l^{\star 2}}. \tag{3.20}$$

Then one has

$$\inf_{u_{\boldsymbol{a},l}} \sup_{\boldsymbol{\Sigma}\in\mathcal{M}_1(\boldsymbol{\Sigma}^\star)} \mathbb{E}\Big[\min\big|u_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{\Sigma})\big|\Big]$$

$$\gtrsim \max_{k:\, k\neq l,\, 1\leq k\leq r} \frac{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}{|\lambda_l^\star-\lambda_k^\star|^2\, n}\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big| + \max_{k:\, k\neq l,\, 1\leq k\leq r} \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big| =: E_{\mathsf{lb1},l};$$

$$\inf_{u_{\boldsymbol{a},l}} \sup_{\boldsymbol{\Sigma}\in\mathcal{M}_2(\boldsymbol{\Sigma}^\star)} \mathbb{E}\Big[\min\big|u_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{\Sigma})\big|\Big]$$

$$\gtrsim \sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2} n}}\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\|_2 =: E_{\mathsf{lb2},l}.$$

Here, the infimum is taken over all estimator $u_{\boldsymbol{a},l}$ for the linear form of the l-th eigenvector.

The proof of this theorem can be found in Appendix E. To interpret this lower bound, let us consider, for simplicity, the scenario where

$$r,\kappa \asymp 1 \qquad \text{and} \qquad |\boldsymbol{a}^\top\boldsymbol{u}_l^\star| \leq (1-\epsilon)\|\boldsymbol{a}\|_2 \tag{3.22}$$

for some arbitrarily small constant $\epsilon > 0$. In this scenario, the statistical error bound (3.14b) derived in Theorem 3 matches the preceding minimax lower bounds in the sense that

$$E_{\mathsf{PCA},l} \asymp E_{\mathsf{lb1},l} + E_{\mathsf{lb2},l}.$$

To verify this relation under the conditions (3.22), it is sufficient to see that

$$\max_{k:k\neq l} \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big| + \sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2} n}}\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\|_2$$

$$\asymp \sum_{k:\, k\neq l} \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big| + \sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2} n}}\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\|_2$$

$$\overset{\text{(i)}}{\asymp} \sum_{k:\, k\neq l} \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big| + \sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2} n}}\Big(\sum_{k:\, k\neq l}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big| + \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\|_2\Big)$$

$$\overset{\text{(ii)}}{\asymp} \sum_{k:\, k\neq l} \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big| + \sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2} n}}\|\boldsymbol{a}\|_2,$$

where (i) holds true since $\max_{k:k\neq l} \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}} \gtrsim \sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2} n}}$, and (ii) holds true as long as $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star| \leq (1-\epsilon)\|\boldsymbol{a}\|_2$ for some constant $\epsilon > 0$. In conclusion, the above calculation unveils the statistical optimality of the proposed de-biased estimator for the scenario specified in (3.22).

Comparison with past works. Estimation for linear forms of eigenvectors in the context of PCA has been investigated in several recent works [Koltchinskii et al., 2016, 2017, 2020], with the bias issue of plug-in estimators first recognized in Koltchinskii et al. [2016]. Among these works, the state-of-the-art result was due to Koltchinskii et al. [2020], which proposed an efficient de-biased estimator and established its asymptotic normality. To better understand our contributions, it is helpful to compare Theorem 3 with the theoretical guarantees in Koltchinskii et al. [2020] under the spiked covariance model with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^\star + \sigma^2\boldsymbol{I}_p$. The theoretical guarantees developed in Koltchinskii et al. [2020, Theorem 3.3] operate under the following conditions (when translated to our setting using our notation)

$$\Delta_l^\star = \Omega(\lambda_{\max}^\star+\sigma^2), \quad \sigma^2 = o(\lambda_{\min}^\star), \quad r,\kappa\asymp 1, \quad \sum_{k:\, k\neq l}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big|^2 + \frac{\sigma^2}{\lambda_l^\star}\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\|_2^2 \asymp \|\boldsymbol{a}\|_2^2. \tag{3.23}$$

In comparison, our results make improvements in the following aspects:

- *Eigen-gap requirement*: our eigen-gap requirement (3.13b) is $\widetilde{O}(\sqrt{r/n})$ times less stringent than the one in (3.23);
- *Requirement on noise variance*: our result (i.e., Theorem 3) allows the noise variance σ^2 to be larger than $\lambda_{\min}^\star$;
- *Requirement on condition number and rank:* our theory permits both κ and r to grow with the dimension.

It is worth noting that Koltchinskii et al. [2020] accommodates a more general class of covariance matrices than the aforementioned spiked covariance. The main purpose of our discussion above is to make clear the inadequacy of prior theories when the eigen-gap is small.

4 Related works

Spectral methods have served as an effective paradigm for a variety of statistical data science problems, examples including matrix completion [Keshavan et al., 2010a,b, Ma et al., 2020, Sun and Luo, 2016], tensor completion [Cai et al., 2020, 2022, Montanari and Sun, 2018, Xia et al., 2021], community detection [Abbe et al., 2020b, Lei, 2019], ranking from pairwise comparisons [Chen and Suh, 2015, Negahban et al., 2017], and so on. The mainstream analysis framework for spectral methods is largely built upon classical matrix perturbation theory [Chen et al., 2021b, Stewart and Sun, 1990]. This set of classical theory typically focuses on deriving ℓ_2 eigenspace or singular subspace perturbation bounds (e.g., the Davis-Kahan theorem [Davis and Kahan, 1970] and the Wedin theorem [Wedin, 1972]), which has been derived for general purposes without incorporating statistical properties of the specific problems of interest. Several useful extensions have been developed tailored to high-dimensional statistical applications, particularly when the perturbation matrix of interest enjoys certain random structure [Cai and Zhang, 2018, O'Rourke et al., 2018, Vu, 2011, Wang, 2015, Xia, 2019, Yu et al., 2015]. In particular, the ℓ_2 perturbation bounds for the eigenvector (or eigenspace) of the sample covariance matrix has been extensively studied in the PCA literature, e.g., [Johnstone and Lu, 2009, Lounici, 2013, 2014, Nadler, 2008, Vu and Lei, 2012, Xia, 2021, Zhang et al., 2022, Zhu et al., 2019]. Another line of works [O'Rourke et al., 2018, Vu, 2011] improved Davis-Kahan's and Wedin's theorems in the matrix denoising setting with small eigen-gaps, which, however, is not tight unless the spectral norm $\|\boldsymbol{H}\|$ of the noise matrix is extremely small.

Moving beyond ℓ_2 perturbation theory, more fine-grained eigenvector perturbation bounds — particularly entrywise eigenvector perturbation or $\ell_{2,\infty}$ eigenspace perturbation — has garnered growing attention over the past few years [Abbe et al., 2020a,b, Cai et al., 2021, 2022, Cape et al., 2019, Chen et al., 2020b, 2021a, Fan et al., 2018, Lei, 2019, Ma et al., 2020, Zhong and Boumal, 2018]. Among these ℓ_∞ or $\ell_{2,\infty}$ theoretical guarantees, the results in Abbe et al. [2020b], Cai et al. [2021, 2022], Chen et al. [2020a, 2019a,b, 2020b], Ma et al. [2020] were established via a powerful leave-one-out analysis framework, while the works [Chen et al., 2021a, Eldridge et al., 2018] invoked a Neumann expansion trick paired with proper control of moments.

In contrast to the rich literature on ℓ_2, ℓ_∞ and/or $\ell_{2,\infty}$ perturbation theory, estimation theory concerning linear functionals of eigenvectors (or singular vectors) are rather scarce and under-explored. While entrywise perturbation can be regarded as a special type of linear functionals of eigenvectors, the analysis techniques mentioned above are typically incapable of analyzing an arbitrary linear form. Only until recently, progress has been made towards addressing this problem. In the matrix denoising setting, effective concentration bounds have been established in Koltchinskii and Xia [2016] for estimating linear forms of singular vectors under i.i.d. Gaussian noise, while Bao et al. [2021] established the limiting distributions of the angle between the singular vectors of the noisy matrix and the corresponding ground-truth singular vectors. In Koltchinskii et al. [2016, 2017, 2020], several bias reduction procedures were developed for the problem of PCA and covariance estimation, which established the asymptotic normality and statistical efficiency of the proposed estimator. The eigen-gap conditions required therein, however, are considerably more stringent than the ones required in our theory. Another line of recent works has studied linear form of eigenvectors was Chen et al. [2021a], Cheng et al. [2021], which, however, tackled a different setting of the matrix denoising problem. Specifically, Chen et al. [2021a], Cheng et al. [2021] focused on the case where the noise matrix $\boldsymbol{H}$ is asymmetric and contains independent entries (so that $H_{i,j}$ and $H_{j,i}$ are two independent copies of noise); in this case, a carefully de-biased estimator proposed based on the eigenvector of the asymmetric data matrix $\boldsymbol{M}$ is shown to be minimax-optimal. Additionally, Fan et al. [2020] pinned down the asymptotic distribution for bilinear forms of eigenvectors for large spiked random matrices, while Xia and Yuan [2021] proposed a de-biasing method to estimate linear forms of the matrix for noisy matrix completion. These are beyond the reach of the current paper.

5 Analysis

In this section, we discuss the analysis ideas for establishing Theorem 1 and Theorem 3. One of the main tools lies in the master theorems stated below, which characterize the principal angle between the perturbed eigenvector and an arbitrary subspace of interest. We shall see momentarily the effectiveness of these master theorems when applied to matrix denoising and PCA.

5.1 Master theorems

For any matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times k}$ obeying $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}_k$ $(1 \leq k \leq n)$, let $\boldsymbol{Q}^\perp \in \mathbb{R}^{n\times(n-k)}$ be an arbitrary matrix whose columns form an orthonormal basis of the complement to the subspace spanned by the columns of $\boldsymbol{Q}$, namely

$$\left[\boldsymbol{Q}, \boldsymbol{Q}^\perp\right]^\top \left[\boldsymbol{Q}, \boldsymbol{Q}^\perp\right] = \boldsymbol{I}_n. \tag{5.1}$$

Our results concern the decomposition of an eigenvector $\boldsymbol{u}_l$ of matrix $\boldsymbol{M}$ taking the following form:

$$\boldsymbol{u}_l = \boldsymbol{u}_{l,\|} \cos\theta + \boldsymbol{u}_{l,\perp} \sin\theta. \tag{5.2}$$

Here, θ denotes the principal angle between $\boldsymbol{u}_l$ and the subspace spanned by $\boldsymbol{Q}$, whereas $\boldsymbol{u}_{l,\|}$ and $\boldsymbol{u}_{l,\perp}$ are two unit vectors (i.e. $\|\boldsymbol{u}_{l,\|}\|_2 = \|\boldsymbol{u}_{l,\perp}\|_2 = 1$) such that

- $\boldsymbol{u}_{l,\|}$ lies in the subspace spanned by $\boldsymbol{Q}$; this means that $\boldsymbol{Q}\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|} = \boldsymbol{u}_{l,\|}$, where $\boldsymbol{Q}\boldsymbol{Q}^\top$ is the projection matrix onto the subspace spanned by $\boldsymbol{Q}$;
- $\boldsymbol{u}_{l,\perp}$ is perpendicular to the subspace spanned by $\boldsymbol{Q}$, so that $\boldsymbol{Q}^\perp(\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp} = \boldsymbol{u}_{l,\perp}$.

When $\boldsymbol{Q}$ is a unit vector. We shall begin with the case when $\boldsymbol{Q}$ is a unit vector. For notational simplicity, let us write $\boldsymbol{q}$ for $\boldsymbol{Q}$ in this case to emphasize that this is a vector, and let $\boldsymbol{q}^\perp \in \mathbb{R}^{n\times(n-1)}$ indicate $\boldsymbol{Q}^\perp$. In this case, we can take $\boldsymbol{u}_{l,\|}$ to be equal to $\boldsymbol{q}$. Our result is this:

Theorem 5. *Consider any vector $\boldsymbol{q} \in \mathbb{R}^n$ with $\|\boldsymbol{q}\|_2 = 1$. Write*

$$\boldsymbol{u}_l = \boldsymbol{q}\cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta \tag{5.3}$$

for some θ as well as some vector $\boldsymbol{u}_{l,\perp}$ obeying $\|\boldsymbol{u}_{l,\perp}\|_2 = 1$ and $\boldsymbol{q}^\top \boldsymbol{u}_{l,\perp} = 0$. Suppose that $\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp$ is invertible. Then one has

$$\cos^2\theta = \frac{1}{1 + \left\|\left(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\right)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\right\|_2^2}, \tag{5.4a}$$

$$\lambda_l = \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q} + \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}^\perp\left(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\right)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}. \tag{5.4b}$$

In addition, when $\sin\theta \neq 0$, the vector $\boldsymbol{u}_{l,\perp}$ satisfies

$$\boldsymbol{u}_{l,\perp} = \pm\frac{\boldsymbol{q}^\perp\left(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\right)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}}{\left\|\left(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\right)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\right\|_2}. \tag{5.4c}$$

Proof. See Appendix A.1. □

In words, Theorem 5 derives closed-form expressions for both $\cos\theta$ and $\boldsymbol{u}_{l,\perp}$ (up to global signs), in terms of simple and direct manipulation of the data matrix $\boldsymbol{M}$ as well as the associated eigenvalue λ_l. While the identities (5.4a) and (5.4c) might seem somewhat complicated at first glance, they often allow for convenient decomposition of the noise into independent components, thus streamlining the analysis. Similarly, while the relation (5.4b) takes the form of a nonlinear equation about λ_l, it often enables convenient decoupling of complicated statistical dependency, as we shall demonstrate momentarily.

When $\boldsymbol{Q}$ is a more general orthonormal matrix. The next theorem extends the relation (5.4b) to the case when $\boldsymbol{Q}$ is a general orthonormal matrix (beyond the vector case), which proves useful in eigenvalue analysis for more general low-rank problems.

Theorem 6. *Assume that $k < n$. Consider the corresponding decomposition (5.2) for any matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times k}$ obeying $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}_k$. Suppose that $\lambda_l \boldsymbol{I}_{n-k} - (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}^\perp$ and $\lambda_l \boldsymbol{I}_k - \boldsymbol{Q}^\top \boldsymbol{M}\boldsymbol{Q}$ are both invertible. Then one has*

$$\cos^2\theta = \frac{1}{1 + \left\| \left(\lambda_l \boldsymbol{I}_{n-k} - (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}^\perp\right)^{-1} (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{u}_{l,\parallel} \right\|_2^2}, \tag{5.5a}$$

$$\left(\lambda_l \boldsymbol{I}_k - \boldsymbol{Q}^\top \boldsymbol{M}\boldsymbol{Q}\right)\boldsymbol{Q}^\top \boldsymbol{u}_{l,\parallel} = \boldsymbol{Q}^\top \boldsymbol{M}\boldsymbol{Q}^\perp \left(\lambda_l \boldsymbol{I}_{n-k} - (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}^\perp\right)^{-1} (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{u}_{l,\parallel}. \tag{5.5b}$$

Proof. See Appendix A.2. □

5.2 Analysis for matrix denoising

Armed with the preceding master theorems, we are now positioned to develop consequences for matrix denoising. As a crucial first step of the analysis, we need to establish an eigenvalue perturbation theory that is tightly connected to the eigenvector perturbation bounds. Recalling that λ_l is the l-th largest eigenvalue (in magnitude) of $\boldsymbol{M}$, we present a theorem that reveals the proximity of λ_l and the ground truth $\lambda_l^\star$.

Theorem 7 (Eigenvalue perturbation for matrix denoising)**.** *Consider the model in Section 2.1. Fix any $1 \le l \le r$, and instate the assumptions of Theorem 1. With probability at least $1 - O(n^{-10})$, one has*

$$|\lambda_l - \gamma(\lambda_l) - \lambda_l^\star| \le C_1 \sigma \sqrt{r} \log n \tag{5.6}$$

for some sufficiently large constant $C_1 > 0$, where $\gamma(\cdot)$ is defined as

$$\gamma(\lambda) := \sigma^2 \mathsf{tr}\left[\left(\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\right)^{-1}\right]. \tag{5.7}$$

Remark 5. Here, we recall that the columns of $\boldsymbol{U}^{\star\perp} \in \mathbb{R}^{n\times(n-r)}$ form an orthonormal basis of the complement to the subspace spanned by $\boldsymbol{U}^\star$.

Remark 6. The error bound (5.6) concerning the empirical eigenvalue λ_l contains a systematic non-negligible term $\gamma(\lambda_l)$. This makes clear the presence of a bias effect, which needs to be properly subtracted if one desires a near-optimal estimate of $\lambda_l^\star$. It is also worth noting that the importance of bias correction in eigenvalue estimation has been recognized in prior literature as well (e.g. [Paul, 2007]).

5.2.1 Proof of eigenvalue perturbation theory (Theorem 7)

We start by demonstrating how to prove the eigenvalue perturbation bound in Theorem 7. Let us fix an arbitrary $1 \le l \le r$. The key ingredient of the analysis is to invoke our master theorem (namely, Theorem 6).

Before proceeding, we first verify a few useful facts. It is well known that if $\sigma\sqrt{n} \le c_0 \lambda_{\min}^\star$ for some sufficiently small constant $c_0 > 0$, then with probability exceeding $1 - O(n^{-20})$ one has (see, e.g., Chen et al. [2021b, Theorem 3.1.4])

$$\|\boldsymbol{H}\| \le \lambda_{\min}^\star / 3. \tag{5.8}$$

Recall that

$$(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{U}^{\star\perp} = (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}^\star \boldsymbol{U}^{\star\perp} + (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp} = (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp},$$

which together with (5.8) implies that

$$\left\|(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{U}^{\star\perp}\right\| = \left\|(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\right\| \le \|\boldsymbol{H}\| \le \lambda_{\min}^\star / 3 \tag{5.9}$$

with probability exceeding $1 - O(n^{-20})$. This means that with high probability: (i) the Weyl inequality yields

$$|\lambda_l| \ge |\lambda_l^\star| - \|\boldsymbol{H}\| \ge 2\,|\lambda_l^\star|/3 \qquad \text{and} \qquad |\lambda_l| \le |\lambda_l^\star| + \|\boldsymbol{H}\| \le 4\,|\lambda_l^\star|/3; \tag{5.10}$$

(2) it holds true that $\lambda_{\min}^\star/3 \geq \|\boldsymbol{H}\| \geq \|(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{U}^{\star\perp}\|$, and hence

$$\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{U}^{\star\perp} \quad \text{is invertible}$$

for any $\lambda \in \mathbb{R}$ obeying $|\lambda| \geq 2\lambda_{\min}^\star/3$.

With the above two observations in mind, take $\boldsymbol{Q} = \boldsymbol{U}^\star$ in Theorem 6 to show that

$$\big(\lambda_l \boldsymbol{I}_r - \boldsymbol{U}^{\star\top}\boldsymbol{M}\boldsymbol{U}^\star\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|} = \boldsymbol{G}(\lambda_l)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|} \tag{5.11}$$

with probability exceeding $1 - O(n^{-20})$, where for any given λ with $2\lambda_{\min}^\star/3 \leq |\lambda| \leq 4\lambda_{\max}^\star/3$, we define

$$\boldsymbol{G}(\lambda) := \boldsymbol{U}^{\star\top}\boldsymbol{M}\boldsymbol{U}^{\star\perp}\Big(\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{U}^{\star\perp}\Big)^{-1}(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{U}^\star. \tag{5.12}$$

Note that $\boldsymbol{U}^{\star\top}$ and $\boldsymbol{u}_l^{\star\perp}$ are not uniquely defined. To avoid ambiguity, here and throughout, we let $\boldsymbol{U}^{\star\perp} \in \mathbb{R}^{n\times(n-r)}$ denote an arbitrary matrix whose columns form an orthonormal basis of the complement to the subspace spanned by $\boldsymbol{U}^\star$, and define

$$\boldsymbol{u}_l^{\star\perp} = [\boldsymbol{u}_1^\star, \boldsymbol{u}_2^\star, \dots, \boldsymbol{u}_{l-1}^\star, \boldsymbol{u}_{l+1}^\star, \dots, \boldsymbol{u}_r^\star, \boldsymbol{U}^{\star\perp}] \in \mathbb{R}^{n\times(n-1)} \tag{5.13}$$

for each $1 \leq l \leq r$.

Recognizing that $\boldsymbol{U}^{\star\top}\boldsymbol{M}^\star\boldsymbol{U}^\star = \boldsymbol{\Lambda}^\star$, $\boldsymbol{M}^\star\boldsymbol{U}^{\star\perp} = \boldsymbol{0}$ and $(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{M}^\star = \boldsymbol{0}$, we can rewrite (5.11) as

$$\big(\lambda_l \boldsymbol{I}_r - \boldsymbol{\Lambda}^\star - \boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^\star\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|} = \boldsymbol{G}(\lambda_l)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|} \tag{5.14a}$$

$$\text{with} \quad \boldsymbol{G}(\lambda) = \boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^{\star\perp}\Big(\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\Big)^{-1}(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^\star. \tag{5.14b}$$

Rearranging terms further gives

$$\big(\lambda_l \boldsymbol{I}_r - \boldsymbol{\Lambda}^\star - \boldsymbol{G}^\perp(\lambda_l)\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|} = \boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|} + \big(\boldsymbol{G}(\lambda_l) - \boldsymbol{G}^\perp(\lambda_l)\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|}, \tag{5.15}$$

where we define

$$\boldsymbol{G}^\perp(\lambda) := \mathbb{E}\left[\boldsymbol{G}(\lambda) \mid (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\right], \tag{5.16}$$

with the (conditional) expectation taken assuming that λ is independent of $\boldsymbol{H}$. Here, we single out the component $\boldsymbol{G}^\perp(\lambda)$ since — as will be seen momentarily — it often contains some non-negligible bias term. Combining (5.15) with the triangle inequality and the fact $\|\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|}\|_2 = 1$ then yields

$$\begin{aligned}\big\|\big(\lambda_l \boldsymbol{I}_r - \boldsymbol{\Lambda}^\star - \boldsymbol{G}^\perp(\lambda_l)\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|}\big\|_2 &\leq \big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^\star\big\| + \big\|\boldsymbol{G}(\lambda_l) - \boldsymbol{G}^\perp(\lambda_l)\big\| &(5.17)\\ &\leq \big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^\star\big\| + \sup_{\lambda:\, |\lambda| \in \left[2|\lambda_l^\star|/3,\, 4|\lambda_l^\star|/3\right]} \big\|\boldsymbol{G}(\lambda) - \boldsymbol{G}^\perp(\lambda)\big\| &(5.18)\end{aligned}$$

with probability at least $1 - O(n^{-20})$, where the last line arises from (5.10).

In order to justify that $\big(\lambda_l \boldsymbol{I}_r - \boldsymbol{\Lambda}^\star - \boldsymbol{G}^\perp(\lambda_l)\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|} \approx \boldsymbol{0}$, it remains to show that the two terms on the right-hand side of (5.18) are both fairly small, which we accomplish through the following lemma.

Lemma 1. *Assume that $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is a symmetric matrix with $H_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), i \geq j$ and $\sigma\sqrt{n} \leq c_0 \lambda_{\min}^\star$ for some sufficiently small constant $c_0 > 0$. Then for any $1 \leq l \leq r$, with probability at least $1 - O(n^{-11})$, one has*

$$\begin{aligned}\big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^\star\big\| &\lesssim \sigma\big(\sqrt{r} + \sqrt{\log n}\big),\\ \sup_{\lambda:\, |\lambda| \in \left[2|\lambda_l^\star|/3,\, 4|\lambda_l^\star|/3\right]} \big\|\boldsymbol{G}(\lambda) - \boldsymbol{G}^\perp(\lambda)\big\| &\lesssim \frac{\sigma^2}{\lambda_{\min}^\star}\big(\sqrt{rn\log n} + r\log n\big).\end{aligned} \tag{5.19}$$

In addition, one has

$$\boldsymbol{G}^\perp(\lambda) = \Big\{\sigma^2 \mathsf{tr}\Big[\big(\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)^{-1}\Big]\Big\}\boldsymbol{I}_r. \tag{5.20}$$

Proof. See Appendix B.1. □

With the above lemma in place, by introducing

$$\boldsymbol{M}_{\lambda} := \boldsymbol{\Lambda}^{\star} + \boldsymbol{G}^{\perp}(\lambda) = \boldsymbol{\Lambda}^{\star} + \gamma(\lambda)\boldsymbol{I}_r \tag{5.21a}$$

$$\gamma(\lambda) := \sigma^2 \mathsf{tr}\Big[\big(\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)^{-1}\Big] \tag{5.21b}$$

for any λ with $2\lambda^{\star}_{\min}/3 \leq |\lambda| \leq 4\lambda^{\star}_{\max}/3$, we can invoke the union bound to show that with probability at least $1 - O(n^{-10})$,

$$\begin{aligned}
\big\|(\lambda_l \boldsymbol{I}_r - \boldsymbol{M}_{\lambda_l})\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|}\big\|_2 &\leq \big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^{\star}\big\| + \sup_{\lambda:\,|\lambda|\in\left[2|\lambda^{\star}_l|/3,\,4|\lambda^{\star}_l|/3\right]} \big\|\boldsymbol{G}(\lambda) - \boldsymbol{G}^{\perp}(\lambda)\big\| \\
&\lesssim \sigma\big(\sqrt{r} + \sqrt{\log n}\big) + \frac{\sigma^2}{\lambda^{\star}_{\min}}\big(\sqrt{rn\log n} + r\log n\big) \\
&\leq C_1 \sigma \sqrt{r}\log n =: \mathcal{E}_{\mathsf{MD}}
\end{aligned} \tag{5.22}$$

holds for all $1 \leq l \leq r$, where $C_1 > 0$ is some sufficiently large constant. Intuitively, this means that $(\lambda_l \boldsymbol{I}_r - \boldsymbol{M}_{\lambda_l})\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\|} \approx \boldsymbol{0}$, and hence λ_l is expected to be close to an eigenvalue of $\boldsymbol{M}_{\lambda_l}$ — which is $\lambda^{\star}_i + \gamma(\lambda_l)$ for some $1 \leq i \leq r$.

With the above bound in place, the only possible range of λ_l is characterized by the following lemma, which in turn establishes Theorem 7.

Lemma 2. *Under the condition (5.22) and the eigen-gap assumption (3.3), with probability at least $1 - O(n^{-10})$ one has*

$$\big|\lambda_l - \lambda^{\star}_l - \gamma(\lambda_l)\big| \leq \mathcal{E}_{\mathsf{MD}}, \qquad 1 \leq l \leq r. \tag{5.23}$$

Proof. See Appendix B.2. □

5.2.2 Proof of eigenvector perturbation theory (Theorem 1)

Let us begin by decomposing $\boldsymbol{u}_l$ along the ground-truth direction $\boldsymbol{u}^{\star}_l$ and its complement subspace as follows

$$\boldsymbol{u}_l = \boldsymbol{u}^{\star}_l \cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta, \tag{5.24}$$

where the vector $\boldsymbol{u}_{l,\perp}$ obeys $\|\boldsymbol{u}_{l,\perp}\|_2 = 1$ and $\boldsymbol{u}_{l,\perp}^{\top}\boldsymbol{u}^{\star}_l = 0$. Writing $\boldsymbol{a} = \boldsymbol{P}_{\boldsymbol{U}^{\star}}\boldsymbol{a} + \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}$ with $\boldsymbol{P}_{\boldsymbol{U}^{\star}} = \sum_{1\leq k\leq r}\boldsymbol{u}^{\star}_k\boldsymbol{u}^{\star\top}_k$ and $\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{U}^{\star}}$, we obtain

$$\begin{aligned}
\boldsymbol{a}^{\top}\boldsymbol{u}_l =& (\boldsymbol{P}_{\boldsymbol{U}^{\star}}\boldsymbol{a})^{\top}\boldsymbol{u}_l + (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^{\top}\boldsymbol{u}_l \\
=& (\boldsymbol{P}_{\boldsymbol{U}^{\star}}\boldsymbol{a})^{\top}(\boldsymbol{u}^{\star}_l\cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta) + \langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{u}_l\rangle \\
=& \sum_{k=1}^{r}\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_k\boldsymbol{u}^{\star\top}_k(\boldsymbol{u}^{\star}_l\cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta) + \langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle \\
=& \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l\cos\theta + \sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_k\boldsymbol{u}^{\star\top}_k\boldsymbol{u}_{l,\perp}\sin\theta + \langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle,
\end{aligned}$$

where the third line relies on the fact $\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} = \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}$. It then follows that

$$\boldsymbol{a}^{\top}\boldsymbol{u}_l \pm \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l = \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l(\cos\theta \pm 1) + \sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_k\boldsymbol{u}^{\star\top}_k\boldsymbol{u}_{l,\perp}\sin\theta + \langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle,$$

allowing us to deduce that

$$\begin{aligned}
\min\big|\boldsymbol{a}^{\top}\boldsymbol{u}_l \pm \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l\big| \leq& \big|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l\big|(1 - |\cos\theta|) + \Big|\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_k\boldsymbol{u}^{\star\top}_k\boldsymbol{u}_{l,\perp}\sin\theta\Big| + \big|\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\big| \\
\leq& \big|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l\big|(1 - \cos^2\theta) + \Big|\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_k\boldsymbol{u}^{\star\top}_k\boldsymbol{u}_{l,\perp}\sin\theta\Big| + \big|\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\big|.
\end{aligned} \tag{5.25}$$

and

$$\min\left|\boldsymbol{a}^\top\boldsymbol{u}_l\sqrt{1+b_l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\le\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\cdot\left|1-\sqrt{1+b_l}|\cos\theta|\right|+\sqrt{1+b_l}\left|\langle\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\,\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\right| + \sqrt{1+b_l}\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_{l,\perp}\sin\theta\Big|. \tag{5.26}$$

As a result, it boils down to bounding the terms

$$1-\cos^2\theta,\quad 1-\sqrt{1+b_l}|\cos\theta|,\quad \sqrt{1+b_l},\quad \sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_{l,\perp}\sin\theta,\quad\text{and}\quad\langle\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\,\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle.$$

We claim that $\lambda_l\boldsymbol{I}_{n-1}-(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}$ is invertible. This can be seen from (5.33) stated in Lemma 3 directly, whose validation is independent with this claim. The invertibility taken together with Theorem 5 reveals that $\cos\theta\neq 0$. If $\sin\theta=0$, then we have $\boldsymbol{u}_l=\pm\boldsymbol{u}_l^\star$ and the conclusion is obvious since $\min\left|\boldsymbol{a}^\top\boldsymbol{u}_l\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|=0$. Therefore, we shall assume $\cos\theta\neq 0$ and $\sin\theta\neq 0$ in the remainder of the proof. Invoking Theorem 5 yields

$$\cos^2\theta=\frac{1}{1+\left\|\left(\lambda_l\boldsymbol{I}_{n-1}-(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^\star\right\|_2^2}, \tag{5.27a}$$

$$\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_{l,\perp}=\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}_{n-1}-(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^\star}{\left\|\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}_{n-1}-(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^\star\right\|_2}. \tag{5.27b}$$

Recognizing that $\boldsymbol{M}^\star\boldsymbol{u}_l^\star=\lambda_l^\star\boldsymbol{u}_l^\star$, we can alternatively write (5.27) as follows

$$\cos^2\theta=\frac{1}{1+\left\|\left(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2}, \tag{5.28a}$$

$$\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_{l,\perp}=\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\left\|\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2}. \tag{5.28b}$$

Here, we define

$$\boldsymbol{M}^{(l)}\coloneqq(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}=(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}^\star\boldsymbol{u}_l^{\star\perp}+(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^{\star\perp}. \tag{5.29}$$

With the above relations in mind, we can demonstrate that

$$\begin{aligned}
&\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_{l,\perp}\sin\theta\Big|=\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_{l,\perp}\Big|\sqrt{1-\cos^2\theta}\\
&=\frac{\left|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right|}{\left\|\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2}\cdot\sqrt{\frac{\left\|\left(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2}{1+\left\|\left(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2}}\\
&\le\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\Big|,
\end{aligned}$$

where the last inequality comes from the fact $\left\|\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2=\left\|\left(\lambda_l\boldsymbol{I}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2$ (since the columns of $\boldsymbol{u}_l^{\star\perp}$ are orthonormal). Substituting this into (5.25) and (5.26) yields

$$\min\left|\boldsymbol{a}^\top\boldsymbol{u}_l\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\le\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\cdot(1-\cos^2\theta)+\left|\langle\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\,\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\right| + \Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\cdot\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\Big|; \tag{5.30}$$

and

$$\min\left|\sqrt{1+b_l}\boldsymbol{a}^\top\boldsymbol{u}_l\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\le\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\cdot\left|1-\sqrt{1+b_l}|\cos\theta|\right|+\sqrt{1+b_l}\left|\langle\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\,\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\right| + \sqrt{1+b_l}\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\cdot\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\Big|. \tag{5.31}$$

In what follows, we shall control these quantities separately.

1. Controlling the spectrum of $\boldsymbol{M}^{(l)}$. Before proceeding, we find it helpful to first study the spectrum of $\boldsymbol{M}^{(l)}$. Let $\{\lambda_i^{(l)}\}_{i=1}^{n-1}$ denote the eigenvalues of $\boldsymbol{M}^{(l)}$ with $|\lambda_1^{(l)}| \geq |\lambda_2^{(l)}| \geq \cdots \geq |\lambda_{n-1}^{(l)}|$ with associate eigenvectors $\{\boldsymbol{u}_i^{(l)}\}_{i=1}^{n-1}$. In addition, we define several matrices as follows

$$\boldsymbol{U}^\star_{\setminus l} \coloneqq [\boldsymbol{u}_1^\star, \cdots, \boldsymbol{u}_{l-1}^\star, \boldsymbol{u}_{l+1}^\star, \cdots, \boldsymbol{u}_r^\star] \in \mathbb{R}^{n\times(r-1)}, \tag{5.32a}$$

$$\boldsymbol{U}^{\star(l)} \coloneqq (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{U}^\star_{\setminus l} = \begin{bmatrix} \boldsymbol{I}_{r-1} \\ \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^{(n-1)\times(r-1)}, \tag{5.32b}$$

$$\boldsymbol{U}^{\star(l)\perp} \coloneqq (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{U}^{\star\perp} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_{n-r} \end{bmatrix} \in \mathbb{R}^{(n-1)\times(n-r)}, \tag{5.32c}$$

$$\boldsymbol{\Lambda}^{\star(l)} \coloneqq \mathsf{diag}\big(\{\lambda_i^\star\}_{i\neq l}\big) \in \mathbb{R}^{(r-1)\times(r-1)}, \tag{5.32d}$$

and define

$$\boldsymbol{u}_{k,\parallel}^{(l)} \coloneqq \frac{1}{\|\boldsymbol{P}_{\boldsymbol{U}^{\star(l)}}\boldsymbol{u}_k^{(l)}\|_2}\boldsymbol{P}_{\boldsymbol{U}^{\star(l)}}\boldsymbol{u}_k^{(l)}$$

for each $k \neq l$.

Armed with this set of notation, we are ready to present Lemma 3, which studies the eigenvalues of $\boldsymbol{M}^{(l)}$.

Lemma 3. *Instate the assumptions of Theorem 1, and recall the definition of $\mathcal{E}_{\mathsf{MD}}$ in Lemma 2. With probability at least $1 - O(n^{-10})$, the following holds:*

1. *For each $1 \leq k < r$, one has $\lambda_k^{(l)} - \gamma(\lambda_k^{(l)}) \in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_i^\star)$ for some $i \neq l$, and*
$$\big\|\big(\lambda_k^{(l)}\boldsymbol{I}_{r-1} - \gamma(\lambda_k^{(l)})\boldsymbol{I}_{r-1} - \boldsymbol{\Lambda}^{\star(l)}\big)\boldsymbol{U}^{\star(l)\top}\boldsymbol{u}_{k,\parallel}^{(l)}\big\|_2 \lesssim \sigma\sqrt{r}\log n;$$
2. *For each $k \geq r$, one has $|\lambda_k^{(l)}| \lesssim \sigma\sqrt{n}$;*
3. *Moreover, one has*
$$\big|\lambda - \lambda_l\big| \gtrsim \begin{cases} \Delta_l^\star, & \text{if } \lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_k^\star) \text{ for some } k \neq l \text{ and } 1 \leq k \leq r; \\ |\lambda_l^\star|, & \text{if } |\lambda| \lesssim \sigma\sqrt{n}. \end{cases}$$
In particular, we have
$$\big|\lambda_k^{(l)} - \lambda_l\big| \gtrsim \begin{cases} \Delta_l^\star, & 1 \leq k < r; \\ |\lambda_l^\star|, & k \geq r. \end{cases} \tag{5.33}$$

Proof. See Appendix C.1. ◻

In words, this lemma tells us that:

- For any $1 \leq k \leq r$, the properly corrected $\lambda_k^{(l)}$ (namely, $\lambda_k^{(l)} - \gamma(\lambda_k^{(l)})$) stays very close to one of the true non-zero eigenvalues excluding $\lambda_l^\star$;
- For any $k \geq r$, the eigenvalue $\lambda_k^{(l)}$ is reasonably small;
- Any eigenvalue of $\boldsymbol{M}^{(l)}$ is sufficiently separated from the l-th eigenvalue λ_l of $\boldsymbol{M}$, where the separation is lower bounded by the order of the associated eigen-gap.

2. Controlling $\cos^2\theta$. We now turn to bounding $\cos^2\theta$. In view of the expression of $\cos^2\theta$ in (5.28a), it suffices to look at $\big\|\big(\lambda_l\boldsymbol{I} - \boldsymbol{M}^{(l)}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big\|_2$. A simple yet crucial observation is that: the matrix $(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^{\star\perp}$ is independent of $(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star$ (which follows from the same argument as in the proof of Lemma 1 in Appendix B.1). Consequently, $\boldsymbol{M}^{(l)}$ (defined in (5.29)) is independent of $(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n-1})$, which is a Gaussian random vector in $\mathbb{R}^{n-1}$. In light of this observation, we can bound $\big\|\big(\lambda_l\boldsymbol{I} - \boldsymbol{M}^{(l)}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big\|_2$ as follows.

Lemma 4. *Instate the assumptions of Theorem 1. The following holds with probability at least $1-O(n^{-10})$:*

$$\left\|\left(\lambda_l \boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2=\sum_{k:\, r<k\le n}\frac{\sigma^2}{(\lambda_l-\lambda_k)^2}+O\left(\frac{\sigma^2 r\log n}{\left(\Delta_l^\star\right)^2}+\frac{\sigma^2\sqrt{n\log n}}{\lambda_l^{\star 2}}\right), \tag{5.34}$$

which further indicates that

$$\left\|\left(\lambda_l \boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2\asymp\frac{\sigma^2 n}{\lambda_l^{\star 2}}+O\left(\frac{\sigma^2 r\log n}{\left(\Delta_l^\star\right)^2}\right)\ll 1. \tag{5.35}$$

Proof. See Appendix C.2. □

Combining this lemma with (5.28a), we reach

$$1-\cos^2\theta=1-\frac{1}{1+\left\|\left(\lambda_l \boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2}\asymp\frac{\sigma^2 n}{\lambda_l^{\star 2}}+O\left(\frac{\sigma^2 r\log n}{\left(\Delta_l^\star\right)^2}\right)\ll 1, \tag{5.36}$$

where the last step arises from the assumption (3.3). In addition, recalling the de-bias parameter b_l

$$b_l=\sum_{k:\, r<k\le n}\frac{\sigma^2}{(\lambda_l-\lambda_k)^2},$$

one arrives at

$$\begin{aligned}
\left|(1+b_l)\cos^2\theta-1\right| &= \left|\frac{1+b_l}{1+\left\|\left(\lambda_l \boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2}-1\right| \\
&= \frac{\left|b_l-\left\|\left(\lambda_l \boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2\right|}{1+\left\|\left(\lambda_l \boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2} \\
&\le \left|b_l-\left\|\left(\lambda_l \boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\right\|_2^2\right| \\
&\overset{\text{(i)}}{\lesssim} \frac{\sigma^2 r\log n}{\left(\Delta_l^\star\right)^2}+\frac{\sigma^2\sqrt{n\log n}}{\lambda_l^{\star 2}}\overset{\text{(ii)}}{\ll} 1,
\end{aligned} \tag{5.37}$$

where (i) follows from (5.34) and (ii) is due to the assumption (3.3). Combined with (5.36), this further allows us to obtain $1+b_l\lesssim 1$ and

$$\left|1-\sqrt{1+b_l}\left|\cos\theta\right|\right|=\left|\frac{1-(1+b_l)\cos^2\theta}{1+\sqrt{1+b_l}\left|\cos\theta\right|}\right|\lesssim\left|1-(1+b_l)\cos^2\theta\right|\lesssim\frac{\sigma^2 r\log n}{\left(\Delta_l^\star\right)^2}+\frac{\sigma^2\sqrt{n\log n}}{\lambda_l^{\star 2}}. \tag{5.38}$$

3. Controlling $\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\cdot\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star$. The key observation is that $(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_{n-1})$ is independent of $\boldsymbol{M}^{(l)}$ (but dependent of λ_l). This term can be bounded via the following lemma, which will be established in Appendix C.3.

Lemma 5. *Instate the assumptions of Theorem 1. With probability at least $1-O(n^{-10})$, one has*

$$\begin{aligned}
&\left|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\cdot\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\left(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\right)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right| \\
&\qquad\lesssim\frac{\sigma}{|\lambda_l^\star|}\sqrt{\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}+\sigma\sqrt{r\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}\sum_{k:k\neq l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star-\lambda_k^\star|}.
\end{aligned} \tag{5.39}$$

4. Controlling $\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle$. When it comes to the last term $\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle$, one can take advantage of the rotational invariance of $\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l$ in the subspace spanned by $\boldsymbol{U}^{\star\perp}$ to upper bound it. This is formalized in Lemma 6, with the proof postponed to Appendix C.4.

Lemma 6. *Instate the assumptions of Theorem 1. With probability at least $1 - O(n^{-10})$,*

$$\left|\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\right| \lesssim \sqrt{\frac{\log n}{n}}\,\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\big\|_2\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2. \tag{5.40}$$

Consequently, it remains to upper bound $\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2$. Recall $\boldsymbol{u}_{l,\|} := \boldsymbol{P}_{\boldsymbol{U}^\star}(\boldsymbol{u})/\|\boldsymbol{P}_{\boldsymbol{U}^\star}(\boldsymbol{u})\|_2$ defined in in Section 5.1. By virtue of Theorem 6, one has

$$\begin{aligned}
\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2^2 &= 1 - \frac{1}{1 + \big\|\big(\lambda_l \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{U}^{\star\perp}\big)^{-1}(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{u}_{l,\|}\big\|_2^2} \\
&\le \big\|\big(\lambda_l \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{U}^{\star\perp}\big)^{-1}(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}\boldsymbol{u}_{l,\|}\big\|_2^2 \\
&\le \big\|\big(\lambda_l \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)^{-1}\big\|^2 \|(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_{l,\|}\|_2^2,
\end{aligned} \tag{5.41}$$

where the last inequality makes use of the fact that

$$(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M} = (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{M}^\star + (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H} = (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}.$$

Additionally, it is easily seen that

$$\begin{gathered}
\|\lambda_l \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\| \ge |\lambda_l| - \|(\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\| \gtrsim |\lambda_l^\star| \\
\|\boldsymbol{U}^{\star\perp\top}\boldsymbol{H}\boldsymbol{u}_{l,\|}\|_2 \le \|\boldsymbol{H}\| \lesssim \sigma\sqrt{n}
\end{gathered}$$

with high probability. These combined with (5.41) lead to

$$\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2 \lesssim \frac{\sigma\sqrt{n}}{|\lambda_l^\star|}. \tag{5.42}$$

Substitution into (5.40) reveals that

$$\left|\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\right| \lesssim \sqrt{\frac{\log n}{n}}\,\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\big\|_2\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2 \lesssim \frac{\sigma\sqrt{\log n}}{|\lambda_l^\star|}\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\big\|_2. \tag{5.43}$$

5. Combining bounds. In view of (5.30), the bounds (5.36), (5.39) and (5.43) taken collectively lead to our advertised result

$$\begin{aligned}
\min\big|\boldsymbol{a}^\top \boldsymbol{u}_l \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\big| &\lesssim \left(\frac{\sigma^2 n}{\lambda_l^{\star 2}} + \frac{\sigma^2 r\log n}{\left(\Delta_l^\star\right)^2}\right)\big|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\big| + \frac{\sigma\sqrt{\log n}}{|\lambda_l^\star|}\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\big\|_2 \\
&\quad + \sigma\sqrt{r\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}\sum_{k:k\neq l}\frac{\big|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\big|}{|\lambda_l^\star - \lambda_k^\star|} + \frac{\sigma}{|\lambda_l^\star|}\sqrt{\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}.
\end{aligned}$$

Regarding the analysis for the de-biased estimate, one can substitute (5.38), (5.39) and (5.43) into (5.31) to obtain

$$\begin{aligned}
\min\big|\boldsymbol{a}^\top \boldsymbol{u}_l\sqrt{1+b_l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\big| &\lesssim \left(\frac{\sigma^2 \sqrt{n\log n}}{\lambda_l^{\star 2}} + \frac{\sigma^2 r\log n}{\left(\Delta_l^\star\right)^2}\right)\big|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\big| + \frac{\sigma\sqrt{\log n}}{|\lambda_l^\star|}\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\big\|_2 \\
&\quad + \sigma\sqrt{r\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}\sum_{k:k\neq l}\frac{\big|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\big|}{|\lambda_l^\star - \lambda_k^\star|} + \frac{\sigma}{|\lambda_l^\star|}\sqrt{\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)} \\
&\lesssim \frac{\sigma^2 r\log n}{\left(\Delta_l^\star\right)^2}\big|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\big| + \sigma\sqrt{r\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}\sum_{k:k\neq l}\frac{\big|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\big|}{|\lambda_l^\star - \lambda_k^\star|} + \frac{\sigma}{|\lambda_l^\star|}\sqrt{\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}
\end{aligned}$$

where the last step holds since $\sigma\sqrt{n} \lesssim \lambda_{\min}^\star$ and $\big|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\big| \le \|\boldsymbol{a}\|_2\|\boldsymbol{u}_l^\star\|_2 = 1$. This concludes the proof.

5.3 Analysis for principal component analysis

Akin to the matrix denoising counterpart, the first step towards establishing the desired eigenvector perturbation bounds lies in the development of a fine-grained eigenvalue perturbation theory. Here and throughout, we let $\boldsymbol{U}^{\star\perp} \in \mathbb{R}^{p\times(p-r)}$ represent a matrix consisting of orthonormal columns perpendicular to the subspace spanned by $\boldsymbol{U}^{\star}$.

Theorem 8 (Eigenvalue perturbation for PCA)**.** *Consider the model in Section 2.2. Fix any $1 \le l \le r$, and instate the assumptions of Theorem 3. Then with probability at least $1 - O(n^{-10})$, one has*

$$\left|\frac{\lambda_l}{1+\beta(\lambda_l)} - \lambda_l^\star - \sigma^2\right| \le C_2(\lambda_{\max}^\star + \sigma^2)\sqrt{\frac{r}{n}}\log n \tag{5.44}$$

for some sufficiently large constant $C_2 > 0$, where we define

$$\beta(\lambda) := \frac{1}{n}\mathsf{tr}\Big[\frac{1}{n}\boldsymbol{S}_\perp^\top\big(\lambda \boldsymbol{I}_{p-r} - \frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\big)^{-1}\boldsymbol{S}_\perp\Big] \qquad \text{with } \boldsymbol{S}_\perp := (\boldsymbol{U}^{\star\perp})^\top\boldsymbol{S}. \tag{5.45}$$

Remark 7. As asserted by Theorem 8, the empirical eigenvalue λ_l exhibits a form of "inflation" in comparison to the corresponding ground-truth value $\lambda_l^\star + \sigma^2$. As a result, it is advisable to properly shrink λ_l when estimating $\lambda_l^\star + \sigma^2$.

In what follows, we shall first outline the proof for Theorem 8 (which is very similar to the analysis for Theorem 7), followed by a proof sketch for the eigenvector perturbation theory in Theorem 3.

5.3.1 Proof of eigenvalue perturbation theory (Theorem 8)

Before embarking on the proof, we shall define

$$\boldsymbol{S}_\parallel := \boldsymbol{U}^{\star\top}\boldsymbol{S} \in \mathbb{R}^{r\times n}, \quad \boldsymbol{S}_\perp := (\boldsymbol{U}^{\star\perp})^\top\boldsymbol{S} \in \mathbb{R}^{(p-r)\times n} \quad \text{and} \quad \boldsymbol{\Lambda} := \boldsymbol{U}^{\star\top}\boldsymbol{\Sigma}\boldsymbol{U}^\star = \boldsymbol{\Lambda}^\star + \sigma^2\boldsymbol{I}_r \tag{5.46}$$

for notional convenience, allowing one to express

$$(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{U}^{\star\perp} = \boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top, \tag{5.47a}$$
$$(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{U}^{\star} = \boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top, \tag{5.47b}$$
$$\boldsymbol{U}^{\star\top}\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{U}^{\star} = \boldsymbol{S}_\parallel\boldsymbol{S}_\parallel^\top. \tag{5.47c}$$

As can be straightforwardly verified:

- The columns of $\boldsymbol{S}_\parallel$ are independent zero-mean Gaussian random vectors with covariance matrix $\boldsymbol{\Lambda}$;
- The columns of $\boldsymbol{S}_\perp$ are i.i.d. zero-mean Gaussian random vectors with covariance matrix $\sigma^2\boldsymbol{I}_{p-r}$;
- $\boldsymbol{S}_\parallel$ is statistically independent of $\boldsymbol{S}_\perp$ (from standard properties for Gaussian random vectors).

In addition, the following lemma controls the distance between $\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top$ and $\boldsymbol{\Sigma}$ when measured by the spectral norm.

Lemma 7. *Assume that $n \ge r$. Then with probability at least $1 - O(n^{-10})$, one has*

$$\Big\|\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{\Sigma}\Big\| \lesssim \lambda_{\max}^\star\sqrt{\frac{r\log n}{n}} + \sqrt{(\lambda_{\max}^\star + \sigma^2)\sigma^2\frac{p}{n}}\log n + \sigma^2\bigg(\sqrt{\frac{p}{n}} + \frac{p}{n} + \sqrt{\frac{\log n}{n}}\bigg). \tag{5.48}$$

Proof. See Appendix D.1. ☐

Remark 8. In particular, under the noise assumption (3.13a), Lemma 7 tells us that $\big\|\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{\Sigma}\big\| \ll \lambda_{\min}^\star$ with probability at least $1 - O(n^{-10})$, which together with Weyl's inequality gives

$$2\lambda_l^\star/3 \le \lambda_l \le 4\lambda_l^\star/3, \qquad 1 \le l \le r. \tag{5.49}$$

We now move on to present the proof of Theorem 8. The key ingredient underlying the analysis is, once again, to invoke our master theorem (namely, Theorem 6), by treating $\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^{\top}$, $\boldsymbol{\Sigma}^{\star}$ and $\boldsymbol{U}^{\star}$ as $\boldsymbol{M}$, $\boldsymbol{M}^{\star}$ and $\boldsymbol{Q}$, respectively. Recalling the definition of

$$\boldsymbol{u}_{l,\parallel} \coloneqq \frac{1}{\|\boldsymbol{P}_{\boldsymbol{U}^{\star}}(\boldsymbol{u})\|_2}\boldsymbol{P}_{\boldsymbol{U}^{\star}}(\boldsymbol{u}) \tag{5.50}$$

as in Section 5.1 (so that $\boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel} = \boldsymbol{u}_{l,\parallel}$), one can invoke (5.5b) in Theorem 6 to derive

$$\Big(\lambda_l \boldsymbol{I}_r - \frac{1}{n}\boldsymbol{S}_{\parallel}\boldsymbol{S}_{\parallel}^{\top}\Big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel} = \boldsymbol{K}(\lambda_l)\,\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel}, \tag{5.51}$$

where we recall the definitions of $\boldsymbol{S}_{\parallel}$ and $\boldsymbol{S}_{\perp}$ in (5.46), and $\boldsymbol{K}(\lambda)$ is given by

$$\boldsymbol{K}(\lambda) \coloneqq \frac{1}{n}\boldsymbol{S}_{\parallel}\cdot\underbrace{\frac{1}{n}\boldsymbol{S}_{\perp}^{\top}\Big(\lambda\boldsymbol{I}_{p-r} - \frac{1}{n}\boldsymbol{S}_{\perp}\boldsymbol{S}_{\perp}^{\top}\Big)^{-1}\boldsymbol{S}_{\perp}}_{=:\,\boldsymbol{C}(\lambda)}\cdot\boldsymbol{S}_{\parallel}^{\top}. \tag{5.52}$$

It is also helpful to define

$$\boldsymbol{K}^{\perp}(\lambda) \coloneqq \mathbb{E}\big[\boldsymbol{K}(\lambda)\mid\boldsymbol{C}(\lambda)\big], \tag{5.53}$$

with λ regarded as a deterministic quantity independent of the data samples. Then rearranging terms in (5.51) yields

$$\big(\lambda_l\boldsymbol{I}_r - \boldsymbol{\Lambda} - \boldsymbol{K}^{\perp}(\lambda_l)\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel} = \Big(\frac{1}{n}\boldsymbol{S}_{\parallel}\boldsymbol{S}_{\parallel}^{\top} - \boldsymbol{\Lambda} + \boldsymbol{K}(\lambda_l) - \boldsymbol{K}^{\perp}(\lambda_l)\Big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel},$$

which together with (5.49) results in the following bound:

$$\big\|\big(\lambda_l\boldsymbol{I}_r - \boldsymbol{\Lambda} - \boldsymbol{K}^{\perp}(\lambda_l)\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel}\big\|_2 \le \Big\|\frac{1}{n}\boldsymbol{S}_{\parallel}\boldsymbol{S}_{\parallel}^{\top} - \boldsymbol{\Lambda}\Big\| + \sup_{\lambda:\,\lambda\in[2\lambda_l^{\star}/3,4\lambda_l^{\star}/3]}\big\|\boldsymbol{K}(\lambda) - \boldsymbol{K}^{\perp}(\lambda)\big\|, \tag{5.54}$$

Akin to the proof of Theorem 7 in Section, our goal is to show $\big(\lambda_l\boldsymbol{I}_r - \boldsymbol{\Lambda} - \boldsymbol{K}^{\perp}(\lambda_l)\big)\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel} \approx \boldsymbol{0}$, which would then imply that λ_l is sufficiently close to some eigenvalue of $\boldsymbol{\Lambda} + \boldsymbol{K}^{\perp}(\lambda_l)$. In light of this, we intend to upper bound the two terms on the right-hand side of (5.54) in the sequel.

- Let us first look at the first term on the right-hand side of (5.54). Since the columns of $\boldsymbol{S}_{\parallel} = \boldsymbol{U}^{\star\top}\boldsymbol{S}$ are independent Gaussian random vectors with distribution $\mathcal{N}(\boldsymbol{0},\boldsymbol{\Lambda})$, we can rewrite

$$\boldsymbol{S}_{\parallel} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}, \tag{5.55}$$

  where $\boldsymbol{Z} = [Z_{i,j}] \in \mathbb{R}^{r\times n}$ is a Gaussian random matrix with i.i.d. entries $Z_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Applying standard Gaussian concentration inequalities reveals that: with probability at least $1 - O(n^{-10})$,

$$\Big\|\frac{1}{n}\boldsymbol{S}_{\parallel}\boldsymbol{S}_{\parallel}^{\top} - \boldsymbol{\Lambda}\Big\| \le \|\boldsymbol{\Lambda}\|\cdot\Big\|\frac{1}{n}\boldsymbol{Z}\boldsymbol{Z}^{\top} - \boldsymbol{I}_r\Big\| \lesssim (\lambda_{\max}^{\star} + \sigma^2)\sqrt{\frac{r\log n}{n}}. \tag{5.56}$$

- As for the second term on the right-hand side of (5.54), we claim for the moment that

$$\sup_{\lambda:\,\lambda\in[2\lambda_l^{\star}/3,4\lambda_l^{\star}/3]}\big\|\boldsymbol{K}(\lambda) - \boldsymbol{K}^{\perp}(\lambda)\big\| \ll (\lambda_{\max}^{\star} + \sigma^2)\sqrt{\frac{r}{n}}\log n, \tag{5.57}$$

$$\sup_{\lambda:\,\lambda\in[2\lambda_l^{\star}/3,4\lambda_l^{\star}/3]}\|\boldsymbol{C}(\lambda)\| \lesssim \frac{\sigma^2}{\lambda_l^{\star}}\Big(1 + \frac{p}{n}\Big), \tag{5.58}$$

  where $\boldsymbol{C}(\lambda)$ is defined in (5.52). The proof of this claim is postponed to the end of the section.

Substituting (5.56) and (5.57) into (5.54) reveals that with probability exceeding $1 - O(n^{-10})$,

$$\big\|\big(\lambda_l \boldsymbol{I}_r - \boldsymbol{\Lambda} - \boldsymbol{K}^{\perp}(\lambda_l)\big)\boldsymbol{u}_{l,\parallel}\big\|_2 \lesssim (\lambda^{\star}_{\max} + \sigma^2)\sqrt{\frac{r}{n}}\log n =: \mathcal{E}_{\mathsf{PCA}}. \tag{5.59}$$

With the preceding inequality in place, we are ready to study the eigenvalues of $\boldsymbol{\Lambda} + \boldsymbol{K}^{\perp}(\lambda_l)$. Similar to the analysis in the proof of Lemma 1 in Appendix B.1, it is straightforward to verify that

$$\boldsymbol{K}^{\perp}(\lambda) = \beta(\lambda)\boldsymbol{\Lambda}, \tag{5.60}$$

where $\beta(\lambda) = \frac{1}{n}\mathsf{tr}\big(\boldsymbol{C}(\lambda)\big)$ has been defined in (5.45). This immediately demonstrates that the l-th eigenvalue of $\boldsymbol{\Lambda} + \boldsymbol{K}^{\perp}(\lambda_l)$ is equal to

$$\big(1 + \beta(\lambda_l)\big)(\lambda^{\star}_l + \sigma^2).$$

Moreover, it is readily seen from (5.58) that $\beta(\lambda)$ satisfies

$$\sup_{\lambda:\,\lambda\in[2\lambda^{\star}_l/3,4\lambda^{\star}_l/3]} \beta(\lambda) \leq \sup_{\lambda:\,\lambda\in[2\lambda^{\star}_l/3,4\lambda^{\star}_l/3]} \frac{n\wedge p}{n}\|\boldsymbol{C}(\lambda)\| \lesssim \frac{n\wedge p}{n}\cdot\frac{\sigma^2}{\lambda^{\star}_l}\left(1+\frac{p}{n}\right) \asymp \frac{\sigma^2 p}{\lambda^{\star}_l n} = o(1) \tag{5.61}$$

as long as the noise level obeys $\sigma^2 p/n \ll \lambda^{\star}_{\min}/\log n$. Finally, combining (5.60) with (5.46) and (5.59), we can repeat the same argument as in the proof for Lemma 2 in Section B.2 to reach

$$\big|\lambda_l - (\lambda^{\star}_l + \sigma^2)\big(1 + \beta(\lambda_l)\big)\big| \lesssim (\lambda^{\star}_{\max} + \sigma^2)\sqrt{\frac{r}{n}}\log n;$$

for conciseness, we omit the details of proof. This inequality establishes the proximity of λ_l and $(\lambda^{\star}_l + \sigma^2)\big(1 + \beta(\lambda_l)\big)$. Taking this collectively with (5.61) (i.e., $1 + \beta(\lambda_l) \asymp 1$), we establish the advertised bound (5.44).

Proof of the inequality (5.57) Recall the definitions of $\boldsymbol{K}(\lambda)$, $\boldsymbol{C}(\lambda)$ as well as $\boldsymbol{K}^{\perp}(\lambda)$ in (5.52) and (5.53). Recognizing that one can express $\boldsymbol{S}_{\parallel} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}$ with $\boldsymbol{Z} \in \mathbb{R}^{r\times n}$ being an i.i.d. standard Gaussian matrix (see (5.55)), we can define

$$\overline{\boldsymbol{K}}(\lambda) := \frac{1}{n}\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^{\top} \qquad \text{and} \qquad \overline{\boldsymbol{K}}^{\perp}(\lambda) := \mathbb{E}\big[\,\overline{\boldsymbol{K}}(\lambda) \mid \boldsymbol{C}(\lambda)\big],$$

which allow us to express

$$\begin{aligned}
\boldsymbol{K}(\lambda) &:= \frac{1}{n}\boldsymbol{S}_{\parallel}\boldsymbol{C}(\lambda)\boldsymbol{S}_{\parallel}^{\top} = \frac{1}{n}\boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^{\top}\boldsymbol{\Lambda}^{1/2} = \boldsymbol{\Lambda}^{1/2}\overline{\boldsymbol{K}}(\lambda)\boldsymbol{\Lambda}^{1/2},\\
\boldsymbol{K}^{\perp}(\lambda) &:= \mathbb{E}\big[\boldsymbol{K}(\lambda) \mid \boldsymbol{C}(\lambda)\big] = \boldsymbol{\Lambda}^{1/2}\mathbb{E}\big[\,\overline{\boldsymbol{K}}(\lambda) \mid \boldsymbol{C}(\lambda)\big]\boldsymbol{\Lambda}^{1/2} = \boldsymbol{\Lambda}^{1/2}\overline{\boldsymbol{K}}^{\perp}(\lambda)\boldsymbol{\Lambda}^{1/2}.
\end{aligned}$$

One can then develop the following upper bound

$$\begin{aligned}
\big\|\boldsymbol{K}(\lambda) - \boldsymbol{K}^{\perp}(\lambda)\big\| &= \big\|\boldsymbol{\Lambda}^{1/2}\big(\overline{\boldsymbol{K}}(\lambda) - \overline{\boldsymbol{K}}^{\perp}(\lambda)\big)\boldsymbol{\Lambda}^{1/2}\big\| \leq \|\boldsymbol{\Lambda}\|\big\|\overline{\boldsymbol{K}}(\lambda) - \overline{\boldsymbol{K}}^{\perp}(\lambda)\big\|\\
&= (\lambda^{\star}_{\max} + \sigma^2)\frac{1}{n}\big\|\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^{\top} - \mathbb{E}[\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^{\top} \mid \boldsymbol{C}(\lambda)]\big\|.
\end{aligned} \tag{5.62}$$

By construction, $\boldsymbol{S}_{\parallel} := \boldsymbol{U}^{\star\top}\boldsymbol{S}$ and $\boldsymbol{S}_{\perp} := (\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{S}$ are mutually statistically independent, thus implying that $\boldsymbol{Z}$ is also independent of $\boldsymbol{C}(\lambda)$ with λ treated as a deterministic quantity.

The remainder of the proof thus comes down to controlling

$$\big\|\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^{\top} - \mathbb{E}[\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^{\top} \mid \boldsymbol{C}(\lambda)]\big\|.$$

By virtue of the rotational invariance of Gaussian random matrices, we can replace $\boldsymbol{C}(\lambda)$ in the quantity above by a diagonal matrix comprised of the eigenvalues of $\boldsymbol{C}(\lambda)$. To see this, we denote by $\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\top}$ the eigen-decomposition of $\boldsymbol{C}(\lambda)$ and find that

$$\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^{\top} = \boldsymbol{Z}\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\top}\boldsymbol{Z}^{\top} \overset{\mathrm{d}}{=} \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^{\top},$$

where the last step arises from the rotational invariance of the Gaussian random matrix, namely $\boldsymbol{Z}\boldsymbol{V} \overset{\mathrm{d}}{=} \boldsymbol{Z}$. In view of Lemma 18, it suffices to control the eigenvalues of $\boldsymbol{C}(\lambda)$.

As can be straightforwardly verified, the rank of $\boldsymbol{C}(\lambda)$ is upper bounded by $(p-r)\wedge n$ and the i-th largest eigenvalue of $\boldsymbol{C}(\lambda)$ (cf. (5.52)) satisfies

$$\lambda_i\big(\boldsymbol{C}(\lambda)\big) = \lambda_i\Big(\frac{1}{n}\boldsymbol{S}_\perp^\top\Big(\lambda \boldsymbol{I}_{p-r} - \frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\Big)^{-1}\boldsymbol{S}_\perp\Big) = \frac{\lambda_i(\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top)}{\lambda - \lambda_i(\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top)}, \qquad 1 \le i \le (p-r)\wedge n.$$

In addition, (D.4) demonstrates that with probability at least $1 - O(n^{-10})$,

$$0 \le \lambda_i\Big(\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\Big) \lesssim \sigma^2\bigg(1 + \sqrt{\frac{p}{n}} + \frac{p}{n} + \sqrt{\frac{\log n}{n}}\bigg) \ll \lambda_{\min}^\star, \qquad 1 \le i \le (p-r)\wedge n,$$

where the last step holds due to the noise assumption (3.13a). Combining these two observations establishes the claim bound (5.58):

$$\sup_{\lambda:\, \lambda\in[2\lambda_l^\star/3, 4\lambda_l^\star/3]} \|\boldsymbol{C}(\lambda)\| \lesssim \frac{\sigma^2}{\lambda_l^\star}\bigg(1 + \sqrt{\frac{p}{n}} + \frac{p}{n} + \sqrt{\frac{\log n}{n}}\bigg) \asymp \frac{\sigma^2}{\lambda_l^\star}\Big(1 + \frac{p}{n}\Big),$$

where the last step arises from the Cauchy-Schwarz inequality. Consequently, one can invoke Lemma 18 and apply the standard epsilon-net argument (similar to the proof of Lemma 1 in Appendix B.1 and hence omitted here) to demonstrate that

$$\begin{aligned}
&\sup_{\lambda:\, \lambda\in[2\lambda_l^\star/3, 4\lambda_l^\star/3]} \frac{1}{n}\big\|\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^\top - \mathbb{E}[\boldsymbol{Z}\boldsymbol{C}(\lambda)\boldsymbol{Z}^\top \mid \boldsymbol{C}(\lambda)]\big\| \\
&\quad\lesssim \sup_{\lambda:\, \lambda\in[2\lambda_l^\star/3, 4\lambda_l^\star/3]} \frac{1}{n}\|\boldsymbol{C}(\lambda)\|_{\mathrm{F}}\sqrt{r\log n} + \sup_{\lambda:\, \lambda\in[2\lambda_l^\star/3, 4\lambda_l^\star/3]} \frac{1}{n}\|\boldsymbol{C}(\lambda)\|\big(r\log n + \log^2 n\big) \\
&\quad\lesssim \sup_{\lambda:\, \lambda\in[2\lambda_l^\star/3, 4\lambda_l^\star/3]} \frac{1}{n}\|\boldsymbol{C}(\lambda)\|\sqrt{r(n\wedge p)}\,\log^2 n \\
&\quad\lesssim \frac{\sigma^2}{\lambda_l^\star}\bigg(\frac{p}{n} + \sqrt{\frac{p}{n}}\bigg)\sqrt{\frac{r}{n}}\,\log^2 n \ll \sqrt{\frac{r}{n}}\log n
\end{aligned}$$

with probability at least $1 - O(n^{-10})$. Here, the last line follows from (5.58) and the noise assumption that $\sigma^2(p/n + \sqrt{p/n}) \ll \lambda_{\min}^\star/\log n$. Combining this with (5.62), we arrive at

$$\sup_{\lambda:\, \lambda\in[2\lambda_l^\star/3, 4\lambda_l^\star/3]} \big\|\boldsymbol{K}(\lambda) - \boldsymbol{K}^\perp(\lambda)\big\| \ll (\lambda_{\max}^\star + \sigma^2)\sqrt{\frac{r}{n}}\log n$$

as claimed.

5.3.2 Proof of eigenvector perturbation theory (Theorem 3)

We now turn to our eigenvector perturbation theory. As before, we find it convenient to decompose the l-th eigenvector $\boldsymbol{u}_l$ of $\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top$ as follows

$$\boldsymbol{u}_l = \boldsymbol{u}_l^\star\cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta, \tag{5.63}$$

where the vector $\boldsymbol{u}_{l,\perp}$ obeys $\|\boldsymbol{u}_{l,\perp}\|_2 = 1$ and $\boldsymbol{u}_{l,\perp}^\top\boldsymbol{u}_l^\star = 0$. We shall employ this decomposition to identify several key quantities that we'd like to control. Specifically, armed with this decomposition, we can derive

$$\begin{aligned}
\boldsymbol{a}^\top\boldsymbol{u}_l &= (\boldsymbol{P}_{\boldsymbol{U}^\star}\boldsymbol{a})^\top\boldsymbol{u}_l + (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top\boldsymbol{u}_l \\
&= (\boldsymbol{P}_{\boldsymbol{U}^\star}\boldsymbol{a})^\top(\boldsymbol{u}_l^\star\cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta) + (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top\boldsymbol{u}_l \\
&= \sum_{1\le k\le r}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}(\boldsymbol{u}_l^\star\cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta) + (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l)
\end{aligned}$$

$$=\boldsymbol{a}^\top \boldsymbol{u}_l^\star \cos\theta + \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top} \boldsymbol{u}_{l,\perp} \sin\theta + (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l),$$

where we use the fact that $\boldsymbol{a} = \boldsymbol{P}_{\boldsymbol{U}^\star}\boldsymbol{a} + \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}$ with

$$\boldsymbol{P}_{\boldsymbol{U}^\star} = \Sigma_{1\leq k\leq r} \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top} \qquad \text{and} \qquad \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{U}^\star}.$$

As a result, we arrive at

$$\boldsymbol{a}^\top \boldsymbol{u}_l \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star = \boldsymbol{a}^\top \boldsymbol{u}_l^\star(\cos\theta \pm 1) + \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top} \boldsymbol{u}_{l,\perp} \sin\theta + (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l),$$

which further implies

$$\begin{aligned}
\min \left|\boldsymbol{a}^\top \boldsymbol{u}_l \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \leq& \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|(1 - |\cos\theta|) + \Big|\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top} \boldsymbol{u}_{l,\perp} \sin\theta\Big| + \left|(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l)\right| \\
\leq& \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|(1 - \cos^2\theta) + \Big|\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top} \boldsymbol{u}_{l,\perp} \sin\theta\Big| + \left|(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l)\right|. \qquad (5.64)
\end{aligned}$$

and

$$\begin{aligned}
\min \left|\boldsymbol{a}^\top \boldsymbol{u}_l \sqrt{1+c_l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \leq \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \cdot \left|1 - \sqrt{1+c_l}\,|\cos\theta|\right| + \sqrt{1+c_l}\left|\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\right| \\
+ \sqrt{1+c_l}\Big|\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top} \boldsymbol{u}_{l,\perp} \sin\theta\Big|. \qquad (5.65)
\end{aligned}$$

Thus, it comes down to bounding the following terms

$$1 - \cos^2\theta, \quad 1 - \sqrt{1+c_l}\,|\cos\theta|, \quad \sqrt{1+c_l}, \quad \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top} \boldsymbol{u}_{l,\perp} \sin\theta, \quad \text{and} \quad \langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle.$$

separately, which forms the main content of the remainder of the proof.

We claim that $\lambda_l \boldsymbol{I}_{p-1} - (\boldsymbol{u}_l^{\star\perp})^\top \frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top \boldsymbol{u}_l^{\star\perp}$ is invertible. This will be seen from (5.75) stated in Lemma 8 directly. The invertibility taken together with Theorem 5 reveals that $\cos\theta \neq 0$. If $\sin\theta = 0$, then we have $\boldsymbol{u}_l = \pm\boldsymbol{u}_l^\star$ and the conclusion is obvious since $\min\left|\boldsymbol{a}^\top \boldsymbol{u}_l \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| = 0$. Therefore, it suffices to focus on the case where $\cos\theta \neq 0$ and $\sin\theta \neq 0$ in the sequel.

1. Identifying several key quantities. Invoke Theorem 5 to show that

$$\cos^2\theta = \frac{1}{1 + \left\|\left(\lambda_l \boldsymbol{I}_{p-1} - (\boldsymbol{u}_l^{\star\perp})^\top \frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top \boldsymbol{u}_l^{\star\perp}\right)^{-1} (\boldsymbol{u}_l^{\star\perp})^\top \frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top \boldsymbol{u}_l^\star\right\|_2^2}, \qquad (5.66a)$$

$$\boldsymbol{u}_k^{\star\top} \boldsymbol{u}_{l,\perp} = \frac{\boldsymbol{u}_k^{\star\top} \boldsymbol{u}_l^{\star\perp}\left(\lambda_l \boldsymbol{I}_{p-1} - (\boldsymbol{u}_l^{\star\perp})^\top \frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top \boldsymbol{u}_l^{\star\perp}\right)^{-1} (\boldsymbol{u}_l^{\star\perp})^\top \frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top \boldsymbol{u}_l^\star}{\left\|\boldsymbol{u}_l^{\star\perp}\left(\lambda_l \boldsymbol{I}_{p-1} - (\boldsymbol{u}_l^{\star\perp})^\top \frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top \boldsymbol{u}_l^{\star\perp}\right)^{-1} (\boldsymbol{u}_l^{\star\perp})^\top \frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top \boldsymbol{u}_l^\star\right\|_2}. \qquad (5.66b)$$

For notational convenience, we shall define

$$\boldsymbol{s}_{l,\parallel} := \boldsymbol{u}_l^{\star\top}\boldsymbol{S} \in \mathbb{R}^{1\times n} \qquad \text{and} \qquad \boldsymbol{S}_{l,\perp} := (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{S} \in \mathbb{R}^{(p-1)\times n}, \qquad (5.67)$$

allowing us to write (5.66) more succinctly as follows

$$\cos^2\theta = \frac{1}{1 + \left\|\left(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\right)^{-1} \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right\|_2^2}, \qquad (5.68a)$$

$$\boldsymbol{u}_k^{\star\top} \boldsymbol{u}_{l,\perp} = \frac{\boldsymbol{u}_k^{\star\top} \boldsymbol{u}_l^{\star\perp}\left(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\right)^{-1} \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top}{\left\|\boldsymbol{u}_l^{\star\perp}\left(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\right)^{-1} \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right\|_2}. \qquad (5.68b)$$

With the above relations in mind, we can demonstrate that

$$
\begin{aligned}
\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_{l,\perp}\sin\theta\Big| &= \Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_{l,\perp}\Big|\sqrt{1-\cos^2\theta}\\
&= \frac{\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\Big|}{\big\|\boldsymbol{u}_l^{\star\perp}\big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2}\cdot\sqrt{\frac{\big\|\big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2^2}{1+\big\|\big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2^2}}\\
&\leq \left|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\Big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right|,
\end{aligned}
$$

where the last step follows since the columns of $\boldsymbol{u}_l^{\star\perp}$ are orthonormal. Substitution into (5.64) then yields

$$
\begin{aligned}
\min\big|\boldsymbol{a}^\top\boldsymbol{u}_l\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big| &\leq \big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big|\cdot(1-\cos^2\theta)+\big|(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l)\big|\\
&\quad+\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\Big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\Big|;
\end{aligned}
\tag{5.69}
$$

and

$$
\begin{aligned}
\min\big|\sqrt{1+c_l}\boldsymbol{a}^\top\boldsymbol{u}_l\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big| &\leq \big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big|\cdot\big|1-\sqrt{1+c_l}\,|\cos\theta|\big|+\sqrt{1+c_l}\big|\langle\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\,\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\rangle\big|\\
&\quad+\sqrt{1+c_l}\Big|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\Big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\Big|.
\end{aligned}
\tag{5.70}
$$

In what follows, we shall control these quantities separately.

2. Controlling the spectrum of $\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top$. Before moving forward to bound the terms mentioned above, we take a moment to first look at the eigenvalues of $\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top$. We first introduce some useful notation as follows:

- Let $\{\gamma_i^{(l)}\}_{i=1}^{p-1}$ denote the eigenvalues of $\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top$ (see the definition of $\boldsymbol{S}_{l,\perp}$ in (5.67)), and we assume that
$$\gamma_1^{(l)}\geq\cdots\geq\gamma_{p-1}^{(l)}. \tag{5.71}$$
- Let $\boldsymbol{u}_i^{(l)}$ be the eigenvector of $\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top$ associated with the eigenvalue $\gamma_i^{(l)}$.

Similar to (5.32), we find it helpful to introduce the following matrices

$$\boldsymbol{U}_{\setminus l}^\star := [\boldsymbol{u}_1^\star,\cdots,\boldsymbol{u}_{l-1}^\star,\boldsymbol{u}_{l+1}^\star,\cdots,\boldsymbol{u}_r^\star]\in\mathbb{R}^{p\times(r-1)}; \tag{5.72a}$$

$$\boldsymbol{U}^{\star(l)} := (\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{U}_{\setminus l}^\star = \begin{bmatrix}\boldsymbol{I}_{r-1}\\ \boldsymbol{0}\end{bmatrix}\in\mathbb{R}^{(p-1)\times(r-1)}; \tag{5.72b}$$

$$\boldsymbol{U}^{\star(l)\perp} := (\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{U}^{\star\perp} = \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{I}_{p-r}\end{bmatrix}\in\mathbb{R}^{(p-1)\times(p-r)}; \tag{5.72c}$$

$$\boldsymbol{\Lambda}^{\star(l)} := \mathsf{diag}\big(\{\lambda_i^\star\}_{i:i\neq l}\big)\in\mathbb{R}^{(r-1)\times(r-1)}. \tag{5.72d}$$

In addition, we define

$$\boldsymbol{u}_{i,\parallel}^{(l)} := \frac{1}{\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star(l)}}\boldsymbol{u}_i^{(l)}\big\|_2}\boldsymbol{P}_{\boldsymbol{U}^{\star(l)}}\boldsymbol{u}_i^{(l)},\qquad i\neq l, \tag{5.72e}$$

where $\boldsymbol{P}_{\boldsymbol{U}^{\star(l)}}=\boldsymbol{U}^{\star(l)}\big(\boldsymbol{U}^{\star(l)}\big)^\top$. Equipped with this set of notation, we are ready to present a lemma that characterizes the eigenvalues of the matrix $\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top$.

Lemma 8. *Instate the assumptions of Theorem 3, and recall the definition of $\beta(\cdot)$ in (5.45). With probability at least $1-O(n^{-10})$, the eigenvalues $\{\gamma_i^{(l)}\}_{i=1}^{p-1}$ of $\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top$ (see (5.71)) satisfy the following properties.*

1. *For each $1 \le i < r$, one has*

$$\frac{\gamma_i^{(l)}}{1+\beta(\gamma_i^{(l)})} \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_k^\star + \sigma^2) \qquad \text{for some } k \neq l \text{ and } 1 \le k \le r$$

and

$$\Big\| \Big(\gamma_i^{(l)} \boldsymbol{I}_{r-1} - \big(1+\beta(\gamma_i^{(l)})\big)\big(\boldsymbol{\Lambda}^{\star(l)} + \sigma^2 \boldsymbol{I}_{r-1}\big)\Big) \boldsymbol{U}^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)} \Big\|_2 \lesssim \mathcal{E}_{\mathsf{PCA}},$$

where $\mathcal{E}_{\mathsf{PCA}}$ is defined in (5.59).

2. *For each $r \le i \le n \wedge (p-1)$, one has*

$$\Big| \gamma_i^{(l)} - \sigma^2 \frac{p \vee n}{n} \Big| \lesssim \sigma^2 \sqrt{\frac{p + \log n}{n}}. \tag{5.73}$$

3. *For each $n \wedge (p-1) < i \le p-1$, we have $\gamma_i^{(l)} = 0$.*

4. *Furthermore, one has*

$$|\lambda - \lambda_l| \gtrsim \begin{cases} \Delta_l^\star, & \text{if } \frac{\lambda}{1+\beta(\lambda)} \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_i^\star + \sigma^2) \quad \text{for some } i \neq l \text{ and } 1 \le i \le r; \\ \lambda_l^\star, & \text{if } \big|\lambda - \sigma^2 \frac{p\vee n}{n}\big| \lesssim \sigma^2 \sqrt{\frac{p+\log n}{n}}. \end{cases} \tag{5.74}$$

In particular, one has

$$\big|\gamma_i^{(l)} - \lambda_l\big| \gtrsim \begin{cases} \Delta_l^\star, & 1 \le i < r; \\ \lambda_l^\star, & i \ge r. \end{cases} \tag{5.75}$$

Proof. See Appendix D.2. □

3. Controlling $\cos^2\theta$**.** In view of the expression of $\cos^2\theta$ in (5.68a), it suffices to control $\|(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\|}^\top\|_2^2$, which is accomplished in the following lemma.

Lemma 9. *Consider any $1 \le l \le r$. Instate the assumptions of Theorem 3, and recall the definition of c_l in (3.12). The following holds with probability at least $1 - O(n^{-10})$:*

$$\Big\| \Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1} \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\|}^\top \Big\|_2^2 \lesssim \frac{(\lambda_{\max}^\star + \sigma^2)(\lambda_l^\star + \sigma^2) r \log n}{(\Delta_l^\star)^2 n} + \frac{(\lambda_l^\star + \sigma^2)\sigma^2 p \log^2 n}{\lambda_l^{\star 2} n} \ll 1. \tag{5.76}$$

Moreover, for the case with $n \ge p$, one has

$$\begin{aligned} &\bigg| \Big\| \Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1} \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\|}^\top \Big\|_2^2 - c_l \bigg| \\ &\quad \lesssim \frac{(\lambda_{\max}^\star + \sigma^2)(\lambda_l^\star + \sigma^2) r \log n}{(\Delta_l^\star)^2 n} + \frac{\sigma^2 p}{\lambda_l^{\star 2} n}\bigg((\lambda_l^\star + \sigma^2)\sqrt{\frac{\log n}{p}} + (\lambda_{\max}^\star + \sigma^2)\sqrt{\frac{r \log n}{n}} \bigg), \end{aligned} \tag{5.77}$$

and for the case with $p > n$, we have

$$\bigg| \Big\| \Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1} \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\|}^\top \Big\|_2^2 - c_l \bigg| \lesssim \frac{(\lambda_{\max}^\star + \sigma^2)(\lambda_l^\star + \sigma^2) r \log n}{(\Delta_l^\star)^2 n} + \frac{\sigma^2 \kappa \sqrt{pr \log n}}{\lambda_l^\star n}. \tag{5.78}$$

Proof. See Appendix D.3. □

This lemma taken collectively with (5.68a) leads to

$$|\cos^2\theta - 1| = \left| \frac{1}{1 + \|(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\|}^\top\|_2^2} - 1 \right|$$

$$
\begin{aligned}
&= \frac{\big\|(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2^2}{1+\big\|(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2^2} \\
&\lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)r\log n}{(\Delta_l^\star)^2 n} + \frac{(\lambda_l^\star+\sigma^2)\sigma^2 p\log^2 n}{\lambda_l^{\star 2} n} \ll 1.
\end{aligned} \tag{5.79}
$$

where the last step follows from the assumptions (3.13a) and (3.13b). In addition, when $n \ge p$, one can combine (5.68a) and (5.77) to demonstrate that

$$
\begin{aligned}
\big|(1+c_l)\cos^2\theta - 1\big| &= \left|\frac{1+c_l}{1+\big\|(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2^2} - 1\right| \\
&= \frac{\Big|c_l - \big\|(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2^2\Big|}{1+\big\|(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2^2} \\
&\le \left|c_l - \left\|\Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right\|_2^2\right| \\
&\lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)r\log n}{(\Delta_l^\star)^2 n} + \frac{\sigma^2 p}{\lambda_l^{\star 2} n}\left((\lambda_{\max}^\star+\sigma^2)\sqrt{\frac{r\log n}{n}} + (\lambda_l^\star+\sigma^2)\frac{\log^2 n}{\sqrt{p}}\right),
\end{aligned} \tag{5.80}
$$

where the first line comes from the definition of $\cos^2\theta$ in (5.68a), and the last inequality holds due to (5.77). Moreover, if $p > n$, putting (5.68a) and (5.78) together reveals that

$$
\begin{aligned}
\big|(1+c_l)\cos^2\theta - 1\big| &= \left|\frac{1+c_l}{1+\big\|(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\big\|_2^2} - 1\right| \\
&\le \left|c_l - \left\|\Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right\|_2^2\right| \\
&\lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)r\log n}{(\Delta_l^\star)^2 n} + \frac{\sigma^2\kappa\sqrt{pr\log n}}{\lambda_l^\star n} \ll 1,
\end{aligned} \tag{5.81}
$$

where the last inequality holds due to the conditions (3.13a) and (3.13b). Taken collectively with (5.79), this leads to $1+c_l \lesssim 1$ and

$$
\begin{aligned}
\big|1-\sqrt{1+c_l}\,|\cos\theta|\big| &= \left|\frac{1-(1+c_l)\cos^2\theta}{1+\sqrt{1+c_l}\,|\cos\theta|}\right| \lesssim \big|1-(1+c_l)\cos^2\theta\big| \\
&\lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)r\log n}{(\Delta_l^\star)^2 n} + \frac{\sigma^2\kappa\sqrt{pr\log n}}{\lambda_l^\star n}.
\end{aligned} \tag{5.82}
$$

4. Controlling $\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top$. Recognizing that the vector $\boldsymbol{s}_{l,\parallel}$ (see (5.67)) obeys

$$
\boldsymbol{s}_{l,\parallel} \sim \mathcal{N}\big(\boldsymbol{0}, (\lambda_l^\star+\sigma^2)\boldsymbol{I}_n\big)
$$

and is independent of $\boldsymbol{S}_{l,\perp}$ (see (5.67)), we can control this quantity through the lemma below.

Lemma 10. *Instate the assumptions of Theorem 3. The following holds with probability at least $1-O(n^{-10})$:*

$$
\begin{aligned}
&\left|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\Big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right| \\
&\qquad \lesssim \sum_{k:k\neq l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}}\sqrt{(\lambda_l^\star+\sigma^2)(\lambda_{\max}^\star+\sigma^2)(\kappa^2+r)\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}
\end{aligned} \tag{5.83}
$$

Proof. See Appendix D.4. ∎

5. Controlling $(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l)$. When it comes to $(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l)$, we attempt to utilize certain rotational invariance property of $\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l$ in the subspace spanned by $\boldsymbol{U}^{\star\perp}$ to upper bound this quantity. This is formalized in Lemma 11.

Lemma 11. *Instate the assumptions of Instate the assumptions of Theorem 3. With probability at least* $1-O(n^{-10})$,

$$\big|(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l)\big| \lesssim \sqrt{\frac{\log n}{p-r}}\,\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\big\|_2\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2. \tag{5.84}$$

Proof. The proof is almost identical to the proof of Lemma 6, and is hence omitted for conciseness of presentation. □

In view of Lemma 11, it suffices to control $\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2$. To this end, it is seen from Theorem 6 that

$$\begin{aligned}
\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2^2 &= 1-\frac{1}{1+\big\|\big(\lambda_l\boldsymbol{I}-\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\big)^{-1}\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel}\big\|_2^2}\\
&\le \left\|\left(\lambda_l\boldsymbol{I}-\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\right)^{-1}\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel}\right\|_2^2\\
&\le \left\|\left(\lambda_l\boldsymbol{I}-\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\right)^{-1}\right\|^2\cdot\left\|\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top\right\|^2\cdot\big\|\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel}\big\|_2^2\\
&= \left\|\left(\lambda_l\boldsymbol{I}-\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\right)^{-1}\right\|^2\cdot\left\|\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top\right\|^2,
\end{aligned}\tag{5.85}$$

where we recall $\boldsymbol{u}_{l,\parallel}$ is defined to be a unit vector $\boldsymbol{u}_{l,\parallel}:=\boldsymbol{P}_{\boldsymbol{U}^\star}(\boldsymbol{u})/\|\boldsymbol{P}_{\boldsymbol{U}^\star}(\boldsymbol{u})\|_2$ and satisfies $\boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\boldsymbol{u}_{l,\parallel}=\boldsymbol{u}_{l,\parallel}$.

The preceding inequality then motivates us to control both $\big\|\big(\lambda_l\boldsymbol{I}-\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\big)^{-1}\big\|$ and $\big\|\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top\big\|$. As shown in the proof of Lemma 7 in Appendix D.1 (cf. (D.4 and (D.6)), we know that

$$\left|\frac{1}{n}\big\|\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\big\|-\sigma^2\right| \lesssim \sigma^2\left(\sqrt{\frac{p}{n}}+\frac{p}{n}+\sqrt{\frac{\log n}{n}}\right)=o(\lambda_{\min}^\star), \tag{5.86}$$

$$\frac{1}{n}\big\|\boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top\big\| \lesssim \sqrt{\frac{(\lambda_{\max}^\star+\sigma^2)\sigma^2(p-r)}{n}}\log n, \tag{5.87}$$

where the relation in (5.86) arises from the noise condition (3.13a). Combining these with Theorem 8, we obtain

$$\begin{aligned}
\lambda_l-\frac{1}{n}\big\|\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\big\| &= \lambda_l-\sigma^2-o(\lambda_{\min}^\star)\\
&\overset{\text{(i)}}{\ge} \big(1+\beta(\lambda_l)\big)(\lambda_l^\star+\sigma^2)-\big(1+\beta(\lambda_l)\big)\cdot O\Big((\lambda_{\max}^\star+\sigma^2)\sqrt{\frac{r}{n}}\log n\Big)-\sigma^2-o(\lambda_{\min}^\star)\\
&= \big(1+\beta(\lambda_l)\big)\lambda_l^\star+\beta(\lambda_l)\sigma^2-O\Big(\big(1+\beta(\lambda_l)\big)(\lambda_{\max}^\star+\sigma^2)\sqrt{\frac{r}{n}}\log n\Big)-o(\lambda_{\min}^\star)\\
&\overset{\text{(ii)}}{\gtrsim} \lambda_l^\star+O\left(\frac{\sigma^4 p}{\lambda_l^\star n}\right)\overset{\text{(iii)}}{\asymp}\lambda_l^\star,
\end{aligned}$$

where (i) is due to the bound developed for λ_l in (5.44) in Theorem 8; (ii) arises from the fact $\beta(\lambda_l)\lesssim\frac{\sigma^2 p}{\lambda_l^\star n}\ll 1$ (as shown in (5.61)) and the noise condition (3.13a) that $(\lambda_{\max}^\star+\sigma^2)\sqrt{r/n}\log n\ll\lambda_{\min}^\star$; (iii) is legal as long as $\sigma^2\sqrt{p/n}\ll\lambda_{\min}^\star$. As a result, we obtain

$$\left\|\left(\lambda_l\boldsymbol{I}-\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\right)^{-1}\right\| \le \frac{1}{\lambda_l-\|\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\|}\lesssim\frac{1}{\lambda_l^\star}; \tag{5.88}$$

Plugging (5.87) and (5.88) into (5.85) immediately reveals that

$$\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\big\|_2 \le \left\|\left(\lambda_l\boldsymbol{I}-\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\right)^{-1}\right\|\cdot\frac{1}{n}\big\|\boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top\big\| \lesssim \frac{\sqrt{(\lambda_{\max}^\star+\sigma^2)\sigma^2}}{\lambda_l^\star}\sqrt{\frac{p-r}{n}}\log n. \tag{5.89}$$

Taken together with Lemma 11, this leads to the bound

$$\left|(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a})^\top(\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l)\right| \lesssim \sqrt{\frac{(\lambda_{\max}^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}}\log^2 n\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2. \tag{5.90}$$

6. Combining bounds. Finally, we can combine (5.79), (5.83) and (5.90) to arrive at the error bound for the plug-in estimator:

$$\begin{aligned}\min\left|\boldsymbol{a}^\top\boldsymbol{u}_l \pm \boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| &\lesssim \left(\frac{\left(\lambda_{\max}^\star+\sigma^2\right)\left(\lambda_l^\star+\sigma^2\right)r\log n}{\Delta_l^{\star 2}n}+\frac{(\lambda_l^\star+\sigma^2)\sigma^2 p}{\lambda_l^{\star 2}n}\right)\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| \\ &\quad+\sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star-\lambda_k^\star\right|\sqrt{n}}\sqrt{(\lambda_l^\star+\sigma^2)(\lambda_{\max}^\star+\sigma^2)(\kappa^2+r)\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)} \\ &\quad+\sqrt{\frac{(\lambda_{\max}^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}}\log^2 n\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2\end{aligned}$$

as claimed.

7. Analyzing the de-biased estimator. To finish up, let us turn to the de-biased estimator.

- Consider first the case with $n \geq p$. We can substitute (5.80), (5.82), (5.83), and (5.90) into (5.70) to obtain

$$\begin{aligned}&\min\left|\boldsymbol{a}^\top\boldsymbol{u}_l\sqrt{1+c_l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| \\ &\lesssim \left(\frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)r\log n}{\Delta_l^{\star 2}n}+\frac{\sigma^2 p}{\lambda_l^{\star 2}n}\left((\lambda_{\max}^\star+\sigma^2)\sqrt{\frac{r\log n}{n}}+(\lambda_l^\star+\sigma^2)\frac{\log n}{\sqrt{p}}\right)\right)\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| \\ &\quad+\sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star-\lambda_k^\star\right|\sqrt{n}}\sqrt{(\lambda_l^\star+\sigma^2)(\lambda_{\max}^\star+\sigma^2)(\kappa^2+r)\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}+\sqrt{\frac{(\lambda_{\max}^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}}\log^2 n\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2 \\ &\lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)\,r\log n}{\Delta_l^{\star 2}n}\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|+\sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star-\lambda_k^\star\right|\sqrt{n}}\sqrt{(\lambda_l^\star+\sigma^2)(\lambda_{\max}^\star+\sigma^2)(\kappa^2+r)\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)} \\ &\quad+\sqrt{\frac{(\lambda_{\max}^\star+\sigma^2)\sigma^2 r}{\lambda_l^{\star 2}n}}\log^2 n,\end{aligned}$$

  where the last step holds due to the noise assumption (3.13a) as well as the facts that $\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| \leq \|\boldsymbol{a}\|_2\|\boldsymbol{u}_l^\star\|_2 \leq 1$ and $\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2 \leq \|\boldsymbol{a}\|_2 = 1$.

- Consider instead the case with $n < p$ (which implies $\sigma^2 \ll \lambda_{\min}^\star$). Then we can substitute (5.81), (5.82), (5.83) and (5.90) into (5.70) to derive

$$\begin{aligned}&\min\left|\boldsymbol{a}^\top\boldsymbol{u}_l\sqrt{1+c_l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| \\ &\lesssim \left(\frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)r\log n}{\Delta_l^{\star 2}n}+\frac{\sigma^2\sqrt{\kappa^2 pr\log n}}{\lambda_l^\star n}\right)\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| \\ &\quad+\sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star-\lambda_k^\star\right|\sqrt{n}}\sqrt{(\lambda_l^\star+\sigma^2)(\lambda_{\max}^\star+\sigma^2)(\kappa^2+r)\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}+\sqrt{\frac{(\lambda_{\max}^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}}\log^2 n\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2 \\ &\lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)r\log n}{\Delta_l^{\star 2}n}\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|+\sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star-\lambda_k^\star\right|\sqrt{n}}\sqrt{(\lambda_l^\star+\sigma^2)(\lambda_{\max}^\star+\sigma^2)(\kappa^2+r)\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)} \\ &\quad+\sqrt{\frac{(\lambda_{\max}^\star+\sigma^2)\sigma^2\kappa^2 r}{\lambda_l^{\star 2}n}}\log^2 n.\end{aligned}$$

Here, we use the noise assumption (3.13a), $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \leq \|\boldsymbol{a}\|_2 \|\boldsymbol{u}_l^\star\|_2 \leq 1$ and $\big\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\big\|_2 \leq \|\boldsymbol{a}\|_2 = 1$ again in the last step.

6 Discussion

This paper has explored estimation of linear functionals of unknown eigenvectors under i.i.d. Gaussian noise, covering the contexts of both matrix denoising and principal component analysis. We have demonstrated a non-negligible bias issue inherent to the naive plug-in estimator, and have proposed more effective estimators that allow for bias correction in a minimax-optimal and data-driven manner. In comparison to prior works, our theory accommodates the scenario in which the associated eigen-gap is substantially smaller than the size of the perturbation, thereby expanding on what generic matrix perturbation theory has to offer in these statistical applications.

Moving forward, there are numerous extensions that are worth pursuing. For example, the present work is likely suboptimal with respect to the dependence on the rank r and the condition number κ, which calls for a more refined analytical framework to achieve optimal estimation for more general scenarios. In addition, our current theory focuses on i.i.d. Gaussian noise, and a natural question arises as to how to accommodate sub-Gaussian noise and/or heteroscedastic data. Furthermore, given the minimax estimation guarantees, an interesting direction lies in developing statistical inference and uncertainty quantification schemes for linear forms of the eigenvectors. Accomplishing this task would require developing distributional guarantees for the proposed de-biased estimators as well as accurate estimation of the error variance, which we leave to future investigation.

Acknowledgements

Y. Chen is supported in part by the Alfred P. Sloan Research Fellowship, the grants AFOSR YIP award FA9550-19-1-0030, ONR N00014-19-1-2120, ARO YIP award W911NF-20-1-0097, ARO W911NF-18-1-0303, NSF CCF-1907661, DMS-2014279, IIS-2218713 and IIS-2218773. H. V. Poor is supported in part by NSF CCF-1908308, and in part by a Princeton Schmidt Data-X Research Award.

A Proofs of master theorems

A.1 Proof of Theorem 5

Given that $\boldsymbol{u}_l$ is an eigenvector of $\boldsymbol{M}$, one has $\boldsymbol{M}\boldsymbol{u}_l = \lambda_l \boldsymbol{u}_l$, which together with the decomposition (5.3) and the condition $\boldsymbol{u}_{l,\perp} = \boldsymbol{q}^\perp (\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}$ gives

$$\begin{aligned} & \boldsymbol{M}(\boldsymbol{q}\cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta) = \lambda_l(\boldsymbol{q}\cos\theta + \boldsymbol{u}_{l,\perp}\sin\theta) \\ \Longleftrightarrow \qquad & \boldsymbol{M}\boldsymbol{q}\cos\theta + \boldsymbol{M}\boldsymbol{q}^\perp(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta = \lambda_l \boldsymbol{q}\cos\theta + \lambda_l \boldsymbol{u}_{l,\perp}\sin\theta. \end{aligned} \tag{A.1}$$

Left-multiplying both sides of this equation by $\boldsymbol{q}^\top$ (resp. $(\boldsymbol{q}^\perp)^\top$) and using the assumptions of $\boldsymbol{u}_{l,\perp}$ (namely, $\boldsymbol{q}^\top \boldsymbol{u}_{l,\perp} = 0$ and $\boldsymbol{q}^\perp (\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp} = \boldsymbol{u}_{l,\perp}$) give

$$\boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}\cos\theta + \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}^\perp (\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta = \lambda_l \cos\theta, \tag{A.2a}$$

$$(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\cos\theta + (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp (\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta = \lambda_l (\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta. \tag{A.2b}$$

Rearrange terms in (A.2b) to arrive at

$$\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta = (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\cos\theta. \tag{A.3}$$

Given the assumption that $\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp$ is invertible and the fact that $\|(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}\|_2 = \|\boldsymbol{q}^\perp (\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}\|_2 = \|\boldsymbol{u}_{l,\perp}\|_2 = 1$, we claim that it is straightforward to verify that $\cos\theta \neq 0$. To see this, suppose instead that

$\cos\theta = 0$, then the right-hand side of (A.3) equals to 0, whereas the left-hand side of (A.3) is non-zero because $\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp} \neq \boldsymbol{0}$ and $\sin\theta = \sqrt{1-\cos^2\theta} = 1$. This leads to contradiction, which in turn reveals that $\cos\theta \neq 0$. In addition, if $\sin\theta = 0$ (or $\cos\theta = 1$), then one has $\boldsymbol{q} = \boldsymbol{u}_l$ and $(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q} = 0$ (see the relation (A.3)), from which the claims (5.4) immediately follow. Hence, we shall focus on the cases where $\cos\theta \neq 0$ and $\sin\theta \neq 0$ in the sequel.

Notice that (A.3) can be rewritten as

$$(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp} = \frac{\cos\theta}{\sin\theta}\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}. \tag{A.4}$$

This together with the unit norm constraint of $\boldsymbol{u}_{l,\perp}$ and $\boldsymbol{u}_{l,\perp} = \boldsymbol{q}^\perp(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}$ implies that

$$\boldsymbol{u}_{l,\perp} = \pm \frac{\boldsymbol{q}^\perp\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}}{\big\|\boldsymbol{q}^\perp\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\big\|_2} \tag{A.5}$$

as claimed in (5.4c). In addition, substitution of (A.4) into (A.2a) with a little algebra yields

$$\begin{aligned}
\big(\lambda_l - \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}\big)\cos\theta &= \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}^\perp(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta \\
&= \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}^\perp\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\cos\theta, \\
\Longrightarrow \qquad \lambda_l - \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q} &= \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}^\perp\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q},
\end{aligned} \tag{A.6}$$

thus establishing the claim (5.4b).

Finally, rearranging terms in (A.2a) yields

$$\frac{\sin\theta}{\cos\theta} = \frac{\lambda_l - \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}}{\boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}^\perp(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}}. \tag{A.7}$$

This taken collectively with the elementary identity $\cos^2\theta + \sin^2\theta = 1$ immediately leads to

$$\begin{aligned}
\cos^2\theta &= \frac{1}{1 + \frac{|\lambda_l - \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}|^2}{|\boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}^\perp(\boldsymbol{q}^\perp)^\top \boldsymbol{u}_{l,\perp}|^2}} \overset{\text{(i)}}{=} \frac{1}{1 + \frac{|\lambda_l - \boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}|^2 \cdot \|\boldsymbol{q}^\perp(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\|_2^2}{|\boldsymbol{q}^\top \boldsymbol{M}\boldsymbol{q}^\perp(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}|^2}} \\
&\overset{\text{(ii)}}{=} \frac{1}{1 + \|\boldsymbol{q}^\perp\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\|_2^2} \\
&\overset{\text{(iii)}}{=} \frac{1}{1 + \|\big(\lambda_l \boldsymbol{I}_{n-1} - (\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}^\perp\big)^{-1}(\boldsymbol{q}^\perp)^\top \boldsymbol{M}\boldsymbol{q}\|_2^2},
\end{aligned}$$

where (i) relies on the expression (A.5), (ii) results from the identity (A.6), and (iii) follows since $(\boldsymbol{q}^\perp)^\top \boldsymbol{q}^\perp = \boldsymbol{I}_{n-1}$. This establishes the claimed relation (5.4a).

A.2 Proof of Theorem 6

Given that $\boldsymbol{M}\boldsymbol{u}_l = \lambda_l \boldsymbol{u}_l$, one can invoke the decomposition (5.2) to obtain

$$\boldsymbol{M}\boldsymbol{u}_{l,\|}\cos\theta + \boldsymbol{M}\boldsymbol{u}_{l,\perp}\sin\theta = \lambda_l \boldsymbol{u}_{l,\|}\cos\theta + \lambda_l \boldsymbol{u}_{l,\perp}\sin\theta, \tag{A.8}$$

which together with the conditions $\boldsymbol{u}_{l,\|} = \boldsymbol{Q}\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|}$ and $\boldsymbol{u}_{l,\perp} = \boldsymbol{Q}^\perp(\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}$ implies that

$$\boldsymbol{M}\boldsymbol{Q}\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|}\cos\theta + \boldsymbol{M}\boldsymbol{Q}^\perp(\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta = \lambda_l \boldsymbol{u}_{l,\|}\cos\theta + \lambda_l \boldsymbol{u}_{l,\perp}\sin\theta. \tag{A.9}$$

Left-multiplying both sides of this relation by $\boldsymbol{Q}^\top$ (resp. $(\boldsymbol{Q}^\perp)^\top$) gives

$$\boldsymbol{Q}^\top \boldsymbol{M}\boldsymbol{Q}\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|}\cos\theta + \boldsymbol{Q}^\top \boldsymbol{M}\boldsymbol{Q}^\perp(\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta = \lambda_l \boldsymbol{Q}^\top \boldsymbol{u}_{l,\|}\cos\theta, \tag{A.10a}$$

$$(\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|}\cos\theta + (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}^\perp(\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta = \lambda_l(\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta, \tag{A.10b}$$

thus indicating that

$$\begin{aligned}\big(\lambda_l \boldsymbol{I}_k - \boldsymbol{Q}^\top \boldsymbol{M}\boldsymbol{Q}\big)\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|}\cos\theta &= \boldsymbol{Q}^\perp \boldsymbol{M}\boldsymbol{Q}^\perp (\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta,\\ \big(\lambda_l \boldsymbol{I}_{n-k} - (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}^\perp\big)(\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta &= (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|}\cos\theta = (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{u}_{l,\|}\cos\theta,\end{aligned} \tag{A.11}$$

where the last identity follows since $\boldsymbol{Q}\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|} = \boldsymbol{u}_{l,\|}$. These two relations taken together demonstrate that

$$\begin{aligned}\big(\lambda_l \boldsymbol{I}_k - \boldsymbol{Q}^\top \boldsymbol{M}\boldsymbol{Q}\big)\boldsymbol{Q}^\top \boldsymbol{u}_{l,\|}\cos\theta &= \boldsymbol{Q}^\perp \boldsymbol{M}\boldsymbol{Q}^\perp \big((\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}\sin\theta\big)\\ &= \boldsymbol{Q}^\perp \boldsymbol{M}\boldsymbol{Q}^\perp \big(\lambda_l \boldsymbol{I}_{n-k} - (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}^\perp\big)^{-1}(\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{u}_{l,\|}\cos\theta.\end{aligned}$$

In addition, in view of the invertibility of $\lambda_l \boldsymbol{I}_{n-k} - (\boldsymbol{Q}^\perp)^\top \boldsymbol{M}\boldsymbol{Q}^\perp$ (due to the assumption) and $\|(\boldsymbol{Q}^\perp)^\top \boldsymbol{u}_{l,\perp}\|_2 = \|\boldsymbol{u}_{l,\perp}\|_2 = 1$, one can deduce from (A.11) that $\cos\theta \neq 0$. To verify this, suppose $\cos\theta = 0$ (or $\sin\theta = 1$), then the left-hand side of (A.11) is non-zero while the right-hand side of (A.11) is zero. This results in contradiction, thus justifying that $\cos\theta \neq 0$. Consequently, dividing both sides of the above identity by $\cos\theta$ concludes the proof for the claim (5.5b).

B Proofs of auxiliary lemmas for Theorem 7

B.1 Proof of Lemma 1

For notational convenience, divide the matrix $\boldsymbol{H}$ as follows

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_{\mathsf{ul}} & \boldsymbol{H}_{\mathsf{ur}} \\ \boldsymbol{H}_{\mathsf{ur}}^\top & \boldsymbol{H}_{\mathsf{lr}} \end{bmatrix}, \quad \boldsymbol{H}_{\mathsf{ul}} \in \mathbb{R}^{r\times r}, \quad \boldsymbol{H}_{\mathsf{ur}} \in \mathbb{R}^{r\times(n-r)}, \quad \boldsymbol{H}_{\mathsf{lr}} \in \mathbb{R}^{(n-r)\times(n-r)}. \tag{B.1}$$

In view of the rotational invariance of a symmetric Gaussian matrix, we know that $\boldsymbol{R}\boldsymbol{H}\boldsymbol{R}^\top$ has the same distribution as $\boldsymbol{H}$ for any fixed orthonormal matrix $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ obeying $\boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{I}_n$. As a result, it is easily seen that the triple

$$\Big(\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^\star, (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}, \boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^{\star\perp}\Big) \overset{\mathrm{d}}{=} (\boldsymbol{H}_{\mathsf{ul}}, \boldsymbol{H}_{\mathsf{lr}}, \boldsymbol{H}_{\mathsf{ur}}), \tag{B.2}$$

where $\overset{\mathrm{d}}{=}$ denotes equivalence in distribution. Equipped with this fact, we are ready to derive the advertised concentration bounds.

Controlling $\big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^\star\big\|$. Apply the standard Gaussian concentration inequalities [Vershynin, 2012] and (B.2) to conclude that with probability at least $1 - O(n^{-10})$,

$$\|\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^\star\| = \|\boldsymbol{H}_{\mathsf{ul}}\| \lesssim \sigma(\sqrt{r} + \sqrt{\log n}).$$

Controlling $\big\|\boldsymbol{G}(\lambda) - \boldsymbol{G}^\perp(\lambda)\big\|$. Consider any fixed λ obeying $2\,|\lambda_l^\star|/3 \le |\lambda| \le 4\,|\lambda_l^\star|/3$. Recalling the expression of $\boldsymbol{G}(\lambda)$ in (5.14b), we have

$$\begin{aligned}\big\|\boldsymbol{G}(\lambda)\big\| &= \big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^\top\big(\lambda\boldsymbol{I}_n - \boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^\top\big)^{-1}\boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^\star\big\|\\ &= \big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^{\star\perp}\big(\lambda\boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)^{-1}(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^\star\big\|.\end{aligned} \tag{B.3}$$

Combining this with the fact (B.2), we see that $\|\boldsymbol{G}(\lambda)\|$ has the same distribution as $\big\|\boldsymbol{H}_{\mathsf{ur}}\big(\lambda\boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}}\big)^{-1}\boldsymbol{H}_{\mathsf{ur}}^\top\big\|$. Repeating the same argument also indicates that $\big\|\boldsymbol{G}(\lambda) - \boldsymbol{G}^\perp(\lambda)\big\|$ has the same distribution as

$$\Big\|\boldsymbol{H}_{\mathsf{ur}}\big(\lambda\boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}}\big)^{-1}\boldsymbol{H}_{\mathsf{ur}}^\top - \mathbb{E}\big[\boldsymbol{H}_{\mathsf{ur}}\big(\lambda\boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}}\big)^{-1}\boldsymbol{H}_{\mathsf{ur}}^\top \mid \boldsymbol{H}_{\mathsf{lr}}\big]\Big\|. \tag{B.4}$$

This allows us to turn attention to $\boldsymbol{H}_{\mathsf{ur}}(\lambda\boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}})^{-1}\boldsymbol{H}_{\mathsf{ur}}^\top$.

As a key observation, $\boldsymbol{H}_{\mathsf{ur}}$ and $\boldsymbol{H}_{\mathsf{lr}}$ are statistically independent, thus enabling convenient decoupling of the randomness. Let $\gamma_1 \ge \dots \ge \gamma_{n-r}$ represent the eigenvalues of $\boldsymbol{H}_{\mathsf{lr}}$. Denote by $\{\boldsymbol{h}_i\}_{i=1}^{n-r}$ the columns of

$\boldsymbol{H}_{\mathsf{ur}}$, i.e. $\boldsymbol{H}_{\mathsf{ur}} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_{n-r}]$, which are independent of $\boldsymbol{H}_{\mathsf{lr}}$ and $\{\gamma_i\}$. Invoking the rotational invariance of Gaussian random matrices once again, we see that

$$\boldsymbol{H}_{\mathsf{ur}}(\lambda \boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}})^{-1}\boldsymbol{H}_{\mathsf{ur}}^{\top} \overset{\mathrm{d}}{=} \sum_{i=1}^{n-r} \frac{1}{\lambda - \gamma_i} \boldsymbol{h}_i \boldsymbol{h}_i^{\top}, \tag{B.5}$$

which is a sum of independent random matrices when conditional on $\boldsymbol{H}_{\mathsf{lr}}$. This can be controlled via Lemma 18. Specifically, conditional on $\boldsymbol{H}_{\mathsf{lr}}$ and assuming that $|\gamma_i| \leq \lambda_{\min}^{\star}/3$ for all i, we have

$$\begin{aligned}
&\Big\| \sum_i \boldsymbol{H}_{\mathsf{ur}}(\lambda \boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}})^{-1}\boldsymbol{H}_{\mathsf{ur}}^{\top} - \mathbb{E}\Big[\sum_i \boldsymbol{H}_{\mathsf{ur}}(\lambda \boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}})^{-1}\boldsymbol{H}_{\mathsf{ur}}^{\top} \mid \boldsymbol{H}_{\mathsf{lr}} \Big] \Big\| \\
&\quad = \Big\| \sum_i \frac{1}{\lambda - \gamma_i} \big(\boldsymbol{h}_i \boldsymbol{h}_i^{\top} - \mathbb{E}[\boldsymbol{h}_i \boldsymbol{h}_i^{\top}] \big) \Big\| \\
&\quad \lesssim \frac{\sigma^2}{\min_i |\lambda - \gamma_i|} \big(\sqrt{rn \log n} + r \log n \big) \\
&\quad \lesssim \frac{\sigma^2}{|\lambda_l^{\star}|} \big(\sqrt{rn \log n} + r \log n \big)
\end{aligned}$$

with probability at least $1 - O(n^{-20})$, where the penultimate line relies on Lemma 18, and the last step follows since $|\lambda - \gamma_i| \geq |\lambda| - \max_i |\gamma_i| \geq 2\,|\lambda_l^{\star}|/3 - \|\boldsymbol{H}\| \geq \lambda_{\min}^{\star}/3$ (see (5.8)). Consequently, we have established that, with probability at least $1 - O(n^{-11})$,

$$\big\| \boldsymbol{G}(\lambda) - \boldsymbol{G}^{\perp}(\lambda) \big\| \lesssim \frac{\sigma^2}{|\lambda_l^{\star}|} \big(\sqrt{rn \log n} + r \log n \big) \leq \frac{\sigma^2}{\lambda_{\min}} \big(\sqrt{rn \log n} + r \log n \big) \tag{B.6}$$

for a given λ.

Finally, we apply the standard epsilon-net argument to establish a uniform bound that holds simultaneously over all λ obeying $2\,|\lambda_l^{\star}|/3 \leq |\lambda| \leq 4\,|\lambda_l^{\star}|/3$. Set $\epsilon_0 = c\,|\lambda_l^{\star}|/n$ for some sufficiently small constant $c > 0$, and let $\mathcal{N}_{\epsilon_0}$ denote an ϵ_0-net for $[-4|\lambda_l^{\star}|/3, -2|\lambda_l^{\star}|/3] \cup [2|\lambda_l^{\star}|/3, 4|\lambda_l^{\star}|/3]$ with cardinality

$$|\mathcal{N}_{\epsilon_0}| \lesssim \lambda_l^{\star}/\epsilon_0 \asymp n; \tag{B.7}$$

see Vershynin [2017] for an introduction of the epsilon-net. This means that for each λ obeying $2\,|\lambda_l^{\star}|/3 \leq |\lambda| \leq 4\,|\lambda_l^{\star}|/3$, one can find a point $\widehat{\lambda} \in \mathcal{N}_{\epsilon_0}$ such that $|\lambda - \widehat{\lambda}| \leq \epsilon_0$.

- Take the union bound to show that: with probability exceeding $1 - O(n^{-11})$,

$$\big\| \boldsymbol{G}(\widehat{\lambda}) - \boldsymbol{G}^{\perp}(\widehat{\lambda}) \big\| \lesssim \frac{\sigma^2}{\lambda_{\min}^{\star}} \big(\sqrt{rn \log n} + r \log n \big), \qquad \forall \widehat{\lambda} \in \mathcal{N}. \tag{B.8}$$

- For any λ of interest, let $\widehat{\lambda}$ be a point in $\mathcal{N}_{\epsilon_0}$ obeying $|\lambda - \widehat{\lambda}| \leq \epsilon_0$. Then conditioned on $\|\boldsymbol{H}\| \leq \lambda_{\min}^{\star}/3$,

$$\begin{aligned}
\|\boldsymbol{G}(\lambda) - \boldsymbol{G}(\widehat{\lambda})\| &\leq \big\| \boldsymbol{H}_{\mathsf{ur}} \big(\lambda \boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}}\big)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top} - \boldsymbol{H}_{\mathsf{ur}} \big(\widehat{\lambda} \boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}}\big)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top} \big\| \\
&\leq \|\boldsymbol{H}_{\mathsf{ur}}\|^2 \cdot \big\| \big(\lambda \boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}}\big)^{-1} - \big(\widehat{\lambda} \boldsymbol{I}_{n-r} - \boldsymbol{H}_{\mathsf{lr}}\big)^{-1} \big\| \\
&\leq \|\boldsymbol{H}_{\mathsf{ur}}\|^2 \max_i \Big| \frac{1}{\lambda - \gamma_i} - \frac{1}{\widehat{\lambda} - \gamma_i} \Big| = \|\boldsymbol{H}_{\mathsf{ur}}\|^2 \max_i \Big| \frac{\lambda - \widehat{\lambda}}{(\lambda - \gamma_i)(\widehat{\lambda} - \gamma_i)} \Big| \\
&\lesssim \sigma^2 n \cdot \max_i \frac{|\lambda - \widehat{\lambda}|}{\lambda_l^{\star 2}} \leq \sigma^2 n \cdot \frac{\epsilon_l}{\lambda_l^{\star 2}} \\
&\lesssim \frac{\sigma^2}{\lambda_{\min}^{\star}}
\end{aligned} \tag{B.9}$$

holds with probability $1 - O(n^{-11})$. Here, the penultimate line has made use of the Gaussian concentration bound $\|\boldsymbol{H}_{\mathsf{ur}}\| \lesssim \sigma\sqrt{n}$, whereas the last inequality results from (B.7).

- Combining the above two facts together, we arrive at

$$\begin{aligned}
&\sup_{\lambda:\,|\lambda|\in[2|\lambda_l^\star|/3,\,4|\lambda_l^\star|/3]}\big\|\boldsymbol{G}(\lambda)-\boldsymbol{G}^\perp(\lambda)\big\| \\
&\quad=\sup_{\lambda:\,|\lambda|\in[2|\lambda_l^\star|/3,\,4|\lambda_l^\star|/3]}\big\|\boldsymbol{G}(\lambda)-\boldsymbol{G}(\widehat{\lambda})+\boldsymbol{G}^\perp(\widehat{\lambda})-\boldsymbol{G}^\perp(\lambda)+\boldsymbol{G}(\widehat{\lambda})-\boldsymbol{G}^\perp(\widehat{\lambda})\big\| \\
&\quad\le\sup_{\lambda:\,|\lambda|\in[2|\lambda_l^\star|/3,\,4|\lambda_l^\star|/3]}\|\boldsymbol{G}(\lambda)-\boldsymbol{G}(\widehat{\lambda})\|+\sup_{\widehat{\lambda}:\,\widehat{\lambda}\in\mathcal{N}_{\epsilon_0}}\big\|\boldsymbol{G}(\widehat{\lambda})-\boldsymbol{G}^\perp(\widehat{\lambda})\big\| \\
&\qquad\quad+\sup_{\lambda:\,|\lambda|\in[2|\lambda_l^\star|/3,\,4|\lambda_l^\star|/3]}\|\boldsymbol{G}^\perp(\lambda)-\boldsymbol{G}^\perp(\widehat{\lambda})\| \\
&\quad\lesssim\frac{\sigma^2}{\lambda_{\min}^\star}\big(\sqrt{rn\log n}+r\log n\big). \qquad\qquad (B.10)
\end{aligned}$$

  Here, the last inequality results from (B.8), (B.9), and the following consequence of Jensen's inequality

$$\|\boldsymbol{G}^\perp(\lambda)-\boldsymbol{G}^\perp(\widehat{\lambda})\|=\|\mathbb{E}\big[\boldsymbol{G}(\lambda)\mid\boldsymbol{H}_{\mathsf{lr}}\big]-\mathbb{E}\big[\boldsymbol{G}(\widehat{\lambda})\mid\boldsymbol{H}_{\mathsf{lr}}\big]\|\le\mathbb{E}\Big[\|\boldsymbol{G}(\lambda)-\boldsymbol{G}(\widehat{\lambda})\|\mid\boldsymbol{H}_{\mathsf{lr}}\Big]\lesssim\frac{\sigma^2}{\lambda_{\min}^\star},$$

  where we have used (B.9) again in the last step.

This concludes the proof of (5.19).

Finally, the above argument also reveals that

$$\begin{aligned}
\boldsymbol{G}^\perp(\lambda)&=\mathbb{E}\Big[\underbrace{\boldsymbol{U}^{\star\top}\boldsymbol{H}\boldsymbol{U}^{\star\perp}\big(\lambda\boldsymbol{I}_{n-r}-(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)^{-1}(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^{\star}}_{=:\,\boldsymbol{A}}\mid(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^{\star\perp}\Big] \\
&=\sigma^2\mathsf{tr}\big[\big(\lambda\boldsymbol{I}_{n-r}-(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)^{-1}\big]\boldsymbol{I}_r,
\end{aligned}$$

which holds since the matrix $\boldsymbol{A}$ obeys $\boldsymbol{A}\overset{\mathrm{d}}{=}\boldsymbol{H}_{\mathsf{ur}}(\lambda\boldsymbol{I}_{n-r}-\boldsymbol{H}_{\mathsf{lr}})^{-1}\boldsymbol{H}_{\mathsf{ur}}^\top$, which has been analyzed in (B.5).

B.2 Proof of Lemma 2

We first claim that: with (5.22) in place, one necessarily has

$$\big|\lambda_l-\lambda_i^\star-\gamma(\lambda_l)\big|\le\mathcal{E}_{\mathsf{MD}},\qquad\text{for some }1\le i\le r \qquad (B.11a)$$

$$\text{or}\qquad\big|\lambda_l\big|\le\mathcal{E}_{\mathsf{MD}} \qquad (B.11b)$$

for any $1\le l\le r$. To see this, we recall that for any symmetric matrix $\boldsymbol{A}$, one has

$$\min_i\big|\lambda_i(\boldsymbol{A})\big|=\sqrt{\lambda_{\min}(\boldsymbol{A}^2)}=\sqrt{\min_{\boldsymbol{x}\in\mathbb{S}^{n-1}}\boldsymbol{x}^\top\boldsymbol{A}^2\boldsymbol{x}}=\min_{\boldsymbol{x}\in\mathbb{S}^{n-1}}\|\boldsymbol{A}\boldsymbol{x}\|_2,$$

where $\mathbb{S}^{n-1}:=\{\boldsymbol{z}\in\mathbb{R}^n\mid\|\boldsymbol{z}\|_2=1\}$ and $\lambda_i(\boldsymbol{A})$ denotes the i-th largest eigenvalue of $\boldsymbol{A}$. Recall the definition of $\boldsymbol{M}_\lambda$ in (5.21a). Given that the eigenvalues of $\lambda_l\boldsymbol{I}-\boldsymbol{M}_{\lambda_l}$ are exactly $\lambda_l-\lambda_i(\boldsymbol{M}_{\lambda_l})$ $(1\le i\le n)$ and that $\boldsymbol{u}_{l,\|}$ is a unit vector, we obtain

$$\min_{1\le i\le n}\big|\lambda_l-\lambda_i(\boldsymbol{M}_{\lambda_l})\big|=\min_{1\le i\le n}\big|\lambda_i\big(\lambda_l\boldsymbol{I}-\boldsymbol{M}_{\lambda_l}\big)\big|\le\big\|(\lambda_l\boldsymbol{I}_n-\boldsymbol{M}_{\lambda_l})\boldsymbol{u}_{l,\|}\big\|_2\le\mathcal{E}_{\mathsf{MD}}.$$

This immediately establishes (B.11), since the set of eigenvalues of $\boldsymbol{M}_{\lambda_l}$ is $\{\lambda_i^\star+\gamma(\lambda_l)\mid1\le i\le r\}\cup\{0\}$ (in view of the definition (5.21a)).

It thus boils down to how to use (B.11) to establish the advertised claim (5.23). Towards this, we find it helpful to define

$$\boldsymbol{M}(t):=\boldsymbol{M}^\star+t\boldsymbol{H}, \qquad (B.12)$$

$$\gamma(\lambda,t):=t^2\sigma^2\mathsf{tr}\Big(\big(\lambda\boldsymbol{I}_{n-r}-t(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)^{-1}\Big). \qquad (B.13)$$

We denote by $\{\lambda_{i,t}\}_{i=1}^n$ the eigenvalues of $\boldsymbol{M}(t)$ obeying $|\lambda_{1,t}| \geq \cdots \geq |\lambda_{n,t}|$; in other words, $\lambda_{1,t}, \cdots, \lambda_{r,t}$ correspond to the r eigenvalues of $\boldsymbol{M}(t)$ with the largest magnitudes. Armed with this notation, we clearly have

$$\lambda_{l,1} = \lambda_l, \qquad 1 \leq l \leq r.$$

The subsequent analysis consists of three steps.

- First, we establish the correspondence between $\{\lambda_{l,t} \mid 1 \leq l \leq r\}$ and $\{\lambda_i^\star \mid 1 \leq i \leq r\}$ through the following lemma; the proof is postponed to Appendix B.3.

  Lemma 12. *Instate the assumptions of Lemma 1. Then with probability exceeding $1 - O(n^{-10})$, for any $1 \leq l \leq r$, one can find $1 \leq i \leq r$ such that*

  $$\sup_{t \in [1/\sqrt{n},\, 1]} \left|\lambda_{l,t} - \gamma(\lambda_{l,t}, t) - \lambda_i^\star\right| \leq \mathcal{E}_t,$$

  where $\mathcal{E}_t := C_1 t\sigma\sqrt{r}\log n$ for some constant $C_1 > 0$ large enough.

  In other words, this lemma reveals that for all $1/\sqrt{n} \leq t \leq 1$, one has

  $$\lambda_{l,t} - \gamma(\lambda_{l,t}, t) \in \cup_{i=1}^r \mathcal{B}_{\mathcal{E}_t}(\lambda_i^\star),$$

  where $\mathcal{B}_\tau(\lambda) := \{z \mid |z - \lambda| \leq \tau\}$ denotes the ball of radius τ centered at λ.

- Secondly, when $0 \leq t \leq 1/\sqrt{n}$, one has $\|t\boldsymbol{H}\| \lesssim c_0/\sqrt{n} \cdot (\sigma\sqrt{n}) \leq \sigma/2$, where $c_0 > 0$ is some sufficiently small constant. In this scenario, Weyl's inequality tells us that $|\lambda_{l,t} - \lambda_l^\star| \leq \|t\boldsymbol{H}\| \leq \sigma/2$. Further, the definition of $\gamma(\cdot, \cdot)$ indicates that

  $$\left|\gamma(\lambda_{l,t}, t)\right| \leq t^2\sigma^2 \frac{n - r}{|\lambda_{l,t}| - \|t\boldsymbol{H}\|} \lesssim \frac{1}{n} \cdot \sigma^2 \frac{n}{\lambda_{\min}^\star} = \frac{\sigma^2}{\lambda_{\min}^\star} \leq \sigma/2,$$

  where the last inequality holds due to the assumption $\sigma\sqrt{n} \lesssim \lambda_{\min}^\star$. As a result,

  $$\begin{aligned} &|\lambda_{l,t} - \gamma(\lambda_{l,t}, t) - \lambda_l^\star| \leq |\lambda_{l,t} - \lambda_l^\star| + |\gamma(\lambda_{l,t}, t)| \leq \sigma \lesssim \mathcal{E}_{1/\sqrt{n}} \\ \Longrightarrow \qquad &\lambda_{l,t} - \gamma(\lambda_{l,t}, t) \in \mathcal{B}(\lambda_l^\star, \mathcal{E}_{1/\sqrt{n}}), \qquad 0 \leq t \leq 1/\sqrt{n}. \end{aligned} \tag{B.14}$$

- Recognizing that the set of eigenvalues $\lambda_{l,t}$ $(1 \leq l \leq r)$ depends continuously on t [Embree and Trefethen, 2001, Theorem 6], we know that $\lambda_{l,t} - \gamma(\lambda_{l,t}, t)$ is also a continuous function in t. In addition, for any $1 \leq l \leq r$, if $\min_{k:k\neq l}\left|\lambda_l^\star - \lambda_k^\star\right| > 2\mathcal{E}_1 \geq 2\mathcal{E}_t$ $(1/\sqrt{n} \leq t \leq 1)$, then one necessarily has

  $$\mathcal{B}_{\mathcal{E}_t}(\lambda_l^\star) \cap \{\cup_{k:k>l}\mathcal{B}_{\mathcal{E}_t}(\lambda_k^\star)\} = \emptyset \qquad \text{and} \qquad \mathcal{B}_{\mathcal{E}_t}(\lambda_l^\star) \cap \{\cup_{k:k<l}\mathcal{B}_{\mathcal{E}_t}(\lambda_k^\star)\} = \emptyset.$$

  In other words, $\mathcal{B}_{\mathcal{E}_t}(\lambda_l^\star)$ remains an isolated region within the set $\cup_{i=1}^r \mathcal{B}_{\mathcal{E}_t}(\lambda_i^\star)$ when we increase t from $1/\sqrt{n}$ to 1. This together with the above two facts (namely, the continuity of $\lambda_{l,t} - \gamma(\lambda_{l,t}, t)$ in t and (B.14)) requires that

  $$\lambda_{l,t} - \gamma(\lambda_{l,t}, t) \in \mathcal{B}_{\mathcal{E}_t}(\lambda_l^\star), \qquad 1/\sqrt{n} \leq t \leq 1,$$

  provided that $\min_{k:k\neq l}\left|\lambda_l^\star - \lambda_k^\star\right| > 2\mathcal{E}_1$.

Given that our notation satisfies $\lambda_{l,1} = \lambda_l$, $\gamma(\lambda_{l,1}, 1) = \gamma(\lambda_l)$, and $\mathcal{E}_1 = \mathcal{E}_{\mathsf{MD}}$, we conclude that with probability at least $1 - O(n^{-10})$,

$$\left|\lambda_l - \gamma(\lambda_l) - \lambda_l^\star\right| \leq \mathcal{E}_{\mathsf{MD}}, \qquad 1 \leq l \leq r. \tag{B.15}$$

B.3 Proof of Lemma 12

Fix an arbitrary $1 \le l \le r$. We have already shown in (B.11) that the claim holds when $t = 1$. An inspection of the proof of (B.11) reveals that: Lemma 12 can be established using the same argument, except that we need to generalize the bound (5.19) into a uniform bound on $\|\boldsymbol{G}(\lambda, t) - \boldsymbol{G}^{\perp}(\lambda, t)\|$, namely,

$$\|\boldsymbol{G}(\lambda, t) - \boldsymbol{G}^{\perp}(\lambda, t)\| \lesssim \frac{t\sigma^2}{\lambda_{\min}^{\star}} \sqrt{rn} \log n$$

holds simultaneously for all $1/\sqrt{n} \le t \le 1$ and λ with $|\lambda| \in [2|\lambda_l^{\star}|/3,\, 4|\lambda_l^{\star}|/3]$. Towards this end, we shall resort to the epsilon-net argument once again. Choose $\epsilon_1 = c/\sqrt{n}$ for some sufficiently small constant $c > 0$, and let $\mathcal{N}_{\epsilon_1}$ be an ϵ-net for $[1/\sqrt{n}, 1]$ such that (1) it has cardinality $|\mathcal{N}_{\epsilon_1}| \lesssim \sqrt{n}$; (2) for any $t \in [1/\sqrt{n}, 1]$, there exists some point $\widehat{t} \in \mathcal{N}_{\epsilon_1}$ obeying $|\widehat{t} - t| \le \epsilon_1$.

- Applying Lemma 1 with the noise matrix chosen as $t\boldsymbol{H}$ and applying the union bound, we see that with probability exceeding $1 - O\left(n^{-11}\right)$, one has

  $$\sup_{\lambda:\,|\lambda| \in [2|\lambda_l^{\star}|/3,\, 4|\lambda_l^{\star}|/3]} \|\boldsymbol{G}(\lambda, \widehat{t}) - \boldsymbol{G}^{\perp}(\lambda, \widehat{t})\| \lesssim \frac{\widehat{t}^2\sigma^2}{\lambda_{\min}^{\star}} \left(\sqrt{rn\log n} + r\log n\right) \le \frac{\widehat{t}\sigma^2}{\lambda_{\min}^{\star}} \sqrt{rn} \log n$$

  simultaneously for all $\widehat{t} \in \mathcal{N}_{\epsilon_1}$, where in the second line we have used $\widehat{t}^2 \le \widehat{t}$ since $\widehat{t} \in [0, 1]$.

- For any $t \in [1/\sqrt{n}, 1]$, let $\widehat{t} \in \mathcal{N}_{\epsilon_1}$ be a point obeying $|\widehat{t} - t| \le \epsilon_1$. Recognizing that $\boldsymbol{G}(\lambda, \widehat{t}) - \boldsymbol{G}(\lambda, t) \overset{\mathrm{d}}{=} t^2 \boldsymbol{H}_{\mathsf{ur}}\left(\lambda \boldsymbol{I}_{n-r} - t\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top} - \widehat{t}^2 \boldsymbol{H}_{\mathsf{ur}}\left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top}$, one can bound

  $$\begin{aligned}
  \|\boldsymbol{G}(\lambda, \widehat{t}) - \boldsymbol{G}(\lambda, t)\| &\le \left\|\widehat{t}^2 \boldsymbol{H}_{\mathsf{ur}}\left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top} - t^2 \boldsymbol{H}_{\mathsf{ur}}\left(\lambda \boldsymbol{I}_{n-r} - t\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top}\right\| \\
  &\le \left\|\widehat{t}^2 \boldsymbol{H}_{\mathsf{ur}}\left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top} - t^2 \boldsymbol{H}_{\mathsf{ur}}\left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top}\right\| \\
  &\quad + \left\|t^2 \boldsymbol{H}_{\mathsf{ur}}\left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top} - t^2 \boldsymbol{H}_{\mathsf{ur}}\left(\lambda \boldsymbol{I}_{n-r} - t\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} \boldsymbol{H}_{\mathsf{ur}}^{\top}\right\| \\
  &\le |t - \widehat{t}| \cdot |t + \widehat{t}| \cdot \|\boldsymbol{H}_{\mathsf{ur}}\|^2 \left\|\left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1}\right\| + t^2 \|\boldsymbol{H}_{\mathsf{ur}}\|^2 \left\|\left(\lambda \boldsymbol{I}_{n-r} - t\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} - \left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1}\right\|.
  \end{aligned}$$

  Recalling the notation that $\gamma_1 \ge \cdots \ge \gamma_{n-r}$ represent the eigenvalues of $\boldsymbol{H}_{\mathsf{lr}}$, we have

  $$\left\|\left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1}\right\| = \max_i \left|\frac{1}{\lambda - \widehat{t}\gamma_i}\right| \lesssim \frac{1}{\lambda_{\min}^{\star}} \qquad \text{and}$$

  $$\begin{aligned}
  \left\|\left(\lambda \boldsymbol{I}_{n-r} - t\boldsymbol{H}_{\mathsf{lr}}\right)^{-1} - \left(\lambda \boldsymbol{I}_{n-r} - \widehat{t}\boldsymbol{H}_{\mathsf{lr}}\right)^{-1}\right\| &= \max_i \left|\frac{1}{\lambda - t\gamma_i} - \frac{1}{\lambda - \widehat{t}\gamma_i}\right| = \max_i \left|\frac{(t - \widehat{t})\gamma_i}{(\lambda - t\gamma_i)(\lambda - \widehat{t}\gamma_i)}\right| \\
  &\lesssim \frac{|t - \widehat{t}|}{\lambda_{\min}^{\star}} \le \frac{\epsilon_1}{\lambda_{\min}^{\star}},
  \end{aligned}$$

  where we have used the bounds $2|\lambda_l^{\star}|/3 \le |\lambda| \le 4|\lambda_l^{\star}|/3$, $|\gamma_i| \le \lambda_{\min}^{\star}/3$ and $\widehat{t} \le t + \epsilon_1 \le 1.1$. Combining these with the high-probability bound $\|\boldsymbol{H}_{\mathsf{ur}}\| \lesssim \sigma\sqrt{n}$, we arrive at

  $$\|\boldsymbol{G}(\lambda, \widehat{t}) - \boldsymbol{G}(\lambda, t)\| \lesssim \epsilon_1 \cdot t \cdot \sigma^2 n \cdot \frac{1}{\lambda_{\min}^{\star}} + t^2 \cdot \sigma^2 n \cdot \frac{\epsilon_1}{\lambda_{\min}^{\star}} \lesssim \frac{t\sigma^2}{\lambda_{\min}^{\star}} \sqrt{n},$$

  where the last step arises since $t^2 \le t$ for any $t \in [0, 1]$. Similarly, this bound holds for $\|\boldsymbol{G}^{\perp}(\lambda, \widehat{t}) - \boldsymbol{G}^{\perp}(\lambda, t)\|$ as well.

Putting these two upper bounds together, we conclude that with probability at least $1 - O(n^{-11})$,

$$\begin{aligned}
\|\boldsymbol{G}(\lambda, t) - \boldsymbol{G}^{\perp}(\lambda, t)\| &\le \|\boldsymbol{G}(\lambda, \widehat{t}) - \boldsymbol{G}^{\perp}(\lambda, \widehat{t})\| + \|\boldsymbol{G}(\lambda, \widehat{t}) - \boldsymbol{G}(\lambda, t)\| + \|\boldsymbol{G}^{\perp}(\lambda, \widehat{t}) - \boldsymbol{G}^{\perp}(\lambda, t)\| \\
&\lesssim \frac{\widehat{t}\sigma^2}{\lambda_{\min}^{\star}} \sqrt{rn} \log n + \frac{t\sigma^2}{\lambda_{\min}^{\star}} \sqrt{n} \asymp \frac{t\sigma^2}{\lambda_{\min}^{\star}} \sqrt{rn} \log n
\end{aligned}$$

holds simultaneously for all λ with $|\lambda| \in [2|\lambda_l^{\star}|/3,\, 4|\lambda_l^{\star}|/3]$ and all $t \in [1/\sqrt{n}, 1]$. Finally, taking a union bound over $1 \le l \le r$ concludes the proof.

C Proofs of auxiliary lemmas for Theorem 1

C.1 Proof of Lemma 3

To begin with, let us first analyze the eigenvalues of $\boldsymbol{M}^{(l)}$, which is accomplished by the following lemma.

Lemma 13. *Instate the assumptions of Theorem 1. With probability at least $1 - O(n^{-10})$, one has*

$$\big|\lambda_k^{(l)} - \gamma(\lambda_k^{(l)}) - \lambda_k^\star\big| \le \mathcal{E}_{\mathsf{MD}}, \qquad 1 \le k < l, \tag{C.1a}$$

$$\big|\lambda_k^{(l)} - \gamma(\lambda_k^{(l)}) - \lambda_{k+1}^\star\big| \le \mathcal{E}_{\mathsf{MD}}, \qquad l \le k < r, \tag{C.1b}$$

$$\big|\lambda_k^{(l)}\big| \le \|\boldsymbol{H}\| \lesssim \sigma\sqrt{n}, \qquad k \ge r, \tag{C.1c}$$

where $\mathcal{E}_{\mathsf{MD}} = C_1\sigma\sqrt{r}\log n$ for some sufficiently large constant $C_1 > 0$ and $\gamma(\cdot)$ is defined in (5.45).

Proof. See Appendix C.1.1. □

Lemma 13 can then be invoked to study Lemma 3. Recalling the fact

$$\begin{aligned} \lambda_k - \gamma(\lambda_k) &\in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_k^\star), \qquad && 1 \le k \le r \\ |\lambda_k| &\le \|\boldsymbol{H}\| \lesssim \sigma\sqrt{n}, \qquad && k > r \end{aligned} \tag{C.2}$$

as shown in Theorem 7, we are positioned to prove the claim (5.33) as follows.

- For any λ such that $|\lambda| \lesssim \sigma\sqrt{n}$, one has

$$\begin{aligned} \big|\lambda_l - \lambda\big| &\ge \big|\lambda_l - \gamma(\lambda_l)\big| - \big|\gamma(\lambda_l)\big| - |\lambda| \overset{\text{(i)}}{\ge} \big|\lambda_l^\star\big| - \mathcal{E}_{\mathsf{MD}} - \big|\gamma(\lambda_l)\big| - O(\sigma\sqrt{n}) \\ &\overset{\text{(ii)}}{\ge} \big|\lambda_l^\star\big| - O(\sigma\sqrt{r}\log n) - O\Big(\frac{\sigma^2 n}{\lambda_{\min}^\star}\Big) - O(\sigma\sqrt{n}) \overset{\text{(iii)}}{\gtrsim} \big|\lambda_l^\star\big|, \end{aligned}$$

  where (i) arises from (C.2) and (C.1c), (ii) follows since

$$\big|\gamma(\lambda_l)\big| = \left|\sum_i \frac{\sigma^2}{\lambda_l - \lambda_i\big((\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)}\right| \lesssim \frac{\sigma^2 n}{|\lambda_l| - \|\boldsymbol{H}\|} \lesssim \frac{\sigma^2 n}{\lambda_{\min}^\star}, \tag{C.3}$$

  and (iii) is valid as long as $\sigma\sqrt{r}\log n \le c_0\lambda_{\min}^\star$ and $\sigma\sqrt{n} \le c_0\lambda_{\min}^\star$ hold for some small constant $c_0 > 0$.

- For any λ satisfying $\lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_k^\star)$ for some $1 \le k \le r$, we define an auxiliary function $f : \pm[2\lambda_{\min}^\star/3,\, 4\lambda_{\max}^\star/3] \to \mathbb{R}$ by

$$f(\lambda) := \lambda - \gamma(\lambda), \tag{C.4}$$

  where we denote $\pm[a, b] = [-b, -a] \cup [a, b]$ for $a < b$ and $\gamma(\cdot)$ is defined in (5.45). To begin with, for any λ with $|\lambda| \in [\lambda_{\min}^\star/3,\, 2\lambda_{\max}^\star]$ one has

$$\begin{aligned} f'(\lambda) &= 1 + \sum_i \frac{\sigma^2}{\big[\lambda - \lambda_i\big((\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H}\boldsymbol{U}^{\star\perp}\big)\big]^2} \ge 1 - \frac{\sigma^2 n}{(\frac{1}{3}\lambda_{\min} - \|\boldsymbol{H}\|)^2} \ge \frac{1}{2}, \\ f'(\lambda) &\le 1 + \frac{\sigma^2 n}{(\frac{1}{3}\lambda_{\min} - \|\boldsymbol{H}\|)^2} \le \frac{3}{2}, \end{aligned}$$

  with the proviso that $\sigma\sqrt{n} \le c_0\lambda_{\min}^\star$ for some constant $c_0 > 0$ small enough. This means that within the range $|\lambda| \in [2\lambda_{\min}^\star/3,\, 4\lambda_{\max}^\star/3]$, the function $f(\cdot)$ is monotonically increasing and continuous. As a result, the inverse of $f(\cdot)$ exists, which is also monotonically increasing and obeys

$$\frac{2}{3} \le \frac{\mathrm{d}f^{-1}(\tau)}{\mathrm{d}\tau} \le 2 \qquad \forall\tau \text{ with } |\tau| \in [\lambda_{\min}^\star/2, 3\lambda_{\max}^\star/2]. \tag{C.5}$$

In view of (5.23) and the condition that $\lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_k^\star)$ for some $1 \le k \le r$, one can invoke (C.5) to reach

$$\begin{aligned}
\left|\lambda_l - \lambda\right| &= \left|f^{-1}(\lambda_l^\star + O(\mathcal{E}_{\mathsf{MD}})) - f^{-1}(\lambda_k^\star + O(\mathcal{E}_{\mathsf{MD}}))\right| \\
&\ge \inf_{\tau:\, |\tau| \in [\lambda_{\min}^\star/2,\, 3\lambda_{\max}^\star/2]} \left|\frac{\mathrm{d}f^{-1}(\tau)}{\mathrm{d}\tau}\right| \left|\lambda_l^\star - \lambda_k^\star + O(\mathcal{E}_{\mathsf{MD}})\right| \\
&\ge \frac{2}{3}\left|\lambda_l^\star - \lambda_k^\star + O(\mathcal{E}_{\mathsf{MD}})\right| \gtrsim \left|\lambda_l^\star - \lambda_k^\star\right|,
\end{aligned}$$

where the last inequality holds due to our eigen-gap assumption (3.3) and the fact that $\mathcal{E}_{\mathsf{MD}} \asymp \sigma\sqrt{r}\log n$.

- With the analysis above, we note that (5.33) is an immediate consequence of (C.1) in Lemma 13.

C.1.1 Proof of Lemma 13

The proof of this lemma follows from the same argument employed to establish Theorem 7. The idea is to invoke Theorem 6 to analyze the spectrum of $\boldsymbol{M}^{(l)}$. Before proceeding, we introduce several notation tailored to this setting as well as a few simple facts. To begin with, we define

$$\begin{aligned}
\boldsymbol{M}^{\star(l)} &:= (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{M}^\star \boldsymbol{u}_l^{\star\perp} = (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{U}^\star \boldsymbol{\Lambda}^\star \boldsymbol{U}^{\star\top} \boldsymbol{u}_l^{\star\perp} = \boldsymbol{U}^{\star(l)} \boldsymbol{\Lambda}^{\star(l)} \boldsymbol{U}^{\star(l)\top}, \\
\boldsymbol{H}^{(l)} &:= (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H} \boldsymbol{u}_l^{\star\perp},
\end{aligned}$$

where we recall the definitions of $\boldsymbol{u}_l^{\star\perp}$ (resp. $\boldsymbol{U}^{\star(l)}$ and $\boldsymbol{\Lambda}^{\star(l)}$) in (5.13) (resp. (5.32)). In addition, denote

$$\begin{aligned}
\boldsymbol{G}^{(l)}(\lambda) &:= \boldsymbol{U}^{\star(l)\top} \boldsymbol{H}^{(l)} \boldsymbol{U}^{\star(l)\perp} \big(\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star(l)\perp})^\top \boldsymbol{H}^{(l)} \boldsymbol{U}^{\star(l)\perp}\big)^{-1} (\boldsymbol{U}^{\star(l)\perp})^\top \boldsymbol{H}^{(l)} \boldsymbol{U}^{\star(l)}, \\
\boldsymbol{G}^{(l)\perp}(\lambda) &:= \mathbb{E}\big[\boldsymbol{G}^{(l)}(\lambda) \mid (\boldsymbol{U}^{\star(l)\perp})^\top \boldsymbol{H}^{(l)} \boldsymbol{U}^{\star(l)\perp}\big],
\end{aligned}$$

where $\boldsymbol{U}^{\star(l)\perp}$ is defined in (5.32) and the expectation is taken assuming that λ is independent of $\boldsymbol{H}$. By construction, one has $\boldsymbol{u}_l^{\star\perp}\boldsymbol{U}^{\star(l)\perp} = \boldsymbol{U}^{\star\perp}$, and consequently

$$\begin{aligned}
(\boldsymbol{U}^{\star(l)\perp})^\top \boldsymbol{H}^{(l)} \boldsymbol{U}^{\star(l)\perp} &= (\boldsymbol{U}^{\star(l)\perp})^\top (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H} \boldsymbol{u}_l^{\star\perp} \boldsymbol{U}^{\star(l)\perp} = (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H} \boldsymbol{U}^{\star\perp}, \\
(\boldsymbol{U}^{\star(l)\perp})^\top \boldsymbol{H}^{(l)} \boldsymbol{U}^{\star(l)} &= (\boldsymbol{U}^{\star(l)\perp})^\top (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H} \boldsymbol{u}_l^{\star\perp} \boldsymbol{U}^{\star(l)} = (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H} \boldsymbol{U}^\star \boldsymbol{P}^{(l)},
\end{aligned} \tag{C.6}$$

where $\boldsymbol{P}^{(l)}$ is obtained by removing the l-th column of $\boldsymbol{I}_r$, namely,

$$\boldsymbol{P}^{(l)} := \begin{bmatrix} \boldsymbol{I}_{l-1} & \boldsymbol{0} \\ 0 & 0 \\ \boldsymbol{0} & \boldsymbol{I}_{r-l} \end{bmatrix} \in \mathbb{R}^{r\times(r-1)}.$$

Therefore, $\boldsymbol{G}^{(l)}(\lambda)$ and $\boldsymbol{G}^{(l)\perp}(\lambda)$ admit the following simplified expressions

$$\begin{aligned}
\boldsymbol{G}^{(l)}(\lambda) &= \boldsymbol{U}^{\star\top} \boldsymbol{u}_l^{\star\perp} (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H} \boldsymbol{U}^{\star\perp} \big(\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H} \boldsymbol{U}^{\star\perp}\big)^{-1} (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H} \boldsymbol{u}_l^{\star\perp} (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{U}^\star \\
&= \boldsymbol{P}^{(l)\top} \boldsymbol{U}^{\star\top} \boldsymbol{H} \boldsymbol{U}^{\star\perp} \big(\lambda \boldsymbol{I}_{n-r} - (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H} \boldsymbol{U}^{\star\perp}\big)^{-1} (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H} \boldsymbol{U}^\star \boldsymbol{P}^{(l)} \\
&= \boldsymbol{P}^{(l)\top} \boldsymbol{G}(\lambda) \boldsymbol{P}^{(l)} \\
\boldsymbol{G}^{(l)\perp}(\lambda) &= \mathbb{E}\big[\boldsymbol{G}^{(l)}(\lambda) \mid (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{H} \boldsymbol{U}^{\star\perp}\big],
\end{aligned}$$

where $\boldsymbol{G}(\lambda)$ is defined in (5.14b).

With the above preparation in place, we can repeat the proof of Theorem 7 to obtain

$$\begin{aligned}
\big\|\boldsymbol{U}^{\star(l)\top} \boldsymbol{H}^{(l)} \boldsymbol{U}^{\star(l)}\big\| &\lesssim \sigma\big(\sqrt{r} + \sqrt{\log n}\big) \\
\sup_{\lambda:\, |\lambda| \in [2\lambda_{\min}^\star/3,\, 4\lambda_{\max}^\star/3]} \big\|\boldsymbol{G}^{(l)}(\lambda) - \boldsymbol{G}^{(l)\perp}(\lambda)\big\| &\lesssim \frac{\sigma^2}{\lambda_{\min}^\star}\big(\sqrt{rn\log n} + r\log n\big) \\
\boldsymbol{G}^{(l)\perp}(\lambda) &= \gamma(\lambda) \boldsymbol{P}^{(l)\top} \boldsymbol{U}^\star \boldsymbol{U}^{\star\top} \boldsymbol{P}^{(l)}
\end{aligned}$$

with probability at least $1 - O(n^{-10})$, where $\gamma(\cdot)$ is defined in (5.45). The above observations reveal that the k-th eigenvalue of $\boldsymbol{M}^{\star(l)} + \boldsymbol{G}^{(l)\perp}(\lambda)$ is given by

$$\lambda_k\big(\boldsymbol{M}^{\star(l)} + \boldsymbol{G}^{(l)\perp}(\lambda)\big) = \begin{cases} \lambda_k^\star + \gamma(\lambda), & 1 \le k \le l-1; \\ \lambda_{k+1}^\star + \gamma(\lambda), & l \le k \le r-1; \\ 0, & r \le k. \end{cases}$$

As a result, repeating the same arguments of Theorem 7 (which we omit for brevity) immediately establishes the claim of this lemma.

C.2 Proof of Lemma 4

Let $\boldsymbol{U}^{(l)}\boldsymbol{\Lambda}^{(l)}\boldsymbol{U}^{(l)\top}$ represent the eigen-decomposition of $\boldsymbol{M}^{(l)}$, where $\boldsymbol{U}^{(l)} = \big[\boldsymbol{u}_1^{(l)}, \cdots, \boldsymbol{u}_{n-1}^{(l)}\big]$. We can derive

$$\begin{aligned} \big\|\big(\lambda_l \boldsymbol{I}_{n-1} - \boldsymbol{M}^{(l)}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\big\|_2^2 &= \big\|\boldsymbol{U}^{(l)}\big(\lambda_l \boldsymbol{I}_{n-1} - \boldsymbol{\Lambda}^{(l)}\big)^{-1}\boldsymbol{U}^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\big\|_2^2 \\ &= \big\|\big(\lambda_l \boldsymbol{I}_{n-1} - \boldsymbol{\Lambda}^{(l)}\big)^{-1}\boldsymbol{U}^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star\big\|_2^2 \\ &= \sum_{1\le k<n}\left(\frac{\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l - \lambda_k^{(l)}}\right)^2. \end{aligned} \tag{C.7}$$

By construction (cf. (5.29)), the matrix $\boldsymbol{M}^{(l)}$ (and hence $\boldsymbol{\Lambda}^{(l)}$ and $\boldsymbol{U}^{(l)}$) is independent of $(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star$, thus indicating that

$$\boldsymbol{U}^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_{n-1}).$$

In addition, notice that the distribution of $\boldsymbol{U}^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star$ is independent with $\boldsymbol{U}^{(l)}$ and $\boldsymbol{\Lambda}^{(l)}$. In what follows, we shall look at (C.7) by controlling the sum over $k < r$ and the sum over $k \ge r$ separately.

- To begin with, let us upper bound $\sum_{1\le k<r}\Big(\frac{\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l - \lambda_k^{(l)}}\Big)^2$. Given a sequence of i.i.d. standard Gaussian random variables $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, one knows from the standard Gaussian concentration inequality that the following holds with probability at least $1 - O(n^{-20})$:

$$\max_{1\le i\le n} |Z_i| \lesssim \sigma\sqrt{\log n}; \tag{C.8a}$$

$$\max_{1\le i\le n} |Z_i^2 - \sigma^2| \le \max_{1\le i\le n} |Z_i - \sigma| \cdot \max_{1\le i\le n} |Z_i + \sigma| \lesssim \sigma^2 \log n. \tag{C.8b}$$

  In addition, Lemma 3 tells us that $\min_{1\le k<r}\big|\lambda_l - \lambda_k^{(l)}\big|^2 \gtrsim \min_{i:i\neq l}|\lambda_l^\star - \lambda_i^\star|^2$. These two bounds taken together give

$$\sum_{1\le k<r}\left(\frac{\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l - \lambda_k^{(l)}}\right)^2 \lesssim \frac{\sigma^2 r \log n}{\min_{i:i\neq l}|\lambda_l^\star - \lambda_i^\star|^2} = \frac{\sigma^2 r\log n}{(\Delta_l^\star)^2}. \tag{C.9}$$

- Next, we move on to the remaining term (the sum over $r \le k < n$). We claim for the moment that: with probability exceeding $1 - O(n^{-10})$,

$$\left|\sum_{r\le k<n}\left(\frac{\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l - \lambda_k^{(l)}}\right)^2 - \sum_{r\le k<n}\frac{\sigma^2}{(\lambda_l - \lambda_k^{(l)})^2}\right| \lesssim \frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{n\log n}, \tag{C.10}$$

  The proof of this claim is deferred to the end of this section. It then suffices to control the term $\sum_{r\le k<n}\sigma^2/(\lambda_l - \lambda_k^{(l)})^2$, which is established in the lemma below (with the proof postponed to Appendix C.2.1).

  Lemma 14. *Instate the assumptions of Theorem 1. With probability at least $1 - O(n^{-10})$,*

$$\left|\sum_{r\le k\le n-1}\frac{1}{(\lambda_l - \lambda_k^{(l)})^2} - \sum_{r+1\le k\le n}\frac{1}{(\lambda_l - \lambda_k)^2}\right| \lesssim \frac{1}{\lambda_l^{\star 2}}. \tag{C.11}$$

As a consequence, we have

$$\Big|\sum_{r\le k\le n-1}\frac{1}{(\lambda_l-\lambda_k^{(l)})^2}\Big|\vee\Big|\sum_{r+1\le k\le n}\frac{1}{(\lambda_l-\lambda_k)^2}\Big|\lesssim\frac{n}{\lambda_l^{\star 2}}. \tag{C.12}$$

Therefore, combining (C.10) and (C.11) gives

$$\begin{aligned}
&\Big|\sum_{r\le k\le n-1}\bigg(\frac{\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l-\lambda_k^{(l)}}\bigg)^2-\sum_{r+1\le k\le n}\frac{\sigma^2}{(\lambda_l-\lambda_k)^2}\Big|\\
&\quad\lesssim\Big|\sum_{r\le k\le n-1}\bigg(\frac{\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l-\lambda_k^{(l)}}\bigg)^2-\sum_{r\le k<n}\frac{\sigma^2}{(\lambda_l-\lambda_k^{(l)})^2}\Big|+\Big|\sum_{r\le k<n}\frac{\sigma^2}{(\lambda_l-\lambda_k^{(l)})^2}-\sum_{r+1\le k\le n}\frac{\sigma^2}{(\lambda_l-\lambda_k)^2}\Big|\\
&\quad\lesssim\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{n\log n}+\frac{\sigma^2}{\lambda_l^{\star 2}}\asymp\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{n\log n}.
\end{aligned}\tag{C.13}$$

- Inserting (C.9) and (C.13) into (C.7), we arrive at the advertised bound:

$$\begin{aligned}
\big\|\big(\lambda_l\boldsymbol{I}_{n-1}-\boldsymbol{M}^{(l)}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big\|_2^2&=\sum_{r+1\le k\le n}\frac{\sigma^2}{(\lambda_l-\lambda_k)^2}+O\bigg(\frac{\sigma^2 r\log n}{(\Delta_l^\star)^2}+\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{n\log n}\bigg)\\
&\asymp\frac{\sigma^2 n}{\lambda_l^{\star 2}}+O\bigg(\frac{\sigma^2 r\log n}{(\Delta_l^\star)^2}\bigg)\ll 1,
\end{aligned}$$

where the last inequality holds due to our noise assumption (3.3).

Proof of the claim (C.10). Note that $\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,\sigma^2)$ is independent of $\boldsymbol{\Lambda}^{(l)}$ but depends on $\boldsymbol{\Lambda}$. Therefore, we shall use the epsilon-net argument (i.e. Lemma 20) to bound it. Before proceeding, observe that from (5.33), (C.3) and the condition $\sigma\sqrt{n}\ll\lambda_{\min}^\star$, the following holds for any λ obeying $\lambda-\gamma(\lambda)\in\mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)$:

$$|\lambda-\lambda_k^{(l)}|\ge|\lambda_l-\lambda_k^{(l)}|-|\lambda-\lambda_l-\gamma(\lambda_l)|-|\gamma(\lambda_l)|\gtrsim|\lambda_l^\star|,\qquad k\ge r. \tag{C.14}$$

Now we begin to check the conditions of Lemma 20. Since $|f(x)-f(y)|\le\sup_x|f'(x)||x-y|$, the following holds with probability at least $1-O(n^{-20})$ for all λ with $\lambda-\gamma(\lambda)\in\mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)$:

$$\begin{aligned}
\Big|\frac{\mathrm{d}}{\mathrm{d}\lambda}\sum_{r\le k<n}\frac{\big(\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big)^2-\sigma^2}{(\lambda-\lambda_k^{(l)})^2}\Big|&=\Big|\sum_{r\le k<n}\frac{\big(\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big)^2-\sigma^2}{(\lambda-\lambda_k^{(l)})^3}\Big|\\
&\le n\cdot\max_{r\le k<n}\frac{1}{\big|\lambda-\lambda_k^{(l)}\big|^3}\cdot\max_{r\le k<n}\big|\big(\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big)^2-\sigma^2\big|\\
&\lesssim\frac{\sigma^2 n\log n}{\lambda_l^{\star 3}},
\end{aligned}$$

where we use (C.8b) and (C.14). In addition, for any fixed λ obeying $\lambda-\gamma(\lambda)\in\mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)$, one has

$$\begin{aligned}
\max_{r\le k<n}\frac{1}{(\lambda-\lambda_k^{(l)})^2}\big\|\big(\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big)^2-\sigma^2\big\|_{\psi_1}&\lesssim\frac{\sigma^2}{\lambda_l^{\star 2}}=:L;\\
\sum_{r\le k<n}\frac{1}{(\lambda-\lambda_k^{(l)})^4}\mathbb{E}\Big[\Big(\big(\boldsymbol{u}_k^{(l)\top}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big)^2-\sigma^2\Big)^2\Big]&\lesssim\frac{\sigma^4 n}{\lambda_l^{\star 4}}=:V
\end{aligned}$$

where $\|\cdot\|_{\psi_1}$ denote the sub-exponential norm. We can then apply the matrix Bernstein inequality [Koltchinskii, 2011, Corollary 2.1] to find: with probability exceeding $1 - O(n^{-20})$,

$$\Bigg| \sum_{r \le k < n} \frac{\big(\boldsymbol{u}_k^{(l)\top} (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H} \boldsymbol{u}_l^\star\big)^2 - \sigma^2}{(\lambda - \lambda_k^{(l)})^2} \Bigg| \lesssim L \log^2 n + \sqrt{V \log n} \lesssim \frac{\sigma^2 \sqrt{n \log n}}{\lambda_l^{\star 2}}.$$

Recognizing the fact that $\{\lambda \colon \lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)\} \subset \big[\lambda_l^\star - |\lambda_l^\star|/3,\ \lambda_l^\star + |\lambda_l^\star|/3\big]$, the claim (C.14) immediately follows from Lemma 20.

C.2.1 Proof of Lemma 14

Let us look at (C.12) first. According to Lemma 3, we have $|\lambda_l - \lambda_k^{(l)}| \gtrsim |\lambda_l^\star|$ for all $k \ge r$, thus leading to

$$\Bigg| \sum_{r \le k \le n-1} \frac{1}{(\lambda_l - \lambda_k^{(l)})^2} \Bigg| \lesssim \frac{n}{\lambda_l^{\star 2}}.$$

In addition, the upper bound for $\sum_{r \le k \le n-1} 1/(\lambda_l - \lambda_k)^2$ is an immediate consequence of (C.11). Therefore, the remainder of the proof amounts to establishing (C.11), which requires us to characterize the relation between the spectrums of $\boldsymbol{M}^{(l)} = (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{M} \boldsymbol{u}_l^{\star\perp}$ and $\boldsymbol{M}$.

Without loss of generality, assume that $\lambda_l^\star > 0$, and that there are m (resp. $r - m$) eigenvalues of $\boldsymbol{M}^\star$ larger (resp. smaller) than 0. By Weyl's inequality (similar to (5.10) in the proof of Theorem 7), it is easily seen that there are m eigenvalues of $\boldsymbol{M}$ larger than $c\sigma\sqrt{n}$ and that there are $r - m$ eigenvalues of $\boldsymbol{M}$ smaller than $-c\sigma\sqrt{n}$, where $c > 0$ is some constant. Recalling that $\{\lambda_k\}_{k=1}^n$ are defined as the eigenvalues of $\boldsymbol{M}$ satisfying $|\lambda_1| \ge \cdots \ge |\lambda_n|$, we further denote by $\{\phi_k\}_{k=1}^n$ the eigenvalues of $\boldsymbol{M}$ so that $\phi_1 \ge \cdots \ge \phi_n$. Consider the set of eigenvalues of $\boldsymbol{M}$ with magnitudes upper bounded by $c\sigma\sqrt{n}$. We have the following relation:

$$\sum_{k \colon r+1 \le k \le n} \frac{1}{(\lambda_l - \lambda_k)^2} = \sum_{k \colon m+1 \le k \le n-r+m} \frac{1}{(\lambda_l - \phi_k)^2}. \tag{C.15}$$

Similarly, for $\boldsymbol{M}^{(l)}$ we can write

$$\sum_{k \colon r \le k \le n-1} \frac{1}{(\lambda_l - \lambda_k^{(l)})^2} = \sum_{k \colon m \le k \le n-r+m-1} \frac{1}{(\lambda_l - \phi_k^{(l)})^2}, \tag{C.16}$$

where $\{\phi_k^{(l)}\}_{k=1}^{n-1}$ denote the eigenvalues of $\boldsymbol{M}^{(l)}$ in descending order. As a result, in order to establish (C.11), it is sufficient to show

$$\Bigg| \sum_{k \colon m \le k \le n-r+m-1} \frac{1}{(\lambda_l - \phi_k^{(l)})^2} - \sum_{k \colon m+1 \le k \le n-r+m} \frac{1}{(\lambda_l - \phi_k)^2} \Bigg| \lesssim \frac{1}{\lambda_l^{\star 2}}. \tag{C.17}$$

In view of an eigenvalue interlacing result stated in Lemma 21, the definition $\boldsymbol{M}^{(l)} := (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{M} \boldsymbol{u}_l^{\star\perp}$ allows us to deduce that

$$\phi_{k+1} \le \phi_k^{(l)} \le \phi_k \qquad 1 \le k < n. \tag{C.18}$$

By the assumption $\lambda_l^\star > 0$, one has $\lambda_l \ge \lambda_l^\star - \|\boldsymbol{H}\| \gtrsim \lambda_l^\star$ and thus $\phi_k \le \lambda_l$ for all $k \ge m$. Consequently, we know from (C.18) that for all $k \ge m$,

$$\phi_{k+1} \le \phi_k^{(l)} \le \phi_k \le \lambda_l,$$

which further implies that

$$\frac{1}{(\lambda_l - \phi_{k+1})^2} \le \frac{1}{(\lambda_l - \phi_k^{(l)})^2} \le \frac{1}{(\lambda_l - \phi_k)^2}, \qquad k \ge m.$$

This enables us to bound

$$\sum_{k \colon m \le k \le n-r+m-1} \frac{1}{(\lambda_l - \phi_k^{(l)})^2} \le \frac{1}{(\lambda_l - \phi_m^{(l)})^2} + \sum_{k \colon m+1 \le k \le n-r+m-1} \frac{1}{(\lambda_l - \phi_k)^2};$$

$$\sum_{k:\, m\le k\le n-r+m-1}\frac{1}{(\lambda_l-\phi_k^{(l)})^2}\ge\sum_{k:\, m\le k\le n-r+m-1}\frac{1}{(\lambda_l-\phi_{k+1})^2}=\sum_{k:\, m+1\le k\le n-r+m}\frac{1}{(\lambda_l-\phi_k)^2}.$$

Consequently, we conclude that

$$\Big|\sum_{k:\, m\le k\le n-r+m-1}\frac{1}{(\lambda_l-\phi_k^{(l)})^2}-\sum_{k:\, m+1\le k\le n-r+m}\frac{1}{(\lambda_l-\phi_k)^2}\Big|\le\frac{1}{(\lambda_l-\phi_m^{(l)})^2}\asymp\frac{1}{\lambda_l^{\star 2}},$$

where the last relation holds since $|\phi_m^{(l)}|\le\|\boldsymbol{H}\|$ and hence $\lambda_l-\phi_m^{(l)}\asymp\lambda_l\asymp\lambda_l^\star$.

The above analysis can be easily adopted to handle the case where $\lambda_l^\star<0$ as well (which we omit here for brevity). Therefore, we have finished the proof.

C.3 Proof of Lemma 5

Since $(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\sim\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I}_{n-1})$ is a Gaussian random vector independent from $\boldsymbol{M}^{(l)}$ but dependent on λ_l, our proof strategy is to apply Lemma 20.

To this end, recall the definition

$$\boldsymbol{u}_k^{\star(l)}:=(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{u}_k^\star$$

and the eigen-decomposition of

$$\boldsymbol{M}^{(l)}=\boldsymbol{U}^{(l)}\boldsymbol{\Lambda}^{(l)}\boldsymbol{U}^{(l)\top}=\sum_{1\le i<n}\lambda_i^{(l)}\boldsymbol{u}_i^{(l)}\boldsymbol{u}_i^{(l)\top}.$$

To begin with, we claim that with probability at least $1-O(n^{-10})$:

$$V:=\sup_{\lambda-\gamma(\lambda)\in\mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)}\Big\|\sum_{k:k\ne l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star(l)\top}\big(\lambda\boldsymbol{I}-\boldsymbol{M}^{(l)}\big)^{-1}\Big\|_2\lesssim\sum_{k:k\ne l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|\sqrt{r}}{|\lambda_l^\star-\lambda_k^\star|}+\frac{1}{|\lambda_l^\star|}.\tag{C.19}$$

whose proof is postponed to the end of the section. Consequently, we can invoke standard Gaussian concentration inequalities to obtain: with probability at least $1-O\big(\kappa^{-10}(\lambda_{\max}/\Delta_l^\star)^{-20}n^{-20}\big)$,

$$\Big|(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\cdot\sum_{k:k\ne l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\big(\lambda\boldsymbol{I}-\boldsymbol{M}^{(l)}\big)^{-1}\Big|\lesssim\sigma\sqrt{\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}\cdot V.$$

In addition, we collect a basic fact regarding the derivatives of matrices: for any invertible matrix $\boldsymbol{A}$,

$$\frac{\mathrm{d}\boldsymbol{A}^{-1}}{\mathrm{d}x}=-\boldsymbol{A}^{-1}\frac{\mathrm{d}\boldsymbol{A}}{\mathrm{d}x}\boldsymbol{A}^{-1}.$$

With this identity in mind, one can derive: with probability at least $1-O(n^{-10})$, for all λ with $\lambda-\gamma(\lambda)\in\mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)$,

$$\begin{aligned}&\Big|\frac{\mathrm{d}}{\mathrm{d}\lambda}\sum_{k:k\ne l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\big(\lambda\boldsymbol{I}-\boldsymbol{M}^{(l)}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\Big|\\&\quad=\Big|\sum_{k:k\ne l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\big(\lambda\boldsymbol{I}-\boldsymbol{M}^{(l)}\big)^{-2}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\Big|\\&\quad\le n\cdot\max_{1\le i<n}\frac{1}{(\lambda-\lambda_i^{(l)})^2}\cdot\Big\|\sum_{k:k\ne l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\Big\|_2\cdot\big\|(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\big\|_2\\&\quad\overset{\text{(i)}}{\lesssim}n\cdot\frac{1}{\Delta_l^{\star2}\wedge\lambda_l^{\star2}}\cdot\sum_{k:k\ne l}|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|\,\big\|\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star\perp}\big\|_2\cdot\sigma\sqrt{n\log n}\end{aligned}$$

$$\overset{\text{(ii)}}{\leq} n^{3/2} \cdot \frac{\max_{i:i\neq l}|\lambda_l^\star - \lambda_i^\star|}{\Delta_l^{\star 2} \wedge \lambda_l^{\star 2}} \cdot \sigma\sqrt{\log n} \sum_{k:k\neq l} \frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{|\lambda_l^\star - \lambda_k^\star|}$$

$$\overset{\text{(iii)}}{\lesssim} n^{3/2} \cdot \left(\frac{\lambda_{\max}^{\star 2}}{\Delta_l^{\star 2}} + \kappa^2\right) \cdot \frac{1}{|\lambda_l^\star|} \cdot \sigma\sqrt{\log n} \sum_{k:k\neq l} \frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{|\lambda_l^\star - \lambda_k^\star|}$$

$$\overset{\text{(iv)}}{\lesssim} n^{3/2} \cdot \frac{\kappa^2 \lambda_{\max}^{\star 2}}{\Delta_l^{\star 2}} \cdot \frac{1}{|\lambda_l^\star|} \cdot V.$$

Here, (i) arises from (5.33) in Lemma 8 and the standard Gaussian concentration inequality; (ii) holds since $\left\|\boldsymbol{u}_k^{\star\top} \boldsymbol{u}_l^{\star\perp}\right\|_2 \leq \|\boldsymbol{u}_k^\star\|_2 \|\boldsymbol{u}_l^{\star\perp}\| \leq 1$; (iii) is due to $\max_{i:i\neq l}|\lambda_l^\star - \lambda_i^\star| \lesssim \lambda_{\max}^\star$; (iv) uses the definition of V in (C.19). Moreover, it is easy to see that $\{\lambda\colon \lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)\} \subset \left[\lambda_l^\star - |\lambda_l^\star|/3, \lambda_l^\star + |\lambda_l^\star|/3\right]$. As a consequence, we invoke Lemma 20 and the union bound to find: with probability at least $1 - O(n^{-10})$:

$$\begin{aligned}
\Big|\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \big(\lambda_l \boldsymbol{I} - \boldsymbol{M}^{(l)}\big)^{-1} (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H} \boldsymbol{u}_l^\star\Big| &\lesssim \sigma\sqrt{r \log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)} \cdot V \\
&\lesssim \sigma\sqrt{r \log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)} \sum_{k:k\neq l} \frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{|\lambda_l^\star - \lambda_k^\star|} + \frac{\sigma}{|\lambda_l^\star|}\sqrt{\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_l^\star}\right)}
\end{aligned}$$

as claimed.

The remainder of this section amounts to establishing (C.19), and we shall work under the event where Lemma 8 holds, which happens with probability at least $1 - O(n^{-10})$. Note that for any λ such that $\lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)$, one can express

$$\begin{aligned}
\Big\|\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \big(\lambda \boldsymbol{I} - \boldsymbol{M}^{(l)}\big)^{-1}\Big\|_2 &= \Big\|\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{U}^{(l)} \big(\lambda \boldsymbol{I} - \boldsymbol{\Lambda}^{(l)}\big)^{-1} \boldsymbol{U}^{(l)\top}\Big\|_2 \\
&= \Big\|\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{U}^{(l)} \big(\lambda \boldsymbol{I} - \boldsymbol{\Lambda}^{(l)}\big)^{-1}\Big\|_2 \\
&= \sqrt{\sum_{1\leq i<n} \left(\frac{1}{\lambda - \lambda_i^{(l)}} \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}\right)^2}. \qquad \text{(C.20)}
\end{aligned}$$

In what follows, we shall control the sum over $i < r$ and the sum over $i \geq r$ separately.

- Let us consider the sum over $i \geq r$ first. According to Lemma 3, we know that $|\lambda - \lambda_i^{(l)}| \gtrsim |\lambda_l^\star|$ for all $i \geq r$. This in turn yields

$$\begin{aligned}
\sqrt{\sum_{r\leq i\leq n-1} \left(\frac{1}{\lambda - \lambda_i^{(l)}} \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}\right)^2} &\lesssim \frac{1}{|\lambda_l^\star|} \sqrt{\sum_{r\leq i\leq n-1} \left(\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}\right)^2} \\
&\leq \frac{1}{|\lambda_l^\star|} \Big\|\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{U}^{(l)}\Big\|_2 \\
&\leq \frac{1}{|\lambda_l^\star|} \|\boldsymbol{a}\|_2 \cdot \Big\|\sum_{k:k\neq l} \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top}\Big\| \cdot \|\boldsymbol{U}^{(l)}\| = \frac{1}{|\lambda_l^\star|}. \qquad \text{(C.21)}
\end{aligned}$$

  Here, we make use of the fact that $\|\boldsymbol{a}\|_2 = 1$, $\|\boldsymbol{U}^{(l)}\| = 1$ as well as $\left\|\sum_{k:k\neq l} \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top}\right\| = 1$, since both $\{\boldsymbol{u}_k^\star\}$ and $\{\boldsymbol{u}_k^{\star(l)}\}_{k:k\neq l}$ form orthonormal bases.

- We then move on to the sum over the range $1 \leq i < r$. Given that $\{\boldsymbol{u}_i^{(l)}\}_i$ are orthonormal, it is straightforward to demonstrate that

$$\sqrt{\sum_{1\leq i<r} \left(\frac{1}{\lambda - \lambda_i^{(l)}} \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}\right)^2} = \Big\|\sum_{1\leq i<r} \left(\frac{1}{\lambda - \lambda_i^{(l)}} \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}\right) \boldsymbol{u}_i^{(l)}\Big\|_2$$

$$
\begin{aligned}
&= \Big\| \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \sum_{1\leq i<r} \frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}} \boldsymbol{u}_i^{(l)} \Big\|_2 \\
&\leq \sum_{k:k\neq l} \big|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\big| \cdot \Big\| \sum_{1\leq i<r} \frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}} \boldsymbol{u}_i^{(l)} \Big\|_2 \\
&= \sum_{k:k\neq l} \big|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\big| \sqrt{\sum_{1\leq i<r} \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}} \right)^2}. \qquad \text{(C.22)}
\end{aligned}
$$

The preceding inequality motivates us to control the quantity $\sum_{1\leq i<r} \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}} \right)^2$. Towards this, let us decompose it as follows

$$
\sum_{1\leq i<r} \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}} \right)^2 = \sum_{i\in\mathcal{A}_1} \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}} \right)^2 + \sum_{i\in\mathcal{A}_2} \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}} \right)^2.
$$

Here, the sets $\mathcal{A}_1$ and $\mathcal{A}_2$ are defined respectively by

$$
\begin{aligned}
\mathcal{A}_1 &:= \{1 \leq i < r \mid \lambda_i^{(l)} - \gamma(\lambda_i^{(l)}) \in \mathcal{B}_{\mathcal{E}_k}(\lambda_k^\star)\}, \\
\mathcal{A}_2 &:= \{1 \leq i < r \mid \lambda_i^{(l)} - \gamma(\lambda_i^{(l)}) \notin \mathcal{B}_{\mathcal{E}_k}(\lambda_k^\star)\},
\end{aligned}
$$

where $\mathcal{E}_k := c\,|\lambda_l^\star - \lambda_k^\star|$ for some sufficiently small constant $c > 0$. In the sequel, we shall control these two sums separately.

- For each $i \in \mathcal{A}_1$, we claim that

$$
|\lambda - \lambda_i^{(l)}| \geq \big|\lambda - f^{-1}(\lambda_k^\star)\big| - \big|f^{-1}(\lambda_k^\star) - \lambda_i^{(l)}\big| \geq \frac{1}{2}\big|\lambda - f^{-1}(\lambda_k^\star)\big| \gtrsim |\lambda_l^\star - \lambda_k^\star|. \qquad \text{(C.23)}
$$

  To see this, arguing similarly as in the proof of Lemma 3, we can use the Lipschitz property of f (cf. (C.4)) to obtain

$$
\begin{aligned}
\big|\lambda - f^{-1}(\lambda_k^\star)\big| &\geq \frac{1}{2}\big|f(\lambda) - f\big(f^{-1}(\lambda_k^\star)\big)\big| = \frac{1}{2}\big|(\lambda - \gamma(\lambda)) - \lambda_k^\star\big| \\
&\geq \frac{1}{2}|\lambda_l^\star - \lambda_k^\star| - \frac{1}{2}\big|\lambda - \gamma(\lambda) - \lambda_l^\star\big| \\
&\geq \frac{1}{2}|\lambda_l^\star - \lambda_k^\star| - \frac{1}{2}\mathcal{E}_{\mathsf{MD}} \\
&\gtrsim |\lambda_l^\star - \lambda_k^\star|.
\end{aligned}
$$

  In a similar manner, we can also derive

$$
\big|f^{-1}(\lambda_k^\star) - \lambda_i^{(l)}\big| \leq 2\big|f\big(f^{-1}(\lambda_k^\star)\big) - f(\lambda_i^{(l)})\big| = 2\big|\lambda_k^\star - \big(\lambda_i^{(l)} - \gamma(\lambda_i^{(l)})\big)\big|.
$$

  Therefore, for any $\lambda_i^{(l)}$ such that $\lambda_i^{(l)} - \gamma(\lambda_i^{(l)}) \in \mathcal{B}_{\mathcal{E}_k}(\lambda_k^\star)$, one has

$$
\big|f^{-1}(\lambda_k^\star) - \lambda_i^{(l)}\big| \leq 2\,\mathcal{E}_k + 2\,\mathcal{E}_{\mathsf{MD}} \leq 2c\,|\lambda_l^\star - \lambda_k^\star| \leq \frac{1}{2}\big|\lambda_l - f^{-1}(\lambda_k^\star)\big|.
$$

  Then the claim is an immediate consequence of these two bounds. As a result, we conclude that

$$
\sum_{i\in\mathcal{A}_1} \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}} \right)^2 \lesssim \frac{1}{(\lambda_l^\star - \lambda_k^\star)^2} \sum_{i\in\mathcal{A}_1} \big(\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}\big)^2 \leq \frac{1}{(\lambda_l^\star - \lambda_k^\star)^2}.
$$

– Turning to the set $\mathcal{A}_2$, we know from Lemma 3 that for any $i \in \mathcal{A}_2$, $|\lambda - \lambda_i^{(l)}| \gtrsim \min_{i:i\neq l} |\lambda_l^\star - \lambda_i^\star|$ and

$$\left\| \left(\lambda_i^{(l)} \boldsymbol{I}_{r-1} - \boldsymbol{\Lambda}^{\star(l)} - \gamma(\lambda_i^{(l)}) \boldsymbol{I}_{r-1}\right) \boldsymbol{U}^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)} \right\|_2 \lesssim \mathcal{E}_{\mathsf{MD}} = \sigma\sqrt{r}\log n.$$

Meanwhile, since $\boldsymbol{u}_{i,\|}^{(l)}$ is the projection of $\boldsymbol{u}_i^{(l)}$ onto the space of $\boldsymbol{U}^{\star(l)}$ followed by normalization, one has

$$\left|\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}\right| \leq \left|\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)}\right|,$$

and therefore,

$$\begin{aligned}
\left\| \left(\lambda_i^{(l)} \boldsymbol{I}_{r-1} - \boldsymbol{\Lambda}^{\star(l)} - \gamma(\lambda_i^{(l)}) \boldsymbol{I}_{r-1}\right) \boldsymbol{U}^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)} \right\|_2 &\geq \left|\lambda_i^{(l)} - \lambda_k^\star - \gamma(\lambda_i^{(l)})\right| \cdot \left|\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)}\right| \\
&\geq \mathcal{E}_k \cdot \left|\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)}\right| \\
&\gtrsim \left|\lambda_l^\star - \lambda_k^\star\right| \cdot \left|\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)}\right|.
\end{aligned}$$

This in turn allows us to derive that

$$\begin{aligned}
\left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}}\right)^2 \leq \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)}}{\lambda - \lambda_i^{(l)}}\right)^2 &\leq \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_{i,\|}^{(l)}}{\left(\Delta_l^\star\right)^2}\right)^2 \\
&\lesssim \frac{\sigma^2 r \log^2 n}{|\lambda_l^\star - \lambda_k^\star|^2 \left(\Delta_l^\star\right)^2}.
\end{aligned}$$

Putting the above two sums together reveals that

$$\sum_{1\leq i<r} \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}}\right)^2 \lesssim \frac{1}{(\lambda_l^\star - \lambda_k^\star)^2} + \frac{\sigma^2 r^2 \log^2 n}{|\lambda_l^\star - \lambda_k^\star|^2 \left(\Delta_l^\star\right)^2} \lesssim \frac{r}{(\lambda_l^\star - \lambda_k^\star)^2},$$

where the last inequality follows since $\left(\Delta_l^\star\right)^2 \gtrsim \sigma^2 r \log^2 n$. Combining this inequality with (C.22), one readily obtains

$$\begin{aligned}
\sqrt{\sum_{1\leq i<r} \left(\frac{1}{\lambda_l - \lambda_i^{(l)}} \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}\right)^2} &\leq \sum_{k:k\neq l} \left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right| \sqrt{\sum_{1\leq i<r} \left(\frac{\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda - \lambda_i^{(l)}}\right)^2} \\
&\lesssim \sum_{k:k\neq l} \frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right| \sqrt{r}}{|\lambda_l^\star - \lambda_k^\star|}. \qquad \text{(C.24)}
\end{aligned}$$

Substituting the above two partial sums (C.21) and (C.24) into (C.20), we conclude that: with probability exceeding $1 - O(n^{-10})$,

$$\left\| \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \left(\lambda_l \boldsymbol{I}_{n-1} - \boldsymbol{M}^{(l)}\right)^{-1} \right\|_2 \lesssim \sum_{k:k\neq l} \frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right| \sqrt{r}}{|\lambda_l^\star - \lambda_k^\star|} + \frac{1}{|\lambda_l^\star|}$$

holds for any λ such that $\lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\mathsf{MD}}}(\lambda_l^\star)$, as claimed in (C.19).

C.4 Proof of Lemma 6

Recall the definitions of $\boldsymbol{u}_l^{\star\perp}$, $\boldsymbol{u}_{l,\perp}$ and $\boldsymbol{U}^{\star(l)\perp}$ in (5.13), (5.24) and (5.32), respectively. Let us rewrite

$$\begin{aligned}
\left|\left\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{u}_l \right\rangle\right| &\overset{\text{(i)}}{=} \left|\left\langle (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{a},\, (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{u}_l \right\rangle\right| \\
&= \left|\left\langle (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{a},\, (\boldsymbol{U}^{\star\perp})^\top (\boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} \boldsymbol{u}_l + \boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}} \boldsymbol{u}_l) \right\rangle\right| \\
&\overset{\text{(ii)}}{=} \left|\left\langle (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{a},\, (\boldsymbol{U}^{\star\perp})^\top \boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}} \boldsymbol{u}_l) \right\rangle\right|
\end{aligned}$$

$$
\begin{aligned}
&\overset{\text{(iii)}}{=} \left|\left\langle (\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{a},\, (\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{u}_{l,\perp}\right\rangle\right| \cdot \left\|\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l\right\|_2 \\
&\overset{\text{(iv)}}{=} \left|\left\langle (\boldsymbol{u}_l^{\star\perp}\boldsymbol{U}^{\star(l)\perp})^{\top}\boldsymbol{a},\, (\boldsymbol{u}_l^{\star\perp}\boldsymbol{U}^{\star(l)\perp})^{\top}\boldsymbol{u}_{l,\perp}\right\rangle\right| \cdot \left\|\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l\right\|_2 \\
&= \left|\left\langle (\boldsymbol{U}^{\star(l)\perp})^{\top}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{a}\big),\, (\boldsymbol{U}^{\star(l)\perp})^{\top}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{u}_{l,\perp}\big)\right\rangle\right| \cdot \left\|\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l\right\|_2 \\
&\overset{\text{(v)}}{=} \left|\left\langle \boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{a}\big),\, \boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{u}_{l,\perp}\big)\right\rangle\right| \cdot \left\|\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l\right\|_2 \\
&\overset{\text{(vi)}}{=} \frac{1}{\left\|\widetilde{\boldsymbol{u}}_{l,\perp}\right\|_2}\left|\left\langle \boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{a}\big),\, \boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\widetilde{\boldsymbol{u}}_{l,\perp}\big)\right\rangle\right| \cdot \left\|\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l\right\|_2. \qquad \text{(C.25)}
\end{aligned}
$$

Here, (i) follows since $(\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{U}^{\star\perp} = \boldsymbol{I}_{n-r}$; (ii) holds true since $(\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{u}_l^{\star} = 0$; (iii) holds due to the definition $\boldsymbol{u}_{l,\perp} := (\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l)/\|\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l\|_2$; (iv) results from the fact $\boldsymbol{u}_l^{\star\perp}\boldsymbol{U}^{\star(l)\perp} = \boldsymbol{U}^{\star\perp}$; (v) holds true since $(\boldsymbol{U}^{\star(l)\perp})^{\top}\boldsymbol{U}^{\star(l)\perp} = \boldsymbol{I}_{n-r}$; (vi) arises from (5.4c) in Theorem 5, where we denote (i.e., $\boldsymbol{u}_{l,\perp}$ is the normalized version of $\widetilde{\boldsymbol{u}}_{l,\perp}$)

$$
\widetilde{\boldsymbol{u}}_{l,\perp} := \boldsymbol{u}_l^{\star\perp}\big(\lambda_l\boldsymbol{I}_{n-1} - (\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{M}\boldsymbol{u}_l^{\star}. \qquad \text{(C.26)}
$$

Our proof strategy is to show that $\frac{\boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\left((\boldsymbol{u}_l^{\star\perp})^{\top}\widetilde{\boldsymbol{u}}_{l,\perp}\right)}{\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\left((\boldsymbol{u}_l^{\star\perp})^{\top}\widetilde{\boldsymbol{u}}_{l,\perp}\right)\right\|_2}$ is a random vector uniformly distributed in the unit sphere of the subspace $\boldsymbol{U}^{\star(l)\perp}$. If this claim were true, then it would follow from standard measure concentration for uniform distributions results [Vershynin, 2017, Theorem 3.4.6] that, with probability at least $1 - O(n^{-10})$,

$$
\begin{aligned}
\left|\left\langle \boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{a}\big),\, \boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\widetilde{\boldsymbol{u}}_{l,\perp}\big)\right\rangle\right| &\lesssim \sqrt{\frac{\log n}{n-r}}\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{a}\big)\right\|_2\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\big((\boldsymbol{u}_l^{\star\perp})^{\top}\widetilde{\boldsymbol{u}}_{l,\perp}\big)\right\|_2 \\
&\asymp \sqrt{\frac{\log n}{n}}\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}(\boldsymbol{u}_l^{\star\perp})^{\top}\widetilde{\boldsymbol{u}}_{l,\perp}\right\|_2,
\end{aligned}
$$

where we use $\boldsymbol{u}_l^{\star\perp}\boldsymbol{U}^{\star(l)\perp} = \boldsymbol{U}^{\star\perp}$ and the rank assumption $r \ll n/\log^2 n$ in the last step. Combining this with (C.25), we arrive at the advertised bound:

$$
\begin{aligned}
\left|\left\langle \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a},\, \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{u}_l\right\rangle\right| &\lesssim \sqrt{\frac{\log n}{n}}\,\frac{\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}(\boldsymbol{u}_l^{\star\perp})^{\top}\widetilde{\boldsymbol{u}}_{l,\perp}\right\|_2}{\left\|\widetilde{\boldsymbol{u}}_{l,\perp}\right\|_2}\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2\left\|\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l\right\|_2 \\
&\le \sqrt{\frac{\log n}{n}}\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2\left\|\boldsymbol{P}_{\boldsymbol{u}_l^{\star\perp}}\boldsymbol{u}_l\right\|_2.
\end{aligned}
$$

To justify the distributional property claimed above, we define — for an arbitrary rotation matrix $\boldsymbol{Q} \in \mathbb{R}^{(n-r)\times(n-r)}$ — a new rotation matrix

$$
\boldsymbol{R} = \boldsymbol{P}_{\boldsymbol{U}^{\star(l)}} + \boldsymbol{U}^{\star(l)\perp}\boldsymbol{Q}\,(\boldsymbol{U}^{\star(l)\perp})^{\top} \in \mathbb{R}^{(n-1)\times(n-1)};
$$

the matrix $\boldsymbol{R}$ rotates vectors in the subspace spanned by $\boldsymbol{U}^{\star(l)\perp}$ according to $\boldsymbol{Q}$, while preserving the part in the subspace spanned by $\boldsymbol{U}^{\star(l)}$. We make note of two important "rotational invariance" properties as follows.

- As shown in the proof of Lemma 3, it is seen that

$$
\begin{aligned}
\boldsymbol{R}\,(\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{u}_l^{\star} &= \boldsymbol{U}^{\star(l)}(\boldsymbol{U}^{\star(l)})^{\top}(\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{u}_l^{\star} + \boldsymbol{U}^{\star(l)\perp}\boldsymbol{Q}\,(\boldsymbol{U}^{\star(l)\perp})^{\top}(\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{u}_l^{\star} \\
&= (\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{U}^{\star}_{\setminus l}\boldsymbol{U}^{\star\top}_{\setminus l}\boldsymbol{H}\boldsymbol{u}_l^{\star} + (\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{U}^{\star\perp}\boldsymbol{Q}\,(\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{u}_l^{\star} \\
&\overset{\mathrm{d}}{=} (\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{U}^{\star}_{\setminus l}\boldsymbol{U}^{\star\top}_{\setminus l}\boldsymbol{H}\boldsymbol{u}_l^{\star} + (\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{u}_l^{\star} \\
&= (\boldsymbol{u}_l^{\star\perp})^{\top}\big(\boldsymbol{I}_n - \boldsymbol{u}_l^{\star}\boldsymbol{u}_l^{\star\top}\big)\boldsymbol{H}\boldsymbol{u}_l^{\star} \\
&= (\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{u}_l^{\star}.
\end{aligned}
$$

  Here, the second line arises from the definitions of $\boldsymbol{U}^{\star(l)}$ and $\boldsymbol{U}^{\star(l)\perp}$ in (5.32); the third line follows because $\boldsymbol{Q}\,(\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{u}_l^{\star} \overset{\mathrm{d}}{=} (\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{H}\boldsymbol{u}_l^{\star}$; the last line holds due to the fact $(\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{u}_l^{\star} = \boldsymbol{0}$.

- In a similar manner, we also know that

$$\begin{aligned}\boldsymbol{R}\,(\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{H}\boldsymbol{u}_l^{\star\perp}\boldsymbol{R}^\top &= (\boldsymbol{u}_l^{\star\perp})^\top\big(\boldsymbol{U}^\star_{\diagdown l}\boldsymbol{U}^{\star\top}_{\diagdown l} + \boldsymbol{U}^{\star\perp}\boldsymbol{Q}\,(\boldsymbol{U}^{\star\perp})^\top\big)\boldsymbol{H}\big(\boldsymbol{U}^\star_{\diagdown l}\boldsymbol{U}^{\star\top}_{\diagdown l} + \boldsymbol{U}^{\star\perp}\boldsymbol{Q}^\top\,(\boldsymbol{U}^{\star\perp})^\top\big)\boldsymbol{u}_l^{\star\perp}\\ &= (\boldsymbol{u}_l^{\star\perp})^\top\big(\boldsymbol{U}^\star\boldsymbol{U}^{\star\top} + \boldsymbol{U}^{\star\perp}\boldsymbol{Q}\,(\boldsymbol{U}^{\star\perp})^\top\big)\boldsymbol{H}\big(\boldsymbol{U}^\star\boldsymbol{U}^{\star\top} + \boldsymbol{U}^{\star\perp}\boldsymbol{Q}^\top\,(\boldsymbol{U}^{\star\perp})^\top\big)\boldsymbol{u}_l^{\star\perp}\\ &\overset{\mathrm{d}}{=} (\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^{\star\perp},\end{aligned}$$

  where the second line comes from $(\boldsymbol{U}^\star\boldsymbol{U}^{\star\top} - \boldsymbol{U}^\star_{\diagdown l}\boldsymbol{U}^{\star\top}_{\diagdown l})\boldsymbol{u}_l^{\star\perp} = \boldsymbol{u}_l^\star\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l^{\star\perp} = 0$, and the last line holds since $\boldsymbol{U}^\star\boldsymbol{U}^{\star\top} + \boldsymbol{U}^{\star\perp}\boldsymbol{Q}^\top\,(\boldsymbol{U}^{\star\perp})^\top$ is a rotation matrix.

Using the statistical independence between these two parts, we reach

$$\begin{aligned}\boldsymbol{R}\,(\boldsymbol{u}_l^{\star\perp})^\top\widetilde{\boldsymbol{u}}_{l,\perp} &= \boldsymbol{R}\,\big(\lambda_l\boldsymbol{I}_{n-1} - (\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^\star = \big(\lambda_l\boldsymbol{I}_{n-1} - \boldsymbol{R}\,(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}\boldsymbol{R}^\top\big)^{-1}\boldsymbol{R}\,(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^\star\\ &= \big(\lambda_l\boldsymbol{I}_{n-1} - \boldsymbol{R}\,(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^{\star\perp}\boldsymbol{R}^\top\big)^{-1}\boldsymbol{R}\,(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\\ &\overset{\mathrm{d}}{=} \big(\lambda_l\boldsymbol{I}_{n-1} - (\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^{\star\perp}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{H}\boldsymbol{u}_l^\star\\ &= \big(\lambda_l\boldsymbol{I}_{n-1} - (\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^{\star\perp}\big)^{-1}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{M}\boldsymbol{u}_l^\star\\ &= (\boldsymbol{u}_l^{\star\perp})^\top\widetilde{\boldsymbol{u}}_{l,\perp},\end{aligned}$$

where the last step replies on the definition of $\widetilde{\boldsymbol{u}}_{l,\perp}$ in (C.26) and the fact $(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{u}_l^{\star\perp} = \boldsymbol{I}_{n-1}$. This enables us to conclude that $\frac{\boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\left((\boldsymbol{u}_l^{\star\perp})^\top\widetilde{\boldsymbol{u}}_{l,\perp}\right)}{\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star(l)\perp}}\left((\boldsymbol{u}_l^{\star\perp})^\top\widetilde{\boldsymbol{u}}_{l,\perp}\right)\right\|_2}$ is uniformly distributed in the unit sphere spanned by $\boldsymbol{U}^{\star(l)\perp}$.

D Proof of auxiliary lemmas in the analysis for Theorem 3

D.1 Proof of Lemma 7

For any matrix $\boldsymbol{A}\in\mathbb{R}^{n\times n}$, we know from the orthogonal invariance of the spectral norm that

$$\begin{aligned}\|\boldsymbol{A}\| &= \|[\boldsymbol{U}^\star,\boldsymbol{U}^{\star\perp}]^\top\boldsymbol{A}[\boldsymbol{U}^\star,\boldsymbol{U}^{\star\perp}]\|\\ &= \left\|\begin{bmatrix}\boldsymbol{U}^{\star\top}\boldsymbol{A}\boldsymbol{U}^\star & \boldsymbol{U}^{\star\top}\boldsymbol{A}\boldsymbol{U}^{\star\perp}\\ (\boldsymbol{U}^{\star\perp})^\top\boldsymbol{A}\boldsymbol{U}^\star & (\boldsymbol{U}^{\star\perp})^\top\boldsymbol{A}\boldsymbol{U}^{\star\perp}\end{bmatrix}\right\|\\ &\le \|\boldsymbol{U}^{\star\top}\boldsymbol{A}\boldsymbol{U}^\star\| + \|\boldsymbol{U}^{\star\top}\boldsymbol{A}\boldsymbol{U}^{\star\perp}\| + \|(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{A}\boldsymbol{U}^\star\| + \|(\boldsymbol{U}^{\star\perp})^\top\boldsymbol{A}\boldsymbol{U}^{\star\perp}\|,\end{aligned}$$

where the last step holds due to the triangle inequality. As a result, one can upper bound

$$\begin{aligned}\Big\|\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{\Sigma}\Big\| &\le \Big\|\boldsymbol{U}^{\star\top}\Big(\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{\Sigma}\Big)\boldsymbol{U}^\star\Big\| + \Big\|(\boldsymbol{U}^{\star\perp})^\top\Big(\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{\Sigma}\Big)\boldsymbol{U}^\star\Big\|\\ &\quad + \Big\|\boldsymbol{U}^{\star\top}\Big(\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{\Sigma}\Big)\boldsymbol{U}^{\star\perp}\Big\| + \Big\|(\boldsymbol{U}^{\star\perp})^\top\Big(\frac{1}{n}\boldsymbol{S}\boldsymbol{S}^\top - \boldsymbol{\Sigma}\Big)\boldsymbol{U}^{\star\perp}\Big\|\\ &= \Big\|\frac{1}{n}\boldsymbol{S}_\parallel\boldsymbol{S}_\parallel^\top - \boldsymbol{\Lambda}\Big\| + 2\Big\|\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\parallel^\top\Big\| + \Big\|\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top - \sigma^2\boldsymbol{I}_{p-r}\Big\| \qquad \text{(D.1)}\end{aligned}$$

where we remind the readers of the notation $\boldsymbol{S}_\parallel := \boldsymbol{U}^{\star\top}\boldsymbol{S}$, $\boldsymbol{S}_\perp := (\boldsymbol{U}^{\star\perp})^\top\boldsymbol{S}$ and $\boldsymbol{\Lambda} := \boldsymbol{U}^{\star\top}\boldsymbol{\Sigma}\boldsymbol{U}^\star$ introduced in (5.46).

Before describing how to control these quantities, we pause to collect a few results regarding a Gaussian random matrix $\boldsymbol{G}\in\mathbb{R}^{p\times n}$ consisting of i.i.d. $\mathcal{N}(0,1)$ entries [Vershynin, 2017, Theorem 4.6.1]: with probability at least $1-O(n^{-10})$,

$$\|\boldsymbol{G}\| \lesssim \sqrt{p}+\sqrt{n},$$

$$\Big\|\frac{1}{n}\boldsymbol{G}\boldsymbol{G}^\top - \boldsymbol{I}_p\Big\| \lesssim \sqrt{\frac{p}{n}} + \frac{p}{n} + \sqrt{\frac{\log n}{n}}, \qquad \text{(D.2)}$$

$$\Big\|\frac{1}{n}\boldsymbol{G}^\top\boldsymbol{G} - \frac{p}{n}\boldsymbol{I}_n\Big\| \lesssim 1 + \sqrt{\frac{p}{n}}. \qquad \text{(D.3)}$$

With these bounds in place, we can start to bound the spectral norms of the quantities in (D.1). Note that the columns of $\boldsymbol{S}_{\|}$ (resp. $\boldsymbol{S}_{\perp}$) are i.i.d. zero-mean Gaussian random vectors with covariance $\boldsymbol{\Lambda}$ (resp. $\sigma^2 \boldsymbol{I}_{p-r}$). Since we can rewrite $\boldsymbol{S}_{\|} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}$ with $\boldsymbol{Z} \in \mathbb{R}^{r\times n}$ being a Gaussian random matrix with i.i.d. $\mathcal{N}(0,1)$ entries, it immediately follows from (D.2) that with probability more than $1 - O(n^{-10})$,

$$\Big\|\frac{1}{n}\boldsymbol{S}_{\perp}\boldsymbol{S}_{\perp}^{\top} - \sigma^2 \boldsymbol{I}_{p-r}\Big\| \lesssim \sigma^2\left(\sqrt{\frac{p}{n}} + \frac{p}{n} + \sqrt{\frac{\log n}{n}}\right) \qquad \text{and} \tag{D.4}$$

$$\begin{aligned}\Big\|\frac{1}{n}\boldsymbol{S}_{\|}\boldsymbol{S}_{\|}^{\top} - \boldsymbol{\Lambda}\Big\| &= \Big\|\frac{1}{n}\boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}\boldsymbol{Z}^{\top}\boldsymbol{\Lambda}^{1/2} - \boldsymbol{\Lambda}\Big\| \le \|\boldsymbol{\Lambda}\|\Big\|\frac{1}{n}\boldsymbol{Z}\boldsymbol{Z}^{\top} - \boldsymbol{I}_r\Big\| \\ &\lesssim (\lambda_{\max}^{\star} + \sigma^2)\left(\sqrt{\frac{r}{n}} + \frac{r}{n} + \sqrt{\frac{\log n}{n}}\right) \\ &\lesssim (\lambda_{\max}^{\star} + \sigma^2)\sqrt{\frac{r\log n}{n}},\end{aligned} \tag{D.5}$$

where the last step arises from the sample size assumption $n \ge r$. As for $\boldsymbol{S}_{\perp}\boldsymbol{S}_{\|}^{\top}$, we can invoke Lemma 19 to show that: with probability at least $1 - O(n^{-10})$,

$$\begin{aligned}\Big\|\frac{1}{n}\boldsymbol{S}_{\perp}\boldsymbol{S}_{\|}^{\top}\Big\| &= \frac{1}{n}\big\|\boldsymbol{S}_{\perp}\boldsymbol{Z}^{\top}\boldsymbol{\Lambda}^{1/2}\big\| \le \frac{1}{n}\big\|\boldsymbol{S}_{\perp}\boldsymbol{Z}^{\top}\big\|\,\big\|\boldsymbol{\Lambda}^{1/2}\big\| \lesssim \frac{\sigma}{n}\big(\sqrt{pn\log n} + \sqrt{pr}\log n\big)\cdot\sqrt{\lambda_{\max}^{\star} + \sigma^2} \\ &\lesssim \sqrt{(\lambda_{\max}^{\star} + \sigma^2)\sigma^2\frac{p}{n}\log n},\end{aligned} \tag{D.6}$$

where the last step holds since $n \ge r$. Putting the bounds above together immediately concludes the proof.

D.2 Proof of Lemma 8

The proof of this lemma is similar to that of Lemma 3. By definition, one can compute

$$\begin{aligned}\boldsymbol{\Sigma}_{l,\perp} &:= \frac{1}{n}\mathbb{E}\big[\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\big] \\ &= \frac{1}{n}(\boldsymbol{u}_l^{\star\perp})^{\top}\mathbb{E}\big[\boldsymbol{S}\boldsymbol{S}^{\top}\big]\boldsymbol{u}_l^{\star\perp} = (\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{\Sigma}\boldsymbol{u}_l^{\star\perp} \\ &= (\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{U}^{\star}\boldsymbol{\Lambda}^{\star}\boldsymbol{U}^{\star\top}\boldsymbol{u}_l^{\star\perp} + \sigma^2\boldsymbol{I}_{p-1} \\ &= \boldsymbol{U}^{\star(l)}\boldsymbol{\Lambda}^{\star(l)}\boldsymbol{U}^{\star(l)\top} + \sigma^2\boldsymbol{I}_{p-1},\end{aligned} \tag{D.7}$$

where $\boldsymbol{U}^{\star(l)}$ and $\boldsymbol{\Lambda}^{\star(l)}$ have been defined in (5.72). Our proof strategy is to invoke Theorem 6 by treating $\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}$ (resp. $\boldsymbol{U}^{\star(l)}$) as $\boldsymbol{M}$ (resp. $\boldsymbol{Q}$).

- We shall start with the first claim. Let us define the following matrices in $\mathbb{R}^{(r-1)\times(r-1)}$:

  $$\begin{aligned}\boldsymbol{K}^{(l)}(\lambda) &:= \boldsymbol{U}^{\star(l)\top}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\boldsymbol{U}^{\star(l)\perp}\Big(\lambda\boldsymbol{I}_{p-r} - (\boldsymbol{U}^{\star(l)\perp})^{\top}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\boldsymbol{U}^{\star(l)\perp}\Big)^{-1}(\boldsymbol{U}^{\star(l)\perp})^{\top}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\boldsymbol{U}^{\star(l)}, \\ \boldsymbol{K}^{(l)\perp}(\lambda) &:= \mathbb{E}\Big[\boldsymbol{G}^{(l)}(\lambda) \mid (\boldsymbol{U}^{\star(l)\perp})^{\top}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\boldsymbol{U}^{\star(l)\perp}\Big].\end{aligned}$$

  Recall the definitions of $\boldsymbol{K}(\lambda)$ (cf. (5.53)) and $\boldsymbol{P}^{(l)}$ (cf. (C.6)), and notice that

  $$\begin{aligned}\boldsymbol{u}_l^{\star\perp}\boldsymbol{U}^{\star(l)\perp} &= \boldsymbol{U}^{\star\perp}, \\ (\boldsymbol{U}^{\star(l)\perp})^{\top}\boldsymbol{S}_{l,\perp} &= (\boldsymbol{U}^{\star(l)\perp})^{\top}(\boldsymbol{u}_l^{\star\perp})^{\top}\boldsymbol{S} = (\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{S} = \boldsymbol{S}_{\perp}.\end{aligned}$$

  Straightforward calculation allows us to simplify the above expressions as follows

  $$\boldsymbol{K}^{(l)}(\lambda) := \frac{1}{n}\boldsymbol{P}^{(l)\top}\boldsymbol{S}_{\|}\cdot\frac{1}{n}\boldsymbol{S}_{\perp}^{\top}\Big(\lambda\boldsymbol{I}_{p-r} - \frac{1}{n}\boldsymbol{S}_{\perp}\boldsymbol{S}_{\perp}^{\top}\Big)^{-1}\boldsymbol{S}_{\perp}\cdot\boldsymbol{S}_{\|}^{\top}\boldsymbol{P}^{(l)} = \boldsymbol{P}^{(l)\top}\boldsymbol{K}(\lambda)\boldsymbol{P}^{(l)},$$

$$\boldsymbol{K}^{(l)\perp}(\lambda) := \mathbb{E}\big[\boldsymbol{G}^{(l)}(\lambda) \mid (\boldsymbol{U}^{\star(l)\perp})^{\top}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\boldsymbol{U}^{\star(l)\perp}\big] = \beta(\lambda)\boldsymbol{P}^{(l)\top}(\boldsymbol{\Lambda}^{\star}+\sigma^{2}\boldsymbol{I}_{r})\boldsymbol{P}^{(l)} = \beta(\lambda)(\boldsymbol{\Lambda}^{\star(l)}+\sigma^{2}\boldsymbol{I}_{r-1}).$$

Theorem 6 then tells us that

$$\begin{aligned}
&\big(\gamma_i^{(l)}\boldsymbol{I}_{r-1} - \boldsymbol{\Lambda}^{\star(l)} - \sigma^2\boldsymbol{I}_{r-1} - \boldsymbol{K}^{(l)\perp}(\gamma_i^{(l)})\big)\boldsymbol{U}^{\star(l)\top}\boldsymbol{u}_{i,\parallel}^{(l)} \\
&\quad = \Big(\frac{1}{n}\boldsymbol{P}^{(l)\top}\boldsymbol{S}_{\parallel}\boldsymbol{S}_{\parallel}^{\top}\boldsymbol{P}^{(l)} - \boldsymbol{\Lambda}^{\star(l)} - \sigma^2\boldsymbol{I}_{r-1} + \boldsymbol{K}(\gamma_i^{(l)}) - \boldsymbol{K}^{\perp}(\gamma_i^{(l)})\Big)\boldsymbol{U}^{\star(l)\top}\boldsymbol{u}_{i,\parallel}^{(l)}.
\end{aligned}$$

One can then adopt a similar argument as in the proof of Theorem 8 to demonstrate that: with probability at least $1-O(n^{-10})$,

$$\begin{aligned}
&\Big\|\Big(\gamma_i^{(l)}\boldsymbol{I}_{r-1} - \big(1+\beta(\gamma_i^{(l)})\big)\big(\boldsymbol{\Lambda}^{\star(l)}+\sigma^2\boldsymbol{I}_{r-1}\big)\Big)\boldsymbol{U}^{\star(l)\top}\boldsymbol{u}_{i,\parallel}^{(l)}\Big\|_2 \\
&\quad \le \Big\|\frac{1}{n}\boldsymbol{P}^{(l)\top}\boldsymbol{S}_{\parallel}\boldsymbol{S}_{\parallel}^{\top}\boldsymbol{P}^{(l)} - \boldsymbol{\Lambda}^{\star(l)} - \sigma^2\boldsymbol{I}_{r-1}\Big\| + \sup_{\lambda:\lambda\in[2\lambda_l^\star/3,\,4\lambda_l^\star/3]}\Big\|\big(\boldsymbol{K}(\gamma_i^{(l)}) - \boldsymbol{K}^{\perp}(\gamma_i^{(l)})\big)\boldsymbol{U}^{\star(l)\top}\boldsymbol{u}_{i,\parallel}^{(l)}\Big\|_2 \\
&\quad \le \Big\|\frac{1}{n}\boldsymbol{S}_{\parallel}\boldsymbol{S}_{\parallel}^{\top} - \boldsymbol{\Lambda}^{\star} - \sigma^2\boldsymbol{I}_{r}\Big\| + \sup_{\lambda:\lambda\in[2\lambda_l^\star/3,\,4\lambda_l^\star/3]}\Big\|\boldsymbol{K}(\gamma_i^{(l)}) - \boldsymbol{K}^{\perp}(\gamma_i^{(l)})\Big\| \lesssim \mathcal{E}_{\mathsf{PCA}}
\end{aligned}$$

and there exists some $k \neq l$ $(1 \le k \le r)$ obeying

$$\Bigg|\frac{\gamma_i^{(l)}}{1+\beta(\gamma_i^{(l)})} - \lambda_k^\star - \sigma^2\Bigg| \lesssim \mathcal{E}_{\mathsf{PCA}}.$$

- Next, we turn to the second claim and we shall prove the upper and lower bounds for $\gamma_i^{(l)}$ separately.
  - For the upper bound, it suffices to upper bound $\gamma_r^{(l)}$ since $\{\gamma_i^{(l)}\}_i$ are defined in descending order. In view of $(\boldsymbol{U}^{\star(l)\perp})^\top\boldsymbol{S}_{l,\perp} = \boldsymbol{S}_\perp$, we can invoke Lemma 21 to see that

    $$\gamma_r^{(l)} \le \lambda_1\Big(\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\Big).$$

    This suggests that we look at the spectrum of $\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top$. From (D.2) and (D.3), we can obtain

    $$\Big|\lambda_i\Big(\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top\Big) - \sigma^2\frac{n\vee(p-r)}{n}\Big| \lesssim \sigma^2\sqrt{\frac{p+\log n}{n}}, \qquad 1\le i\le (p-r)\wedge n,$$

    where we use the fact that the non-zero eigenvalues of $\frac{1}{n}\boldsymbol{S}_\perp\boldsymbol{S}_\perp^\top$ and $\frac{1}{n}\boldsymbol{S}_\perp^\top\boldsymbol{S}_\perp$ are identical. Therefore, one has

    $$\gamma_i^{(l)} \le \gamma_r^{(l)} \le \sigma^2(n\vee p)/n + O(\sigma^2\sqrt{(p+\log n)/n})$$

    for all $r\le i\le n\wedge(p-1)$.
  - Next, we move on to consider the lower bound. Observe that the matrix $\boldsymbol{\Sigma}_{l,\perp}$ defined in (D.7) satisfies the following properties: (i) $\boldsymbol{\Sigma}_{l,\perp} \succeq \sigma^2\boldsymbol{I}_{p-1}$; (ii) $\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}$ is a Gaussian random matrix composed of i.i.d. standard Gaussian entries. Then we can lower bound the eigenvalues $\gamma_i^{(l)}$ for any $1\le i\le (p-1)\wedge n$ as follows

    $$\begin{aligned}
    \gamma_i^{(l)} &= \lambda_i\Big(\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big) \overset{\text{(i)}}{=} \lambda_i\Big(\frac{1}{n}\boldsymbol{S}_{l,\perp}^\top\boldsymbol{S}_{l,\perp}\Big) \\
    &= \lambda_i\Big(\frac{1}{n}\big(\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}\big)^\top\boldsymbol{\Sigma}_{l,\perp}\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}\Big) \\
    &\overset{\text{(ii)}}{\ge} \sigma^2\lambda_i\Big(\frac{1}{n}\big(\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}\big)^\top\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}\Big) \\
    &\overset{\text{(iii)}}{=} \sigma^2\lambda_i\Big(\frac{1}{n}\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}\big(\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}\big)^\top\Big),
    \end{aligned}$$

where (i) and (iii) hold because $i \leq (p-1) \wedge n$; (ii) follows since for any matrix $\boldsymbol{A}$, one has $\boldsymbol{A}^{\top}(\boldsymbol{\Sigma}_{l,\perp} - \sigma^2 \boldsymbol{I})\boldsymbol{A} \succeq 0$ and hence $\lambda_i(\boldsymbol{A}^{\top}\boldsymbol{\Sigma}_{l,\perp}\boldsymbol{A}) \geq \sigma^2 \lambda_i(\boldsymbol{A}^{\top}\boldsymbol{A})$. By invoking (D.2) and (D.3) once again, we arrive at

$$\Big|\lambda_i\Big(\frac{1}{n}\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}\big(\boldsymbol{\Sigma}_{l,\perp}^{-1/2}\boldsymbol{S}_{l,\perp}\big)^{\top}\Big) - \sigma^2\frac{n \vee (p-1)}{n}\Big| \lesssim \sigma^2\sqrt{\frac{p+\log n}{n}}, \qquad 1 \leq i \leq n \wedge (p-1).$$

This immediately establishes the claimed lower bound.

- The third claim is an immediate consequence of the fact that

$$\mathsf{rank}\Big(\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\Big) \leq n \wedge (p-1).$$

- Finally, let us consider the last claim. In view of Theorem 8, we have

$$\frac{\lambda_l}{1+\beta(\lambda_l)} \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^{\star} + \sigma^2).$$

  In addition, the first claim asserts that for each $1 \leq i < r$, one has

$$\frac{\gamma_i^{(l)}}{1+\beta(\gamma_i^{(l)})} \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_k^{\star} + \sigma^2)$$

  for some $k \neq l$. In view of the Lipschitz property of the function $f(\lambda) := \frac{\lambda}{1+\beta(\lambda)}$ (so that $|f'(\lambda)| \lesssim 1$), applying a similar argument as in the proof of Lemma 3 (see Appendix C.1) immediately allows us to establish the claim.

D.3 Proof of Lemma 9

Before continuing, we introduce several useful notation that will be used throughout.

- Let $\boldsymbol{U}^{(l)}\sqrt{\boldsymbol{\Gamma}^{(l)}}\boldsymbol{V}^{(l)\top}$ denote the SVD of $\frac{1}{\sqrt{n}}\boldsymbol{S}_{l,\perp}$, where $\boldsymbol{\Gamma}^{(l)}$ is a diagonal matrix consisting of the singular values of interest. Here, we recall that $\boldsymbol{S}_{l,\perp}$ has been defined in (5.67).
- Let $\boldsymbol{u}_i^{(l)}$ (resp. $\boldsymbol{v}_i^{(l)}$) indicate the i-th column of $\boldsymbol{U}^{(l)}$ (resp. $\boldsymbol{V}^{(l)}$), and let $\gamma_i^{(l)}$ represent the i-th diagonal entry of $\boldsymbol{\Gamma}^{(l)}$.

In addition, we note that the vector $\boldsymbol{s}_{l,\|}$ (see (5.67)) obeys

$$\boldsymbol{s}_{l,\|}^{\top} \sim \mathcal{N}\big(\boldsymbol{0}, (\boldsymbol{u}_l^{\star\top}\boldsymbol{\Sigma}\boldsymbol{u}_l^{\star})\boldsymbol{I}_n\big) = \mathcal{N}\big(\boldsymbol{0}, (\lambda_l^{\star} + \sigma^2)\boldsymbol{I}_n\big)$$

and is independent of $\boldsymbol{S}_{l,\perp}$ (and thus $\boldsymbol{\Gamma}^{(l)}$ and $\boldsymbol{V}^{(l)}$). This implies that condition on $\boldsymbol{V}^{(l)}$, one has

$$\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^{\top} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \lambda_l^{\star} + \sigma^2\right), \qquad 1 \leq i < p. \tag{D.8}$$

Moreover, by virtue of the rotational invariance of i.i.d. Gaussian random matrices, it is readily seen that $\boldsymbol{\Gamma}^{(l)}$ is independent of $\boldsymbol{V}^{(l)}$.

Now, we can begin to present the proof, towards which we start with the following decomposition

$$\begin{aligned}
&\Big\|\Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\|}^{\top}\Big\|_2^2 \\
&\quad = \frac{1}{n}\Big\|\big(\lambda_l \boldsymbol{I}_{p-1} - \boldsymbol{U}^{(l)}\boldsymbol{\Gamma}^{(l)}\boldsymbol{U}^{(l)\top}\big)^{-1}\boldsymbol{U}^{(l)}\sqrt{\boldsymbol{\Gamma}^{(l)}}\boldsymbol{V}^{(l)\top}\boldsymbol{s}_{l,\|}^{\top}\Big\|_2^2 \\
&\quad = \frac{1}{n}\Big\|\boldsymbol{U}^{(l)}\big(\lambda_l \boldsymbol{I}_{p-1} - \boldsymbol{\Gamma}^{(l)}\big)^{-1}\boldsymbol{U}^{(l)\top}\boldsymbol{U}^{(l)}\sqrt{\boldsymbol{\Gamma}^{(l)}}\boldsymbol{V}^{(l)\top}\boldsymbol{s}_{l,\|}^{\top}\Big\|_2^2 \\
&\quad = \frac{1}{n}\Big\|\big(\lambda_l \boldsymbol{I}_{p-1} - \boldsymbol{\Gamma}^{(l)}\big)^{-1}\sqrt{\boldsymbol{\Gamma}^{(l)}}\boldsymbol{V}^{(l)\top}\boldsymbol{s}_{l,\|}^{\top}\Big\|_2^2 \\
&\quad = \frac{1}{n}\sum_{1 \leq i \leq n \wedge (p-1)}\frac{\gamma_i^{(l)}}{(\lambda_l - \gamma_i^{(l)})^2}\big(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^{\top}\big)^2.
\end{aligned} \tag{D.9}$$

In what follows, we shall control the sum over $i < r$ and the sum over $i \geq r$ separately.

Controlling the sum over $i < r$. According to Lemma 8, one has

$$\gamma_i^{(l)} \lesssim \lambda_{\max}^\star + \sigma^2 \qquad \text{and} \qquad (\gamma_i^{(l)} - \lambda_l)^2 \gtrsim \min_{i:i\neq l} (\lambda_l^\star - \lambda_i^\star)^2$$

for all $1 \leq i < r$. In addition, recall that $\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\parallel}^\top$ $(1 \leq i < p)$ are i.i.d. zero-mean Gaussian random variables with variance $\lambda_l^\star + \sigma^2$ (see (D.8)). Invoking standard Gaussian inequalities shows that with probability at least $1 - O(n^{-10})$,

$$\max_{1\leq i<r} \big(\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\parallel}^\top\big)^2 \lesssim (\lambda_l^\star + \sigma^2) \log n. \tag{D.10}$$

As a result, we obtain

$$\sum_{1\leq i<r} \frac{\gamma_i^{(l)} \big(\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\parallel}^\top\big)^2}{(\lambda_l - \gamma_i^{(l)})^2} \lesssim \frac{(\lambda_{\max}^\star + \sigma^2)(\lambda_l^\star + \sigma^2)\, r \log n}{(\Delta_l^\star)^2} \tag{D.11}$$

with probability at least $1 - O(n^{-10})$.

Controlling the sum over $i \geq r$. Now, let us control the sum over $i \geq r$, and we shall consider the case with $n \geq p$ and the case with $n < p$ separately.

- *Case I: $n \geq p$.* Note that $\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\parallel}^\top \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \lambda_l^\star + \sigma^2)$. This suggests that we decompose

$$\sum_{r\leq i\leq n\wedge(p-1)} \frac{\gamma_i^{(l)} (\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\parallel}^\top)^2}{(\lambda_l - \gamma_i^{(l)})^2} = \underbrace{\sum_{r\leq i\leq n\wedge(p-1)} \frac{\gamma_i^{(l)} (\lambda_l^\star + \sigma^2)}{(\lambda_l - \gamma_i^{(l)})^2}}_{=:\,\alpha_1} + \underbrace{\sum_{r\leq i\leq n\wedge(p-1)} \frac{\gamma_i^{(l)} \big((\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\parallel}^\top)^2 - (\lambda_l^\star + \sigma^2)\big)}{(\lambda_l - \gamma_i^{(l)})^2}}_{=:\,\alpha_2}, \tag{D.12}$$

  and control α_1 as well as α_2 individually.

  – To begin with, let us consider α_1, which requires estimating $\lambda_l^\star + \sigma^2$ and $\sum_{r\leq i\leq n\wedge(p-1)} \gamma_i^{(l)}/(\lambda_l - \gamma_i^{(l)})^2$. This task is accomplished in Lemma 15 and Lemma 16 stated below.

  Lemma 15. *Instate the assumptions of Theorem 3. With probability at least $1 - O(n^{-10})$, we have*

$$\underbrace{\left|\sum_{i\geq r} \frac{\gamma_i^{(l)}}{(\lambda_l - \gamma_i^{(l)})^2} - \sum_{i>r} \frac{\lambda_i}{(\lambda_l - \lambda_i)^2}\right|}_{=:\,\epsilon_1} \lesssim \frac{\sigma^2}{\lambda_l^{\star 2}} \Big(1 + \frac{p}{n}\Big). \tag{D.13}$$

  and

$$\left|\sum_{i>r} \frac{\lambda_i}{(\lambda_l - \lambda_i)^2}\right| \vee \left|\sum_{i\geq r} \frac{\gamma_i^{(l)}}{(\lambda_l - \gamma_i^{(l)})^2}\right| \lesssim \frac{\sigma^2 p}{\lambda_l^{\star 2}}, \tag{D.14}$$

  Proof. See Appendix D.3.1. □

  Lemma 16. *Instate the assumptions of Theorem 3. With probability at least $1 - O(n^{-10})$, one has*

$$\underbrace{\left|\frac{\lambda_l}{1 + \frac{1}{n}\sum_{i>r} \frac{\lambda_i}{\lambda_l - \lambda_i}} - (\lambda_l^\star + \sigma^2)\right|}_{=:\,\epsilon_2} \lesssim \underbrace{(\lambda_{\max}^\star + \sigma^2)\sqrt{\frac{r \log n}{n}}}_{=\,\mathcal{E}_{\mathsf{PCA}}}. \tag{D.15}$$

  and

$$\frac{\lambda_l}{1 + \frac{1}{n}\sum_{i>r} \frac{\lambda_i}{\lambda_l - \lambda_i}} \asymp \lambda_l^\star. \tag{D.16}$$

  Proof. See Appendix D.3.2. □

With these two lemmas in place, we are ready to control α_1. According to (D.14), we have

$$\alpha_1 \lesssim \frac{(\lambda_l^\star + \sigma^2)\sigma^2 p}{\lambda_l^{\star 2}}. \tag{D.17}$$

In addition, recall the definition of c_l (cf. (15)). We can upper bound

$$\begin{aligned}
|\alpha_1 - c_l \cdot n| &= \left| (\lambda_l^\star + \sigma^2) \sum_{i \geq r} \frac{\gamma_i^{(l)}}{(\lambda_l - \gamma_i^{(l)})^2} - c_l \cdot n \right| \\
&= \left| (\lambda_l^\star + \sigma^2) \left(\sum_{i > r} \frac{\lambda_i}{(\lambda_l - \lambda_i)^2} + \epsilon_1 \right) - (\lambda_l^\star + \sigma^2 + \epsilon_2) \sum_{i > r} \frac{\lambda_i}{(\lambda_l - \lambda_i)^2} \right| \\
&= \left| \epsilon_1 (\lambda_l^\star + \sigma^2) - \epsilon_2 \sum_{i > r} \frac{\lambda_i}{(\lambda_l - \lambda_i)^2} \right| \\
&\overset{\text{(i)}}{\lesssim} (\lambda_l^\star + \sigma^2) |\epsilon_1| + \frac{\sigma^2 p}{\lambda_l^{\star 2}} |\epsilon_2| \\
&\overset{\text{(ii)}}{\lesssim} (\lambda_l^\star + \sigma^2) \cdot \frac{\sigma^2}{\lambda_l^{\star 2}} \left(1 + \frac{p}{n} \right) + \frac{\sigma^2 p}{\lambda_l^{\star 2}} \cdot (\lambda_{\max}^\star + \sigma^2) \sqrt{\frac{r \log n}{n}} \\
&\asymp \frac{(\lambda_l^\star + \sigma^2)\sigma^2}{\lambda_l^{\star 2}} + \frac{(\lambda_{\max}^\star + \sigma^2)\sigma^2 p}{\lambda_l^{\star 2}} \sqrt{\frac{r \log n}{n}}.
\end{aligned} \tag{D.18}$$

Here, (i) makes use of (D.14) and (D.13); (ii) relies on (D.13) and (D.15).

– Next, we move on to look at α_2. Observe that $\big\{ (\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\|}^\top)^2 - (\lambda_l^\star + \sigma^2) \big\}_{i \geq r}$ is a sequence of zero-mean sub-exponential random variables, which is independent of $\boldsymbol{\Gamma}^{(l)}$ but depends on λ_l. Hence, we shall apply the epsilon-net argument (cf. Lemma 20) to bound α_2. To do so, let us first verify the conditions required therein. With probability exceeding $1 - O(n^{-20})$, one has

$$\begin{aligned}
&\left| \frac{\mathrm{d}}{\mathrm{d}\lambda} \sum_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} \big((\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\|}^\top)^2 - (\lambda_l^\star + \sigma^2) \big) \right| \\
&\quad = \left| \sum_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^3} \big((\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\|}^\top)^2 - (\lambda_l^\star + \sigma^2) \big) \right| \\
&\quad \leq (p \wedge n) \cdot \max_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^3} \cdot \max_{r \leq i \leq n \wedge (p-1)} \big| (\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\|}^\top)^2 - (\lambda_l^\star + \sigma^2) \big| \\
&\quad \lesssim (p \wedge n) \cdot \frac{\sigma^2 (p \vee n)}{\lambda_l^{\star 3} n} \cdot (\lambda_l^\star + \sigma^2) \log n \\
&\quad = \frac{(\lambda_l^\star + \sigma^2)\sigma^2 p \log n}{\lambda_l^{\star 3}}
\end{aligned}$$

for all λ with $\lambda / \big(1 + \beta(\lambda)\big) \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star + \sigma^2)$. In addition, it is seen that

$$\begin{aligned}
\big\| (\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\|}^\top)^2 - (\lambda_l^\star + \sigma^2) \big\|_{\psi_1} &\lesssim \lambda_l^\star + \sigma^2, \\
\mathbb{E}\Big[\big((\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\|}^\top)^2 - (\lambda_l^\star + \sigma^2) \big)^2 \Big] &\lesssim (\lambda_l^\star + \sigma^2)^2,
\end{aligned}$$

where $\|\cdot\|_{\psi_1}$ denotes the sub-exponential norm. Invoke the matrix Bernstein inequality [Koltchinskii, 2011, Corollary 2.1] to show that: for any fixed λ obeying $\lambda / \big(1 + \beta(\lambda)\big) \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star + \sigma^2)$, one has

$$\left| \sum_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} \big((\boldsymbol{v}_i^{(l)\top} \boldsymbol{s}_{l,\|}^\top)^2 - (\lambda_l^\star + \sigma^2) \big) \right|$$

$$\lesssim \max_{r\leq i\leq n\wedge(p-1)} \frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^2}\cdot(\lambda_l^\star+\sigma^2)(\log n+\sqrt{p\wedge n}\log n)$$
$$\lesssim \frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2}}\frac{p\vee n}{n}\sqrt{p\wedge n}\log n$$
$$\lesssim \frac{(\lambda_l^\star+\sigma^2)\sigma^2 p}{\lambda_l^{\star 2}}\frac{\log n}{\sqrt{p\wedge n}}$$

with probability at least $1-O(n^{-10})$, where the last inequalities hold since $|\lambda-\gamma_i^{(l)}|\gtrsim\lambda_l^\star$ and $\gamma_i^{(l)}\lesssim\sigma^2(1+p/n)$. In addition, we make the observation that

$$\{\lambda\colon \lambda/\big(1+\beta(\lambda)\big)\in\mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star+\sigma^2)\}\subseteq[2\lambda_l^\star/3,\,4\lambda_l^\star/3].$$

With these in place, one can readily invoke Lemma 20 to derive

$$\alpha_2\lesssim\frac{(\lambda_l^\star+\sigma^2)\sigma^2 p}{\lambda_l^{\star 2}}\frac{\log n}{\sqrt{p\wedge n}} \tag{D.19}$$

with probability at least $1-O(n^{-10})$.

– Combining (D.17) and (D.19), we conclude

$$\sum_{r\leq i\leq n\wedge(p-1)}\frac{\gamma_i^{(l)}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{n\,(\lambda_l-\gamma_i^{(l)})^2}\leq\frac{|\alpha_1|+|\alpha_2|}{n}\lesssim\frac{(\lambda_l^\star+\sigma^2)\sigma^2 p}{\lambda_l^{\star 2}n}+\frac{(\lambda_l^\star+\sigma^2)\sigma^2 p}{\lambda_l^{\star 2}n}\frac{\log n}{\sqrt{p\wedge n}}$$
$$\lesssim\frac{(\lambda_l^\star+\sigma^2)\sigma^2 p\log n}{\lambda_l^{\star 2}n}. \tag{D.20}$$

Moreover, putting (D.18) and (D.19) together gives

$$\left|\sum_{i\geq r}\frac{\gamma_i^{(l)}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{n\,(\lambda_l-\gamma_i^{(l)})^2}-c_l\right|\leq\frac{|\alpha_1-c_l\cdot n|+|\alpha_2|}{n}$$
$$\lesssim\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}+\frac{(\lambda_{\max}^\star+\sigma^2)\sigma^2 p}{\lambda_l^{\star 2}n}\sqrt{\frac{r\log n}{n}}+\frac{(\lambda_l^\star+\sigma^2)\sigma^2 p}{\lambda_l^{\star 2}n}\frac{\log n}{\sqrt{p\wedge n}}$$
$$\asymp\frac{\sigma^2 p}{\lambda_l^{\star 2}n}\left((\lambda_{\max}^\star+\sigma^2)\sqrt{\frac{r\log n}{n}}+(\lambda_l^\star+\sigma^2)\frac{\log n}{\sqrt{p\wedge n}}\right), \tag{D.21}$$

where c_l is defined in (3.12).

- *Case II: $n<p$.* As it turns out, the above analysis for (D.21) is not tight when it comes to the case $n<p$. To remedy the issue, we provide a more precise estimate for terms (D.12) in the following lemma.

Lemma 17. *Instate the assumptions of Theorem 3. Suppose that $p>n$, then the following holds with probability at least $1-O(n^{-10})$:*

$$\sum_{i\geq r}\frac{\gamma_i^{(l)}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{(\lambda_l-\gamma_i^{(l)})^2}=\underbrace{\frac{\sigma^2 p}{\lambda_l-\sigma^2 p/n}+\frac{\lambda_l}{\lambda_l-\sigma^2 p/n}\frac{\lambda_l}{1+\frac{1}{n}\sum_{i>r}\frac{\lambda_i}{\lambda_l-\lambda_i}}\sum_{r<i\leq n}\frac{\lambda_i-\sigma^2 p/n}{(\lambda_l-\lambda_i)^2}}_{=c_l\cdot n}$$
$$+O\Big(\frac{\sigma^2 pr\log n}{\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star|\,n}+\frac{\sigma^2\kappa\sqrt{pr\log n}}{\lambda_l^\star}\Big). \tag{D.22}$$

Proof. See Appendix D.3.3. ☐

Here, the quantity c_l introduced above is precisely the one defined in (3.12). Consequently, we arrive at

$$\left|\sum_{i\geq r}\frac{\gamma_i^{(l)}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{n\,(\lambda_l-\gamma_i^{(l)})^2}-c_l\right|\lesssim\frac{\sigma^2 pr\log n}{\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star|\,n}+\frac{\sigma^2\kappa\sqrt{pr\log n}}{\lambda_l^\star n}. \tag{D.23}$$

Combining two sums. Substituting (D.11) and (D.20) (which holds universally for any n) into (D.9), we reach the first claim (5.76):

$$\left\|\Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right\|_2^2 \lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)\,r\log n}{\min_{i:i\neq l}\left|\lambda_l^\star-\lambda_i^\star\right|^2 n} + \frac{(\lambda_l^\star+\sigma^2)\sigma^2 p\log^2 n}{\lambda_l^{\star 2} n} \ll 1,$$

where the last step holds due to the assumptions (3.13a) and (3.13b).

We now turn attention to the estimation error. Regarding the case with $n \geq p$, we can combine (D.11) and (D.21) to conclude that

$$\begin{aligned}&\left|\left\|\Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right\|_2^2 - c_l\right| \\ &\qquad\lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2)\,r\log n}{\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star|^2 n} + \frac{\sigma^2 p}{\lambda_l^{\star 2} n}\left((\lambda_l^\star+\sigma^2)\frac{\log^2 n}{\sqrt{p\wedge n}} + (\lambda_{\max}^\star+\sigma^2)\sqrt{\frac{r\log n}{n}}\right).\end{aligned}$$

As for the case with $n < p$, substituting (D.11) and (D.23) into (D.9) yields

$$\begin{aligned}\left|\left\|\Big(\lambda_l \boldsymbol{I}_{p-1} - \frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^\top\right\|_2^2 - c_l\right| &\lesssim \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2) r\log n}{\min_{i:i\neq l}\left|\lambda_l^\star-\lambda_i^\star\right|^2 n} + \frac{\sigma^2\kappa\sqrt{pr\log n}}{\lambda_l^\star n} + \frac{\sigma^2 pr\log n}{\min_{i:i\neq l}\left|\lambda_l^\star-\lambda_i^\star\right| n^2} \\ &\asymp \frac{(\lambda_{\max}^\star+\sigma^2)(\lambda_l^\star+\sigma^2) r\log n}{\min_{i:i\neq l}\left|\lambda_l^\star-\lambda_i^\star\right|^2 n} + \frac{\sigma^2\kappa\sqrt{pr\log n}}{\lambda_l^\star n},\end{aligned}$$

where in the last line we use the conditions $\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star| \lesssim \lambda_{\max}^\star$ and $\sigma^2 p/n \ll \lambda_l^\star$ (according to the noise assumption (3.13a)).

D.3.1 Proof of Lemma 15

To begin with, let us consider (D.13). By construction, we have $\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top = (\boldsymbol{u}_l^{\star\perp})^\top \boldsymbol{S}\boldsymbol{S}^\top \boldsymbol{u}_l^{\star\perp}$, and it follows from Lemma 21 that $\lambda_{i+1} \leq \gamma_i^{(l)} \leq \lambda_i$ for all $1 \leq i < p$. Simple calculation yields

$$\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2} - \frac{\lambda_i}{(\lambda_l-\lambda_i)^2} = \frac{(\gamma_i^{(l)}-\lambda_i)(\lambda_l^2-\lambda_i\gamma_i^{(l)})}{(\lambda_l-\gamma_i^{(l)})^2(\lambda_l-\lambda_i)^2},$$

and consequently

$$\frac{\lambda_{i+1}}{(\lambda_l-\lambda_{i+1})^2} \leq \frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2} \leq \frac{\lambda_i}{(\lambda_l-\lambda_i)^2}, \qquad i \geq r.$$

We can then invoke a similar argument used in the proof of Lemma 14 to bound

$$\sum_{r<i\leq n}\frac{\lambda_i}{(\lambda_l-\lambda_i)^2} \leq \sum_{r\leq i\leq n}\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2} \leq \sum_{r<i\leq n}\frac{\lambda_i}{(\lambda_l-\lambda_i)^2} + \frac{\gamma_r^{(l)}}{(\lambda_l-\gamma_r^{(l)})^2}.$$

Hence, the conclusion immediately follows since $\gamma_r^{(l)}/(\lambda_l-\gamma_r^{(l)})^2 \lesssim \sigma^2(1+p/n)/\lambda_l^{\star 2}$ by Lemma 8.

We proceed to consider (D.14). According to Lemma 8, the following holds for eigenvalues $\{\gamma_i^{(l)}\}_{i\geq r}$: (i) $|\lambda_l-\gamma_i^{(l)}| \gtrsim \lambda_l^\star$ for all $i \geq r$; (ii) $|\gamma_i^{(l)}| \lesssim \sigma^2(p\vee n)/n$ for all $r \leq i \leq n\wedge(p-1)$; (iii) $\gamma_i^{(l)} = 0$ for $n\wedge(p-1) < i < p$. Therefore, we can upper bound

$$\left|\sum_{i:\, i\geq r}\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}\right| \lesssim (p\wedge n)\frac{\sigma^2(p\vee n)}{\lambda_l^{\star 2} n} = \frac{\sigma^2 p}{\lambda_l^{\star 2}},$$

and the upper bound for $\sum_{i>r}\lambda_i/(\lambda_l-\lambda_i)^2$ immediately follows from the triangle inequality.

D.3.2 Proof of Lemma 16

By the definition of $\beta(\cdot)$ in (5.45), we can express

$$\beta(\lambda_l) = \frac{1}{n} \sum_{1\leq i\leq p-r} \frac{\lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top)}{\lambda_l - \lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top)}.$$

From Lemma 21, we know that $\lambda_{i+r} \leq \lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top) \leq \lambda_i$ for each $1 \leq i \leq p-r$, and thus

$$\frac{\lambda_{i+r}}{\lambda_l - \lambda_{i+r}} \leq \frac{\lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top)}{\lambda_l - \lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top)} \leq \frac{\lambda_i}{\lambda_l - \lambda_i}, \qquad i > r.$$

Hence, we have

$$0 \leq \beta(\lambda_l) - \frac{1}{n} \sum_{r<i\leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} \leq \frac{1}{n} \sum_{1\leq i\leq r} \frac{\lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top)}{\lambda_l - \lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top)}. \tag{D.24}$$

As shown in ((5.44)), the eigenvalue λ_l satisfies $\lambda_l/\big(1+\beta(\lambda_l)\big) = \lambda_l^\star + \sigma^2 + O(\mathcal{E}_{\mathsf{PCA}})$ where $\mathcal{E}_{\mathsf{PCA}}$ is defined in (5.59). In particular, we note that for the case with $n < p$, the assumption (3.13a) guarantees that

$$\sigma^2 = o(\lambda_{\min}^\star) \qquad \text{and} \qquad \mathcal{E}_{\mathsf{PCA}} \coloneqq (\lambda_{\max}^\star + \sigma^2)\sqrt{\frac{r}{n}}\log n = o(\lambda_{\min}^\star). \tag{D.25}$$

Combined with (5.61), this also implies that

$$\lambda_l \asymp \lambda_l^\star. \tag{D.26}$$

With these estimates in place, one can use (D.24) and the high-probability bound $\|\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top\| \lesssim \sigma^2(1 + p/n) \ll \lambda_l^\star$ in (D.4) to derive

$$\left|\frac{1}{n} \sum_{r<i\leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} - \beta(\lambda_l)\right| \leq \frac{r}{n} \cdot \max_{1\leq i\leq r} \left|\frac{\lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top)}{\lambda_l - \lambda_i(\frac{1}{n}\boldsymbol{S}_\perp \boldsymbol{S}_\perp^\top)}\right| \lesssim \frac{r}{n} \cdot \frac{\sigma^2}{\lambda_l^\star}\Big(1 + \frac{p}{n}\Big) = o\Big(\frac{r}{n}\Big) \tag{D.27}$$

where the last step holds due the noise condition (3.13a). Meanwhile, we can combine (5.61) with (D.27) to find

$$\left|\frac{1}{n} \sum_{r<i\leq p} \frac{\lambda_i}{\lambda_l - \lambda_i}\right| \ll 1 \tag{D.28}$$

as long as $n \gg r$. Plugging this into ((5.44)) reveals that

$$\begin{aligned}
\lambda_l^\star + \sigma^2 &= \frac{\lambda_l}{1 + \beta(\lambda_l)} + O(\mathcal{E}_{\mathsf{PCA}}) \\
&\overset{\text{(i)}}{=} \frac{\lambda_l}{1 + \frac{1}{n}\sum_{r<i\leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} + o(\frac{r}{n})} + O(\mathcal{E}_{\mathsf{PCA}}) \\
&= \frac{\lambda_l}{1 + \frac{1}{n}\sum_{r<i\leq p} \frac{\lambda_i}{\lambda_l - \lambda_i}} + o\Big(\frac{\lambda_l^\star r}{n}\Big) + O(\mathcal{E}_{\mathsf{PCA}}) \\
&= \frac{\lambda_l}{1 + \frac{1}{n}\sum_{r<i\leq p} \frac{\lambda_i}{\lambda_l - \lambda_i}} + O(\mathcal{E}_{\mathsf{PCA}}),
\end{aligned} \tag{D.29}$$

where (i) holds due to (D.27); (ii) follows from (D.26) and (D.28); (iii) holds as long as $r \ll n$. This completes the proof for (D.15).

In addition, the claim (D.16) is an immediate consequence of (D.25) and (D.28).

D.3.3 Proof of Lemma 17

In view of (5.73) and the fact that $\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^{\top} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \lambda_l^\star + \sigma^2)$ (see (D.8)), we are motivated to first decompose

$$
\begin{aligned}
\sum_{r\le i\le n} \frac{\gamma_i^{(l)}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^{\top})^2}{(\lambda_l-\gamma_i^{(l)})^2} &= \underbrace{\sum_{r\le i\le n} \frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^{\top})^2}_{=:\,\alpha_1} \\
&\quad + \underbrace{\sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}-\frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)(\lambda_l^\star+\sigma^2)}_{=:\,\alpha_2} \\
&\quad + \underbrace{\sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}-\frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)\big\{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^{\top})^2-(\lambda_l^\star+\sigma^2)\big\}}_{=:\,\alpha_3}.
\end{aligned}
\tag{D.30}
$$

In what follows, we shall control α_1, α_2 and α_3 separately in a reverse order.

Controlling α_3. We intend to apply Lemma 20 in Section G to control α_3. Before proceeding, we pause to make a few observations. It is straightforward to compute that $f'(x) = \frac{k}{(\lambda-x)^{k+1}}$ for the function $f(x) := \frac{1}{(\lambda-x)^k}$ and $g'(x) = \frac{\lambda+(k-1)x}{(\lambda-x)^{k+1}}$ for the function $g(x) := \frac{x}{(\lambda-x)^k}$. Since $|f(x)-f(y)| \le \{\sup_z |f'(z)|\}\,|x-y|$ for any function $f(\cdot)$, one can demonstrate that: for all λ satisfying $\lambda/\big(1+\beta(\lambda)\big) \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star+\sigma^2)$, we claim that the following holds

$$
\begin{aligned}
\max_{1\le i\le n}\left|\frac{1}{\lambda-\gamma_i^{(l)}}-\frac{1}{\lambda-\sigma^2 p/n}\right| &\lesssim \left\{\max_{\gamma:\,|\gamma-\sigma^2 p/n|\lesssim\sigma^2\sqrt{p/n}} \frac{1}{(\lambda-\gamma)^2}\right\}\cdot\max_{1\le i\le n}|\gamma_i^{(l)}-\sigma^2 p/n| \\
&\lesssim \frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}},
\end{aligned}
\tag{D.31a}
$$

$$
\begin{aligned}
\max_{1\le i\le n}\left|\frac{1}{(\lambda-\gamma_i^{(l)})^2}-\frac{1}{(\lambda-\sigma^2 p/n)^2}\right| &\lesssim \left\{\max_{\gamma:\,|\gamma-\sigma^2 p/n|\lesssim\sigma^2\sqrt{p/n}} \frac{1}{(\lambda-\gamma)^3}\right\}\cdot\max_{1\le i\le n}|\gamma_i^{(l)}-\sigma^2 p/n| \\
&\lesssim \frac{\sigma^2}{\lambda_l^{\star 3}}\sqrt{\frac{p}{n}},
\end{aligned}
\tag{D.31b}
$$

$$
\begin{aligned}
\max_{1\le i\le n}\left|\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^2}-\frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^2}\right| &\lesssim \left\{\max_{\gamma:\,|\gamma-\sigma^2 p/n|\lesssim\sigma^2\sqrt{p/n}} \frac{|\lambda+\gamma|}{|\lambda-\gamma|^3}\right\}\cdot\max_{1\le i\le n}|\gamma_i^{(l)}-\sigma^2 p/n| \\
&\lesssim \frac{\lambda_l^\star}{\lambda_l^{\star 3}}\cdot\sigma^2\sqrt{\frac{p}{n}} = \frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}},
\end{aligned}
\tag{D.31c}
$$

and

$$
\begin{aligned}
\max_{1\le i\le n}\left|\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^3}-\frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^3}\right| &\lesssim \left\{\max_{\gamma:\,|\gamma-\sigma^2 p/n|\lesssim\sigma^2\sqrt{p/n}} \frac{|\lambda+2\gamma|}{(\lambda-\gamma)^4}\right\}\cdot\max_{1\le i\le n}|\gamma_i^{(l)}-\sigma^2 p/n| \\
&\lesssim \frac{\lambda_l^\star}{\lambda_l^{\star 4}}\cdot\sigma^2\sqrt{\frac{p}{n}} = \frac{\sigma^2}{\lambda_l^{\star 3}}\sqrt{\frac{p}{n}}.
\end{aligned}
\tag{D.31d}
$$

To justify the inequalities above, we have taken advantage of the following conditions:

- It is seen from (5.73) in Lemma 8 that $|\gamma_i^{(l)}-\sigma^2 p/n| \lesssim \sigma^2\sqrt{p/n}$.

- We have used the condition that for any λ, γ satisfying $\lambda/\big(1+\beta(\lambda)\big) \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star + \sigma^2)$ and $|\gamma - \sigma^2 p/n| \lesssim \sigma^2\sqrt{p/n}$, one has
$$|\lambda - \gamma| \gtrsim \lambda_l^\star;$$
this can be established via almost the same argument for justifying (5.74) in Lemma 8 (which we omit here for brevity).

With the preceding upper bounds in place, one can begin to verify the conditions required to invoke Lemma 20. First, one can derive: with probability at least $1 - O(n^{-20})$, for all λ satisfying $\lambda/\big(1+\beta(\lambda)\big) \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star + \sigma^2)$,

$$\begin{aligned}
&\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^2}\Big)\big((\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^\top)^2 - (\lambda_l^\star+\sigma^2)\big)\right| \\
&\qquad = \left|\sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^3} - \frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^3}\Big)\big((\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^\top)^2 - (\lambda_l^\star+\sigma^2)\big)\right| \\
&\qquad \le n\cdot\max_{1\le i\le n}\left|\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^3} - \frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^3}\right|\cdot\max_{1\le i\le n}\left|(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^\top)^2 - (\lambda_l^\star+\sigma^2)\right| \\
&\qquad \lesssim n\cdot\frac{\sigma^2}{\lambda_l^{\star 3}}\sqrt{\frac{p}{n}}\cdot(\lambda_l^\star+\sigma^2)^2\log n \asymp \frac{\sigma^2\sqrt{pn}\log n}{\lambda_l^\star},
\end{aligned}$$

where the last line holds due to the upper bound (D.31c), the fact $\sigma^2 \le \sigma^2 p/n \ll \lambda_l^\star$ (from the noise assumption (3.13a)), as well as the the standard Gaussian concentration inequality. In addition, one can then apply the matrix Bernstein inequality [Koltchinskii, 2011, Corollary 2.1] to conclude: for any fixed λ satisfying $\lambda/\big(1+\beta(\lambda)\big) \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star + \sigma^2)$, with probability at least $1 - O(n^{-10})$,

$$\begin{aligned}
&\left|\sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^2}\Big)\big((\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^\top)^2 - (\lambda_l^\star+\sigma^2)\big)\right| \\
&\qquad \lesssim \max_{1\le i\le n}\left|\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^2}\right|\cdot(\lambda_l^\star+\sigma^2)\cdot(\log^2 n + \sqrt{n\log n}) \\
&\qquad \lesssim \frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}}\cdot\lambda_l^\star(\log^2 n + \sqrt{n\log n}) \asymp \frac{\sigma^2\sqrt{p\log n}}{\lambda_l^\star},
\end{aligned}$$

where the second line arises from the the matrix Bernstein inequality, and the last line follows from (D.31d) and the facts $\lambda_l^\star + \sigma^2 \asymp \lambda_l^\star$ (given the assumption that $\sigma^2 \le \sigma^2 p/n \ll \lambda_l^\star$). Taking this together with the fact

$$\big\{\lambda\colon \lambda/\big(1+\beta(\lambda)\big) \in \mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star+\sigma^2)\big\} \subseteq [2\lambda_l^\star/3,\, 4\lambda_l^\star/3],$$

we can apply Lemma 20 to show that

$$|\alpha_3| \lesssim \frac{\sigma^2}{\lambda_l^\star}\sqrt{p\log n} \tag{D.32}$$

with probability exceeding $1 - O(n^{-10})$.

Controlling α_2. With regards to α_2, we claim that the following upper bound holds, whose proof is deferred to the end of this section.

$$\left|\sum_{r<i\le n}\Big(\frac{\lambda_i}{(\lambda_l-\lambda_i)^2} - \frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big) - \sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)\right| \lesssim \frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}}. \tag{D.33}$$

Combing this with $\mathcal{E}_{\mathsf{PCA}}$ defined in (5.59), we arrive at

$$\alpha_2 = \sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)(\lambda_l^\star+\sigma^2)$$

$$
\begin{aligned}
&\overset{\text{(i)}}{=} \sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}-\frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)\Big(\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}+O(\mathcal{E}_{\mathsf{PCA}})\Big)\\
&\le \sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}-\frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}\\
&\quad + n\cdot\max_{1\le i\le n}\Big|\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^2}-\frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^2}\Big|\cdot O(\mathcal{E}_{\mathsf{PCA}})\\
&\overset{\text{(ii)}}{=} \left(\sum_{r<i\le n}\Big(\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}-\frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)+O\Big(\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}}\Big)\right)\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}\\
&\quad + n\cdot\max_{1\le i\le n}\Big|\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^2}-\frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^2}\Big|\cdot O(\mathcal{E}_{\mathsf{PCA}})\\
&= \sum_{r\le i\le n}\Big(\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}-\frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}\\
&\quad + O\Big(\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}}\Big)\cdot\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}} + n\cdot\max_{1\le i\le n}\Big|\frac{\gamma_i^{(l)}}{(\lambda-\gamma_i^{(l)})^2}-\frac{\sigma^2 p/n}{(\lambda-\sigma^2 p/n)^2}\Big|\cdot O(\mathcal{E}_{\mathsf{PCA}})\\
&\overset{\text{(iii)}}{=} \sum_{r\le i\le n}\Big(\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}-\frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}\\
&\quad + O\Big(\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}}\cdot\lambda_l^\star\Big)+O\Big(\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}}\cdot n\cdot(\lambda_{\max}^\star+\sigma^2)\sqrt{\frac{r\log n}{n}}\Big)\\
&\overset{\text{(iv)}}{=} \sum_{r\le i\le n}\Big(\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}-\frac{\sigma^2 p/n}{(\lambda_l-\sigma^2 p/n)^2}\Big)\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}+O\Big(\frac{\sigma^2}{\lambda_l^\star}\kappa\sqrt{pr\log n}\Big). \qquad \text{(D.34)}
\end{aligned}
$$

Here, (i) arises from (D.15); (ii) is due to the claim (D.33); (iii) follows from (D.16), (D.31c) and the definition of $\mathcal{E}_{\mathsf{PCA}}$; (iv) holds true under the condition $\sigma^2 \ll \lambda_{\max}^\star$ (see (D.25)).

Controlling α_1. Regarding α_1, the key step lies in controlling $\sum_{r\le i\le n}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2$. Recall that $\boldsymbol{U}^{(l)}\sqrt{\boldsymbol{\Gamma}^{(l)}}\boldsymbol{V}^{(l)\top}$ is the SVD of $\frac{1}{\sqrt{n}}\boldsymbol{S}_{l,\perp}$ with $\boldsymbol{V}^{(l)} \coloneqq [\boldsymbol{v}_1^{(l)},\cdots,\boldsymbol{v}_n^{(l)}]\in\mathbb{R}^{n\times n}$. Towards this, we invoke Theorem 5 to derive the following identity:

$$
\begin{aligned}
\lambda_l &\overset{\text{(i)}}{=} \frac{1}{n}\boldsymbol{u}_l^{\star\top}\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{u}_l^\star+\frac{1}{n}\boldsymbol{u}_l^{\star\top}\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{u}_l^{\star\perp}\Big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{u}_l^{\star\perp}\Big)^{-1}\frac{1}{n}(\boldsymbol{u}_l^{\star\perp})^\top\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{u}_l^{\star\top}\\
&\overset{\text{(ii)}}{=} \frac{1}{n}\|\boldsymbol{s}_{l,\|}^\top\|_2^2+\frac{1}{n}\boldsymbol{s}_{l,\|}^\top\boldsymbol{S}_{l,\perp}^\top\Big(\lambda_l\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\|}^\top\\
&\overset{\text{(iii)}}{=} \frac{1}{n}\sum_{1\le i\le n}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2+\frac{1}{n}\sum_{1\le i\le n}\frac{\gamma_i^{(l)}}{\lambda_l-\gamma_i^{(l)}}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2\\
&= \frac{1}{n}\sum_{1\le i\le n}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2\Big(1+\frac{\gamma_i^{(l)}}{\lambda_l-\gamma_i^{(l)}}\Big)\\
&= \frac{\lambda_l}{n}\sum_{1\le i\le n}\frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{\lambda_l-\gamma_i^{(l)}},
\end{aligned}
$$

where (i) arises from (5.4b); (ii) relies on the definitions of $\boldsymbol{s}_{l,\|}$ and $\boldsymbol{S}_{l,\perp}$ in (D.15); (iii) follows since $\{\boldsymbol{v}_i^{(l)}\}_{1\le i\le n}$ forms a set of orthonormal bases in $\mathbb{R}^n$. Rearranging terms, we are left with

$$
n=\sum_{1\le i\le n}\frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{\lambda_l-\gamma_i^{(l)}}=\sum_{1\le i<r}\frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{\lambda_l-\gamma_i^{(l)}}+\sum_{r\le i\le n}\frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{\lambda_l-\gamma_i^{(l)}}
$$

$$= \sum_{1\le i<r} \frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^{\top})^2}{\lambda_l-\gamma_i^{(l)}} + \sum_{r\le i\le n} \frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^{\top})^2}{\lambda_l-\sigma^2 p/n} + \sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^{\top})^2.$$

As a result, we obtain the following decomposition:

$$\sum_{r\le i\le n} \frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^{\top})^2}{\lambda_l-\sigma^2 p/n} = n - \underbrace{\sum_{1\le i<r} \frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^{\top})^2}{\lambda_l-\gamma_i^{(l)}}}_{=:\varphi_1} - \underbrace{(\lambda_l^\star+\sigma^2)\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)}_{=:\varphi_2}$$
$$- \underbrace{\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)\Big((\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^{\top})^2-(\lambda_l^\star+\sigma^2)\Big)}_{=:\varphi_3}. \tag{D.35}$$

In what follows, we shall control φ_1, φ_2 and φ_3 separately.

- We start with φ_1. Given that $\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^{\top} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\lambda_l^\star+\sigma^2)$, we can develop an upper bound as follows: with probability at least $1-O(n^{-10})$,

$$|\varphi_1| \overset{\text{(i)}}{\le} \frac{\sum_{1\le i<r}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\parallel}^{\top})^2}{\min_{i:i\ne l}|\lambda_l^\star-\lambda_i^\star|} \overset{\text{(ii)}}{\lesssim} \frac{(\lambda_l^\star+\sigma^2) r\log n}{\min_{i:i\ne l}|\lambda_l^\star-\lambda_i^\star|} \overset{\text{(iii)}}{\asymp} \frac{\lambda_l^\star r\log n}{\min_{i:i\ne l}|\lambda_l^\star-\lambda_i^\star|}. \tag{D.36}$$

  Here, (i) utilizes the condition $|\lambda_l-\gamma_i^{(l)}| \gtrsim \min_{i:i\ne l}|\lambda_l^\star-\lambda_i^\star|$ (in view of Lemma 8), (ii) holds due to (D.10), whereas (iii) arises from the noise assumption $\sigma^2 \lesssim \lambda_l^\star$.

- As for φ_2, recall from Lemma 21 that $\lambda_{i+1}\le\gamma_i^{(l)}\le\lambda_i$ for all $1\le i<p$. This in turn leads to

$$\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big) - \sum_{r< i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big) \ge \frac{1}{\lambda_l-\gamma_n^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n};$$
$$\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big) - \sum_{r< i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big) \le \frac{1}{\lambda_l-\gamma_r^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}.$$

  This taken collectively with (D.31a) yields

$$\left|\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big) - \sum_{r< i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)\right| \lesssim \frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}}. \tag{D.37}$$

  In addition, it is also seen from (D.31a) that

$$\left|\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)\right| \vee \left|\sum_{r< i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)\right| \lesssim \frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{pn}. \tag{D.38}$$

  Therefore, we can obtain

$$\left|(\lambda_l^\star+\sigma^2)\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big) - \frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}\sum_{r< i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)\right|$$
$$\le \left|(\lambda_l^\star+\sigma^2) - \frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}\right| \cdot \left|\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)\right|$$
$$+ \frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}} \cdot \left|\sum_{r\le i\le n}\Big(\frac{1}{\lambda_l-\gamma_i^{(l)}}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big) - \sum_{r< i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)\right|$$
$$\lesssim \lambda_{\max}^\star\sqrt{\frac{r\log n}{n}}\cdot\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{pn} + \lambda_l^\star\cdot\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{\frac{p}{n}} \asymp \frac{\sigma^2}{\lambda_l^\star}\kappa\sqrt{pr\log n},$$

where the last step uses (D.15), (D.16), (D.37) and (D.38). This reveals that

$$\varphi_2 = \frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le p}\frac{\lambda_i}{\lambda_l-\lambda_i}}\sum_{r<i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2 p/n}\Big)+O\Big(\frac{\sigma^2}{\lambda_l^\star}\kappa\sqrt{pr\log n}\Big). \tag{D.39}$$

- Turning to φ_3, we shall apply Lemma 20 to bound it. Similar to the analysis above for bounding α_3, one can check that the following holds with probability at least $1-O(n^{-10})$: for all λ satisfying $\lambda/\big(1+\beta(\lambda)\big)\in\mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star+\sigma^2)$,

$$\begin{aligned}
&\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\sum_{r\le i\le n}\Big(\frac{1}{\lambda-\gamma_i^{(l)}}-\frac{1}{\lambda-\sigma^2p/n}\Big)\big((\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2-(\lambda_l^\star+\sigma^2)\big)\right|\\
&\quad=\left|\sum_{r\le i\le n}\Big(\frac{1}{(\lambda-\gamma_i^{(l)})^2}-\frac{1}{(\lambda-\sigma^2p/n)^2}\Big)\big((\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2-(\lambda_l^\star+\sigma^2)\big)\right|\\
&\quad\lesssim n\cdot\max_{r\le i\le n}\left|\frac{1}{(\lambda-\gamma_i^{(l)})^2}-\frac{1}{(\lambda-\sigma^2p/n)^2}\right|\cdot\max_{r\le i\le n}\left|(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2-(\lambda_l^\star+\sigma^2)\right|\\
&\quad\lesssim n\cdot\frac{\sigma^2}{\lambda_l^{\star 3}}\sqrt{\frac{p}{n}}\cdot(\lambda_l^\star+\sigma^2)\log n\asymp\frac{\sigma^2}{\lambda_l^{\star 2}}\sqrt{pn}\log n,
\end{aligned}$$

where the last line comes from (D.31b). In addition, for any fixed λ such that $\lambda/\big(1+\beta(\lambda)\big)\in\mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star+\sigma^2)$, we can use the matrix Bernstein inequality [Koltchinskii, 2011, Corollary 2.1] to demonstrate that: with probability at least $1-O(n^{-10})$,

$$\begin{aligned}
&\left|\sum_{r\le i\le n}\Big(\frac{1}{\lambda-\gamma_i^{(l)}}-\frac{1}{\lambda-\sigma^2p/n}\Big)\big((\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2-(\lambda_l^\star+\sigma^2)\big)\right|\\
&\quad\lesssim\max_{r\le i\le n}\left|\frac{1}{\lambda-\gamma_i^{(l)}}-\frac{1}{\lambda-\sigma^2p/n}\right|\cdot(\lambda_l^\star+\sigma^2)(\log^2 n+\sqrt{n\log n})\\
&\quad\lesssim\frac{\sigma^2}{\lambda_l^\star}\sqrt{p\log n},
\end{aligned}$$

where the last line comes from (D.31a). With these in place, we invoke Lemma 20 to conclude that

$$|\varphi_3|\lesssim\frac{\sigma^2}{\lambda_l^\star}\sqrt{p\log n} \tag{D.40}$$

with probability at least $1-O(n^{-10})$.

- Substituting (D.36), (D.39) and (D.40) into (D.35) reveals that: with probability exceeding $1-O(n^{-10})$,

$$\begin{aligned}
\sum_{r\le i\le n}\frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{\lambda_l-\sigma^2p/n}&=n-\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le n}\frac{\lambda_i}{\lambda_l-\lambda_i}}\sum_{r<i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2p/n}\Big)\\
&\quad+O\Big(\frac{\sigma^2\kappa\sqrt{pr\log n}}{\lambda_l^\star}+\frac{\lambda_l^\star r\log n}{\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star|}\Big).
\end{aligned}\tag{D.41}$$

As a consequence, we arrive at

$$\begin{aligned}
\alpha_1&=\frac{\sigma^2p/n}{\lambda_l-\sigma^2p/n}\sum_{r\le i\le n}\frac{(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^\top)^2}{\lambda_l-\sigma^2p/n}\\
&=\frac{\sigma^2p/n}{\lambda_l-\sigma^2p/n}\Big(n-\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le n}\frac{\lambda_i}{\lambda_l-\lambda_i}}\sum_{r<i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2p/n}\Big)\Big)\\
&\quad+o\Big(\frac{\sigma^2}{\lambda_l^\star}\kappa\sqrt{pr\log n}\Big)+O\Big(\frac{\sigma^2pr\log n}{\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star|\,n}\Big),
\end{aligned}\tag{D.42}$$

where we have made use of the bound $\lambda_l-\sigma^2p/n\gtrsim\lambda_l^\star$ and $\sigma^2p/n=o(\lambda_l^\star)$.

Combining the bounds on α_1, α_2 and α_3. Putting (D.30), (D.32), (D.34) and (D.42) together, we conclude

$$\begin{aligned}\sum_{r\le i\le n}\frac{\gamma_i^{(l)}(\boldsymbol{v}_i^{(l)\top}\boldsymbol{s}_{l,\|}^{\top})^2}{(\lambda_l-\gamma_i^{(l)})^2}&=\alpha_1+\alpha_2+\alpha_3\\&=\frac{\sigma^2p/n}{\lambda_l-\sigma^2p/n}\Big(n-\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le n}\frac{\lambda_i}{\lambda_l-\lambda_i}}\sum_{r<i\le n}\Big(\frac{1}{\lambda_l-\lambda_i}-\frac{1}{\lambda_l-\sigma^2p/n}\Big)\Big)\\&\quad+\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le n}\frac{\lambda_i}{\lambda_l-\lambda_i}}\sum_{r\le i\le n}\Big(\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}-\frac{\sigma^2p/n}{(\lambda_l-\sigma^2p/n)^2}\Big)\\&\quad+O\Big(\frac{\sigma^2}{\lambda_l^\star}\kappa\sqrt{pr\log n}+\frac{\sigma^2pr\log n}{\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star|\,n}\Big)\\&=\frac{\sigma^2p}{\lambda_l-\sigma^2p/n}+\frac{\lambda_l}{1+\frac{1}{n}\sum_{r<i\le n}\frac{\lambda_i}{\lambda_l-\lambda_i}}\sum_{r\le i\le n}\Big(\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}-\frac{\sigma^2p/n}{(\lambda_l-\lambda_i)(\lambda_l-\sigma^2p/n)}\Big)\\&\quad+O\Big(\frac{\sigma^2}{\lambda_l^\star}\kappa\sqrt{pr\log n}+\frac{\sigma^2pr\log n}{\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star|\,n}\Big)\end{aligned}$$

as claimed.

Proof of the inequality (D.33). Given that $\lambda_{i+1}\le\gamma_i^{(l)}\le\lambda_i$ for all $1\le i<p$, we can bound

$$\sum_{r<i\le n}\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}+\frac{\gamma_n^{(l)}}{(\lambda_l-\gamma_n^{(l)})^2}\le\sum_{r\le i\le n}\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}\le\sum_{r<i\le n}\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}+\frac{\gamma_r^{(l)}}{(\lambda_l-\gamma_r^{(l)})^2}.$$

By subtracting $\sum_{r\le i\le n}\frac{\sigma^2p/n}{(\lambda_l-\sigma^2p/n)^2}$ from both sides and rearranging terms, we have

$$\begin{aligned}&\Big|\sum_{r\le i\le n}\Big(\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}-\frac{\sigma^2p/n}{(\lambda_l-\sigma^2p/n)^2}\Big)-\sum_{r<i\le n}\Big(\frac{\lambda_i}{(\lambda_l-\lambda_i)^2}-\frac{\sigma^2p/n}{(\lambda_l-\sigma^2p/n)^2}\Big)\Big|\\&\quad\le\Big|\frac{\gamma_n^{(l)}}{(\lambda_l-\gamma_n^{(l)})^2}-\frac{\sigma^2p/n}{(\lambda_l-\sigma^2p/n)^2}\Big|\vee\Big|\frac{\gamma_r^{(l)}}{(\lambda_l-\gamma_r^{(l)})^2}-\frac{\sigma^2p/n}{(\lambda_l-\sigma^2p/n)^2}\Big|.\end{aligned}$$

In view of the basic property $|f(x)-f(y)|\le\{\sup_z|f'(z)|\}\,|x-y|$, we can upper bound

$$\begin{aligned}\Big|\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}-\frac{\sigma^2p/n}{(\lambda_l-\sigma^2p/n)^2}\Big|&\le\max_{\gamma:\,|\gamma-\sigma^2p/n|\lesssim\sigma^2\sqrt{p/n}}\Big|\frac{\lambda_l+\gamma}{(\lambda_l-\gamma)^3}\Big|\cdot|\gamma_i^{(l)}-\sigma^2p/n|\\&\lesssim\frac{\lambda_l^\star}{\lambda_l^{\star3}}\sigma^2\sqrt{\frac{p}{n}}=\frac{\sigma^2}{\lambda_l^{\star2}}\sqrt{\frac{p}{n}}\end{aligned}$$

for any $r\le i\le n$. Here, the last line holds because (i) $|\lambda_l-\lambda_l^\star|\vee\sigma^2(p/n+\sqrt{p/n})\ll\lambda_l^\star$ holds due to the assumption (3.13a), and hence $|\lambda_l-\gamma|\gtrsim\lambda_l^\star$ and $\lambda_l+\gamma\lesssim\lambda_l^\star$; (ii) $|\gamma_i^{(l)}-\sigma^2p/n|\lesssim\sigma^2\sqrt{p/n}$ holds according to Lemma 8. This finishes the proof for the inequality (D.33).

D.4 Proof of Lemma 10

Our proof strategy is to utilize the Gaussian concentration inequality and the epsilon-net argument.

To apply Lemma 20, we shall first check its conditions. To begin with, we claim that the following holds with probability at least $1-O(n^{-20})$:

$$V:=\sup_{\lambda:\,\frac{\lambda}{1+\beta(\lambda)}\in\mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_l^\star+\sigma^2)}\Big\|\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star(l)\top}\Big(\lambda\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^\top\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\Big\|_2$$

$$\lesssim \sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\right|}{\left|\lambda_{l}^{\star}-\lambda_{k}^{\star}\right|}\sqrt{\frac{\left(\lambda_{\max}^{\star}+\sigma^{2}\right)\left(\kappa^{2}+r\right)}{n}}. \tag{D.43}$$

Consequently, we can apply the Gaussian concentration inequality to show that: for any fixed λ such that $\lambda/(1+\beta(\lambda))\in\mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_{l}^{\star}+\sigma^{2})$, one has

$$\Big|\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\boldsymbol{u}_{k}^{\star(l)\top}\Big(\lambda\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\cdot\boldsymbol{s}_{l,\parallel}^{\top}\Big|\lesssim\sqrt{(\lambda_{l}^{\star}+\sigma^{2})\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_{l}^{\star}}\right)}\cdot V$$

with probability exceeding $1-O\big(\kappa^{-10}(\lambda_{\max}/\Delta_{l}^{\star})^{-20}n^{-20}\big)$.

In addition, one can derive: for all λ such that $\lambda/(1+\beta(\lambda))\in\mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_{l}^{\star}+\sigma^{2})$,

$$\begin{aligned}
&\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\cdot\boldsymbol{u}_{k}^{\star\top}\boldsymbol{u}_{l}^{\star\perp}\Big(\lambda\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\cdot\boldsymbol{s}_{l,\parallel}^{\top}\right|\\
&\quad=\left|\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\cdot\boldsymbol{u}_{k}^{\star\top}\boldsymbol{u}_{l}^{\star\perp}\Big(\lambda\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\Big)^{-2}\frac{1}{n}\boldsymbol{S}_{l,\perp}\cdot\boldsymbol{s}_{l,\parallel}^{\top}\right|\\
&\quad\leq n\cdot\max_{1\leq i<n}\frac{1}{(\lambda-\gamma_{i}^{(l)})^{2}}\cdot\Big\|\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\boldsymbol{u}_{k}^{\star\top}\boldsymbol{u}_{l}^{\star\perp}\Big\|_{2}\cdot\big\|\boldsymbol{s}_{l,\parallel}\big\|_{2}\\
&\quad\overset{\text{(i)}}{\leq} n\cdot\frac{1}{\min_{i:i\neq l}\left|\lambda_{l}^{\star}-\lambda_{i}^{\star}\right|^{2}\wedge\lambda_{l}^{\star2}}\cdot\sum_{k:k\neq l}\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\right|\big\|\boldsymbol{u}_{k}^{\star\top}\boldsymbol{u}_{l}^{\star\perp}\big\|_{2}\cdot(\lambda_{l}^{\star}+\sigma^{2})\sqrt{n\log n}\\
&\quad\overset{\text{(ii)}}{\lesssim} n^{3/2}\cdot\frac{\max_{i:i\neq l}\left|\lambda_{l}^{\star}-\lambda_{i}^{\star}\right|}{\min_{i:i\neq l}\left|\lambda_{l}^{\star}-\lambda_{i}^{\star}\right|^{2}\wedge\lambda_{l}^{\star2}}\cdot(\lambda_{l}^{\star}+\sigma^{2})\sqrt{\log n}\cdot\sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\right|}{\left|\lambda_{l}^{\star}-\lambda_{k}^{\star}\right|}\\
&\quad\overset{\text{(iii)}}{\lesssim} n^{3/2}\cdot\frac{\kappa^{2}\lambda_{\max}^{\star2}}{\Delta_{l}^{\star2}}\cdot\frac{1}{\lambda_{l}^{\star}}\cdot V
\end{aligned}$$

holds with probability at least $1-O(n^{-20})$. Here, (i) uses Lemma 8 and the high-probability fact that $\|\boldsymbol{s}_{l,\parallel}\|_{2}\lesssim(\lambda_{l}^{\star}+\sigma^{2})\sqrt{n\log n}$, (ii) holds since $\big\|\boldsymbol{u}_{k}^{\star\top}\boldsymbol{u}_{l}^{\star\perp}\big\|_{2}\leq\|\boldsymbol{u}_{k}^{\star}\|_{2}\|\boldsymbol{u}_{l}^{\star\perp}\|\leq1$, whereas (iii) arises from the definition of V in (D.43).

Combining the above two bounds, we are ready to invoke Lemma 20 and the union bound to arrive at the advertised bound

$$\begin{aligned}
\Big\|\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\boldsymbol{u}_{k}^{\star(l)\top}\Big(\lambda\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\Big)^{-1}\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{s}_{l,\parallel}^{\top}\Big\|_{2}&\lesssim\sqrt{(\lambda_{l}^{\star}+\sigma^{2})\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_{l}^{\star}}\right)}\cdot V\\
&\lesssim\sum_{k:k\neq l}\frac{\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\right|}{\left|\lambda_{l}^{\star}-\lambda_{k}^{\star}\right|\sqrt{n}}\sqrt{(\lambda_{l}^{\star}+\sigma^{2})(\lambda_{\max}^{\star}+\sigma^{2})(\kappa^{2}+r)\log\left(\frac{n\kappa\lambda_{\max}}{\Delta_{l}^{\star}}\right)}
\end{aligned}$$

with probability at least $1-O(n^{-10})$.

Therefore, the remainder of the proof amounts to establishing (D.43). Let us work under the event where the claims in Lemma 8 holds, which holds with probability exceeding $1-O(n^{-10})$. Recall the SVD of $\frac{1}{\sqrt{n}}\boldsymbol{S}_{l,\perp}=\boldsymbol{U}^{(l)}\sqrt{\boldsymbol{\Gamma}^{(l)}}\boldsymbol{V}^{(l)\top}$. Similar to (C.20), any λ such that $\lambda/(1+\beta(\lambda))\in\mathcal{B}_{\mathcal{E}_{\mathsf{PCA}}}(\lambda_{l}^{\star}+\sigma^{2})$, one can rewrite

$$\begin{aligned}
&\Big\|\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\boldsymbol{u}_{k}^{\star(l)\top}\Big(\lambda\boldsymbol{I}_{p-1}-\frac{1}{n}\boldsymbol{S}_{l,\perp}\boldsymbol{S}_{l,\perp}^{\top}\Big)^{-1}\frac{1}{\sqrt{n}}\boldsymbol{S}_{l,\perp}\Big\|_{2}\\
&\quad=\Big\|\sum_{k:k\neq l}\boldsymbol{a}^{\top}\boldsymbol{u}_{k}^{\star}\boldsymbol{u}_{k}^{\star(l)\top}\boldsymbol{U}^{(l)}\big(\lambda\boldsymbol{I}_{p-1}-\boldsymbol{\Gamma}^{(l)}\big)^{-1}\boldsymbol{U}^{(l)\top}\boldsymbol{U}^{(l)}\sqrt{\boldsymbol{\Gamma}^{(l)}}\boldsymbol{V}^{(l)\top}\Big\|_{2}
\end{aligned}$$

$$= \Big\| \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{U}^{(l)} \big(\lambda \boldsymbol{I}_{p-1} - \boldsymbol{\Gamma}^{(l)}\big)^{-1} \sqrt{\boldsymbol{\Gamma}^{(l)}} \Big\|_2$$

$$= \sqrt{\sum_{1\le i\le n\wedge(p-1)} \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} \Big(\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)} \Big)^2 }.$$

- With regards to the sum over the range $i \ge r$, it is seen from Lemma 8 and the assumption (3.13a) that for all $i \ge r$, $\gamma_i^{(l)} \lesssim \sigma^2(1+p/n) \lesssim \lambda_{\max}^\star + \sigma^2$ and $|\lambda_l - \gamma_i^{(l)}| \gtrsim \lambda_l^\star \ge |\lambda_l^\star - \lambda_k^\star|/\kappa$ for any $k \neq i$. This enables us to derive

$$\begin{aligned}
&\sqrt{\sum_{r\le i\le n\wedge(p-1)} \frac{\gamma_i^{(l)}}{(\lambda_l - \gamma_i^{(l)})^2} \Big(\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)} \Big)^2 } \\
&\quad \lesssim \sqrt{(\lambda_{\max}^\star + \sigma^2) \sum_{r\le i\le n\wedge(p-1)} \Big(\sum_{k:k\neq l} \frac{\kappa}{|\lambda_l^\star - \lambda_k^\star|} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)} \Big)^2 } \\
&\quad \le \sqrt{\lambda_{\max}^\star + \sigma^2} \Big\| \sum_{k:k\neq l} \frac{\kappa}{|\lambda_l^\star - \lambda_k^\star|} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{U}^{(l)} \Big\| \\
&\quad \le \sqrt{(\lambda_{\max}^\star + \sigma^2)\kappa^2} \sum_{k:k\neq l} \frac{|\boldsymbol{a}^\top \boldsymbol{u}_k^\star|}{|\lambda_l^\star - \lambda_k^\star|} \big\| \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{U}^{(l)} \big\|_2 \\
&\quad \le \sqrt{(\lambda_{\max}^\star + \sigma^2)\kappa^2} \sum_{k:k\neq l} \frac{|\boldsymbol{a}^\top \boldsymbol{u}_k^\star|}{|\lambda_l^\star - \lambda_k^\star|},
\end{aligned}$$

  where the last line holds since $\big\|\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{U}^{(l)}\big\|_2 \le \big\|\boldsymbol{u}_k^{\star(l)}\big\|_2 \big\|\boldsymbol{U}^{(l)}\big\| \le 1$.

- Turning to the sum over the range $i < r$, we can control

$$\begin{aligned}
\sqrt{\sum_{1\le i<r} \frac{\gamma_i^{(l)}}{(\lambda_l - \gamma_i^{(l)})^2} \Big(\sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)} \Big)^2 } &= \Big\| \sum_{1\le i<r} \Big(\frac{(\gamma_i^{(l)})^{1/2}}{\lambda_l - \gamma_i^{(l)}} \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)} \Big) \boldsymbol{u}_i^{(l)} \Big\|_2 \\
&= \Big\| \sum_{k:k\neq l} \boldsymbol{a}^\top \boldsymbol{u}_k^\star \sum_{1\le i<r} \frac{(\gamma_i^{(l)})^{1/2} \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda_l - \gamma_i^{(l)}} \boldsymbol{u}_i^{(l)} \Big\|_2 \\
&\le \sum_{k:k\neq l} |\boldsymbol{a}^\top \boldsymbol{u}_k^\star| \Big\| \sum_{1\le i<r} \frac{(\gamma_i^{(l)})^{1/2} \boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)}}{\lambda_l - \gamma_i^{(l)}} \boldsymbol{u}_i^{(l)} \Big\|_2 \\
&= \sum_{k:k\neq l} |\boldsymbol{a}^\top \boldsymbol{u}_k^\star| \sqrt{\sum_{1\le i<r} \frac{\gamma_i^{(l)} (\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)})^2}{(\lambda_l - \gamma_i^{(l)})^2} }.
\end{aligned}$$

  This leads us to control $\sum_{1\le i\le r-1} \gamma_i^{(l)} (\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)})^2/(\lambda_l - \gamma_i^{(l)})^2$ for each $k \neq l$, which can be decomposed as follows

$$\sum_{1\le i<r} \frac{\gamma_i^{(l)} (\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)})^2}{(\lambda_l - \gamma_i^{(l)})^2} = \sum_{i\in\mathcal{C}_1} \frac{\gamma_i^{(l)} (\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)})^2}{(\lambda_l - \gamma_i^{(l)})^2} + \sum_{i\in\mathcal{C}_2} \frac{\gamma_i^{(l)} (\boldsymbol{u}_k^{\star(l)\top} \boldsymbol{u}_i^{(l)})^2}{(\lambda_l - \gamma_i^{(l)})^2}.$$

  Here, the sets $\mathcal{C}_1$ and $\mathcal{C}_2$ are defined respectively as follows

$$\mathcal{C}_1 := \{1 \le i < r \mid \gamma_i^{(l)}/(1+\gamma(\gamma_i^{(l)})) \in \mathcal{B}_{\mathcal{E}_k}(\lambda_k^\star)\},$$
$$\mathcal{C}_2 := \{1 \le i < r \mid \gamma_i^{(l)}/(1+\gamma(\gamma_i^{(l)})) \notin \mathcal{B}_{\mathcal{E}_k}(\lambda_k^\star)\},$$

  where we take $\mathcal{E}_k := c\,|\lambda_l^\star - \lambda_k^\star|$ for some sufficiently small constant $c > 0$. In the sequel, we shall control the above two sums separately.

– With respect to the sum over $\mathcal{C}_1$, one can apply a similar argument for (C.23) to show $|\lambda_l - \gamma_i^{(l)}| \gtrsim |\lambda_l^\star - \lambda_k^\star|$ for $i \in \mathcal{C}_1$. This enables us to bound

$$\sum_{i\in\mathcal{C}_1} \frac{\gamma_i^{(l)}\big(\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_i^{(l)}\big)^2}{(\lambda_l-\gamma_i^{(l)})^2} \lesssim \frac{\lambda_{\max}^\star+\sigma^2}{|\lambda_l^\star-\lambda_k^\star|^2}\sum_{i\in\mathcal{C}_1}\big(\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_i^{(l)}\big)^2 \le \frac{\lambda_{\max}^\star+\sigma^2}{|\lambda_l^\star-\lambda_k^\star|^2}\|\boldsymbol{u}_k^{\star(l)}\|_2^2\|\boldsymbol{U}^{(l)}\|^2 \le \frac{\lambda_{\max}^\star+\sigma^2}{|\lambda_l^\star-\lambda_k^\star|^2}.$$

– Next, we move on to look at the sum over $\mathcal{C}_2$. According to Lemma 8, we have

$$\begin{aligned}
\mathcal{E}_{\mathsf{PCA}} &\gtrsim \Big\|\big(\gamma_i^{(l)}\boldsymbol{I}_{r-1} - (1+\beta(\gamma_i^{(l)}))\boldsymbol{\Lambda}^{(l)}\big)\boldsymbol{U}^{\star(l)\top}\boldsymbol{u}_{i,\parallel}^{(l)}\Big\|_2\\
&\ge \big|\gamma_i^{(l)} - (1+\beta(\gamma_i^{(l)}))\lambda_k^{(l)}\big| \cdot \big|\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_{i,\parallel}^{(l)}\big|\\
&\gtrsim \mathcal{E}_k \cdot \big|\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_{i,\parallel}^{(l)}\big|\\
&\gtrsim |\lambda_l^\star - \lambda_k^\star| \cdot \big|\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_i^{(l)}\big|,
\end{aligned}$$

where we use the fact that $\big|\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_i^{(l)}\big| \le \big|\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_{i,\parallel}^{(l)}\big|$ and $\big|\gamma_i^{(l)} - (1+\beta(\gamma_i^{(l)}))\lambda_k^{(l)}\big| \gtrsim \mathcal{E}_k$ for all $i \in \mathcal{C}_2$. Therefore, we arrive at the upper bound

$$\frac{\gamma_i^{(l)}(\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_i^{(l)})^2}{(\lambda_l-\gamma_i^{(l)})^2} \lesssim \frac{(\lambda_{\max}^\star+\sigma^2)\mathcal{E}_{\mathsf{PCA}}^2}{|\lambda_l^\star-\lambda_k^\star|^2\min_{i:i\neq l}|\lambda_l^\star-\lambda_i^\star|^2} \lesssim \frac{\lambda_{\max}^\star+\sigma^2}{|\lambda_l^\star-\lambda_k^\star|^2}, \qquad i\in\mathcal{C}_2,$$

where we invoke the condition $\min_{i:i\neq l}|\lambda_l^\star - \lambda_i^\star| \gtrsim \mathcal{E}_{\mathsf{PCA}}$. Taking these two bounds collectively, we reach

$$\sum_{1\le i<r} \frac{\gamma_i^{(l)}\big(\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_i^{(l)}\big)^2}{(\lambda_l-\gamma_i^{(l)})^2} \lesssim \frac{\lambda_{\max}^\star+\sigma^2}{|\lambda_l^\star-\lambda_k^\star|^2} + \frac{\big(\lambda_{\max}^\star+\sigma^2\big)\,r}{|\lambda_l^\star-\lambda_k^\star|^2} \asymp \frac{\big(\lambda_{\max}^\star+\sigma^2\big)\,r}{|\lambda_l^\star-\lambda_k^\star|^2},$$

and hence

$$\sqrt{\sum_{1\le i<r}\frac{\gamma_i^{(l)}}{(\lambda_l-\gamma_i^{(l)})^2}\Big(\sum_{k:k\neq l}\boldsymbol{a}^\top\boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star(l)\top}\boldsymbol{u}_i^{(l)}\Big)^2} \lesssim \sqrt{(\lambda_{\max}^\star+\sigma^2)\,r}\sum_{k:k\neq l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star-\lambda_k^\star|}.$$

- Combining the preceding two bounds, we finish the proof for (D.43).

E Proof for minimax lower bounds (Theorem 4)

Fix an arbitrary $1 \le l \le r$ and an arbitrary $k \neq l$ and $1 \le k \le r$. In what follows, we intend to prove the following two claims:

$$\inf_{u_{\boldsymbol{a},l}}\sup_{\boldsymbol{\Sigma}\in\mathcal{M}_1(\boldsymbol{\Sigma}^\star)}\mathbb{E}\Big[\min\big|u_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l(\boldsymbol{\Sigma})\big|\Big] \gtrsim \frac{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}{|\lambda_l^\star-\lambda_k^\star|^2\,n}\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big| + \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big|, \tag{E.1}$$

$$\inf_{u_{\boldsymbol{a},l}}\sup_{\boldsymbol{\Sigma}\in\mathcal{M}_2(\boldsymbol{\Sigma}^\star)}\mathbb{E}\Big[\min\big|u_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l(\boldsymbol{\Sigma})\big|\Big] \gtrsim \sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star2}n}}\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\|_2, \tag{E.2}$$

where the infimum is over all estimators, and $\mathcal{M}_1(\boldsymbol{\Sigma}^\star)$ and $\mathcal{M}_2(\boldsymbol{\Sigma}^\star)$ are defined right before the statement of Theorem 4. It is self-evident that Theorem 4 follows from these two claims by taking the maximum over all $k \neq l$.

E.1 Proof of the lower bound (E.1)

Step 1: constructing a collection of hypotheses. Let us consider the following two hypotheses:

$$\mathcal{H}_0 \,:\, \boldsymbol{s}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}^\star + \sigma^2 \boldsymbol{I}_p), \quad 1 \le i \le n;$$
$$\mathcal{H}_k \,:\, \boldsymbol{s}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_k + \sigma^2 \boldsymbol{I}_p), \quad 1 \le i \le n.$$

Here, the covariance matrix $\boldsymbol{\Sigma}_k$ is defined as follows:

$$\boldsymbol{\Sigma}_k \coloneqq \lambda_l^\star \boldsymbol{u}_l \boldsymbol{u}_l^\top + \lambda_k^\star \boldsymbol{u}_k \boldsymbol{u}_k^\top + \sum_{i:\, i \neq k,l,\, 1 \le i \le r} \lambda_i^\star \boldsymbol{u}_i^\star \boldsymbol{u}_i^{\star\top}.$$

where $\boldsymbol{u}_l$ and $\boldsymbol{u}_k$ are defined as

$$[\boldsymbol{u}_l,\, \boldsymbol{u}_k] \coloneqq [\boldsymbol{u}_l^\star,\, \boldsymbol{u}_k^\star] \begin{bmatrix} \cos\theta_n & -\sin\theta_n \\ \sin\theta_n & \cos\theta_n \end{bmatrix} \tag{E.3}$$

for some $\theta_n \in [-\pi/2, \pi/2]$ to be specified later. Straightforward calculation yields

$$\boldsymbol{u}_l \boldsymbol{u}_l^\top + \boldsymbol{u}_k \boldsymbol{u}_k^\top = \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} + \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top}.$$

This identity further leads to

$$\begin{aligned} \boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}^\star &= \lambda_l^\star \boldsymbol{u}_l \boldsymbol{u}_l^\top + \lambda_k^\star \boldsymbol{u}_k \boldsymbol{u}_k^\top - (\lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} + \lambda_k^\star \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top}) \\ &= \lambda_l^\star (\boldsymbol{u}_l \boldsymbol{u}_l^\top + \boldsymbol{u}_k \boldsymbol{u}_k^\top) + (\lambda_k^\star - \lambda_l^\star) \boldsymbol{u}_k \boldsymbol{u}_k^\top - \big(\lambda_l^\star (\boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} + \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top}) + (\lambda_k^\star - \lambda_l^\star) \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top}\big) \\ &= (\lambda_k^\star - \lambda_l^\star)(\boldsymbol{u}_k \boldsymbol{u}_k^\top - \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top}). \end{aligned}$$

In addition, it is also seen that

$$\begin{aligned} \|\boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}^\star\|_{\mathrm{F}} &= |\lambda_k^\star - \lambda_l^\star| \cdot \|\boldsymbol{u}_k \boldsymbol{u}_k^\top - \boldsymbol{u}_k^\star \boldsymbol{u}_k^{\star\top}\|_{\mathrm{F}} \le |\lambda_k^\star - \lambda_l^\star| \cdot \big(\|\boldsymbol{u}_k (\boldsymbol{u}_k - \boldsymbol{u}_k^\star)^\top\|_{\mathrm{F}} + \|(\boldsymbol{u}_k - \boldsymbol{u}_k^\star) \boldsymbol{u}_k^{\star\top}\|_{\mathrm{F}}\big) \\ &= |\lambda_k^\star - \lambda_l^\star| \cdot \big(\|\boldsymbol{u}_k\|_2 \|\boldsymbol{u}_k - \boldsymbol{u}_k^\star\|_2 + \|\boldsymbol{u}_k - \boldsymbol{u}_k^\star\|_2 \|\boldsymbol{u}_k^\star\|_2\big) \\ &= 2\,|\lambda_k^\star - \lambda_l^\star| \cdot \|\boldsymbol{u}_k - \boldsymbol{u}_k^\star\|_2 \\ &\overset{\text{(i)}}{=} 2\,|\lambda_k^\star - \lambda_l^\star| \cdot \| - \boldsymbol{u}_l^\star \sin\theta_n + \boldsymbol{u}_k^\star \cos\theta_n - \boldsymbol{u}_k^\star\|_2 \\ &\le 2\,|\lambda_k^\star - \lambda_l^\star| \cdot \big(\sin\theta_n + 2\sin^2(\theta_n/2)\big) \\ &\overset{\text{(ii)}}{\le} 4\,|\lambda_k^\star - \lambda_l^\star| \cdot |\theta_n|, \end{aligned} \tag{E.4}$$

where (i) arises from the definition of $\boldsymbol{u}_l$ in (E.3); (ii) holds since $\sin\theta \le |\theta|$.

In what follows, we denote by $\mathbb{P}^0$ and $\mathbb{P}^k$ the distribution of $\boldsymbol{S}$ under the hypothesis $\mathcal{H}_0$ and $\mathcal{H}_k$, respectively, and let $\mathbb{P}_i^0$ and $\mathbb{P}_i^k$ denote the distribution of $\boldsymbol{s}_i$ (i-th column of $\boldsymbol{S}$) under $\mathcal{H}_0$ and $\mathcal{H}_k$, respectively.

Step 2: bounding the KL divergence between hypotheses. Recall the elementary fact that the KL divergence of multivariate Gaussians is given by [Kullback et al., 1952]

$$\mathsf{KL}\big(\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_1) \,\|\, \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_0)\big) = \frac{1}{2}\bigg(\mathsf{tr}\big(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1\big) - p + \log\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|}\bigg).$$

Since the KL divergence is additive over independent distributions [Tsybakov, 2009], one has

$$\mathsf{KL}\big(\mathbb{P}^k \,\|\, \mathbb{P}^0\big) = \sum_{i=1}^n \mathsf{KL}\big(\mathbb{P}_i^k \,\|\, \mathbb{P}_i^0\big) = \frac{1}{2} \sum_{i=1}^n \big(\mathsf{tr}\big((\boldsymbol{\Sigma}^\star + \sigma^2 \boldsymbol{I}_p)^{-1} (\boldsymbol{\Sigma}_k + \sigma^2 \boldsymbol{I}_p)\big) - p\big). \tag{E.5}$$

This suggests that we need to compute $\mathsf{tr}\big((\boldsymbol{\Sigma}^\star + \sigma^2 \boldsymbol{I}_p)^{-1} (\boldsymbol{\Sigma}_k + \sigma^2 \boldsymbol{I}_p)\big)$. By construction in (E.3), we know that $\boldsymbol{u}_l$ and $\boldsymbol{u}_k$ span the same subspace as $\boldsymbol{u}_l^\star$ and $\boldsymbol{u}_k^\star$, and are orthogonal to $\{\boldsymbol{u}_i^\star\}_{i: i \neq k,l}$. Denote by

$\boldsymbol{U}^{\star\perp} \in \mathbb{R}^{p\times(p-r)}$ the matrix whose columns form an orthonormal basis of the complement to the subspace spanned by $\boldsymbol{U}^{\star}$. One can then derive

$$
\begin{aligned}
(\boldsymbol{\Sigma}^{\star}+\sigma^2\boldsymbol{I}_p)^{-1}(\boldsymbol{\Sigma}_k+\sigma^2\boldsymbol{I}_p) &= \Bigg(\sum_{1\le i\le r}\frac{1}{\lambda_i^{\star}+\sigma^2}\boldsymbol{u}_i^{\star}\boldsymbol{u}_i^{\star\top}+\frac{1}{\sigma^2}\boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^{\top}\Bigg)\\
&\quad\cdot\Bigg((\lambda_l^{\star}+\sigma^2)\boldsymbol{u}_l\boldsymbol{u}_l^{\top}+(\lambda_k^{\star}+\sigma^2)\boldsymbol{u}_k\boldsymbol{u}_k^{\top}+\sum_{i:\,i\ne k,l,\,1\le i\le r}(\lambda_i^{\star}+\sigma^2)\boldsymbol{u}_i^{\star}\boldsymbol{u}_i^{\star\top}+\sigma^2\boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^{\top}\Bigg)\\
&= \Big(\frac{1}{\lambda_l^{\star}+\sigma^2}\boldsymbol{u}_l^{\star}\boldsymbol{u}_l^{\star\top}+\frac{1}{\lambda_k^{\star}+\sigma^2}\boldsymbol{u}_k^{\star}\boldsymbol{u}_k^{\star\top}\Big)\Big((\lambda_l^{\star}+\sigma^2)\boldsymbol{u}_l\boldsymbol{u}_l^{\top}+(\lambda_k^{\star}+\sigma^2)\boldsymbol{u}_k\boldsymbol{u}_k^{\top}\Big)\\
&\quad+\sum_{i:\,i\ne k,l,\,1\le i\le r}\boldsymbol{u}_i^{\star}\boldsymbol{u}_i^{\star\top}+\boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^{\top}.
\end{aligned}
$$

As a result, we find

$$
\begin{aligned}
&\mathsf{tr}\big((\boldsymbol{\Sigma}^{\star}+\sigma^2\boldsymbol{I}_p)^{-1}(\boldsymbol{\Sigma}_k+\sigma^2\boldsymbol{I}_p)\big)\\
&\overset{\text{(i)}}{=}\mathsf{tr}\Bigg(\Big(\frac{1}{\lambda_l^{\star}+\sigma^2}\boldsymbol{u}_l^{\star}\boldsymbol{u}_l^{\star\top}+\frac{1}{\lambda_k^{\star}+\sigma^2}\boldsymbol{u}_k^{\star}\boldsymbol{u}_k^{\star\top}\Big)\Big((\lambda_l^{\star}+\sigma^2)\boldsymbol{u}_l\boldsymbol{u}_l^{\top}+(\lambda_k^{\star}+\sigma^2)\boldsymbol{u}_k\boldsymbol{u}_k^{\top}\Big)\Bigg)+p-2\\
&=|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l|^2+\frac{\lambda_k^{\star}+\sigma^2}{\lambda_l^{\star}+\sigma^2}|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_k|^2+\frac{\lambda_l^{\star}+\sigma^2}{\lambda_k^{\star}+\sigma^2}|\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l|^2+|\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_k|^2+p-2\\
&\overset{\text{(ii)}}{=}\cos^2\theta_n+\frac{\lambda_k^{\star}+\sigma^2}{\lambda_l^{\star}+\sigma^2}\sin^2\theta_n+\frac{\lambda_l^{\star}+\sigma^2}{\lambda_k^{\star}+\sigma^2}\sin^2\theta_n+\cos^2\theta_n+p-2\\
&=\frac{(\lambda_l^{\star}+\sigma^2)^2+(\lambda_k^{\star}+\sigma^2)^2}{(\lambda_l^{\star}+\sigma^2)(\lambda_k^{\star}+\sigma^2)}\sin^2\theta_n-2\sin^2\theta_n+p\\
&=\frac{(\lambda_l^{\star}-\lambda_k^{\star})^2}{(\lambda_l^{\star}+\sigma^2)(\lambda_k^{\star}+\sigma^2)}\sin^2\theta_n+p.
\end{aligned}
$$

Here, (i) holds since $\mathsf{tr}(\boldsymbol{u}_i^{\star}\boldsymbol{u}_i^{\star\top})=1$ and $\mathsf{tr}\big(\boldsymbol{U}^{\star\perp}(\boldsymbol{U}^{\star\perp})^{\top}\big)=\mathsf{tr}\big((\boldsymbol{U}^{\star\perp})^{\top}\boldsymbol{U}^{\star\perp}\big)=\mathsf{tr}(\boldsymbol{I}_{p-r})=p-r$; (ii) follows from the following observations:

$$
\begin{aligned}
\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l&=\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l^{\star}\cos\theta_n+\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_k^{\star}\sin\theta_n=\cos\theta_n;\\
\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l&=\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star}\cos\theta_n+\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_k^{\star}\sin\theta_n=\sin\theta_n;\\
\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_k&=-\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l^{\star}\sin\theta_n+\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_k^{\star}\cos\theta_n=-\sin\theta_n;\\
\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_k&=-\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l^{\star}\sin\theta_n+\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_k^{\star}\cos\theta_n=\cos\theta_n;
\end{aligned}
$$

where we have used the construction (E.3) and the fact that $\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_k^{\star}=0$. Therefore, combining the above identities allows us to conclude that

$$
\mathsf{KL}\big(\mathbb{P}^k\,\|\,\mathbb{P}^0\big)=\frac{n(\lambda_l^{\star}-\lambda_k^{\star})^2}{2(\lambda_l^{\star}+\sigma^2)(\lambda_k^{\star}+\sigma^2)}\sin^2\theta_n. \tag{E.6}
$$

Step 3: invoking Fano's inequality. Suppose that we choose θ_n

$$
|\theta_n|=c_n\sqrt{\frac{(\lambda_l^{\star}+\sigma^2)(\lambda_k^{\star}+\sigma^2)}{(\lambda_l^{\star}-\lambda_k^{\star})^2 n}} \tag{E.7}
$$

where $c_n\asymp 1$ is a sequence that depends on n and obeys $c_n\in\{1/64,\,1/16,\,1/4\}$ (which we shall discuss momentarily). Then we can see from (E.4) that

$$
\|\boldsymbol{\Sigma}_k-\boldsymbol{\Sigma}^{\star}\|_{\mathrm{F}}\le\sqrt{\frac{(\lambda_l^{\star}+\sigma^2)(\lambda_k^{\star}+\sigma^2)}{n}}.
$$

In other words, $\boldsymbol{\Sigma}_k \in \mathcal{M}_1(\boldsymbol{\Sigma}^\star)$. Moreover, plugging the value (E.7) of θ_n into (E.6) and using the facts $|\sin\theta| \le |\theta|$ as well as $\max_n c_n = 1/4$ yields

$$\mathsf{KL}(\mathbb{P}^k \parallel \mathbb{P}^0) \le 1/16.$$

It then follows from Fano's inequality [Tsybakov, 2009, Theorem 2] that

$$p_{e,k} := \inf_\psi \max\Big\{\mathbb{P}\{\psi \text{ rejects } \mathcal{H}_0 \mid \mathcal{H}_0\}, \mathbb{P}\{\psi \text{ rejects } \mathcal{H}_k \mid \mathcal{H}_k\}\Big\} \ge 1/5,$$

where the infimum is taken over all tests. One can then apply the standard reduction scheme in [Tsybakov, 2009, Chapter 2.2] to show that

$$\inf_{u_{\boldsymbol{a},l}} \sup_{\boldsymbol{\Sigma}\in\mathcal{M}_1(\boldsymbol{\Sigma}^\star)} \mathbb{E}\Big[\min\big|u_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{\Sigma})\big|\Big] \gtrsim p_{e,k} \min\big|\boldsymbol{a}^\top\boldsymbol{u}_l \pm \boldsymbol{a}^\top\boldsymbol{u}_l^\star\big| \gtrsim \min\big|\boldsymbol{a}^\top\boldsymbol{u}_l \pm \boldsymbol{a}^\top\boldsymbol{u}_l^\star\big|.$$

Observe that once we prove

$$\min\big|\boldsymbol{a}^\top\boldsymbol{u}_l \pm \boldsymbol{a}^\top\boldsymbol{u}_l^\star\big| \ge \frac{1}{8\pi^2}\big(\theta_n^2\cdot\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big| + |\theta_n|\cdot\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big|\big), \tag{E.8}$$

then (E.7) would immediately lead to the advertised bound

$$\begin{aligned}\inf_{u_{\boldsymbol{a},l}} \sup_{\boldsymbol{\Sigma}\in\mathcal{M}_1(\boldsymbol{\Sigma}^\star)} \mathbb{E}\Big[\min\big|u_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{\Sigma})\big|\Big] &\gtrsim c_n \frac{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big|}{(\lambda_l^\star-\lambda_k^\star)^2 n} + c_n \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big|}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}} \\ &\gtrsim \frac{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big|}{(\lambda_l^\star-\lambda_k^\star)^2 n} + \frac{\sqrt{(\lambda_k^\star+\sigma^2)(\lambda_l^\star+\sigma^2)}\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big|}{|\lambda_l^\star-\lambda_k^\star|\sqrt{n}}\end{aligned}$$

where the last step holds since $\min_n c_n = 1/64$. As a consequence, the remainder of the proof amounts to establishing the claim (E.8). In view of (E.3), we know that

$$\begin{aligned}\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star - \boldsymbol{a}^\top\boldsymbol{u}_l\big| &= \big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star - \boldsymbol{a}^\top(\boldsymbol{u}_l^\star\cos\theta_n + \boldsymbol{u}_k^\star\sin\theta_n)\big| \\ &= \big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star(1-\cos\theta_n) - \boldsymbol{a}^\top\boldsymbol{u}_k^\star\sin\theta_n\big| \\ &= \big|2\boldsymbol{a}^\top\boldsymbol{u}_l^\star\sin^2(\theta_n/2) - \boldsymbol{a}^\top\boldsymbol{u}_k^\star\sin\theta_n\big| \\ &\overset{\text{(i)}}{=} \big|2\boldsymbol{a}^\top\boldsymbol{u}_l^\star\sin^2(\theta_n/2)\big| + \big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\sin\theta_n\big| \\ &\overset{\text{(ii)}}{\ge} \frac{2}{\pi^2}\big(\theta_n^2\cdot|\boldsymbol{a}^\top\boldsymbol{u}_l^\star| + |\theta_n|\cdot\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big|\big)\end{aligned} \tag{E.9}$$

where (i) holds true as long as we choose $\mathsf{sign}(\theta_n) = -\mathsf{sign}(\boldsymbol{a}^\top\boldsymbol{u}_l^\star/\boldsymbol{a}^\top\boldsymbol{u}_k^\star)$; (ii) relies on the fact $|\sin\theta| \ge \frac{2}{\pi}|\theta|$ for $\theta\in[-\frac{\pi}{2},\frac{\pi}{2}]$. In addition, we can derive

$$\begin{aligned}\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star + \boldsymbol{a}^\top\boldsymbol{u}_l\big| &= \big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star + \boldsymbol{a}^\top(\boldsymbol{u}_l^\star\cos\theta_n + \boldsymbol{u}_k^\star\sin\theta_n)\big| \\ &= \big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star(1+\cos\theta_n) + \boldsymbol{a}^\top\boldsymbol{u}_k^\star\sin\theta_n\big| \\ &= \big|2\boldsymbol{a}^\top\boldsymbol{u}_l^\star\cos^2(\theta_n/2) + \boldsymbol{a}^\top\boldsymbol{u}_k^\star\sin(\theta_n/2)\cos(\theta_n/2)\big| \\ &= \cos(\theta_n/2)\big|2\boldsymbol{a}^\top\boldsymbol{u}_l^\star\cos(\theta_n/2) + \boldsymbol{a}^\top\boldsymbol{u}_k^\star\sin(\theta_n/2)\big| \\ &\overset{\text{(i)}}{\ge} \frac{1}{2}\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\cos(\theta_n/2) + \boldsymbol{a}^\top\boldsymbol{u}_k^\star\sin(\theta_n/2)\big| \\ &\overset{\text{(ii)}}{=} \frac{1}{2}\Big|\sqrt{(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2 + (\boldsymbol{a}^\top\boldsymbol{u}_k^\star)^2}\sin(\theta_n/2+\omega_k)\Big| \\ &\ge \frac{1}{4}\left(|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\sin|\theta_n/2+\omega_k| + |\boldsymbol{a}^\top\boldsymbol{u}_k^\star|\sin|\theta_n/2+\omega_k|\right)\end{aligned} \tag{E.10}$$

where (i) holds due to $|\theta_n| \le 1/4$ by the choice of c_n in (E.7) and the sample size condition (3.20); $\omega_k \in [-\frac{\pi}{2},\frac{\pi}{2}]$ in (ii) is defined such that $\tan\omega_k = \boldsymbol{a}^\top\boldsymbol{u}_l^\star/\boldsymbol{a}^\top\boldsymbol{u}_k^\star$. In particular, recall that the sign of θ_n is chosen such that $\mathsf{sign}(\theta_n) = -\mathsf{sign}(\boldsymbol{a}^\top \boldsymbol{u}_l^\star / \boldsymbol{a}^\top \boldsymbol{u}_k^\star)$, one has $\mathsf{sign}(\theta_n) = -\mathsf{sign}(\omega_k)$. Next, our goal is to show if we choose $c_n \in \{1/64,\ 1/16,\ 1/4\}$ of θ_n in (E.7) suitably, one has

$$\sin|\theta_n/2 + \omega_k| \geq \frac{1}{2\pi}|\theta_n|. \tag{E.11}$$

To this end, for each n, we choose c_n of θ_n in (E.7) to be $c_n = 1/16$ temporarily, and consider the following three scenarios:

- If $|\theta_n|/2 \geq 2\,|\omega_k|$, then one has $\pi/2 \geq |\theta_n/2 + \omega_k| \geq |\theta_n|/2 - |\omega_k| \geq |\theta_n/4|$ where the first inequality holds since the signs of θ_n and ω_k are different. Combined with the inequality $|\sin\theta| \geq \frac{2}{\pi}|\theta|$ for $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, this leads to $\sin|\theta_n/2 + \omega_k| \geq \sin|\theta_n/4| \geq \frac{1}{2\pi}|\theta_n|$;
- If $|\theta_n|/2| \leq |\omega_k|/2$, then we know $\pi/2 \geq |\theta_n/2 + \omega_k| \geq |\omega_k| - |\theta_n|/2 \geq |\omega_k/2| \geq |\theta_n/2|$. This implies that $\sin|\theta_n/2 + \omega_k| \geq \sin|\theta_n/2| \geq \frac{1}{\pi}|\theta_n|$.
- Otherwise, (i.e. $|\omega_k|/2 < |\theta_n|/2 < 2\,|\omega_k|$), one can adjust c_n to be either $1/4$ or $1/64$ (namely, increasing it or decreasing it by 4 times). After doing so, it is easily seen that θ_n must satisfy one of the two conditions above, thereby guaranteeing that $\sin|\theta_n/2 + \omega_k| \geq \frac{1}{2\pi}|\theta_n|$.

This completes the proof for the claim (E.11). Combining (E.11) with (E.10), we arrive at

$$\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \boldsymbol{a}^\top \boldsymbol{u}_l\right| \geq \frac{1}{8\pi}\big(|\theta_n| \cdot |\boldsymbol{a}^\top \boldsymbol{u}_l^\star| + |\theta_n| \cdot |\boldsymbol{a}^\top \boldsymbol{u}_k^\star|\big) \geq \frac{1}{8\pi^2}\big(\theta_n^2 \cdot |\boldsymbol{a}^\top \boldsymbol{u}_l^\star| + |\theta_n| \cdot |\boldsymbol{a}^\top \boldsymbol{u}_k^\star|\big),$$

where the last step holds since $|\theta_n| \leq 1$. Combining this with (E.9) finishes the proof of the claim (E.8).

E.2 Proof of the lower bound (E.2)

Step 1: constructing a collection of hypotheses. Consider the following hypotheses regarding the eigen-decomposition of the covariance matrix:

$$\begin{aligned}
\mathcal{H}_0 &: \boldsymbol{s}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}^\star + \sigma^2 \boldsymbol{I}_p), \quad 1 \leq i \leq n;\\
\mathcal{H}_1 &: \boldsymbol{s}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}} + \sigma^2 \boldsymbol{I}_p), \quad 1 \leq i \leq n.
\end{aligned}$$

Here, the covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ is defined to be

$$\widetilde{\boldsymbol{\Sigma}} \coloneqq \lambda_l^\star \widetilde{\boldsymbol{u}}_l \widetilde{\boldsymbol{u}}_l^\top + \sum_{i:i\neq l} \lambda_i^\star \boldsymbol{u}_i^\star \boldsymbol{u}_i^{\star\top},$$

where $\widetilde{\boldsymbol{u}}_l$ is defined as

$$\widetilde{\boldsymbol{u}}_l \coloneqq \frac{\boldsymbol{u}_l^\star + \delta_n \boldsymbol{a}_\perp}{\|\boldsymbol{u}_l^\star + \delta_n \boldsymbol{a}_\perp\|_2} = \frac{\boldsymbol{u}_l^\star + \delta_n \boldsymbol{a}_\perp}{\sqrt{1+\delta_n^2}} \qquad \text{with} \quad \boldsymbol{a}_\perp \coloneqq \frac{\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}}{\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2}$$

for some $0 < \delta_n < 1$ to be specified later. We note that $\boldsymbol{u}_i^{\star\top} \boldsymbol{a}_\perp = 0$ for all $1 \leq i \leq r$ and $\boldsymbol{u}_i^{\star\top} \widetilde{\boldsymbol{u}}_l = 0$ for all $i \neq l$. As can be straightforwardly verified, one has

$$\|\widetilde{\boldsymbol{u}} - \boldsymbol{u}_l^\star\|_2 \leq \Big(1 - \frac{1}{\sqrt{1+\delta_n^2}}\Big)\|\boldsymbol{u}_l^\star\|_2 + \frac{\delta_n}{\sqrt{1+\delta_n^2}}\|\boldsymbol{a}_\perp\|_2 = \frac{\sqrt{1+\delta_n^2} - 1 + \delta_n}{\sqrt{1+\delta_n^2}} \leq 2\delta_n, \tag{E.12}$$

where the last step holds since $\sqrt{1+\delta_n^2} \leq 1 + \delta_n$ for $\delta_n > 0$.

In the sequel, we denote by $\mathbb{P}^0$ and $\mathbb{P}^1$ the distribution of $\boldsymbol{S}$ under the hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. We also let $\mathbb{P}_i^0$ and $\mathbb{P}_i^1$ denote the distribution of $\boldsymbol{s}_i$ (i-th column of $\boldsymbol{S}$) under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively.

Step 2: bounding the KL divergence between hypotheses. Let us define vector $\widehat{\boldsymbol{u}}_l$ as

$$\widehat{\boldsymbol{u}}_l := \frac{\boldsymbol{u}_l^\star - \frac{1}{\delta_n}\boldsymbol{a}_\perp}{\|\boldsymbol{u}_l^\star - \frac{1}{\delta_n}\boldsymbol{a}_\perp\|_2} = \frac{\boldsymbol{u}_l^\star - \frac{1}{\delta_n}\boldsymbol{a}_\perp}{\sqrt{1+\frac{1}{\delta_n^2}}}.$$

where the last step holds since $\boldsymbol{u}_l^\star$ is orthogonal to $\boldsymbol{a}_\perp$. Note that $\widehat{\boldsymbol{u}}_l$ is a unit vector orthogonal to the subspace spanned by $\widetilde{\boldsymbol{u}}_l$ and $\{\boldsymbol{u}_i^\star\}_{i\neq l}$, namely, $\widehat{\boldsymbol{u}}_l^\top \widetilde{\boldsymbol{u}}_l = 0$ and $\boldsymbol{u}_i^{\star\top}\widehat{\boldsymbol{u}}_l = 0$ for all $i \neq l$. Similar to the proof for the claim (E.1), one can derive

$$\begin{aligned}
\mathsf{tr}\big((\boldsymbol{\Sigma}^\star+\sigma^2\boldsymbol{I}_p)^{-1}(\widetilde{\boldsymbol{\Sigma}}+\sigma^2\boldsymbol{I}_p)\big) &= \mathsf{tr}\bigg(\Big(\frac{1}{\lambda_l^\star+\sigma^2}\boldsymbol{u}_l^\star\boldsymbol{u}_l^{\star\top} + \frac{1}{\sigma^2}\boldsymbol{a}_\perp\boldsymbol{a}_\perp^\top\Big)\Big((\lambda_l^\star+\sigma^2)\widetilde{\boldsymbol{u}}_l\widetilde{\boldsymbol{u}}_l^\top + \sigma^2\widehat{\boldsymbol{u}}_l\widehat{\boldsymbol{u}}_l^\top\Big)\bigg) + p - 2 \\
&= (\widetilde{\boldsymbol{u}}_l^\top\boldsymbol{u}_l^\star)^2 + \frac{\sigma^2}{\lambda_l^\star+\sigma^2}(\widehat{\boldsymbol{u}}_l^\top\boldsymbol{u}_l^\star)^2 + \frac{\lambda_l^\star+\sigma^2}{\sigma^2}(\widetilde{\boldsymbol{u}}_l^\top\boldsymbol{a}_\perp)^2 + (\widehat{\boldsymbol{u}}_l^\top\boldsymbol{a}_\perp)^2 + p - 2 \\
&= \frac{1}{1+\delta_n^2}\left(1+\frac{\lambda_l^\star+\sigma^2}{\sigma^2}\delta_n^2\right) + \frac{1}{1+\delta_n^2}\left(\frac{\sigma^2}{\lambda_l^\star+\sigma^2}\delta_n^2+1\right) + p - 2 \\
&= \frac{\lambda_l^{\star 2}}{(\lambda_l^\star+\sigma^2)\sigma^2}\frac{\delta_n^2}{1+\delta_n^2} + p,
\end{aligned}$$

where the second step is due to $\boldsymbol{u}_i^{\star\top}\boldsymbol{a}_\perp = 0$ and the third line follows from the following facts:

$$\begin{aligned}
\widetilde{\boldsymbol{u}}_l^\top\boldsymbol{u}_l^\star &= \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l^\star + \delta_n\boldsymbol{a}_\perp^\top\boldsymbol{u}_l^\star}{\|\boldsymbol{u}_l^\star+\delta_n\boldsymbol{a}_\perp\|_2} = \frac{1}{\sqrt{1+\delta_n^2}}; \\
\widetilde{\boldsymbol{u}}_l^\top\boldsymbol{a}_\perp &= \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{a}_\perp + \delta_n\boldsymbol{a}_\perp^\top\boldsymbol{a}_\perp}{\|\boldsymbol{u}_l^\star+\delta_n\boldsymbol{a}_\perp\|_2} = \frac{\delta_n}{\sqrt{1+\delta_n^2}}; \\
\widehat{\boldsymbol{u}}_l^\top\boldsymbol{u}_l^\star &= \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l^\star - \frac{1}{\delta_n}\boldsymbol{a}_\perp^\top\boldsymbol{u}_l^\star}{\|\boldsymbol{u}_l^\star-\frac{1}{\delta_n}\boldsymbol{a}_\perp\|_2} = \frac{\delta_n}{\sqrt{1+\delta_n^2}}; \\
\widehat{\boldsymbol{u}}_l^\top\boldsymbol{a}_\perp &= \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{a}_\perp - \frac{1}{\delta_n}\boldsymbol{a}_\perp^\top\boldsymbol{a}_\perp}{\|\boldsymbol{u}_l^\star-\frac{1}{\delta_n}\boldsymbol{a}_\perp\|_2} = -\frac{1}{\sqrt{1+\delta_n^2}}.
\end{aligned}$$

As a consequence, we can upper bound the KL divergence as follows

$$\begin{aligned}
\mathsf{KL}\big(\mathbb{P}^1 \,\|\, \mathbb{P}^0\big) &= \sum_{i=1}^n \mathsf{KL}\big(\mathbb{P}_i^1 \,\|\, \mathbb{P}_i^0\big) = \frac{1}{2}\sum_{i=1}^n \Big(\mathsf{tr}\big((\boldsymbol{\Sigma}^\star+\sigma^2\boldsymbol{I}_p)^{-1}(\widetilde{\boldsymbol{\Sigma}}+\sigma^2\boldsymbol{I}_p)\big) - p\Big) \\
&= \frac{n\lambda_l^{\star 2}}{2(\lambda_l^\star+\sigma^2)\sigma^2}\frac{\delta_n^2}{1+\delta_n^2} \leq \frac{\delta_n^2 n\lambda_l^{\star 2}}{2(\lambda_l^\star+\sigma^2)\sigma^2}.
\end{aligned}$$

Step 3: invoking Fano's inequality. From the preceding upper bound on the KL divergence, it is easy to see that $\mathsf{KL}(\mathbb{P}^1 \,\|\, \mathbb{P}^0) \leq 1/16$ if we choose

$$\delta_n = c_n\sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}} \leq 1, \tag{E.13}$$

where $c_n \asymp 1$ obeys $c_n \in \{1/64,\, 1/16,\, 1/4\}$ and the last step holds due to the assumption (3.20). It follows from Fano's inequality [Tsybakov, 2009, Theorem 2] that

$$p_e := \inf_\psi \max\big\{\mathbb{P}\{\psi \text{ rejects } \mathcal{H}_0 \mid \mathcal{H}_0\},\, \mathbb{P}\{\psi \text{ rejects } \mathcal{H}_1 \mid \mathcal{H}_1\}\big\} \geq 1/5,$$

where the infimum is taken over all tests. Further, we know from (E.12), (E.13) and $c_n \leq 1/4$ that

$$\|\widetilde{\boldsymbol{u}} - \boldsymbol{u}_l^\star\|_2 \leq \sqrt{\frac{(\lambda_l^\star+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}},$$

namely, $\boldsymbol{\Sigma}_1 \in \mathcal{M}_2(\boldsymbol{\Sigma}^\star)$.

Next, let us continue to control $\min \left|\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|$. Our goal is to show

$$\min \left|\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \gtrsim \delta_n \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2,$$

and we shall use the same argument as for (E.8) to prove it. Towards this, let us first consider $\left|\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l - \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l\right|$. By construction, one can derive

$$\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l - \boldsymbol{a}^\top \boldsymbol{u}_l^\star = \frac{\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \delta_n \boldsymbol{a}^\top \boldsymbol{a}_\perp}{\sqrt{1+\delta_n^2}} - \boldsymbol{a}^\top \boldsymbol{u}_l^\star = \underbrace{\frac{\delta_n}{\sqrt{1+\delta_n^2}} \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2}_{=:\,\eta_1} - \underbrace{\left(1 - \frac{1}{\sqrt{1+\delta_n^2}}\right) \boldsymbol{a}^\top \boldsymbol{u}_l^\star}_{=:\,\eta_2} \tag{E.14}$$

where the last step holds because

$$\boldsymbol{a}^\top \boldsymbol{a}_\perp = \boldsymbol{a}^\top \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a} / \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2 = (\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a})^\top \boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a} / \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2 = \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2.$$

Moreover, it is straightforward to verify that

$$\frac{1}{4}\delta_n^2 \le 1 - \frac{1}{\sqrt{1+\delta_n^2}} \le \sqrt{1+\delta_n^2} - 1 \le \frac{1}{2}\delta_n^2 \tag{E.15}$$

for $0 < \delta_n < 1$. With these basic facts in place, let us first choose the pre-factor $c_n \asymp 1$ in E.13 to be $c_n = 1/16$ for the moment, and compare the two terms on the right-hand side of (E.14).

- If $|\eta_1| \ge 2\,|\eta_2|$, then one has

  $$\left|\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l - \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \ge |\eta_1| - |\eta_2| \ge \frac{|\eta_1|}{2} \ge \frac{\delta_n}{4} \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2,$$

  where we have used the fact that $\delta_n \le 1$.

- If $|\eta_1| \le |\eta_2|/2$, then we know that

  $$\left|\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l - \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \ge |\eta_2| - |\eta_1| \ge |\eta_1| \ge \frac{\delta_n}{2} \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2$$

  as long as $\delta_n \le 1$.

- Otherwise, consider the case where $|\eta_2|/2 < |\eta_1| < 2\,|\eta_2|$. In this case, we can adjust the pre-factor c_n to be $1/4$. By doing so, $|\eta_1|$ increases by at most 4 times, while $|\eta_2|$ increases by at least 8 times (according to (E.15)). As a result, the new values of η_1 and η_2 satisfy $|\eta_1| \le |\eta_2|/2$, thus belonging to the second case discussed above and hence $\left|\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l - \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \ge \frac{\delta_n}{4} \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2$. Clearly, we can also adjust c_n to be $1/64$ so as to meet the condition of the first case discussed above.

To sum up, the above analysis reveals that: by properly choosing the constants $\{c_n\}$ in (E.13), one can guarantee that

$$\left|\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l - \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \ge \frac{\delta_n}{4} \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2.$$

Similarly, we can also derive

$$\begin{aligned}
\left|\boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l + \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| &= \left|\frac{\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \delta_n \boldsymbol{a}^\top \boldsymbol{a}_\perp}{\sqrt{1+\delta_n^2}} + \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| = \left|\frac{\delta_n}{\sqrt{1+\delta_n^2}} \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2 + \left(1 + \frac{1}{\sqrt{1+\delta_n^2}}\right) \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \\
&\ge \left|\frac{\delta_n}{\sqrt{1+\delta_n^2}} \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2 - \left(1 + \frac{1}{\sqrt{1+\delta_n^2}}\right) \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\right| \\
&\gtrsim \frac{\delta_n}{\sqrt{1+\delta_n^2}} \|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}} \boldsymbol{a}\|_2
\end{aligned}$$

Taking these two relations collectively yields the advertised bound:

$$\min \left|\boldsymbol{a}^{\top}\boldsymbol{u}_l \pm \boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right| \gtrsim \delta_n \left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2.$$

As a consequence, one can readily apply the standard reduction scheme in [Tsybakov, 2009, Chapter 2.2] again to arrive that

$$\begin{aligned}\inf_{u_{\boldsymbol{a},l}} \sup_{\boldsymbol{\Sigma}\in\mathcal{M}_2(\boldsymbol{\Sigma}^{\star})} \mathbb{E}\Big[\min \left|u_{\boldsymbol{a},l} \pm \boldsymbol{a}^{\top}\boldsymbol{u}_l(\boldsymbol{\Sigma})\right|\Big] &\gtrsim p_e \min\left|\boldsymbol{a}^{\top}\widetilde{\boldsymbol{u}}_l \pm \boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right| \gtrsim \min\left|\boldsymbol{a}^{\top}\widetilde{\boldsymbol{u}}_l \pm \boldsymbol{a}^{\top}\boldsymbol{u}_l^{\star}\right| \\ &\gtrsim c_n\sqrt{\frac{(\lambda_l^{\star}+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}}\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2 \gtrsim \sqrt{\frac{(\lambda_l^{\star}+\sigma^2)\sigma^2}{\lambda_l^{\star 2}n}}\left\|\boldsymbol{P}_{\boldsymbol{U}^{\star\perp}}\boldsymbol{a}\right\|_2\end{aligned}$$

where the last step holds since $\min_n c_n = 1/64$.

F Proof for the lower bound of the plug-in estimator (Theorem 2)

Evidently, it is sufficient to establish the lower bound for the rank-1 case (i.e. $r=1$), which forms the content of this section. To begin with, let us decompose the leading eigenvector $\boldsymbol{u}$ of $\boldsymbol{M}$ as follows:

$$\boldsymbol{u} = \boldsymbol{u}^{\star}\cos\theta + \boldsymbol{u}_{\perp}\sin\theta, \qquad \text{with } \theta\in\big[0,\pi/2\big],$$

where as before, $\boldsymbol{u}_{\perp}$ denotes some unit vector perpendicular to $\boldsymbol{u}^{\star}$. Denote by $s := \mathrm{sign}(\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp})$ the sign of $\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}$. Armed with these, one can express the plug-in estimator $u_{\boldsymbol{a}}^{\mathsf{plugin}}$ as

$$u_{\boldsymbol{a}}^{\mathsf{plugin}} = \boldsymbol{a}^{\top}\boldsymbol{u} = \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta + \boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}\sin\theta = \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta + s\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}\right|\sin\theta,$$

which in turn leads to

$$\begin{aligned}\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{plugin}}, \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right) &= \min\left|\boldsymbol{a}^{\top}\boldsymbol{u} \pm \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right| = \big|\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right| - \left|\boldsymbol{a}^{\top}\boldsymbol{u}\right|\big| \\ &= \Big|\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right| - \left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta + s\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}\right|\sin\theta\right|\Big|.\end{aligned}$$

In view of the analysis in Appendix C.4, one can see that: conditioned on θ, $\boldsymbol{u}_{\perp}$ is uniformly distributed over the unit sphere when restricted to the subspace spanned by the columns of $\boldsymbol{u}^{\star\perp}$. Consequently, we have

$$\mathbb{P}\big\{s=1 \mid |\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}|,\theta\big\} = \mathbb{P}\big\{s=-1 \mid |\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}|,\theta\big\} = \frac{1}{2}.$$

This implies the independence between s and $|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}|$ as well as θ, which further reveals that

$$s = \underset{s'=\pm1}{\mathrm{argmax}}\Big|\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right| - \left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta + s'\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}\right|\sin\theta\right|\Big|$$

with probability 1/2. Therefore, it is seen that

$$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{plugin}}, \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right) = \max_{s'=\pm1}\Big|\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right| - \left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta + s'\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}\right|\sin\theta\right|\Big|$$

with probability 1/2. In the following, we seek to lower bound $\max_{s'=\pm1}\big|\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right| - \left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta + s'\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}\right|\sin\theta\right|\big|$, dividing into two cases based on the relative size of $\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta$ compared to $|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}|\sin\theta$. Without loss of generality, let us assume that $\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\geq 0$ in the sequel, and recall that $\theta\in[0,\pi/2]$.

- In the case where $\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta \geq |\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}|\sin\theta$, one can demonstrate that

$$\begin{aligned}\max_{s'=\pm1}\Big|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star} - \left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\cos\theta + s'\left|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}\right|\sin\theta\right|\Big| &= \boldsymbol{a}^{\top}\boldsymbol{u}^{\star}(1-\cos\theta) + \left|\boldsymbol{a}^{\top}\boldsymbol{u}_{\perp}\right|\sin\theta \\ &\geq (1-\cos\theta)\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right| \geq \frac{1-\cos^2\theta}{2}\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right| \\ &\gtrsim \frac{\sigma^2 n}{\lambda_1^{\star 2}}\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right|\end{aligned}$$

  with high probability, where the last step results from (5.36).

- On the other hand, if instead $\boldsymbol{a}^\top \boldsymbol{u}^\star \cos\theta < |\boldsymbol{a}^\top \boldsymbol{u}_\perp| \sin\theta$, then one can deduce that

$$\begin{aligned}
\max_{s'=\pm 1} \Big| \boldsymbol{a}^\top \boldsymbol{u}^\star - \big| \boldsymbol{a}^\top \boldsymbol{u}^\star \cos\theta + s' \left| \boldsymbol{a}^\top \boldsymbol{u}_\perp \right| \sin\theta \big| \Big| &= \max_{s'=\pm 1} \Big| \boldsymbol{a}^\top \boldsymbol{u}^\star - \big| s' \boldsymbol{a}^\top \boldsymbol{u}^\star \cos\theta + |\boldsymbol{a}^\top \boldsymbol{u}_\perp| \sin\theta \big| \Big| \\
&\overset{\text{(i)}}{=} \max_{s'=\pm 1} \Big| \boldsymbol{a}^\top \boldsymbol{u}^\star - |\boldsymbol{a}^\top \boldsymbol{u}_\perp| \sin\theta - s' \boldsymbol{a}^\top \boldsymbol{u}^\star \cos\theta \Big| \\
&\overset{\text{(ii)}}{\geq} \big| \boldsymbol{a}^\top \boldsymbol{u}^\star \big| \cos\theta \overset{\text{(iii)}}{\asymp} \big| \boldsymbol{a}^\top \boldsymbol{u}^\star \big| \\
&\overset{\text{(iv)}}{\gtrsim} \frac{\sigma^2 n}{\lambda_1^{\star 2}} \big| \boldsymbol{a}^\top \boldsymbol{u}^\star \big| .
\end{aligned}$$

  Here, (i) is valid since $\boldsymbol{a}^\top \boldsymbol{u}^\star \cos\theta < |\boldsymbol{a}^\top \boldsymbol{u}_\perp| \sin\theta$; (ii) holds given that $\max |a \pm b| \geq |b|$ for any a, b; (iii) follows since, according to (5.36), $\cos\theta \asymp 1$ holds with high probability; and (iv) arises from the assumption (3.3).

Combining the preceding two cases, we can readily conclude that

$$\mathsf{dist}\left(u_{\boldsymbol{a}}^{\mathsf{plugin}}, \boldsymbol{a}^\top \boldsymbol{u}^\star\right) \gtrsim \frac{\sigma^2 n}{\lambda_1^{\star 2}} \left| \boldsymbol{a}^\top \boldsymbol{u}^\star \right|$$

with probability at least $1/3$.

G Technical lemmas

This section collects a few technical lemmas that prove useful in the analysis of our main results. In what follows, we shall start by stating the precise statements of these lemmas, followed by the proofs for each of them.

Lemma 18. *Let $\{\boldsymbol{h}_i\}_{i=1}^n$ be a sequence of independent zero-mean Gaussian random vectors in $\mathbb{R}^r$ with covariance matrix $\sigma^2 \boldsymbol{I}_r$, and let $\boldsymbol{a} = [a_i]_{1\leq i\leq n} \in \mathbb{R}^n$ be a fixed vector. Then with probability at least $1 - O\left(n^{-10}\right)$, one has*

$$\begin{aligned}
\Big\| \sum_{1\leq i\leq n} a_i \big(\boldsymbol{h}_i \boldsymbol{h}_i^\top - \sigma^2 \boldsymbol{I}_r \big) \Big\| &\leq C_1 \sigma^2 \big(\|\boldsymbol{a}\|_2 \sqrt{r \log n} + \|\boldsymbol{a}\|_\infty (r \log n + \log^2 n) \big) \\
&\leq C_2 \sigma^2 \|\boldsymbol{a}\|_\infty \big(\sqrt{rn \log n} + r \log n \big) \qquad \text{(G.1)}
\end{aligned}$$

for some sufficiently large constants $C_1, C_2 > 0$. Here, $\|\boldsymbol{a}\|_\infty := \max_{1\leq i\leq n} |a_i|$.

Lemma 19. *Let $\{\boldsymbol{h}_i\}_{i=1}^n$ and $\{\boldsymbol{g}_i\}_{i=1}^n$ be two independent sequences of standard Gaussian random vectors in $\mathbb{R}^r$ and $\mathbb{R}^p$, respectively. Then with probability at least $1 - O\left(n^{-10}\right)$, the following holds:*

$$\Big\| \sum_{1\leq i\leq n} \boldsymbol{h}_i \boldsymbol{g}_i^\top \Big\| \leq C_3 \big(\sqrt{pn \log n} + \sqrt{pr} \log n \big)$$

where $C_3 > 0$ is some sufficiently large constant.

Lemma 20. *Let $\{X_i\}_{i=1}^n$ be a sequence of independent random variables in $\mathbb{R}$, and let $\mathcal{I}$ be an interval in $\mathbb{R}$. Consider a collection of functions $\{f_i\}_{i=1}^n$ from $\mathbb{R} \times \mathcal{I}$ to $\mathbb{R}$, and we suppose that*

1. *for any fixed $\lambda \in \mathcal{I}$, with probability at least $1 - \delta_1$,*

$$\Big| \sum_{1\leq i\leq n} f_i(X_i, \lambda) \Big| \leq \frac{\varepsilon}{2};$$

2. *with probability at least $1 - \delta_2$,*

$$\sup_{\lambda\in\mathcal{I}} \Big| \frac{\mathrm{d}}{\mathrm{d}\lambda} \sum_{1\leq i\leq n} f_i(X_i, \lambda) \Big| \leq L.$$

Then with probability exceeding $1 - \frac{8L|\mathcal{I}|}{\varepsilon}\delta_1 - \delta_2$*, one has*

$$\sup_{\lambda\in\mathcal{I}} \Big| \sum_{1\le i\le n} f_i(X_i,\lambda) \Big| \le \varepsilon.$$

Next, we record an eigenvalue interlacing lemma, which has been documented in Horn and Johnson [2012, Corollary 4.3.37].

Lemma 21 (Poincaré separation theorem)**.** *Let* $\boldsymbol{M}$ *be a symmetric matrix in* $\mathbb{R}^{n\times n}$ *and* $\boldsymbol{U}$ *be an orthonormal matrix in* $\mathbb{R}^{n\times r}$ *satisfying* $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r$*. Then one has*

$$\lambda_{n-r+i}(\boldsymbol{M}) \le \lambda_i(\boldsymbol{U}^\top\boldsymbol{M}\boldsymbol{U}) \le \lambda_i(\boldsymbol{M}), \qquad 1\le i\le r,$$

where $\lambda_i(\boldsymbol{A})$ *denote the* i*-th largest eigenvalue of matrix* $\boldsymbol{A}$*.*

G.1 Proof of Lemma 18

As can be easily seen, the second inequality in (G.1) follows immediately from the elementary bound $\|\boldsymbol{a}\|_2 \le \|\boldsymbol{a}\|_\infty\sqrt{n}$. Hence, the proof boils down to justifying the first inequality in (G.1).

We shall invoke the truncated matrix Bernstein inequality [Hopkins et al., 2016, Proposition A.7] to control the spectral norm of $\sum_i a_i\big(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\big)$, which is a sum of independent zero-mean random matrices. To do so, we need to bound three quantities: (1) the covariance of the sum $\sum_i a_i\big(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\big)$, (2) a high-probability upper bound L on $\max_i \big\|a_i\big(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\big)\big\|$, (3) the expectation of the truncated summand $\max_i \mathbb{E}\big[\|a_i\big(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\big)\|\,\mathbb{1}\{\|a_i\big(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\big)\| \ge L\}\big]$. We shall look at each of them separately.

1. Straightforward computation gives

$$\boldsymbol{\Sigma} := \sum_{i=1}^n a_i^2\mathbb{E}\big[(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r)^2\big] = (r+1)\sigma^4\sum_{i=1}^n a_i^2\boldsymbol{I}_r = (r+1)\sigma^4\|\boldsymbol{a}\|_2^2\boldsymbol{I}_r. \tag{G.2}$$

2. We now turn to bounding the spectral norm of each summand $a_i\big(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\big)$, which clearly satisfies

$$\|a_i(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r)\| \le |a_i|\cdot(\|\boldsymbol{h}_i\|_2^2 + \sigma^2).$$

   By virtue of the Gaussian concentration inequality [Hsu et al., 2012, Proposition 1.1], we obtain

$$\mathbb{P}\big\{\|\boldsymbol{h}_i\|_2^2 - \sigma^2 r \ge t\big\} \le \exp\left(-\frac{1}{16}\min\left\{\frac{t^2}{r\sigma^4}, \frac{t}{\sigma^2}\right\}\right). \tag{G.3}$$

   In particular, this implies that with probability at least $1 - O\left(n^{-20}\right)$, one has

$$\|\boldsymbol{h}_i\|_2^2 \lesssim \sigma^2\big(r + \log n\big). \tag{G.4}$$

   In what follows, we shall set

$$L := C\sigma^2\big(r + \log n\big) \tag{G.5}$$

   for some sufficiently large constant $C > 0$.

3. We then look the truncated mean. To this end, we observe that

$$\begin{aligned}\mathbb{E}\big[\|\boldsymbol{h}_i\|_2^2\,\mathbb{1}\{\|\boldsymbol{h}_i\|_2^2 \ge L\}\big] &\le L\mathbb{P}\big\{\|\boldsymbol{h}_i\|_2^2 \ge L\big\} + \int_L^\infty \mathbb{P}\big\{\|\boldsymbol{h}_i\|_2^2 \ge t\big\}\,\mathrm{d}t \\ &\le O\left(n^{-20}\right)L + \int_L^\infty \mathbb{P}\big\{\|\boldsymbol{h}_i\|_2^2 \ge t\big\}\,\mathrm{d}t.\end{aligned}$$

   For any $t \ge L/2$, it is seen that $\min\big\{t^2/(r\sigma^4),\, t/\sigma^2\big\} \ge t/\sigma^2$, and hence

$$\int_L^\infty \mathbb{P}\big\{\|\boldsymbol{h}_i\|_2^2 \ge t\big\}\,\mathrm{d}t \le \int_{L/2}^\infty \mathbb{P}\big\{\|\boldsymbol{h}_i\|_2^2 - \sigma^2 r \ge t\big\}\,\mathrm{d}t \le \int_{L/2}^\infty \exp\left(-\frac{t}{16\sigma^2}\right)\,\mathrm{d}t$$

$$\lesssim \sigma^2 \exp\left(-\frac{C(r+\log n)}{32}\right) \lesssim \frac{L}{n^2},$$

provided that $C > 0$ is sufficiently large. As a result, taking this together with $\|\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\| \le \|\boldsymbol{h}_i\|_2^2 + \sigma^2$, we arrive at

$$R := \mathbb{E}\big[\|\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\|\,\mathbb{1}\{\|\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\| \ge L\}\big] \le \mathbb{E}\big[\big(\|\boldsymbol{h}_i\|_2^2 + \sigma^2\big)\,\mathbb{1}\{\|\boldsymbol{h}_i\|_2^2 + \sigma^2 \ge L\}\big] \lesssim \frac{L}{n^2}. \tag{G.6}$$

With the preceding bounds in place, we can invoke the truncated matrix Bernstein inequality [Hopkins et al., 2016, Proposition A.7] to obtain that: with probability at least $1 - O\left(n^{-11}\right)$,

$$\begin{aligned}
\Big\|\sum_i a_i\big(\boldsymbol{h}_i\boldsymbol{h}_i^\top - \sigma^2\boldsymbol{I}_r\big)\Big\| &\lesssim \sqrt{\|\boldsymbol{\Sigma}\|\log n} + \|\boldsymbol{a}\|_\infty(nR + L\log n) \\
&\overset{\text{(i)}}{\asymp} \sqrt{\|\boldsymbol{\Sigma}\|\log n} + \|\boldsymbol{a}\|_\infty L\log n \\
&\overset{\text{(ii)}}{\lesssim} \sqrt{\sum_{i=1}^n a_i^2(r+1)\sigma^4\log n} + \|\boldsymbol{a}\|_\infty\sigma^2(r\log n + \log^2 n) \\
&\lesssim \sigma^2\big(\|\boldsymbol{a}\|_2\sqrt{r\log n} + \|\boldsymbol{a}\|_\infty(r\log n + \log^2 n)\big),
\end{aligned}$$

where (i) arises from (G.6), and (ii) relies on (G.2) and (G.5). This completes the proof.

G.2 Proof of Lemma 19

The proof strategy here is almost identical to that for Lemma 18 — we shall apply the truncated matrix Bernstein inequality [Hopkins et al., 2016, Proposition A.7] to upper bound the spectral norm of $\sum_{1\le i\le n}\boldsymbol{h}_i\boldsymbol{g}_i^\top$, which is a sum of independent zero-mean random matrices. Towards this, we start by estimating several key quantities.

- In view of the independence between $\{\boldsymbol{h}_i\}_i$ and $\{\boldsymbol{g}_i\}_i$, the covariance matrices can be computed as

$$\boldsymbol{\Sigma}_1 := \sum_{i=1}^n \mathbb{E}\big[\boldsymbol{h}_i\boldsymbol{g}_i^\top\boldsymbol{g}_i\boldsymbol{h}_i^\top\big] = \sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{g}_i\|_2^2\big]\mathbb{E}\big[\boldsymbol{h}_i\boldsymbol{h}_i^\top\big] = np\boldsymbol{I}_r; \tag{G.7}$$

$$\boldsymbol{\Sigma}_2 := \sum_{i=1}^n \mathbb{E}\big[\boldsymbol{g}_i\boldsymbol{h}_i^\top\boldsymbol{h}_i\boldsymbol{g}_i^\top\big] = \sum_{i=1}^n \mathbb{E}\big[\|\boldsymbol{h}_i\|_2^2\big]\mathbb{E}\big[\boldsymbol{g}_i\boldsymbol{g}_i^\top\big] = nr\boldsymbol{I}_p. \tag{G.8}$$

- As for the spectral norm of each summand $\boldsymbol{h}_i\boldsymbol{g}_i^\top$, we know from (G.4) that with probability at least $1 - O\left(n^{-20}\right)$,

$$\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\| = \|\boldsymbol{h}_i\|_2\|\boldsymbol{g}_i\|_2 \lesssim \sqrt{(r+\log n)(p+\log n)} \asymp \sqrt{pr} + \sqrt{p\log n} + \log n.$$

  Therefore, this suggests that we define

$$L := C\big(\sqrt{pr} + \sqrt{p\log n} + \log n\big) \tag{G.9}$$

  for some sufficiently large constant $C > 0$.

- Next, we turn to the truncated mean. Observe that

$$\begin{aligned}
\mathbb{E}\big[\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\|\,\mathbb{1}\{\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\| \ge L\}\big] &\le L\mathbb{P}\big\{\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\| \ge L\big\} + \int_L^\infty \mathbb{P}\big\{\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\| \ge t\big\}\,\mathrm{d}t \\
&\le O\left(n^{-20}\right)L + \int_L^\infty \mathbb{P}\big\{\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\| \ge t\big\}\,\mathrm{d}t
\end{aligned}$$

$$\leq O\left(n^{-20}\right)L + \int_L^\infty \mathbb{P}\{\|\boldsymbol{h}_i\|_2^2 \geq t\}\,\mathrm{d}t + \int_L^\infty \mathbb{P}\{\|\boldsymbol{g}_i\|_2^2 \geq t\}\,\mathrm{d}t,$$

where the last holds arises from the following bound due to the union bound:

$$\mathbb{P}\{\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\| \geq t\} = \mathbb{P}\{\|\boldsymbol{h}_i\|_2^2\|\boldsymbol{g}_i\|_2^2 \geq t^2\} \leq \mathbb{P}\{\|\boldsymbol{h}_i\|_2^2 \geq t\} + \mathbb{P}\{\|\boldsymbol{g}_i\|_2^2 \geq t\}.$$

In addition, since $\min\{t^2/r, t\} \geq t$ for all $t \geq L/2 \geq 2r$, we can use (G.3) to bound

$$\begin{aligned}
\int_L^\infty \mathbb{P}\{\|\boldsymbol{h}_i\|_2^2 \geq t\}\,\mathrm{d}t &\leq \int_L^\infty \mathbb{P}\Big\{\|\boldsymbol{h}_i\|_2^2 - r \geq \frac{t}{2}\Big\}\,\mathrm{d}t \lesssim \int_{L/2}^\infty \mathbb{P}\{\|\boldsymbol{h}_i\|_2^2 - r \geq t\}\,\mathrm{d}t \\
&\leq \int_{L/2}^\infty \exp\left(-\frac{1}{16}\min\left\{\frac{t^2}{r}, t\right\}\right)\mathrm{d}t \leq \int_{L/2}^\infty \exp\left(-\frac{t}{16}\right)\mathrm{d}t \\
&\lesssim \exp\left(-\frac{C}{32}\sqrt{pr + p\log n + \log^2 n}\right) \lesssim \frac{L}{n^2}.
\end{aligned}$$

Clearly, the same bound also holds for $\int_L^\infty \mathbb{P}\{\|\boldsymbol{g}_i\|_2^2 \geq t\}\,\mathrm{d}t$. Therefore, combining these estimates yields

$$R := \mathbb{E}\big[\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\|\,\mathbb{1}\{\|\boldsymbol{h}_i\boldsymbol{g}_i^\top\| \geq L\}\big] \lesssim \frac{L}{n^2}. \tag{G.10}$$

With these parameters in place, one can apply the truncated matrix Bernstein inequality [Hopkins et al., 2016, Proposition A.7] to demonstrate that: with probability at least $1 - O\left(n^{-10}\right)$,

$$\begin{aligned}
\Big\|\sum_{1\leq i\leq n}\boldsymbol{h}_i\boldsymbol{g}_i^\top\Big\| &\lesssim \sqrt{(\|\boldsymbol{\Sigma}_1\| + \|\boldsymbol{\Sigma}_2\|)\log n} + nR + L\log n \\
&\overset{\text{(i)}}{\asymp} \sqrt{\|\boldsymbol{\Sigma}_1\|\log n} + L\log n \\
&\overset{\text{(ii)}}{\lesssim} \sqrt{pn\log n} + \sqrt{pr}\log n + \sqrt{p\log^3 n} + \log^2 n \\
&\asymp \sqrt{pn\log n} + \sqrt{pr}\log n.
\end{aligned}$$

Here, (i) uses (G.7), (G.7) and (G.10); (ii) arises from (G.9). The proof is thus complete.

G.3 Proof of Lemma 20

Let $\mathcal{N}$ be a $\frac{\varepsilon}{2L}$-covering of $\mathcal{I}$ with cardinality $|\mathcal{N}| \leq \frac{4L|\mathcal{I}|}{\varepsilon}$, and let $\mathcal{E}$ denote an event such that

$$\begin{aligned}
\sup_{\lambda\in\mathcal{N}}\Big|\sum_{1\leq i\leq n} f_i(X_i,\lambda)\Big| &\leq \frac{\varepsilon}{2}, \\
\sup_{\lambda\in\mathcal{I}}\Big|\frac{\mathrm{d}}{\mathrm{d}\lambda}\sum_{1\leq i\leq n} f_i(X_i,\lambda)\Big| &\leq L,
\end{aligned}$$

which holds with probability at least $1 - \frac{8L|\mathcal{I}|}{\varepsilon}\delta_1 - \delta_2$ (according to the assumptions and the union bound).

For any $\lambda \in \mathcal{I}$, let $\hat{\lambda} \in \mathcal{N}$ such that $|\lambda - \hat{\lambda}| \leq \frac{\varepsilon}{2L}$. One can easily check that on the event $\mathcal{E}$, one has

$$\begin{aligned}
\sup_{\lambda\in\mathcal{I}}\Big|\sum_{1\leq i\leq n} f_i(X_i,\lambda)\Big| &= \sup_{\lambda\in\mathcal{I}}\Big|\sum_{1\leq i\leq n}\big(f_i(X_i,\lambda) - f_i(X_i,\hat{\lambda}) + f_i(X_i,\hat{\lambda})\big)\Big| \\
&\leq \sup_{\lambda\in\mathcal{I}}\Big|\frac{\mathrm{d}}{\mathrm{d}\lambda}\sum_{1\leq i\leq n} f_i(X_i,\lambda)\Big| \cdot |\lambda - \hat{\lambda}| + \sup_{\hat{\lambda}\in\mathcal{N}}\Big|\sum_{1\leq i\leq n} f_i(X_i,\hat{\lambda})\Big| \\
&\leq L\cdot\frac{\varepsilon}{2L} + \sup_{\hat{\lambda}\in\mathcal{N}}\Big|\sum_{1\leq i\leq n} f_i(X_i,\hat{\lambda})\Big| \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

thus concluding the proof.

References

E. Abbe, J. Fan, and K. Wang. An l_p theory of pca and spectral clustering. arXiv preprint arXiv:2006.14062, 2020a.

E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics*, 48(3):1452–1474, 2020b.

Z. Bao, X. Ding, and K. Wang. Singular vector and singular subspace distribution for the matrix denoising model. *The Annals of Statistics*, 49(1):370–392, 2021.

Z. Bao, X. Ding, J. Wang, and K. Wang. Statistical inference for principal components of spiked covariance matrices. *The Annals of Statistics*, 50(2):1144–1169, 2022.

C. Cai, H. V. Poor, and Y. Chen. Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. arXiv preprint arXiv:2006.08580, 2020.

C. Cai, G. Li, Y. Chi, H. V. Poor, and Y. Chen. Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics*, 49(2):944–967, 2021.

C. Cai, G. Li, H. V. Poor, and Y. Chen. Nonconvex low-rank tensor completion from noisy data. *Operations Research*, 70(2):1219–1237, 2022.

T. T. Cai and A. Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89, 2018.

J. Cape, M. Tang, C. E. Priebe, et al. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439, 2019.

P. Chen, C. Gao, and A. Y. Zhang. Partial recovery for top-k ranking: Optimality of MLE and sub-optimality of spectral method. arXiv preprint arXiv:2006.16485, 2020a.

Y. Chen and C. Suh. Spectral MLE: Top-k rank aggregation from pairwise comparisons. In *Proceedings of the International Conference on Machine Learning*, pages 371–380. PMLR, 2015.

Y. Chen, J. Fan, C. Ma, and K. Wang. Spectral method and regularized MLE are both optimal for top-K ranking. *The Annals of Statistics*, 47(4):2204–2235, August 2019a.

Y. Chen, J. Fan, C. Ma, and Y. Yan. Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences of the U.S.A.*, 116(46):22931–22937, 2019b.

Y. Chen, Y. Chi, J. Fan, C. Ma, and Y. Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM Journal on Optimization*, 30(4):3098–3121, 2020b.

Y. Chen, C. Cheng, and J. Fan. Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices. *The Annals of Statistics*, 49(1):435–458, 2021a.

Y. Chen, Y. Chi, J. Fan, C. Ma, et al. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806, 2021b.

C. Cheng, Y. Wei, and Y. Chen. Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *IEEE Transactions on Information Theory*, 67(11):7380–7419, 2021.

C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

J. Eldridge, M. Belkin, and Y. Wang. Unperturbed: spectral analysis beyond Davis-Kahan. In *Proceedings of the Twenty-ninth Algorithmic Learning Theory*, pages 321–358, 2018.

M. Embree and L. N. Trefethen. Generalizing eigenvalue theorems to pseudospectra theorems. *SIAM Journal on Scientific Computing*, 23(2):583–590, 2001.

J. Fan, W. Wang, and Y. Zhong. An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning*, 18:1–42, 2018.

J. Fan, Y. Fan, X. Han, and J. Lv. Asymptotic theory of eigenvectors for random matrices with diverging spikes. *Journal of the American Statistical Association*, pages 1–14, 2020.

X. Han, Q. Yang, and Y. Fan. Universal rank inference via residual subsampling with application to large networks. *arXiv preprint arXiv:1912.11583*, 2019.

S. B. Hopkins, T. Schramm, J. Shi, and D. Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proceedings of the Forty-eighth Annual ACM symposium on Theory of Computing*, pages 178–191, 2016.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

D. Hsu, S. M. Kakade, and T. Zhang. A tail inequality for quadratic forms of sub-Gaussian random vectors. *Electronic Communications in Probability*, 17(52), 2012.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, 29(2):295–327, 2001.

I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693, 2009.

R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *IEEE transactions on information theory*, 56(6):2980–2998, 2010a.

R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. *Journal of Machine Learning Research*, 11(Jul):2057–2078, 2010b.

A. Knowles and J. Yin. The isotropic semicircle law and deformation of wigner matrices. *Communications on Pure and Applied Mathematics*, 66(11):1663–1749, 2013.

V. Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*, volume 2033 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. ISBN 978-3-642-22146-0.

V. Koltchinskii and D. Xia. Perturbation of linear forms of singular vectors under Gaussian noise. In *High Dimensional Probability VII*, pages 397–423. Springer, 2016.

V. Koltchinskii, K. Lounici, et al. Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 52, pages 1976–2013. Institut Henri Poincaré, 2016.

V. Koltchinskii, K. Lounici, et al. Normal approximation and concentration of spectral projectors of sample covariance. *The Annals of Statistics*, 45(1):121–157, 2017.

V. Koltchinskii, M. Löffler, R. Nickl, et al. Efficient estimation of linear functionals of principal components. *The Annals of Statistics*, 48(1):464–490, 2020.

S. Kullback et al. An application of information theory to multivariate analysis. *The Annals of Mathematical Statistics*, 23(1):88–102, 1952.

J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, 43(1):215–237, 2015.

L. Lei. Unified $\ell_{2\to\infty}$ eigenspace perturbation theory for symmetric random matrices. arXiv preprint arXiv:1909.04798, 2019.

M. Löffler, A. Y. Zhang, and H. H. Zhou. Optimality of spectral clustering in the Gaussian mixture model. arXiv preprint arXiv:1911.00538, 2019.

K. Lounici. Sparse principal component analysis with missing observations. In *High Dimensional Probability VI*, pages 327–356. Springer, 2013.

K. Lounici. High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3): 1029–1058, 2014.

C. Ma, K. Wang, Y. Chi, and Y. Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. *Foundations of Computational Mathematics*, 20(3):451–632, 2020.

A. Montanari and N. Sun. Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics*, 71(11):2381–2425, 2018.

B. Nadler. Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics*, 36(6):2791–2817, 2008.

S. Negahban, S. Oh, and D. Shah. Rank centrality: Ranking from pairwise comparisons. *Operations Research*, 65(1):266–287, 2017.

S. O'Rourke, V. Vu, and K. Wang. Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra and its Applications*, 540:26–59, 2018.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

A. Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1):20–36, 2011.

G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

R. Sun and Z.-Q. Luo. Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory*, 62(11):6535–6579, 2016.

T. Tao. *Topics in Random Matrix Theory*. Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island, 2012.

A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics, 2009.

R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing, Theory and Applications*, pages 210 – 268, 2012.

R. Vershynin. *High Dimensional Probability*. Cambridge University Press, 2017.

V. Vu. Singular vectors under random perturbation. *Random Structures & Algorithms*, 39(4):526–538, 2011.

V. Vu and J. Lei. Minimax rates of estimation for sparse pca in high dimensions. In *Artificial intelligence and statistics*, pages 1278–1286. PMLR, 2012.

R. Wang. Singular vector perturbation under Gaussian noise. *SIAM Journal on Matrix Analysis and Applications*, 36(1):158–177, 2015.

P. Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972.

D. Xia. Confidence region of singular subspaces for low-rank matrix regression. *IEEE Transactions on Information Theory*, 65(11):7437–7459, 2019.

D. Xia. Normal approximation and confidence region of singular subspaces. *Electronic Journal of Statistics*, 15(2):3798–3851, 2021.

D. Xia and M. Yuan. Statistical inferences of linear forms for noisy matrix completion. *Journal of the Royal Statistical Society Series B*, 83(1):58–77, 2021.

D. Xia, M. Yuan, and C.-H. Zhang. Statistically optimal and computationally efficient low rank tensor completion from noisy entries. *The Annals of Statistics*, 49(1), 2021.

Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.

A. R. Zhang, T. T. Cai, and Y. Wu. Heteroskedastic PCA: Algorithm, optimality, and applications. *The Annals of Statistics*, 50(1):53–80, 2022.

Y. Zhong and N. Boumal. Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016, 2018.

Z. Zhu, T. Wang, and R. J. Samworth. High-dimensional principal component analysis with heterogeneous missingness. arXiv preprint arXiv:1906.12125, 2019.