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

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

April 7, 2021

Eigenvector perturbation analysis plays a vital role in various statistical data science applications. A large body of prior works, however, focused on establishing $\ell_2$ 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 theory.

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

Contents

1 Introduction 2

2 Problem formulation 4
  2.1 Matrix denoising .......................................................... 4
  2.2 Principal component analysis and covariance estimation .......... 4

3 Main results 5
  3.1 Matrix denoising .......................................................... 5
  3.2 Principal component analysis ......................................... 8

4 Related works 12

5 Analysis 13
  5.1 Master theorems .......................................................... 13
  5.2 Analysis for matrix denoising ......................................... 14
    5.2.1 Proof of eigenvalue perturbation theory (Theorem 6) ........ 14
    5.2.2 Proof of eigenvector perturbation theory (Theorem 1) ....... 16
  5.3 Analysis for principal component analysis ......................... 21
    5.3.1 Proof of eigenvalue perturbation theory (Theorem 7) ....... 21
    5.3.2 Proof of eigenvector perturbation theory (Theorem 2) ....... 24

*Department of Electronic Engineering, Tsinghua University, Beijing, China 100086, USA; email: {g-li16@mails.tsinghua.edu.cn,gyt@tsinghua.edu.cn}.
†Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA; email: {ccai,poor,yuxin.chen}@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 et al., 2015], to name just a few. In reality, it is often the case that one only observes a randomly corrupted version of the 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., 2020b]. 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 \( M^* \in \mathbb{R}^{n \times n} \) be an unknown rank-\( r \) symmetric matrix whose \( l \)-th eigenvector (resp. eigenvalue) is \( u_l^* \) (resp. \( \lambda_l^* \)). What we have observed is a corrupted version \( M = M^* + H \) of \( M^* \), where \( H = [H_{i,j}]_{1 \leq i,j \leq n} \) represents a symmetric Gaussian random matrix with \( H_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), i \geq j \). The aim is to estimate \( u_l^* \) based on the \( l \)-th eigenvector of the data matrix \( M \).
• Principal component analysis (PCA) / covariance estimation. Imagine that we have collected $n$ independent sample vectors $s_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$, $1 \leq i \leq n$. Suppose that the underlying covariance matrix enjoys a “spiked” structure $\Sigma = \Sigma^* + \sigma^2 I_p$, where $\Sigma^* \succeq 0$ is an unknown rank-$r$ matrix whose $l$-th eigenvector (resp. eigenvalue) is given by $u_l^*$ (resp. $\lambda_l^*$). We seek to estimate $u_l^*$ by examining the $l$-th eigenvector of the sample covariance matrix $\frac{1}{n} \sum_{i=1}^n s_i s_i^\top$.

While a large body of prior literature has investigated eigenvector perturbation theory for the aforementioned two models, the majority of past works focused on $\ell_2$ statistical analysis, namely, quantifying the $\ell_2$ estimation error of $u_i$ when it is employed to estimate $u_l^*$. Such $\ell_2$ 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 $u_l^*$ (e.g., the Fourier transform of or any given entry of $u_l^*$). Motivated by the inadequacy of existing $\ell_2$ theory, we seek to investigate how to faithfully estimate linear functionals of the eigenvectors — that is, $a^\top u_l^*$ for some vector $a \in \mathbb{R}^n$ 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 $a^\top u_l^*$ is to invoke the naïve “plug-in” estimator $a^\top u_i$. 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.

• Enabling faithful estimation in the presence of small eigen-gaps. When estimating the eigenvector $u_l^*$, most prior works require the associated eigen-gap $\min_{i \neq l} |\lambda_i^* - \lambda_l^*|$ to exceed the spectral norm of the perturbation matrix (i.e., $H$ in the matrix denoising case and $\frac{1}{n} \sum_i s_i s_i^\top - \Sigma$ in the PCA setting). 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 $a^\top u_l^*$ 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 $\min_{i \neq l} |\lambda_i^* - \lambda_l^*|$ 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 $a^\top u_l^*$ 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 $v$, we denote by $\|v\|_2$ and $\|v\|_\infty$ its $\ell_2$ norm and $\ell_\infty$ norm, respectively; for any vectors $v$ and $u$, we use $\langle v, u \rangle$ to represent their inner product. For any matrix $M$, we let $\|M\|$ and $\|M\|_F$ denote the spectral norm and the Frobenius norm of $M$, respectively. For any matrix $U$ whose columns are orthonormal, we use $U^\perp$ to denote a matrix whose columns form an orthonormal basis of the orthogonal complement of $U$, and let $P_U(M) = UU^\top M$ be the Euclidean projection of a matrix $M$ onto the column space of $U$. For any two random matrices $Z$ and $X$, the notation $Z \overset{d}{=} X$ means $Z$ and $X$ 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 
\[B_r(z) := \{x \mid \|x - z\|_2 \leq r\}\]
the ball of radius \(r\) and center \(z\). Throughout the paper, we denote by 
\(f(n) \leq g(n)\) 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 \(f(n) \geq Cg(n)\) 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, \(O(g(n))\) (resp. \(\Omega(g(n))\)) is similar to \(O(g(n))\) (resp. \(\Omega(g(n))\)) except that it hides any logarithmic term. 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 a sufficiently 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 take the notation \(\sum_{k \neq l, 1 \leq k \leq r} g(k)\) to be zero for any \(g(\cdot)\) if \(r = 1\) (namely, 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 \(M^* = [M_{i,j}]_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}\) with eigen-decomposition

\[
M^* = \sum_{i=1}^{r} \lambda_i^* u_i^* u_i^\top =: U^* \Lambda^* U^{*\top}.
\]

Here, \(\{\lambda_i^*\}\) denotes the set of non-zero eigenvalues of \(M^*\), and \(u_i^*\) indicates the (normalized) eigenvector associated with \(\lambda_i^*\). It is assumed throughout that

\[
\lambda_{\min}^* = |\lambda_1^*| \leq \cdots \leq |\lambda_r^*| = \lambda_{\max}^*,
\]

and the condition number of \(M^*\) is defined as

\[
\kappa := \lambda_{\max}^*/\lambda_{\min}^*.
\]

In addition, we introduce an eigen-gap (or eigenvalue separation) metric that quantifies the distance between the \(l\)-th eigenvalue and the remaining spectrum:

\[
\Delta_l^* := \begin{cases} 
\min_{k: k \neq l, 1 \leq k \leq r} |\lambda_l^* - \lambda_k^*|, & \text{if } r > 1, \\
+\infty, & \text{if } r = 1,
\end{cases}
\]

which plays a crucial role in our perturbation theory.

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

\[
M = M^* + H,
\]

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

\[
H_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad i \geq j.
\]

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

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

\[
s_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma), \quad 1 \leq i \leq n,
\]

While the text is expository, the notation and definitions are crucial for understanding the problem formulation. For matrix denoising, the eigen-decomposition is central, with implications for the condition number and eigen-gap metrics. In PCA and covariance estimation, the spiked model is foundational, enabling the estimation of linear functionals of eigenvectors.
where 
\[ \Sigma = \Sigma^* + \sigma^2 I_p \in \mathbb{R}^{p \times p} \]
denotes the covariance matrix. Here and throughout, we assume that the "spiked component" \( \Sigma^* \) of \( \Sigma \) is an unknown rank-\( r \) matrix with eigen-decomposition
\[ \Sigma^* = U^* \Lambda^* U^{*\top} = \sum_{i=1}^{r} \lambda_i^* u_i^* u_i^{*\top} \geq 0, \]
where \( \lambda_i^* \) denotes the \( i \)-th largest eigenvalue of \( \Sigma^* \), with \( u_i^* \) representing the associated eigenvector. Akin to the matrix denoising case, we assume
\[ 0 < \lambda_{\min}^* = \lambda_1^* \leq \cdots \leq \lambda_r^* = \lambda_{\max}^*, \]
and introduce the condition number \( \kappa := \lambda_{\max}^*/\lambda_{\min}^* \) and the eigen-separation metric
\[ \Delta_l^* := \begin{cases} \min_{k: k \neq l, 1 \leq k \leq r} |\lambda_l^* - \lambda_k^*|, & \text{if } r > 1, \\ \infty, & \text{if } r = 1. \end{cases} \] (2.6)

Given a fixed vector \( a \in \mathbb{R}^p \), our aim is to develop a faithful estimate of the linear functional \( a^\top u_l^* \) of an eigenvector \( u_l^* \) \( (1 \leq l \leq r) \), on the basis of the sample vectors \( \{s_i\}_{1 \leq i \leq n} \) (or the sample covariance matrix \( \frac{1}{n} \sum_{i=1}^{n} s_is_i^\top \)).

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 \( u_l^* \) and \(-u_l^*\) based on the observed matrix, the error of an estimator \( u_a \) for estimating \( a^\top u_l^* \) shall be measured via the following metric that accounts for such a global ambiguity issue:
\[ \text{dist}(u_a, a^\top u_l^*) := \min \{ |u_a \pm a^\top u_l^*| \} = \min \{ |u_a - a^\top u_l^*|, |u_a + a^\top u_l^*| \}, \] (3.1)

3.1 Matrix denoising

We begin with the matrix denoising problem introduced in Section 2.1. Recalling that \( u_l \) is the eigenvector of \( M \) associated with \( \lambda_l \) \( (1 \leq l \leq n) \), we investigate the following two estimators when estimating the linear function \( a^\top u_l^* \).

- A plug-in estimator:
  \[ u_a^{\text{plug-in}} := a^\top u_l; \] (3.2a)

- A modified estimator that we propose (which we shall refer to as a \textit{de-biased estimator} from now on):
  \[ u_a^{\text{debiased}} := \sqrt{1 + b_l a^\top u_l} \quad \text{with} \quad b_l := \sum_{i: r < i \leq n} \frac{\sigma^2}{(\lambda_l - \lambda_i)^2}, \] (3.2b)
  where \( b_l \) can be computed directly using the eigenvalues of \( M \) 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 \( 1 \leq l \leq r \), and suppose that
\[ \sigma \sqrt{n} \leq c_0 \lambda_{\min}^*, \quad r \leq c_1 n/\log^2 n \quad \text{and} \quad \Delta_l^* > C_0 \sigma \sqrt{r} \log n \] (3.3)
for some sufficiently small (resp. large) constants $c_0, c_1 > 0$ (resp. $C_0 > 0$). Let $a \in \mathbb{R}^n$ be any fixed vector with $\|a\|_2 = 1$. With probability at least $1 - O(n^{-10})$, the estimators in (3.2) satisfy

$$\text{dist} \left( u_{a}^{\text{plug in}}, a^\top u_i^* \right) \lesssim E_{\text{md},l} + \frac{\sigma^2 n}{\lambda_i^2} |a^\top u_i^*|,$$

(3.4a)

$$\text{dist} \left( u_{a}^{\debiased}, a^\top u_i^* \right) \lesssim E_{\text{md},l},$$

(3.4b)

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

$$E_{\text{md},l} := \frac{\sigma^2 r \log n}{(\Delta_i^*)^2} |a^\top u_i^*| + \sigma \sqrt{r \log \left( \frac{n \kappa \lambda_{\text{max}}}{\Delta_i^*} \right)} \sum_{k \neq i, 1 \leq k \leq r} |\lambda_i^* - \lambda_k^*| + \frac{\sigma \sqrt{\log \left( \frac{n \kappa \lambda_{\text{max}}}{\Delta_i^*} \right)}}{|\lambda_i^*|}. \quad (3.5)$$

Implications. Theorem 1 develops statistical performance guarantees for the aforementioned two estimators when estimating the linear form $a^\top u_i^*$ for a prescribed vector $a \in \mathbb{R}^n$. 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 $\Delta_i^*$ is allowed to be substantially smaller than the spectral norm $\|H\|$ of the perturbation matrix. This is in stark contrast to, and significantly improves upon, the celebrated Davis-Kahan sin $\Theta$ theorem that requires $\Delta_i^* \gtrsim \|H\|$ [Chen et al., 2020b, Davis and Kahan, 1970]. To be more precise, recalling from standard random matrix theory [Tao, 2012] that $\|H\| \asymp \sigma \sqrt{n}$ with high probability, one can compare our result with classical matrix perturbation theory as follows

our eigen-gap requirement: $\Delta_i^* = \tilde{\Omega}(\sigma \sqrt{r})$;

eigen-gap requirement in classical theory: $\Delta_i^* = \tilde{\Omega}(\sigma \sqrt{n})$.

- Bias reduction. The statistical error (3.4a) for the plug-in estimator $a^\top u_i^*$ contains an additional “bias” term

$$E_{\text{md},l}^{\text{bias}} := \frac{\sigma^2 n}{\lambda_i^2} |a^\top u_i^*|$$

(3.7)

when compared to that of the de-biased estimator (cf. (3.4b)). The influence of this extra term becomes increasingly large and non-negligible as the correlation of $a$ and $u_i^*$ increases. To demonstrate the possibly severe impact incurred by this additional term, let us look at a simple scenario as follows.

- Example. Suppose that $r = 2$, $\lambda_1^* = 2 \lambda_2^*$ (so that $\lambda_1^* - \lambda_2^* \asymp \lambda_1^*$), $|a^\top u_i^*| \asymp 1$ and $\sigma \sqrt{n} \asymp |\lambda_i^*|$. As can be straightforwardly verified, the main term (3.5) and the addition term (3.7) in this example satisfy

$$E_{\text{md},1} = \tilde{O} \left( \frac{\sigma^2}{\lambda_i^2} |a^\top u_i^*| + \frac{\sigma}{|\lambda_i|} |a^\top u_i^*| + \frac{\sigma}{|\lambda_i|} \right) = \tilde{O} \left( \frac{\sigma^2}{\lambda_i^2} + \frac{\sigma}{|\lambda_i|} \right) = \tilde{O} \left( \frac{1}{\sqrt{n}} \right);$$

$$E_{\text{md},1}^{\text{bias}} \asymp \frac{\sigma^2 n}{\lambda_i^2} |a^\top u_i^*| \asymp 1.$$

In other words, the additional bias term $E_{\text{md},1}^{\text{bias}}$ could be a factor of $\sqrt{n}$ times larger than the main term $E_{\text{md},1}$ (modulo some logarithmic factor).

The above discussion reveals the necessity of proper bias correction in order to mitigate the undesired effect of the bias term $E_{\text{md},l}^{\text{bias}}$. Aimed at addressing this issue, our de-biased estimator $u_{a}^{\text{debiased}}$ compensates for the bias term $E_{\text{md},l}^{\text{bias}}$ by properly rescaling the plug-in estimator by a data-driven correction factor $\sqrt{1 + b_i}.

- 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 $|a^\top u_i^*| \leq (1 - \epsilon)\|a\|_2$ for any small non-zero constant $\epsilon > 0$ (so that $a$ is not perfectly
aligned with \( u_i^* \). Note that Cheng et al. [2020, Theorem 3] established the following instance-dependent minimax lower bound:

\[
\inf_{u_{a,i}} \sup_{A \in M_0(M^*)} \mathbb{E} \left[ \text{dist} \left( u_{a,i}, a^\top u_i(A) \right) \right] \geq \frac{\sigma^2}{(\Delta_i^*)^2} |a^\top u_i^*| + \sigma \max_{k:k \neq l} \left| \frac{a^\top u_k^*}{\lambda_i^* - \lambda_k^*} \right|
\]

\[
\inf_{u_{a,i}} \sup_{A \in M_1(M^*)} \mathbb{E} \left[ \text{dist} \left( u_{a,i}, a^\top u_i(A) \right) \right] \geq \frac{\sigma^2}{(\Delta_i^*)^2} |a^\top u_i^*| + \sigma \max_{k:k \neq l} \left| \frac{a^\top u_k^*}{\lambda_i^* - \lambda_k^*} \right|
\]

where the infimum is taken over all estimator \( u_{a,i} \) based on an observed matrix \( M = A + H \), and the estimation target is \( a^\top u_i(A) \) with \( u_i(A) \) denoting the \( l \)-th eigenvector of the matrix \( A \). Here, the sets \( M_0 \) and \( M_1 \) are defined respectively by

\[
M_0(M^*) := \left\{ A \mid \text{rank}(A) = r, \lambda_i(A) = \lambda_i^* (1 \leq i \leq r), \|A - M^*\|_F \leq \frac{\sigma}{2} \right\}
\]

\[
M_1(M^*) := \left\{ A \mid \text{rank}(A) = r, \lambda_i(A) = \lambda_i^* (1 \leq i \leq r), \|u_i(A) - u_i^*\|_2 \leq \frac{\sigma}{4|\lambda_i^*|} \right\}
\]

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

\[
\text{dist} \left( u_{a,\text{debiased}}, a^\top u_i^* \right) \leq \tilde{O} \left( \frac{\sigma^2}{(\Delta_i^*)^2} |a^\top u_i^*| + \sigma \sum_{k:k \neq l} \left| \frac{a^\top u_k^*}{\lambda_i^* - \lambda_k^*} \right| + \frac{\sigma}{|\lambda_i^*|} \right)
\]

in this scenario, which matches the minimax lower bound \( 3.8 \) (modulo some logarithmic factor). This confirms the near optimality of our de-biased estimator.

**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

\[
\Delta_i^* = \Omega (\mathbb{E}[|H|]) = \Omega(\sqrt{n})
\]

which is \( \tilde{O}(\sqrt{n/r}) \) times more stringent than the eigen-gap condition imposed in our theory (see \( 3.6 \)).

- 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 \( M \) 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 \( u_{a,\text{debiased},\text{KD}} \) proposed in Koltchinskii and Xia [2016]. When \( r \approx 1 \), Koltchinskii and Xia [2016, Theorem 1.3] asserts that

\[
\text{dist} \left( u_{a,\text{debiased},\text{KD}}, a^\top u_i^* \right) \leq \tilde{O} \left( \frac{\sigma}{\Delta_i^*} \right) = E_{\text{KD},l}
\]

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

\[
\frac{E_{\text{KD},l}}{E_{\text{md},l}} = \tilde{O} \left( \frac{\Delta_i^*}{\sigma |a^\top u_i^*|} \wedge \sum_{k:k \neq l} \frac{1}{|a^\top u_k^*|} \wedge \frac{|\lambda_i^*|}{\Delta_i^*} \right)
\]


For concreteness, consider the case with \( r = 1 \), \( \|a^\top u_1^*\| \approx 1/\sqrt{n} \) for all \( k \neq l \), \( \Delta_l^* \approx |\lambda_l^*|/\sqrt{n} \), and \( \Delta_l^* \approx \sigma \sqrt{n} \), thus leading to the gain
\[
\frac{E_{KD,l}}{E_{md,l}} = \tilde{\Theta}(\sqrt{n}).
\]

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

### 3.2 Principal component analysis

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

\[
3.2 \text{ Principal component analysis}
\]

the data matrix whose columns consist of i.i.d. samples \( s_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma) \), and let \( \lambda_l \) represent the \( l \)-th largest eigenvalue of \( \frac{1}{n}SS^\top \) with associated eigenvector \( u_l \). Our focus is the following two estimators aimed at estimating the linear form \( a^\top u_l^* \) (\( 1 \leq l \leq r \)).

- A plug-in estimator:
  \[
  \hat{u}_a^{\text{plugin}} := a^\top u_l;
  \quad (3.9a)
  \]

- A “de-biased” estimator:
  \[
  \hat{u}_a^{\text{debiased}} := \sqrt{1+c_l a^\top u_l};
  \quad (3.9b)
  \]

Here, \( c_l \) is a quantity that can be directly computed using the spectrum of \( \frac{1}{n}SS^\top \) as follows:

\[
c_l := \begin{cases} 
\frac{\lambda_l}{n + \sum_{i \neq l \leq n} \frac{\lambda_i}{\lambda_l - \lambda_i}} \sum_{i \neq l \leq n} \frac{\lambda_i}{(\lambda_l - \lambda_i)^2}, & \text{if } n \geq p, \\
\frac{\sigma_p^2}{n} + \frac{\lambda_l}{n} \sum_{i \neq l \leq n} \frac{\lambda_i}{\lambda_l - \lambda_i} \sum_{i \neq l \leq n} \frac{\lambda_i - \sigma_p^2}{(\lambda_l - \lambda_i)^2}, & \text{if } n < p,
\end{cases}
\]

(3.10)

without any need of using sample splitting.

Akin to the matrix denoising counterpart, the plug-in estimator (3.9a) often incurs some non-negligible estimation bias, which motivates the design of the adjusted estimator (3.9b) to compensate for the bias.

We are now ready to present our statistical guarantees for the two estimators introduced in (3.9).

**Theorem 2.** Consider any \( 1 \leq l \leq r \), and assume that

\[
\lambda_{\max}^* \sqrt{\frac{r}{n}} + \sqrt{\lambda_{\max}^* \sigma^2 \frac{p}{n}} + \sigma \left( \frac{p}{n} + \sqrt{\frac{p}{n}} \right) \leq C_0 \frac{\lambda_{\min}^*}{\log^2 n}
\]

and

\[
\Delta_l^* > C_1 (\lambda_{\max}^* + \sigma^2) \sqrt{\frac{r}{n}} \log n
\]

(3.11a) hold for some sufficiently small (resp. large) constant \( C_0 > 0 \) (resp. \( C_1 > 0 \)). Consider any fixed vector \( a \in \mathbb{R}^p \) with \( \|a\|_2 = 1 \). Then with probability at least \( 1 - O(n^{-10}) \), the estimators in (3.9) satisfy

\[
\text{dist} \left( \hat{u}_a^{\text{plugin}}, a^\top u_l^* \right) \lesssim E_{PCA,l} + \frac{(\lambda_l^* + \sigma^2) \sigma_p^2}{\lambda_l^* n} \|a^\top u_l^*\|,
\]

(3.12a)

\[
\text{dist} \left( \hat{u}_a^{\text{debiased}}, a^\top u_l^* \right) \lesssim E_{PCA,l},
\]

(3.12b)

where the quantity \( E_{PCA,l} \) is defined as

\[
E_{PCA,l} := \frac{(\lambda_{\max}^* + \sigma^2)(\lambda_l^* + \sigma^2) r \log n}{(\Delta_l^*)^2 n} \|a^\top u_l^*\|
\]

\[
+ \sum_{k: k \neq l} \frac{\|a^\top u_k^*\|}{\lambda_l^* - \lambda_k^*} \left( \frac{\lambda_l^* + \sigma^2}{\lambda_{\max}^* + \sigma^2} \right) \log \left( \frac{n k \lambda_{\max}^*}{\Delta_l^*} \right).
\]

(3.13)
Implications. In short, Theorem 2 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, $E_{\text{PCA},l}$ in (3.13)). For simplicity of presentation, we shall focus on the case with $r, \kappa \asymp 1$, where the error term $E_{\text{PCA},l}$ admits the following simpler expression

$$E_{\text{PCA},l} = \tilde{\Theta} \left( \frac{(\lambda_1^* + \sigma^2)^2}{(\Delta_1^*)^2 n} \right) |a^T u_1^*| + (\lambda_1^* + \sigma^2) \max_{k: k \neq l} \frac{|a^T u_k^*|}{|\lambda_1^* - \lambda_k^*| \sqrt{n}} + \frac{\sigma}{\lambda_1^* \sqrt{n}} \sqrt{\lambda_1^* + \sigma^2}. \tag{3.14}$$

In particular, the first term on the right-hand side of (3.14) quantifies the role of the ground truth $a^T u_1^*$ on the estimation error, which scales inverse quadratically in the eigen-gap $\Delta_1^*$; the second term on the right-hand side of (3.14) can be understood as the additional interference resulting from the linear form of other eigenvectors (namely, $a^T u_k^*$ for $k \neq l$), which is inversely proportional to the corresponding eigen-gap $|\lambda_1^* - \lambda_k^*|$.

• 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 $\sin \Theta$ theorem [Davis and Kahan, 1970]) requires the eigen-gap to exceed the size of perturbation, namely,

$$\Delta_1^* \gtrsim \frac{1}{n} \| S S^T - \Sigma \|.$$  

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)

$$\Delta_1^* \gtrsim \lambda_1^* + \frac{\sqrt{\lambda_1^* \sigma^2 p}}{\sqrt{n}} + \sigma \left( \sqrt{\frac{p}{n}} + \frac{p}{n} \right) := \text{gap}_{\text{DK}}.$$  

In comparison, the eigen-gap condition (3.11b) in Theorem 2 reads

$$\Delta_1^* \gtrsim \frac{\lambda_1^* + \sigma^2}{\sqrt{n}} := \text{gap}.$$  

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

- If $\sigma^2 (\sqrt{\frac{p}{n}} + \frac{p}{n}) \lesssim \lambda_1^* \lesssim \sigma^2$ (the sample size needs to satisfy $n \geq p$ by the assumption (3.11a)), the eigen-gap conditions above simplify to

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

$$\Rightarrow \quad \frac{\text{gap}_{\text{DK}}}{\text{gap}} \asymp \sqrt{p}.$$  

- If $\sigma^2 \lesssim \lambda_1^* \lesssim \sigma^2 p$, then one has

$$\text{gap} \asymp \frac{\lambda_1^*}{\sqrt{n}} \quad \text{and} \quad \text{gap}_{\text{DK}} \asymp \sqrt{\frac{\lambda_1^* \sigma^2 p}{n} + \frac{\sigma^2 p}{n}}.$$  

Comparing these two terms reveals that

$$\frac{\text{gap}_{\text{DK}}}{\text{gap}} \asymp \sqrt{\frac{\sigma^2 p}{\lambda_1^*} \left( 1 + \frac{\sigma^2 p}{\lambda_1^* n} \right)} \asymp \sqrt{\frac{\sigma^2 p}{\lambda_1^*} \gtrsim 1},$$

where (i) holds due to the assumption (3.11a) and (ii) follows from the condition $\lambda_1^* \lesssim \sigma^2 p$.  

Theorem 3. Consider any fixed vector $a \in \mathbb{R}^p$. For any given $\Sigma$, let $\{s_i\}_{i=1}^n$ be independent samples satisfying $s_i \overset{i.i.d.}\sim \mathcal{N}(0, \Sigma + \sigma^2 I_p)$. Assume that the sample size obeys
\[
 n \geq \left\{ \max_{k: k \neq l} \frac{(\lambda_k^* + \sigma^2)(\lambda_k^* + \sigma^2)}{|\lambda_k^* - \lambda_l^*|^2} \right\} \vee \frac{(\lambda_k^* + \sigma^2)^2}{\lambda_k^*}.
\]
Then one has

\[
\inf_{u_{a,l}} \sup_{\Sigma \in M_2(\Sigma^*)} E \left[ \min \left\{ |u_{a,l} \pm \alpha^T u_l(\Sigma)| \right\} \right] 
\geq \max_{k: k \neq l, \ 1 \leq k \leq r} \frac{(\lambda_k^* + \sigma^2)(\lambda_l^* + \sigma^2)}{|\lambda_k^* - \lambda_l^*|^2 n} |a^T u_k^*| + \max_{k: k \neq l, \ 1 \leq k \leq r} \frac{\sqrt{(\lambda_k^* + \sigma^2)(\lambda_l^* + \sigma^2)}}{|\lambda_k^* - \lambda_l^*|\sqrt{n}} |a^T u_k^*| =: E_{ib1,l};
\]

\[
\inf_{u_{a,l}} \sup_{\Sigma \in M_2(\Sigma^*)} E \left[ \min \left\{ |u_{a,l} \pm \alpha^T u_l(\Sigma)| \right\} \right] 
\geq \sqrt{\frac{(\lambda_k^* + \sigma^2)^2}{\lambda_l^* n}} \|P_{U^*} a\|_2 =: E_{ib2,l}.
\]

Here, the infimum is taken over all estimator \(u_{a,l}\) 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

\[
r, \kappa \asymp 1 \quad \text{and} \quad |a^T u_l^*| \leq (1-\epsilon)\|a\|_2
\]

for some arbitrarily small constant \(\epsilon > 0\). In this scenario, the statistical error bound (3.12b) derived in Theorem 2 matches the preceding minimax lower bounds in the sense that

\[
E_{PCA,l} \asymp E_{ib1,l} + E_{ib2,l}.
\]

To verify this relation under the conditions (3.18), it is sufficient to see that

\[
\max_{k: k \neq l} \frac{\sqrt{(\lambda_k^* + \sigma^2)(\lambda_l^* + \sigma^2)}}{|\lambda_k^* - \lambda_l^*|\sqrt{n}} |a^T u_k^*| + \sum_{k: k \neq l} \frac{\sqrt{(\lambda_k^* + \sigma^2)^2}}{|\lambda_l^* - \lambda_k^*|\sqrt{n}} |a^T u_k^*| + \max_{k: k \neq l} \frac{\sqrt{(\lambda_k^* + \sigma^2)^2}}{|\lambda_l^* - \lambda_k^*|\sqrt{n}} \|P_{U^*} a\|_2
\]

where (i) holds true since \(\max_{k: k \neq l} \frac{\sqrt{(\lambda_k^* + \sigma^2)(\lambda_l^* + \sigma^2)}}{|\lambda_k^* - \lambda_l^*|\sqrt{n}} \geq \sqrt{\frac{(\lambda_k^* + \sigma^2)(\lambda_l^* + \sigma^2)}}{|\lambda_k^* - \lambda_l^*|\sqrt{n}}\), and (ii) holds true as long as \(|a^T u_l^*| \leq (1-\epsilon)\|a\|_2\) 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 (3.18).

**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 2 with the theoretical guarantees in Koltchinskii et al. [2020] under the spiked covariance model with \(\Sigma = \Sigma^* + \sigma^2 I_p\).

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)

\[
\Delta_l^* = \Omega(\lambda_{\text{max}}^* + \sigma^2), \quad \sigma^2 = o(\lambda_{\text{max}}^*) \quad \text{and} \quad \sum_{k: k \neq l} |a^T u_k^*|^2 + \frac{\sigma^2}{\lambda_l^*} \|P_{U^*} a\|_2^2 \asymp \|a\|_2^2.
\]

In comparison, our results make improvements in the following aspects:
• **Eigen-gap requirement**: our eigen-gap requirement (3.11b) is $\tilde{O}(\sqrt{r/n})$ times less stringent than the one in (3.19);

• **Requirement on noise variance**: our result (i.e., Theorem 2) allows the noise variance $\sigma^2$ to be larger than $\lambda_{\min}^*$;

• **Requirement on condition number and rank**: our theory permits both $\kappa$ 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, Montanari and Sun, 2018, Xia et al., 2017], 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., 2020b, Stewart and Sun, 1990]. This set of classical theory typically focuses on deriving $\ell_2$ 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 $\ell_2$ 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, Zhang et al., 2018, 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 $\|H\|$ of the noise matrix is extremely small.

Moving beyond $\ell_2$ 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., 2020, 2021, Cape et al., 2019, Chen et al., 2020c, 2021, Fan et al., 2018, Lei, 2019, Ma et al., 2020, Zhong and Boumal, 2018]. Among these $\ell_\infty$ or $\ell_2,\infty$ theoretical guarantees, the results in Abbe et al. [2020b], Cai et al. [2020, 2021], Chen et al. [2020a, 2019a,b, 2020c], Ma et al. [2020] were established via a powerful leave-one-out analysis framework, while the works [Chen et al., 2021, Eldridge et al., 2018] invoked a Neumann expansion trick paired with proper control of moments.

In contrast to the rich literature on $\ell_2, \ell_\infty$ 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. [2021], Cheng et al. [2020], which, however, tackled a different setting of the matrix denoising problem. Specifically, Chen et al. [2021], Cheng et al. [2020] focused on the case where the noise matrix $H$ 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 $M$ is shown to be minimax-optimal. Additionally, Fan et al. [2019] pinned down the asymptotic distribution for bilinear
forms of eigenvectors for large spiked random matrices, while Xia and Yuan [2019] 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 2. 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 $Q \in \mathbb{R}^{n \times k}$ obeying $Q^\top Q = I_k$ ($1 \leq k \leq n$), let $Q^\perp \in \mathbb{R}^{n \times (n-k)}$ be an arbitrary matrix whose columns form an orthonormal basis of the complement to the subspace spanned by $Q$, namely

$$[Q, Q^\perp]^\top [Q, Q^\perp] = I_n. \quad (5.1)$$

Our results concern the decomposition of an eigenvector $u_l$ taking the following form:

$$u_l = u_{l,\parallel} \cos \theta + u_{l,\perp} \sin \theta. \quad (5.2)$$

Here, $\theta$ denotes the principal angle between $u_l$ and the subspace spanned by $Q$, whereas $u_{l,\parallel}$ and $u_{l,\perp}$ are two unit vectors (i.e. $\|u_{l,\parallel}\|_2 = \|u_{l,\perp}\|_2 = 1$) such that

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

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

**Theorem 4.** Consider any vector $q \in \mathbb{R}^n$ with $\|q\|_2 = 1$. Write

$$u_l = q \cos \theta + u_{l,\perp} \sin \theta \quad (5.3)$$

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

$$\cos^2 \theta = \frac{1}{1 + \|((\lambda_l I_{n-1} - (q^\perp)^\top M q^\perp)^{-1}(q^\perp)^\top M q)^\perp\|_2^2}, \quad (5.4a)$$

$$\lambda_l = q^\top M q + q^\top M q^\perp(\lambda_l I_{n-1} - (q^\perp)^\top M q^\perp)^{-1}(q^\perp)^\top M q. \quad (5.4b)$$

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

$$u_{l,\perp} = \pm \frac{q^\perp(\lambda_l I_{n-1} - (q^\perp)^\top M q^\perp)^{-1}(q^\perp)^\top M q}{\|((\lambda_l I_{n-1} - (q^\perp)^\top M q^\perp)^{-1}(q^\perp)^\top M q)^\perp\|_2}. \quad (5.4c)$$

**Proof.** See Appendix A.1. \qed

In words, Theorem 4 derives closed-form expressions for both $\cos \theta$ and $u_{l,\perp}$ (up to global signs), in terms of simple and direct manipulation of the data matrix $M$ as well as the associated eigenvalue $\lambda_l$. 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 $\lambda_l$, it often enables convenient decoupling of complicated statistical dependency, as we shall demonstrate momentarily.
When $Q$ is a more general orthonormal matrix. The next theorem extends the relation (5.4b) to the case when $Q$ is a general orthonormal matrix (beyond the vector case), which proves useful in eigenvalue analysis for more general low-rank problems.

**Theorem 5.** Assume that $k < n$. Consider the corresponding decomposition (5.2) for any matrix $Q \in \mathbb{R}^{n \times k}$ obeying $Q^\top Q = I_k$. Suppose that $\lambda_l I_{n-k} - (Q^\perp)^\top MQ^\perp$ and $\lambda_l I_k - Q^\top MQ$ are both invertible. Then one has

\[
\cos^2 \theta = \frac{1}{1 + \|((\lambda_l I_{n-k} - (Q^\perp)^\top MQ^\perp)^{-1}(Q^\perp)^\top MU_l\|_2^2}
\]

\[
(\lambda_l I_k - Q^\top MQ)Q^\top U_l = Q^\top MQ^\perp(\lambda_l I_{n-k} - (Q^\perp)^\top MQ^\perp)^{-1}(Q^\perp)^\top MU_l.
\]

**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 $\lambda_l$ is the $l$-th largest eigenvalue (in magnitude) of $M$, we present a theorem that reveals the proximity of $\lambda_l$ and the ground truth $\lambda_l^*$.

**Theorem 6** (Eigenvalue perturbation for matrix denoising). Consider the model in Section 2.1. Fix any $1 \leq l \leq r$, and instate the assumptions of Theorem 1. With probability at least $1 - O(n^{-10})$, one has

\[
|\lambda_l - \gamma(\lambda_l) - \lambda_l^*| \leq C_1 \sigma \sqrt{r \log n}
\]

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

\[
\gamma(\lambda) := \sigma^2 \text{tr}\left[\left(\lambda I_{n-r} - (U^*^\perp)^\top HU^*^\perp\right)^{-1}\right].
\]

**Remark 1.** Here, we recall that the columns of $U^*^\perp \in \mathbb{R}^{n \times (n-r)}$ form an orthonormal basis of the complement to the subspace spanned by $U^*$.

**Remark 2.** The error bound (5.6) concerning the empirical eigenvalue $\lambda_l$ 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 $\lambda_l^*$.

#### 5.2.1 Proof of eigenvalue perturbation theory (Theorem 6)

We start by demonstrating how to prove the eigenvalue perturbation bound in Theorem 6. Let us fix an arbitrary $1 \leq l \leq r$. The key ingredient of the analysis is to invoke our master theorem (namely, Theorem 5).

Before proceeding, we first verify a few useful facts. It is well known that if $\sigma \sqrt{n} \leq c_0 \lambda_{\min}^*$ for some sufficiently small constant $c_0 > 0$, then with probability exceeding $1 - O(n^{-10})$ one has (see, e.g., Chen et al. [2020b, Theorem 3.1.4])

\[
\|H\| \leq \lambda_{\min}/3.
\]

Recall that

\[
(U^*^\perp)^\top MU^*^\perp = (U^*^\perp)^\top M^* U^*^\perp + (U^*^\perp)^\top HU^*^\perp = (U^*^\perp)^\top HU^*^\perp,
\]

which together with (5.8) implies that

\[
\|(U^*^\perp)^\top MU^*^\perp\| = \|(U^*^\perp)^\top HU^*^\perp\| \leq \|H\| \leq \lambda_{\min}/3
\]

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

\[
|\lambda_l| \geq |\lambda_l^*| - \|H\| \geq 2|\lambda_l^*|/3 \quad \text{and} \quad |\lambda_l| \leq |\lambda_l^*| + \|H\| \leq 4|\lambda_l^*|/3;
\]

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

\[
\lambda I_{n-r} - (U^\perp)^\top M U^\perp
\]

is invertible

for any \(\lambda \in \mathbb{R}\) obeying \(|\lambda| \geq 2\lambda^*/3\).

With the above two observations in mind, take \(Q = U^\ast\) in Theorem 5 to show that

\[
(\lambda I_r - U^\ast^\top M U^\ast) U^\ast^\top u_{l,\|} = G(\lambda) U^\ast^\top u_{l,\|}
\]

(5.11)

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

\[
G(\lambda) := U^\ast^\top M U^\ast \left(\lambda I_{n-r} - (U^\perp)^\top M U^\perp\right)^{-1} (U^\perp)^\top M U^\ast.
\]

(5.12)

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

\[
u_l^\perp = [u_1^\perp, u_2^\perp, \ldots, u_{l-1}^\perp, u_{l+1}^\perp, \ldots, u_r^\perp, U^\perp] \in \mathbb{R}^{n \times (n-1)}
\]

(5.13)

for each \(1 \leq l \leq r\).

Recognizing that \(U^\ast^\top M^\ast U^\ast = \Lambda^\ast\), \(M^\ast U^\perp = 0\) and \((U^\perp)^\top M^\ast = 0\), we can rewrite (5.11) as

\[
(\lambda I_r - \lambda^\ast - U^\perp H U^\ast) U^\ast^\top u_{l,\|} = G(\lambda) U^\ast^\top u_{l,\|}
\]

(5.14a)

with

\[
G(\lambda) = U^\ast^\top H U^\ast \left(\lambda I_{n-r} - (U^\perp)^\top H U^\perp\right)^{-1} (U^\perp)^\top H U^\ast.
\]

(5.14b)

Rearranging terms further gives

\[
(\lambda I_r - \lambda^\ast - G^\perp(\lambda_1)) U^\ast^\top u_{l,\|} = U^\ast^\top H U^\ast U^\ast^\top u_{l,\|} + (G(\lambda) - G^\perp(\lambda_1)) U^\ast^\top u_{l,\|},
\]

(5.15)

where we define

\[
G^\perp(\lambda) := \mathbb{E}[G(\lambda) | \{U^\perp\}^\top H U^\perp],
\]

(5.16)

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

\[
\begin{align*}
\|(\lambda I_r - \lambda^\ast - G^\perp(\lambda_1)) U^\ast^\top u_{l,\|}\|_2 & \leq \|U^\ast^\top H U^\ast\| + \|G(\lambda_1) - G^\perp(\lambda_1)\|_2 \\
& \leq \|U^\ast^\top H U^\ast\| + \sup_{\lambda: |\lambda| \in [2\lambda^*/3, 4\lambda^*/3]} \|G(\lambda) - G^\perp(\lambda)\|_2
\end{align*}
\]

(5.17)

(5.18)

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

In order to justify that \(\lambda I_r - \lambda^\ast - G^\perp(\lambda_1)\|U^\ast^\top u_{l,\|}\|_2 \approx 0\), 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 \(H \in \mathbb{R}^{n \times n}\) is a symmetric matrix with \(H_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), i \geq j\) and \(\sigma\sqrt{n} \leq c_0\lambda^*/\min\) 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

\[
\sup_{\lambda: |\lambda| \in [2\lambda^*/3, 4\lambda^*/3]} \|G(\lambda) - G^\perp(\lambda)\|_2 \lesssim \frac{\sigma^2}{\lambda_{\min}} \left(\sqrt{r n \log n + r \log n}\right).
\]

(5.19)

In addition, one has

\[
G^\perp(\lambda) = \left\{\sigma^2 \text{tr}\left[(\lambda I_{n-r} - (U^\perp)^\top H U^\perp)^{-1}\right]\right\} I_r.
\]

(5.20)
Lemma 2. Under the condition (5.22) and the eigen-gap assumption (3.3), with probability at least $1 - O(n^{-10})$ one has

$$|\lambda_l - \lambda^*_l - \gamma(\lambda_l)| \leq \varepsilon_{MD}, \quad 1 \leq l \leq r. \quad (5.23)$$

Proof. See Appendix B.2.

5.2.2 Proof of eigenvector perturbation theory (Theorem 1)

Let us begin by decomposing $u_l$ along the ground-truth direction $u_l^*$ and its complement subspace as follows

$$u_l = u_l^* \cos \theta + u_{l,\perp} \sin \theta, \quad (5.24)$$

where the vector $u_{l,\perp}$ obeys $\|u_{l,\perp}\|_2 = 1$ and $u_{l,\perp} u_l^* = 0$. Writing $a = P_{U,*} a + P_{U,\perp} a$ with $P_{U,*} = \sum_{1 \leq k \leq r} u_k^* u_k^\top$ and $P_{U,\perp} = I - P_{U,*}$, we obtain

$$a^\top u_l = (P_{U,*} a)^\top u_l + (P_{U,\perp} a)^\top u_l$$

$$= (P_{U,*} a)^\top (u_l^* \cos \theta + u_{l,\perp} \sin \theta) + (P_{U,\perp} a, u_l)$$

$$= \sum_{k=1}^r a^\top u_k^* u_k^\top (u_l^* \cos \theta + u_{l,\perp} \sin \theta) + (P_{U,*} a, P_{U,\perp} u_l)$$

$$= a^\top u_l^* \cos \theta + \sum_{k:k \neq l} a^\top u_k^* u_k^\top u_{l,\perp} \sin \theta + (P_{U,*} a, P_{U,\perp} u_l),$$

where the third line relies on the fact $P_{U,*} P_{U,\perp} = P_{U,\perp}$. It then follows that

$$a^\top u_l \pm a^\top u_l^* = a^\top u_l^* (\cos \theta \pm 1) + \sum_{k:k \neq l} a^\top u_k^* u_k^\top u_{l,\perp} \sin \theta + (P_{U,*} a, P_{U,\perp} u_l),$$

allowing us to deduce that

$$\min |a^\top u_l \pm a^\top u_l^*| \leq |a^\top u_l^*| (1 - |\cos \theta|) + \sum_{k:k \neq l} a^\top u_k^* u_k^\top u_{l,\perp} \sin \theta + |(P_{U,*} a, P_{U,\perp} u_l)|$$

$$\leq |a^\top u_l^*| (1 - \cos^2 \theta) + \sum_{k:k \neq l} a^\top u_k^* u_k^\top u_{l,\perp} \sin \theta + |(P_{U,*} a, P_{U,\perp} u_l)|. \quad (5.25)$$
Here, we define

\[
\min |a^\top u_l \sqrt{1+b_l} \pm a^\top u_l^\perp| \leq |a^\top u_l^\perp| \cdot (1 - \sqrt{1+b_l} |\cos \theta|) + \sqrt{1+b_l} |\langle P_{U_l} a, P_{U_l} u_l \rangle|
\]

\[
+ \sqrt{1+b_l} |\sum_{k:k \neq l} a^\top u_k^\perp u_k^\top u_{l,\perp}^\top \sin \theta|.
\]

As a result, it boils down to bounding the terms

\[
1 - \cos^2 \theta, \quad 1 - \sqrt{1+b_l} |\cos \theta|, \quad \sqrt{1+b_l}, \quad \sum_{k:k \neq l} a^\top u_k^\perp u_k^\top u_{l,\perp}^\top \sin \theta, \quad \langle P_{U_l} a, P_{U_l} u_l \rangle.
\]

We claim that \(\lambda_l I_{n-1} - (u_i^\perp)^\top M u_i^\perp\) 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 4 reveals that \(\cos \theta \neq 0\). If \(\sin \theta = 0\), then we have \(u_l = \pm u_l^\star\) and the conclusion is obvious since \(\min |a^\top u_l \pm a^\top u_l^\perp| = 0\). Therefore, we shall assume \(\cos \theta \neq 0\) and \(\sin \theta \neq 0\) in the remainder of the proof. Invoking Theorem 4 yields

\[
\cos^2 \theta = \frac{1}{1 + \left\| (\lambda_l I_{n-1} - M^{(l)})^{-1}(u_l^\perp)^\top M u_l^\perp \right\|^2_2},
\]

\[
u_k^{\perp} u_{l,\perp} = u_k^{\perp} u_{l,\perp} \left(\frac{u_l^\perp + (u_l^\perp)^\top M u_l^\perp}{\left\| (u_l^\perp + (u_l^\perp)^\top M u_l^\perp) \right\|^2_2} \right).
\]

Recognizing that \(M^* u_l^\star = \lambda_l^* u_l^\star\), we can alternatively write (5.27) as follows

\[
\cos^2 \theta = \frac{1}{1 + \left\| (\lambda_l I_{n-1} - M^{(l)})^{-1}(u_l^\perp)^\top M u_l^\perp \right\|^2_2},
\]

\[
u_k^{\perp} u_{l,\perp} = u_k^{\perp} u_{l,\perp} \left(\frac{u_l^\perp + (u_l^\perp)^\top M u_l^\perp}{\left\| (u_l^\perp + (u_l^\perp)^\top M u_l^\perp) \right\|^2_2} \right).
\]

Here, we define

\[
M^{(l)} := (u_l^\perp)^\top M u_l^\perp = (u_l^\perp)^\top M^* u_l^\perp + (u_l^\perp)^\top H u_l^\perp.
\]

With the above relations in mind, we can demonstrate that

\[
\left| \sum_{k:k \neq l} a^\top u_k^\perp u_k^\top u_{l,\perp}^\top \sin \theta \right| = \left| \sum_{k:k \neq l} a^\top u_k^\perp u_k^\top u_{l,\perp}^\top \sqrt{1 - \cos^2 \theta} \right|
\]

\[
= \left| \sum_{k:k \neq l} a^\top u_k^\perp u_k^\top u_{l,\perp}^\top u_l^\perp \left(\lambda_l I - M^{(l)}\right)^{-1}(u_l^\perp)^\top M u_l^\perp \right| \cdot \left| \left\| (\lambda_l I_{n-1} - M^{(l)})^{-1}(u_l^\perp)^\top M u_l^\perp \right\|^2_2 \right|
\]

\[
\leq \left| \sum_{k:k \neq l} a^\top u_k^\perp u_k^\top u_{l,\perp}^\top u_l^\perp \left(\lambda_l I - M^{(l)}\right)^{-1}(u_l^\perp)^\top M u_l^\perp \right|
\]

where the last inequality comes from the fact \(\left\| u_l^\perp (\lambda_l I - M^{(l)})^{-1}(u_l^\perp)^\top M u_l^\perp \right\|^2_2 \leq \left\| (\lambda_l I - M^{(l)})^{-1}(u_l^\perp)^\top M u_l^\perp \right\|^2_2 \) (since the columns of \(u_l^\perp\) are orthonormal). Substituting this into (5.25) and (5.26) yields

\[
\min |a^\top u_l \pm a^\top u_l^\perp| \leq |a^\top u_l^\perp| \cdot (1 - \cos^2 \theta) + \left| \langle P_{U_l} a, P_{U_l} u_l \rangle \right|
\]

\[
+ \left| \sum_{k:k \neq l} a^\top u_k^\perp u_k^\top u_{l,\perp}^\top (\lambda_l I - M^{(l)})^{-1}(u_l^\perp)^\top M u_l^\perp \right|.
\]

\[
\min |\sqrt{1+b_l a^\top u_l \pm a^\top u_l^\perp| \leq |a^\top u_l^\perp| \cdot |1 - \sqrt{1+b_l} |\cos \theta|) + \sqrt{1+b_l} |\langle P_{U_l} a, P_{U_l} u_l \rangle|
\]

\[
+ \sqrt{1+b_l} \left| \sum_{k:k \neq l} a^\top u_k^\perp u_k^\top u_{l,\perp}^\top (\lambda_l I - M^{(l)})^{-1}(u_l^\perp)^\top M u_l^\perp \right|.
\]

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

\[
U_{-l}^{(l)} := [u_1^{(l)}, \cdots, u_{l-1}^{(l)}, u_{l+1}^{(l)}, \cdots, u_{n-1}^{(l)}] \in \mathbb{R}^{n \times (r-1)},
\]

\[
U^{(l)} := (u_1^{(l)})^\top U_{-l}^{(l)} = \begin{bmatrix} I_{l-1} & 0 \end{bmatrix} \in \mathbb{R}^{(n-1) \times (r-1)},
\]

\[
U^{(l)\perp} := (u_1^{(l)})^\top U^{(l)\perp} = \begin{bmatrix} 0 & I_{n-r} \end{bmatrix} \in \mathbb{R}^{(n-1) \times (n-r)},
\]

\[
\Lambda^{(l)} := \text{diag} \{\{\lambda_i^{(l)}\}_{i \neq l}\} \in \mathbb{R}^{(r-1) \times (r-1)},
\]

and define

\[
u_k^{(l)} := \frac{1}{\|P_{U^{(l)\perp}} u_k^{(l)}\|_2} P_{U^{(l)\perp}} u_k^{(l)}
\]

for each $k \neq l$.

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

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

1. For each $1 \leq k < r$, one has $\lambda_k^{(l)} - \gamma(\lambda_k^{(l)}) \in \mathcal{E}_{\text{MD}}(\lambda_k^{*})$ for some $i \neq l$, and

\[
\|((\lambda_k^{(l)} I_{r-1} - \gamma(\lambda_k^{(l)}) I_{r-1} - \Lambda^{(l)}) U^{(l)\top} u_k^{(l)}\|_2 \lesssim \sqrt{r} \log n.
\]

2. For each $k \geq r$, one has $|\lambda_k^{(l)}| \lesssim \sigma \sqrt{n}$;

3. Moreover, one has

\[
|\lambda - \lambda_k| \gtrsim \begin{cases} \Delta_i^*, & \text{if } \lambda - \gamma(\lambda) \in \mathcal{E}_{\text{MD}}(\lambda_k^*) \text{ for some } k \neq l \text{ and } 1 \leq k \leq r; \\
|\lambda_k^*|, & \text{if } |\lambda| \lesssim \sigma \sqrt{n}. \end{cases}
\]

In particular, we have

\[
|\lambda_k^{(l)} - \lambda_l| \gtrsim \begin{cases} \Delta_i^*, & 1 \leq k < r; \\
|\lambda_k^*|, & k \geq r. \end{cases}
\]

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 $\lambda_l^*$;

- For any $k \geq r$, the eigenvalue $\lambda_k^{(l)}$ is reasonably small;

- Any eigenvalue of $M^{(l)}$ is sufficiently separated from the $l$-th eigenvalue $\lambda_l$ of $M$, 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 $\|((\lambda_l I - M^{(l)})^{-1}(u_i^{(l)})^\top H u_i^*)\|_2$. A simple yet crucial observation is that: the matrix $(u_i^{(l)})^\top H u_i^*$ is independent of $(u_i^{(l)})^\top H u_i^*$ (which follows from the same argument as in the proof of Lemma 1 in Appendix B.1). Consequently, $M^{(l)}$ (defined in (5.29)) is independent of $(u_i^{(l)})^\top H u_i^* \sim \mathcal{N}(0, \sigma^2 I_{n-1})$, which is a Gaussian random vector in $\mathbb{R}^{n-1}$. In light of this observation, we can bound $\|((\lambda_l I - M^{(l)})^{-1}(u_i^{(l)})^\top H u_i^*)\|_2$ as follows.

18
Lemma 4. Instate the assumptions of Theorem 1. The following holds with probability at least $1 - O(n^{-10})$:

$$
\| (\lambda_1 I_{n-1} - M^{(i)})^{-1} (u^*_i)^\top H u^*_i \|^2_2 = \sum_{k : r < k \leq n} \frac{\sigma^2}{(\lambda_k - \lambda_k)^2} + O\left( \frac{\sigma^2 r \log n}{(\Delta^*_i)^2} + \frac{\sigma^2 \sqrt{n \log n}}{\lambda^2_i} \right),
$$

(5.34)

which further indicates that

$$
\| (\lambda_1 I_{n-1} - M^{(i)})^{-1} (u^*_i)^\top H u^*_i \|^2_2 \lesssim \frac{\sigma^2 r \log n}{(\Delta^*_i)^2} + \frac{\sigma^2 \sqrt{n \log n}}{\lambda^2_i} \ll 1.
$$

(5.35)

Proof. See Appendix C.2.

Combining this lemma with (5.28a), we reach

$$
\cos^2 \theta = \frac{1}{1 + \| (\lambda_1 I_{n-1} - M^{(i)})^{-1} (u^*_i)^\top H u^*_i \|^2_2} = 1 + O\left( \frac{\sigma^2 r \log n}{(\Delta^*_i)^2} + \frac{\sigma^2 \sqrt{n \log n}}{\lambda^2_i} \right) \asymp 1,
$$

(5.36)

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

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

one arrives at

$$
\left| (1 + b_i) \cos^2 \theta - 1 \right| = \left| \frac{1 + b_i}{1 + \| (\lambda_1 I_{n-1} - M^{(i)})^{-1} (u^*_i)^\top H u^*_i \|^2_2} - 1 \right|
\leq \left| b_i - \| (\lambda_1 I_{n-1} - M^{(i)})^{-1} (u^*_i)^\top H u^*_i \|^2_2 \right|
\leq \left| b_i - \| (\lambda_1 I_{n-1} - M^{(i)})^{-1} (u^*_i)^\top H u^*_i \|^2_2 \right|
\lesssim \frac{(\sigma^2 r \log n)}{(\Delta^*_i)^2} + \frac{\sigma^2 \sqrt{n \log n}}{\lambda^2_i} \ll 1.
$$

(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 $1 + b_i \asymp 1$ and

$$
|1 - \sqrt{1 + b_i} \cos \theta| = \left| \frac{1 - (1 + b_i) \cos^2 \theta}{1 + \sqrt{1 + b_i} \cos \theta} \right| \leq |1 - (1 + b_i) \cos^2 \theta| \lesssim \frac{\sigma^2 r \log n}{(\Delta^*_i)^2} + \frac{\sigma^2 \sqrt{n \log n}}{\lambda^2_i}.
$$

(5.38)

3. Controlling $\sum_{k : k \neq i} a^\top u_k^* \cdot u_k^\top (\lambda_1 I - M^{(i)})^{-1} (u^*_i)^\top H u^*_i$. The key observation is that $(u^*_i)^\top H u^*_i \sim \mathcal{N}(0, I_{n-1})$ is independent of $M^{(i)}$ (but dependent of $\lambda_i$). 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

$$
\left| \sum_{k : k \neq i} a^\top u_k^* \cdot u_k^\top (\lambda_1 I_{n-1} - M^{(i)})^{-1} (u^*_i)^\top H u^*_i \right|
\lesssim \frac{\sigma}{|\lambda^*_i|} \sqrt{\log \left( \frac{n \kappa_{\max}}{\Delta^*_i} \right)} + \sqrt{r \log \left( \frac{n \kappa_{\max}}{\Delta^*_i} \right)} \sum_{k : k \neq i} |a^\top u^*_i| \frac{1}{|\lambda^*_i - \lambda_k^*|}.
$$

(5.39)
4. Controlling $\langle P_{U^\perp} a, P_{U^\perp} u_i \rangle$. When it comes to the last term $\langle P_{U^\perp} a, P_{U^\perp} u_i \rangle$, one can take advantage of the rotational invariance of $P_{U^\perp} u_i$ in the subspace spanned by $U^\perp$ 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 $1 - O(\lambda^{10})$,

$$\|\langle P_{U^\perp} a, P_{U^\perp} u_i \rangle\| \leq \sqrt{\frac{\log n}{n}} \|P_{U^\perp} a\|_2 \|P_{U^\perp} u_i\|_2,$$  

(5.40)

Consequently, it remains to upper bound $\|P_{U^\perp} u_i\|_2$. Recalling $u_i := P_{U^\perp}(u)/\|P_{U^\perp}(u)\|_2$ defined in in Section 5.1. By virtue of Theorem 5, one has

$$\|P_{U^\perp} u_i\|_2^2 = 1 - \frac{1}{1 + \|\langle \lambda_i I_{n-r} - (U^\perp)^T MU^\perp \rangle^{-1}(U^\perp)^T M u_i, \|_2^2}
\leq \|\langle \lambda_i I_{n-r} - (U^\perp)^T MU^\perp \rangle^{-1}(U^\perp)^T M u_i, \|_2^2
\leq \|\langle \lambda_i I_{n-r} - (U^\perp)^T HU^\perp \rangle^{-1}\|_2^2\|((U^\perp)^T H u_i, \|_2^2,$$  

(5.41)

where the last inequality makes use of the fact that

$$(U^\perp)^T M = (U^\perp)^T M^* + (U^\perp)^T H = (U^\perp)^T H.$$  

Additionally, it is easily seen that

$$\|\lambda_i I_{n-r} - (U^\perp)^T HU^\perp \| \geq |\lambda_i| - \|((U^\perp)^T HU^\perp) \| \simeq |\lambda_i|$$

$$\|U^\perp^T H u_i, \|_2 \leq \|H\| \simeq \sigma \sqrt{n}$$

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

$$\|P_{U^\perp} u_i\|_2 \simeq \frac{\sigma \sqrt{n}}{|\lambda_i|}.$$  

(5.42)

Substitution into (5.40) reveals that

$$\|\langle P_{U^\perp} a, P_{U^\perp} u_i \rangle\| \leq \sqrt{\frac{\log n}{n}} \|P_{U^\perp} a\|_2 \|P_{U^\perp} u_i\|_2 \simeq \frac{\sigma \sqrt{\log n}}{|\lambda_i|} \|P_{U^\perp} a\|_2.$$  

(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

$$\min |a^T u_i \pm a^T u_i^*| \simeq \left(\frac{\sigma^2 n}{\lambda_i^2} + \frac{\sigma^2 r \log n}{\lambda_i^2}\right) |a^T u_i^*| + \sigma \sqrt{r \log \left(\frac{n k \lambda_{\max}}{\Delta_i^*}\right) \sum_{k : k \neq i} |a^T u_k^*|}
+ \sigma \frac{\sqrt{\log n}}{|\lambda_i^*|} \|P_{U^\perp} a\|_2.$$  

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

$$\min |a^T u_i \sqrt{1 + b_i} \pm a^T u_i^*| \simeq \left(\frac{\sigma^2 \sqrt{n \log n}}{\lambda_i^2} + \frac{\sigma^2 r \log n}{\lambda_i^2}\right) |a^T u_i^*| + \sigma \sqrt{r \log \left(\frac{3 n k \lambda_{\max}}{\Delta_i^*}\right) \sum_{k : k \neq i} |a^T u_k^*|}
+ \sigma \frac{\sqrt{\log n}}{|\lambda_i^*|} \|P_{U^\perp} a\|_2
\leq \frac{\sigma^2 r \log n}{\Delta_i^*} |a^T u_i^*| + \sigma \frac{\sqrt{\log n}}{|\lambda_i^*|} \sum_{k : k \neq i} |a^T u_k^*|
+ \frac{\sigma}{|\lambda_i^*|} \sqrt{\log \left(\frac{3 n k \lambda_{\max}}{\Delta_i^*}\right)},$$

where the last step holds since $\sigma \sqrt{n} \simeq \lambda_{\min}^*$ and $|a^T u_i^*| \leq \|a\|_2 \|u_i^*\|_2 = 1$. 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 $U^\perp \in \mathbb{R}^{p \times (p-r)}$ represent a matrix consisting of orthonormal columns perpendicular to the subspace spanned by $U^*$.

**Theorem 7** (Eigenvalue perturbation for PCA). Consider the model in Section 2.2. Fix any $1 \leq l \leq r$, and instate the assumptions of Theorem 2. Then with probability at least $1 - O(n^{-10})$, one has

$$\left| \frac{\lambda_l}{1 + \beta(\lambda_l)} - \lambda^*_l - \sigma^2 \right| \leq C_2(\lambda_{\max}^* + \sigma^2)\left(\frac{r}{n}\log n\right)$$

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

$$\beta(\lambda) := \frac{1}{n} \text{tr}\left[ \frac{1}{n}SS^T (\lambda I_{p-r} - \frac{1}{n}S_\perp S_\perp^T)^{-1}S_\perp \right] \quad \text{with} \quad S_\perp := (U^\perp)^T S.$$  

**Remark 3.** As asserted by Theorem 7, the empirical eigenvalue $\lambda_l$ exhibits a form of “inflation” in comparison to the corresponding ground-truth value $\lambda^*_l + \sigma^2$. As a result, it is advisable to properly shrink $\lambda_l$ when estimating $\lambda_l^* + \sigma^2$.

In what follows, we shall first outline the proof for Theorem 7 (which is very similar to the analysis for Theorem 6), followed by a proof sketch for the eigenvector perturbation theory in Theorem 2.

5.3.1 Proof of eigenvalue perturbation theory (Theorem 7)

Before embarking on the proof, we shall define

$$S_\parallel := U^*S \in \mathbb{R}^{r \times n}, \quad S_\perp := (U^\perp)^T S \in \mathbb{R}^{(p-r) \times n} \quad \text{and} \quad \Lambda := U^*\Sigma U^* = \Lambda^* + \sigma^2 I_r$$

for notional convenience, allowing one to express

$$(U^\perp)^T SS^T U^\perp = S_\perp S_\perp^T, \quad (U^\perp)^T SS^T U^* = S_\perp S_\parallel^T, \quad U^*SS^T U^* = S_\parallel S_\parallel^T.$$  

As can be straightforwardly verified:

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

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

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

$$\left\| \frac{1}{n}SS^T - \Sigma \right\| \lesssim \lambda_{\max}^* \sqrt{\frac{r\log n}{n}} + \sqrt{(\lambda_{\max}^* + \sigma^2)\sigma^2} \frac{p}{n} \log n + \sigma^2 \left( \sqrt{\frac{p}{n}} + \frac{p}{n} + \sqrt{\frac{\log n}{n}} \right).$$  

**Proof.** See Appendix D.1. □

**Remark 4.** In particular, under the noise assumption (3.11a), Lemma 7 tells us that $\left\| \frac{1}{n}SS^T - \Sigma \right\| \ll \lambda_{\min}^*$ with probability at least $1 - O(n^{-10})$, which together with Weyl’s inequality gives

$$2\lambda_l^* / 3 \leq \lambda_l \leq 4\lambda_l^*/3, \quad 1 \leq l \leq r.$$  

21
We now move on to present the proof of Theorem 7. The key ingredient underlying the analysis is, once again, to invoke our master theorem (namely, Theorem 5), by treating $\frac{1}{n}SS^T$, $\Sigma^*$ and $U^*$ as $M$, $M^*$ and $Q$, respectively. Recalling the definition of

$$u_{i,\|} := \frac{1}{\|P_{U^*}(u)\|_2}P_{U^*}(u)$$  \hspace{1cm} (5.50)$$
as in Section 5.1 (so that $U^*U^TU_{i,\|} = u_{i,\|}$), one can invoke (5.5b) in Theorem 5 to derive

$$\left(\lambda I_r - \frac{1}{n}S_{\|}S_{\|}^T\right)U^TU_{i,\|} = K(\lambda I)U^TU_{i,\|},$$  \hspace{1cm} (5.51)$$
where we recall the definitions of $S_{\|}$ and $S_{\perp}$ in (5.46), and $K(\lambda)$ is given by

$$K(\lambda) := \frac{1}{n}S_{\|} - \frac{1}{n}S_{\|}^T \left(\lambda I_p - \frac{1}{n}S_{\perp}S_{\perp}^T\right)^{-1}S_{\perp}S_{\perp}^T.$$  \hspace{1cm} (5.52)$$
It is also helpful to define

$$K^\perp(\lambda) := \mathbb{E}[K(\lambda) \mid C(\lambda)],$$  \hspace{1cm} (5.53)$$
with $\lambda$ regarded as a deterministic quantity independent of the data samples. Then rearranging terms in (5.51) yields

$$\left(\lambda I_r - \Lambda - K^\perp(\lambda I)\right)U^TU_{i,\|} = \left(\frac{1}{n}S_{\|}S_{\|}^T - \Lambda + K(\lambda I) - K^\perp(\lambda I)\right)U^TU_{i,\|},$$

which together with (4.49) results in the following bound:

$$\| \left(\lambda I_r - \Lambda - K^\perp(\lambda I)\right)U^TU_{i,\|} \|_2 \leq \| \frac{1}{n}S_{\|}S_{\|}^T - \Lambda \| + \sup_{\lambda: \lambda \in [2\lambda^*_l/3,4\lambda^*_l/3]} \| K(\lambda) - K^\perp(\lambda) \|,$$  \hspace{1cm} (5.54)$$
Akin to the proof of Theorem 6 in Section, our goal is to show $\left(\lambda I_r - \Lambda - K^\perp(\lambda I)\right)U^TU_{i,\|} \approx 0$, which would then imply that $\lambda_1$ is sufficiently close to some eigenvalue of $\Lambda + K^\perp(\lambda I)$. 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 $S_{\|} = U^TS$ are independent Gaussian random vectors with distribution $\mathcal{N}(0, \Lambda)$, we can rewrite

$$S_{\|} = \Lambda^{1/2}Z,$$  \hspace{1cm} (5.55)$$
where $Z = [Z_{i,j}] \in \mathbb{R}^{r \times n}$ is a Gaussian random matrix with i.i.d. entries $Z_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Applying standard Gaussian concentration inequalities reveals that: with probability at least $1 - O(n^{-10})$,

$$\| \frac{1}{n}S_{\|}S_{\|}^T - \Lambda \| \leq \| \Lambda \| \cdot \| \frac{1}{n}ZZ^T - I_r \| \lesssim (\lambda^*_{\text{max}} + \sigma^2) \sqrt{\frac{r \log n}{n}}.$$  \hspace{1cm} (5.56)$$
- As for the second term on the right-hand side of (5.54), we claim for the moment that

$$\sup_{\lambda: \lambda \in [2\lambda^*_l/3,4\lambda^*_l/3]} \| K(\lambda) - K^\perp(\lambda) \| \ll (\lambda^*_{\text{max}} + \sigma^2) \sqrt{\frac{r \log n}{n}},$$  \hspace{1cm} (5.57)$$
$$\sup_{\lambda: \lambda \in [2\lambda^*_l/3,4\lambda^*_l/3]} \| C(\lambda) \| \ll \frac{\sigma^2}{\lambda^*_l} \left(1 + \frac{p}{n}\right),$$  \hspace{1cm} (5.58)$$
where $C(\lambda)$ 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 $1 - O(n^{-10})$,

$$
\| (\lambda_l I - \Lambda - K^\perp(\lambda_l)) u_l \|_2 \lesssim (\lambda_{\text{max}}^* + \sigma^2) \sqrt{\frac{r}{n} \log n} =: \epsilon_{\text{PCA}}.
$$

(5.59)

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

$$
K^\perp(\lambda) = \beta(\lambda) \Lambda,
$$

(5.60)

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

$$
(1 + \beta(\lambda_l))(\lambda_l^* + \sigma^2).
$$

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

$$
\sup_{\lambda : \lambda \in [2\lambda_l^{\ast}, 4\lambda_l^{\ast}/3]} \beta(\lambda) \leq \sup_{\lambda : \lambda \in [2\lambda_l^{\ast}, 4\lambda_l^{\ast}/3]} \frac{n \wedge \text{p} \wedge \lambda}{n} \| C(\lambda) \| \lesssim \frac{n \wedge \text{p} \wedge \lambda^2}{n} \left( 1 + \frac{\text{p}}{\lambda_l^*} \right) \approx \frac{\text{p} \lambda^2}{\lambda_l^* n} = o(1)
$$

(5.61)

as long as the noise level obeys $\sigma^2 p/n \ll \lambda_{\text{min}}^*/\log n$. 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

$$
\left| \lambda_l - (\lambda_l^* + \sigma^2)(1 + \beta(\lambda_l)) \right| \lesssim (\lambda_{\text{max}}^* + \sigma^2) \sqrt{\frac{r}{n} \log n};
$$

for conciseness, we omit the details of proof. This inequality establishes the proximity of $\lambda_l$ and $(\lambda_l^* + \sigma^2)(1 + \beta(\lambda_l))$. Taking this collectively with (5.61) (i.e., $1 + \beta(\lambda_l) \approx 1$), we establish the advertised bound (5.44).

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

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

which allow us to express

$$
K(\lambda) := \frac{1}{n} S \parallel C(\lambda) S^\top \parallel = \frac{1}{n} \Lambda^{1/2} Z C(\lambda) Z^\top \Lambda^{1/2} = \Lambda^{1/2} \overline{K}(\lambda) \Lambda^{1/2},
$$

$$
K^\perp(\lambda) := \mathbb{E} \left[ K(\lambda) \mid C(\lambda) \right] = \Lambda^{1/2} \mathbb{E} \left[ \overline{K}(\lambda) \mid C(\lambda) \right] \Lambda^{1/2} = \Lambda^{1/2} \overline{K}^\perp(\lambda) \Lambda^{1/2}.
$$

One can then develop the following upper bound

$$
\| K(\lambda) - K^\perp(\lambda) \| = \| \Lambda^{1/2} (K(\lambda) - \overline{K}^\perp(\lambda)) \Lambda^{1/2} \| \leq \| \Lambda \| \| K(\lambda) - \overline{K}^\perp(\lambda) \| = (\lambda_{\text{max}}^* + \sigma^2) \frac{1}{n} \| Z C(\lambda) Z^\top - \mathbb{E} [Z C(\lambda) Z^\top \mid C(\lambda)] \|.
$$

(5.62)

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

The remainder of the proof thus comes down to controlling

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

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

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

23
where the last step arises from the rotational invariance of the Gaussian random matrix, namely \( ZV = Z \).

In view of Lemma 18, it suffices to control the eigenvalues of \( C(\lambda) \).

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

\[
\lambda_i(C(\lambda)) = \lambda_i \left( \frac{1}{n} S^\top \left( \lambda I_{p-r} - \frac{1}{n} S S^\top \right)^{-1} S \right) = \frac{\lambda_i \left( \frac{1}{n} S S^\top \right)}{\lambda - \lambda_i \left( \frac{1}{n} S S^\top \right)}, \quad 1 \leq i \leq (p - r) \wedge n.
\]

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

\[
0 \leq \lambda_i \left( \frac{1}{n} S S^\top \right) \lesssim \sigma^2 \left( 1 + \sqrt{\frac{p}{n}} + \sqrt{\frac{p}{n} + \sqrt{\frac{\log n}{n}}} \right) \lesssim \lambda_{\min}^*, \quad 1 \leq i \leq (p - r) \wedge n,
\]

where the last step holds due to the noise assumption (3.11a). Combining these two observations establishes the claim bound (5.58):

\[
\sup_{\lambda : \lambda \in [2\lambda^*/3, 3\lambda^*/3]} \| C(\lambda) \| \lesssim \sigma^2 \lambda_i^* \left( 1 + \sqrt{\frac{p}{n}} + \frac{p}{n} + \sqrt{\frac{\log n}{n}} \right) \lesssim \sigma^2 \left( 1 + \frac{p}{n} \right),
\]

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

\[
\sup_{\lambda : \lambda \in [2\lambda^*/3, 3\lambda^*/3]} \frac{1}{n} \| ZC(\lambda) Z^\top - E[ZC(\lambda) Z^\top | C(\lambda)] \| \lesssim \sup_{\lambda : \lambda \in [2\lambda^*/3, 3\lambda^*/3]} \frac{1}{n} \| C(\lambda) \| \sqrt{r \log n} + \sup_{\lambda : \lambda \in [2\lambda^*/3, 3\lambda^*/3]} \frac{1}{n} \| C(\lambda) \| (r \log n + \log^2 n)
\]

\[
\lesssim \sup_{\lambda : \lambda \in [2\lambda^*/3, 3\lambda^*/3]} \frac{1}{n} \| C(\lambda) \| \sqrt{r(n \wedge p)} \log^2 n
\]

\[
\lesssim \frac{\sigma^2}{\lambda_i^*} \left( \frac{p}{n} + \sqrt{\frac{p}{n}} \right) \sqrt{\frac{r}{n}} \log^2 n \ll \sqrt{\frac{r}{n}} \log n
\]

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

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

as claimed.

### 5.3.2 Proof of eigenvector perturbation theory (Theorem 2)

We now turn to our eigenvector perturbation theory. As before, we find it convenient to decompose the \( l \)-th eigenvector \( u_l \) of \( \frac{1}{n} SS^\top \) as follows

\[
u_l = u_l^* \cos \theta + u_{l,\perp} \sin \theta,
\]

where the vector \( u_{l,\perp} \) obeys \( \| u_{l,\perp} \|_2 = 1 \) and \( u_{l,\perp}^T u_l^* = 0 \). We shall employ this decomposition to identify several key quantities that we’d like to control. Specifically, armed with this decomposition, we can derive

\[
a^\top u_l = (P_U \cdot a)^\top u_l + (P_{U^\perp} \cdot a)^\top u_l
\]

\[
= (P_U \cdot a)^\top (u_l^* \cos \theta + u_{l,\perp} \sin \theta) + (P_{U^\perp} \cdot a)^\top u_l
\]

\[
= \sum_{1 \leq k \leq r} a^\top u_k^* u_k^T (u_l^* \cos \theta + u_{l,\perp} \sin \theta) + (P_{U^\perp} \cdot a)^\top (P_{U^\perp} u_l)
\]
\[ a^\top u_1^* \cos \theta + \sum_{k:k \neq l} a^\top u_k^* u_k^\top u_{l,\perp} \sin \theta + (P_{U_{\perp}^*}a)^\top (P_{U_{\perp}^*}u_l), \]

where we use the fact that \( a = P_{U_{\perp}^*}a + P_{U_{\perp}}a \) with

\[ P_{U_{\perp}^*} = \Sigma_{1 \leq k \leq n} u_k^* u_k^\top \quad \text{and} \quad P_{U_{\perp}} = I - P_{U_{\perp}^*}. \]

As a result, we arrive at

\[ a^\top u_l + a^\top u_1^* (\cos \theta \pm 1) + \sum_{k:k \neq l} a^\top u_k^* u_k^\top u_{l,\perp} \sin \theta + (P_{U_{\perp}^*}a)^\top (P_{U_{\perp}^*}u_l), \]

which further implies

\[
\min |a^\top u_l \pm a^\top u_1^*| \leq |a^\top u_1^*| (1 - |\cos \theta|) + \left| \sum_{k:k \neq l} a^\top u_k^* u_k^\top u_{l,\perp} \sin \theta \right| + \| (P_{U_{\perp}^*}a)^\top (P_{U_{\perp}^*}u_l) \|.
\]

and

\[
\min |a^\top u_l \sqrt{1 + c_l} \pm a^\top u_1^*| \leq |a^\top u_1^*| \left| 1 - \sqrt{1 + c_l} |\cos \theta| \right| + \sqrt{1 + c_l} \| (P_{U_{\perp}^*}a, P_{U_{\perp}^*}u_l) \|
\]

Thus, it comes down to bounding the following terms

\[
1 - \cos^2 \theta, \quad 1 - \sqrt{1 + c_l} |\cos \theta|, \quad \sqrt{1 + c_l}, \quad \sum_{k:k \neq l} a^\top u_k^* u_k^\top u_{l,\perp} \sin \theta, \quad \text{and} \quad (P_{U_{\perp}^*}a, P_{U_{\perp}^*}u_l).
\]

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

We claim that \( \lambda_1 I_{p-1} - (u_1^\perp)^\top \frac{1}{n} S S^\top u_1^\perp \) is invertible. This will be seen from (5.75) stated in Lemma 8 directly. The invertibility taken together with Theorem 4 reveals that \( \cos \theta \neq 0 \). If \( \sin \theta = 0 \), then we have \( u_l = \pm u_1^* \) and the conclusion is obvious since \( \min |a^\top u_l \pm a^\top u_1^*| = 0 \). 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 4 to show that

\[
\cos^2 \theta = \frac{1}{1 + \| (\lambda_1 I_{p-1} - (u_1^\perp)^\top \frac{1}{n} S S^\top u_1^\perp)^{-1} (u_1^\perp)^\top \frac{1}{n} S S^\top u_1^\perp \|_2^2}, \quad (5.66a)
\]

\[
u_k^\top u_{l,\perp} = \frac{u_k^\top u_{l,\perp}^\perp (\lambda_1 I_{p-1} - (u_1^\perp)^\top \frac{1}{n} S S^\top u_1^\perp)^{-1} (u_1^\perp)^\top \frac{1}{n} S S^\top u_1^\perp}{\|u_1^\top (\lambda_1 I_{p-1} - (u_1^\perp)^\top \frac{1}{n} S S^\top u_1^\perp)^{-1} (u_1^\perp)^\top \frac{1}{n} S S^\top u_1^\perp \|_2}. \quad (5.66b)
\]

For notational convenience, we shall define

\[
s_{l,\|} := u_1^\top S \in \mathbb{R}^{1 \times n} \quad \text{and} \quad S_{l,\perp} := (u_1^\perp)^\top S \in \mathbb{R}^{(p-1) \times n}, \quad (5.67)
\]

allowing us to write (5.66) more succinctly as follows

\[
\cos^2 \theta = \frac{1}{1 + \| (\lambda_1 I_{p-1} - \frac{1}{n} S_{l,\|} S_{l,\perp}^\top)^{-1} \frac{1}{n} S_{l,\perp} S_{l,\|} \|_2^2}, \quad (5.68a)
\]

\[
u_k^\top u_{l,\perp} = \frac{u_k^\top u_{l,\perp}^\perp (\lambda_1 I_{p-1} - \frac{1}{n} S_{l,\|} S_{l,\perp}^\top)^{-1} \frac{1}{n} S_{l,\perp} S_{l,\|}}{\|u_1^\top (\lambda_1 I_{p-1} - \frac{1}{n} S_{l,\|} S_{l,\perp}^\top)^{-1} \frac{1}{n} S_{l,\perp} S_{l,\|} \|_2}. \quad (5.68b)
\]
With the above relations in mind, we can demonstrate that

\[
\left| \sum_{k \neq l} a^\top u_k^* u_k^\top u_{i,l} \sin \theta \right| = \left| \sum_{k \neq l} a^\top u_k^* u_k^\top u_{i,l} \right| \sqrt{1 - \cos^2 \theta}
\]

\[
= \left| \sum_{k \neq l} a^\top u_k^* u_k^\top u_{i,l}^\perp \left( \lambda I_{p-1} - \frac{1}{n} S_{i,l} S_{i,l}^\top \right)^{-1} \frac{1}{n} S_{i,l} s_{i,l}^\top \right| \cdot \sqrt{1 + \frac{\left\| \left( \lambda I_{p-1} - \frac{1}{n} S_{i,l} S_{i,l}^\top \right)^{-1} \frac{1}{n} S_{i,l} s_{i,l}^\top \right\|_2^2}{2}}
\]

\[
\leq \left| \sum_{k \neq l} a^\top u_k^* u_k^\top u_{i,l}^\perp \left( \lambda I_{p-1} - \frac{1}{n} S_{i,l} S_{i,l}^\top \right)^{-1} \frac{1}{n} S_{i,l} s_{i,l}^\top \right|,
\]

where the last step follows since the columns of \(u_i^\perp\) are orthonormal. Substitution into (5.64) then yields

\[
\min |a^\top u_i + a^\top u_i^\perp| \leq |a^\top u_i| \cdot (1 - \cos^2 \theta) + \left| \sum_{k \neq l} a^\top u_k^* u_k^\top u_{i,l}^\perp \left( \lambda I_{p-1} - \frac{1}{n} S_{i,l} S_{i,l}^\top \right)^{-1} \frac{1}{n} S_{i,l} s_{i,l}^\top \right|
\]

\[
+ \left| \left( P_{U,l} \cdot a^\top \right) (P_{U,l} \cdot u_l) \right|,
\]

\[
\min |\sqrt{1 + c} a^\top u_i + a^\top u_i^\perp| \leq |a^\top u_i| \cdot |1 - \sqrt{1 + c} \cos \theta| + \sqrt{1 + c} |\left( P_{U,l} \cdot a^\top \right) (P_{U,l} \cdot u_l) |
\]

\[
+ \sqrt{1 + c} \left| \sum_{k \neq l} a^\top u_k^* u_k^\top u_{i,l}^\perp \left( \lambda I_{p-1} - \frac{1}{n} S_{i,l} S_{i,l}^\top \right)^{-1} \frac{1}{n} S_{i,l} s_{i,l}^\top \right|.
\]  

(5.69)

(5.70)

In what follows, we shall control these quantities separately.

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

- Let \(\{\gamma_i^{(l)}\}_{i=1}^{p-1}\) denote the eigenvalues of \(\frac{1}{n} S_{i,l} S_{i,l}^\top\) (see the definition of \(S_{i,l}\) in (5.67)), and we assume that

\[
\gamma_1^{(l)} \geq \cdots \geq \gamma_{p-1}^{(l)}.
\]  

(5.71)

- Let \(u_i^{(l)}\) be the eigenvector of \(\frac{1}{n} S_{i,l} S_{i,l}^\top\) associated with the eigenvalue \(\gamma_i^{(l)}\).

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

\[
U_{\perp,l}^\top := [u_1^{\perp}, \cdots, u_{i-1}^{\perp}, u_{i+1}^{\perp}, \cdots, u_p^{\perp}] \in \mathbb{R}^{p \times (r-1)};
\]

\[
U_{\perp,l}^{(l)} := (u_{i,l}^{\perp})^\top U_{\perp,l}^\top = \begin{bmatrix} I_{r-1} \ 0 \end{bmatrix} \in \mathbb{R}^{(p-1) \times (r-1)};
\]

\[
U_{\perp,l}^{(l)\perp} := (u_{i,l}^{\perp})^\top U_{\perp,l}^{(l)} = \begin{bmatrix} 0 \\ I_{p-r} \end{bmatrix} \in \mathbb{R}^{(p-1) \times (p-r)};
\]

\[
\Lambda_{\perp,l}^{(l)} := \text{diag}\{\lambda_i^{(l)}\}_{i \neq l} \in \mathbb{R}^{(r-1) \times (r-1)}.
\]  

(5.72a)

(5.72b)

(5.72c)

(5.72d)

In addition, we define

\[
u_i^{(l)} := \frac{1}{\| P_{U,(l)} u_i^{(l)} \|_2} P_{U,(l)} u_i^{(l)}, \quad i \neq l,
\]  

(5.72e)

where \(P_{U,(l)} = U_{\perp,l}^{(l)} (U_{\perp,l}^{(l)})^\top\). Equipped with this set of notation, we are ready to present a lemma that characterizes the eigenvalues of the matrix \(\frac{1}{n} S_{i,l} S_{i,l}^\top\).

**Lemma 8.** Instate the assumptions of Theorem 2, and recall the definition of \(\beta(\cdot)\) in (5.44). With probability at least \(1 - O(n^{-10})\), the eigenvalues \(\{\gamma_i^{(l)}\}_{i=1}^{p-1}\) of \(\frac{1}{n} S_{i,l} S_{i,l}^\top\) (see (5.71)) satisfy the following properties.
1. For each $1 \leq i < r$, one has
\[
\frac{\gamma_i^{(l)}}{1 + \beta(\gamma_i^{(l)})} \in B_{\mathcal{E}_{\text{PCA}}}(\lambda_i^* + \sigma^2) \quad \text{for some } k \neq l \text{ and } 1 \leq k \leq r
\]
and
\[
\left\| \left(\gamma_i^{(l)} I_{r-1} - (1 + \beta(\gamma_i^{(l)}))(A^{*l} + \sigma^2 I_{r-1})\right) U^{*l}_i u_i^{(l)} \right\|_2 \lesssim \mathcal{E}_{\text{PCA}},
\]
where $\mathcal{E}_{\text{PCA}}$ is defined in (5.59).

2. For each $r \leq i \leq n \cap (p-1)$, one has
\[
\left| \gamma_i^{(l)} - \sigma^2 \frac{p \vee n}{n} \right| \lesssim \sigma^2 \sqrt{\frac{p + \log n}{n}}.
\tag{5.73}
\]

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

4. Furthermore, one has
\[
|\lambda - \lambda_i| \gtrsim \begin{cases} \Delta_i^*, & \text{if } \frac{\lambda}{1 + \beta(\lambda)} \in B_{\mathcal{E}_{\text{PCA}}}(\lambda_i^* + \sigma^2) \text{ for some } i \neq l \text{ and } 1 \leq i \leq r; \\ \lambda_i^*, & \text{if } |\lambda - \sigma^2 \frac{p \vee n}{n}| \lesssim \sigma^2 \left(\frac{\log n}{n}\right)^{1/2}. \end{cases}
\tag{5.74}
\]
In particular, one has
\[
|\gamma_i^{(l)} - \lambda_i| \gtrsim \begin{cases} \Delta_i^*, & 1 \leq i < r; \\ \lambda_i^*, & i \geq r. \end{cases}
\tag{5.75}
\]

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 $\left\| (\lambda_l I_{p-1} - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T) - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T \right\|_2$, which is accomplished in the following lemma.

**Lemma 9.** Consider any $1 \leq l \leq r$. Instate the assumptions of Theorem 2, and recall the definition of $c_l$ in (3.10). The following holds with probability at least $1 - O(n^{-10})$:
\[
\left\| (\lambda_l I_{p-1} - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T) - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T \right\|_2 \lesssim \frac{(\lambda_i^* + \sigma^2)(\lambda_i^* + \sigma^2)r \log n}{\Delta_i^* n} + \frac{(\lambda_i^* + \sigma^2)^2 \sigma^2 r \log^2 n}{\lambda_i^* n} \ll 1.
\tag{5.76}
\]
Moreover, for the case with $n \geq p$, one has
\[
\left\| (\lambda_l I_{p-1} - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T) - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T \right\|_2 \lesssim \frac{(\lambda_{\text{max}}^* + \sigma^2)(\lambda_i^* + \sigma^2)r \log n}{(\Delta_i^*)^2 n} + \frac{\sigma^2 p}{\lambda_i^* n} \left(\frac{\log n}{p} + \frac{(\lambda_i^* + \sigma^2)^2 \sqrt{r \log n}}{\lambda_i^* n}\right),
\tag{5.77}
\]
and for the case with $p > n$, we have
\[
\left\| (\lambda_l I_{p-1} - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T) - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T \right\|_2 \lesssim \frac{(\lambda_{\text{max}}^* + \sigma^2)(\lambda_i^* + \sigma^2)r \log n}{(\Delta_i^*)^2 n} + \frac{\sigma^2 \sqrt{r \log n}}{\lambda_i^* n}.
\tag{5.78}
\]

Proof. See Appendix D.3.

This lemma taken collectively with (5.68a) leads to
\[
\left| \cos^2 \theta - 1 \right| = \left| \frac{1}{1 + \left\| (\lambda_l I_{p-1} - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T) - \frac{1}{n} S_{l,\perp} s_{l,\perp}^T \right\|_2} - 1 \right|
\]
\[4. \text{Controlling combine (5.68a) and (5.77) to demonstrate that where the last step follows from the assumptions (3.11a) and (3.11b). In addition, when } n \geq p, \text{ one can combine (5.68a) and (5.77) to demonstrate that}
\]
\[
\begin{align*}
&|(1 + c_l) \cos^2 \theta - 1| = \left| \frac{1 + c_l}{1 + \| (\lambda_l I_{p-1} - \frac{1}{n} S_{t, l} S_{t, l}^T )^{-1} \|_2} - 1 \right| \\
&\leq c_l - \left| \left( \lambda_l + \frac{1}{n} S_{t, l} S_{t, l}^T \right)^{-1} \|_2 \right| \\
&\lesssim \frac{(\lambda^*_\max + \sigma^2)(\lambda^*_l + \sigma^2)r \log n}{(\Delta^*_l)^2 n} + \frac{\sigma^2 \kappa \sqrt{p} \log n}{\lambda^*_n} \ll 1, \quad (5.79)
\end{align*}
\]

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 \( p \geq n \), putting (5.68a) and (5.78) together reveals that

\[
|1 - \sqrt{1 + c_l} \cos \theta| = \left| \frac{1 - (1 + c_l) \cos^2 \theta}{1 + \sqrt{1 + c_l} \cos \theta} \right| \lesssim |1 - (1 + c_l) \cos^2 \theta| \lesssim \frac{(\lambda^*_\max + \sigma^2)(\lambda^*_l + \sigma^2)r \log n}{(\Delta^*_l)^2 n} + \frac{\sigma^2 \kappa \sqrt{p} \log n}{\lambda^*_n}. \quad (5.80)
\]

\[4. \text{Controlling } \sum_{k \neq l} a_k^T u_k^* u_k^T u_k^* \left( \lambda_l I_{p-1} - \frac{1}{n} S_{t, l} S_{t, l}^T \right)^{-1} \|_2 S_{t, l} \|_2. \text{ Recognizing that the vector } s_{t, l} \text{ (see (5.67)) obeys}
\]
\[s_{t, l} \sim N(0, (\lambda^*_t + \sigma^2)I_n)\]

and is independent of \( S_{t, l} \) (see (5.67)), we can control this quantity through the lemma below.

**Lemma 10.** Instate the assumptions of Theorem 2. The following holds with probability at least \( 1 - O(n^{-10}) \):

\[
\begin{align*}
&\left| \sum_{k \neq l} a_k^T u_k^* u_k^T u_k^* \left( \lambda_l I_{p-1} - \frac{1}{n} S_{t, l} S_{t, l}^T \right)^{-1} \|_2 S_{t, l} \|_2 \right| \\
&\lesssim \sum_{k \neq l} \frac{|a_k^T u_k^*|}{\lambda^*_l - \lambda^*_k} \sqrt{n} \sqrt{(\lambda^*_l + \sigma^2)(\lambda^*_\max + \sigma^2)(\kappa^2 + r) \log \left( \frac{n \kappa \lambda^*_\max}{\Delta^*_l} \right)} \quad (5.83)
\end{align*}
\]

**Proof.** See Appendix D.4. □
5. Controlling \((P_{U\perp} a)^\top (P_{U\perp} u_i)\). When it comes to \((P_{U\perp} a)^\top (P_{U\perp} u_i)\), we attempt to utilize certain rotational invariance property of \(P_{U\perp} u_i\) in the subspace spanned by \(U^{\perp}\) to upper bound this quantity. This is formalized in Lemma 11.

Lemma 11. Instate the assumptions of Theorem 2. With probability at least \(1 - O(n^{-10})\),

\[
\| (P_{U\perp} a)^\top (P_{U\perp} u_i) \| \lesssim \sqrt{\frac{\log n}{p-r}} \| P_{U\perp} a \|_2 \| P_{U\perp} u_i \|_2.
\]  

(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 \(\| P_{U\perp} u_i \|_2\). To this end, it is seen from Theorem 5 that

\[
\| P_{U\perp} u_i \|_2^2 = 1 - \| (\lambda_i I - \frac{1}{n} S_{\perp} S_{\perp}^\top)^{-1} \frac{1}{n} S_{\perp} S_{\perp}^\top U^\top u_i \|_2^2 \\
\leq \| (\lambda_i I - \frac{1}{n} S_{\perp} S_{\perp}^\top)^{-1} \frac{1}{n} S_{\perp} S_{\perp}^\top U^\top u_i \|_2^2 \\
\leq \| (\lambda_i I - \frac{1}{n} S_{\perp} S_{\perp}^\top)^{-1} \|_2^2 \| \frac{1}{n} S_{\perp} S_{\perp}^\top \|_2^2 \| U^\top u_i \|_2^2 \\
= \| (\lambda_i I - \frac{1}{n} S_{\perp} S_{\perp}^\top)^{-1} \|_2^2 \| \frac{1}{n} S_{\perp} S_{\perp}^\top \|_2^2 ,
\]  

(5.85)

where we recall \(u_i\) is defined to be a unit vector \(u_i := P_{U\perp}(u)/\| P_{U\perp}(u) \|_2\) and satisfies \(U^* U^\top u_i = u_i\).

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

\[
\left| \frac{1}{n} \| S_{\perp} S_{\perp}^\top \| - \sigma^2 \right| \lesssim \sigma^2 \sqrt{\frac{p}{n} + \frac{p}{n} + \frac{\log n}{n}} = o(\lambda_i^{*\min}),
\]  

(5.86)

\[
\frac{1}{n} \| S_{\perp} S_{\perp}^\top \| \lesssim \sqrt{\frac{(\lambda_i^{*\max} + \sigma^2) \sigma^2 (p-r)}{n}} \log n,
\]  

(5.87)

where the relation in (5.86) arises from the noise condition (3.11a). Combining these with Theorem 7, we obtain

\[
\lambda_i - \frac{1}{n} \| S_{\perp} S_{\perp}^\top \| = \lambda_i - \sigma^2 - o(\lambda_i^{*\min}) \\
\stackrel{(i)}{\geq} (1 + \beta(\lambda_i)) (\lambda_i^{*\max} + \sigma^2) - (1 + \beta(\lambda_i)) \cdot O \left( (\lambda_i^{*\max} + \sigma^2) \sqrt{\frac{p}{n} \log n} \right) - \sigma^2 - o(\lambda_i^{*\min}) \\
= (1 + \beta(\lambda_i)) \lambda_i^{*\max} + \beta(\lambda_i) \sigma^2 - O \left( (1 + \beta(\lambda_i)) (\lambda_i^{*\max} + \sigma^2) \sqrt{\frac{p}{n} \log n} \right) - o(\lambda_i^{*\min}) \\
\stackrel{(ii)}{\geq} \lambda_i^{*\max} + O \left( \frac{\sigma^4 p}{\lambda_i^{*\min}} \right) \leq \lambda_i^{*\min},
\]

where (i) is due to the bound developed for \(\lambda_i\) in (5.44) in Theorem 7; (ii) arises from the fact \(\beta(\lambda_i) \lesssim \frac{\sigma^2 p}{\lambda_i^{*\min}} \ll 1\) (as shown in (5.61)) and the noise condition (3.11a) that \((\lambda_i^{*\max} + \sigma^2) \sqrt{\frac{p}{n} \log n} \ll \lambda_i^{*\min}\); (iii) is legal as long as \(\sigma^2 \sqrt{\frac{p}{n}} \ll \lambda_i^{*\min}\). As a result, we obtain

\[
\left\| (\lambda_i I - \frac{1}{n} S_{\perp} S_{\perp}^\top)^{-1} \right\| \leq \frac{1}{\lambda_i - \| \frac{1}{n} S_{\perp} S_{\perp}^\top \|} \ll \frac{1}{\lambda_i^{*\min}},
\]  

(5.88)

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

\[
\| P_{U\perp} u_i \|_2 \leq \| (\lambda_i I - \frac{1}{n} S_{\perp} S_{\perp}^\top)^{-1} \| \cdot \frac{1}{n} \| S_{\perp} S_{\perp}^\top \| \ll \frac{\sqrt{(\lambda_i^{*\max} + \sigma^2) \sigma^2}}{\lambda_i^{*\min}} \sqrt{\frac{p-r}{n}} \log n.
\]  

(5.89)
Taken together with Lemma 11, this leads to the bound
\[
|\langle P U^\perp a \rangle^\top (P U^\perp u_i)\rangle \lesssim \sqrt{\frac{(\lambda_{\max}^* + \sigma^2)\sigma^2}{\lambda_i^2 n}} \log^2 n \|P U^\perp a\|_2, \tag{5.90}
\]

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:

\[
\begin{align*}
\min |a^\top u_i \pm a^\top u_i^*| & \lesssim \left(\frac{(\lambda_{\max}^* + \sigma^2)(\lambda_i^* + \sigma^2)r\log n}{\min_{k:k\neq l}|\lambda_i^* - \lambda_k^*|^2 n} + \frac{\sigma^2 p}{\lambda_i^2 n} \left(\frac{r\log n}{n} + (\lambda_i^* + \sigma^2)\sqrt{\frac{\log n}{n}}\right)\right) |a^\top u_i^*| \\
& \quad + \sum_{k:k\neq l} \left| \frac{a^\top u_i^*}{|\lambda_i^* - \lambda_k^*|\sqrt{n}} \right| (\lambda_i^* + \sigma^2)(\lambda_{\max}^* + \sigma^2)(\kappa^2 + r) \log \left(\frac{n\kappa\lambda_{\max}}{\Delta_i^*}\right) \\
& \quad + \sqrt{\frac{(\lambda_{\max}^* + \sigma^2)\sigma^2}{\lambda_i^2 n}} \log^2 n \|P U^\perp a\|_2
\end{align*}
\]

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

\[
\begin{align*}
\min |a^\top u_i \pm a^\top u_i^*| & \lesssim \left(\frac{(\lambda_{\max}^* + \sigma^2)(\lambda_i^* + \sigma^2)r\log n}{\min_{k:k\neq l}|\lambda_i^* - \lambda_k^*|^2 n} + \frac{\sigma^2 p}{\lambda_i^2 n} \left(\frac{r\log n}{n} + (\lambda_i^* + \sigma^2)\sqrt{\frac{\log n}{n}}\right)\right) |a^\top u_i^*| \\
& \quad + \sum_{k:k\neq l} \left| \frac{a^\top u_i^*}{|\lambda_i^* - \lambda_k^*|\sqrt{n}} \right| (\lambda_i^* + \sigma^2)(\lambda_{\max}^* + \sigma^2)(\kappa^2 + r) \log \left(\frac{n\kappa\lambda_{\max}}{\Delta_i^*}\right) \\
& \quad + \sqrt{\frac{(\lambda_{\max}^* + \sigma^2)\sigma^2}{\lambda_i^2 n}} \log^2 n
\end{align*}
\]

where the last step holds due to the noise assumption (3.11a) as well as the facts that \(|a^\top u_i^*| \leq \|a\|_2\|u_i^*\|_2 \leq 1\) and \(\|P U^\perp a\|_2 \leq \|a\|_2 = 1\).

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

\[
\begin{align*}
\min |a^\top u_i \pm a^\top u_i^*| & \lesssim \left(\frac{(\lambda_{\max}^* + \sigma^2)(\lambda_i^* + \sigma^2)r\log n}{\min_{k:k\neq l}|\lambda_i^* - \lambda_k^*|^2 n} + \frac{\sigma^2 \sqrt{\kappa^2 p r \log n}}{\lambda_i^2 n}\right) |a^\top u_i^*| \\
& \quad + \sum_{k:k\neq l} \left| \frac{a^\top u_i^*}{|\lambda_i^* - \lambda_k^*|\sqrt{n}} \right| \sqrt{\lambda_i^* + \sigma^2}(\lambda_{\max}^* + \sigma^2)(\kappa^2 + r) \log \left(\frac{n\kappa\lambda_{\max}}{\Delta_i^*}\right) \\
& \quad + \sqrt{\frac{(\lambda_{\max}^* + \sigma^2)\sigma^2}{\lambda_i^2 n}} \log^2 n
\end{align*}
\]
Here, we use the noise assumption (3.11a), $|a^T u_j^*| \leq \|a\|_2 \|u_j^*\|_2 \leq 1$ and $\|P_{\mathcal{U}_{2,n}} a\|_2 \leq \|a\|_2 = 1$ 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 $\kappa$, 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

G. Li and Y. Gu are supported in part by the grant NSFC-61971266. Y. Chen is supported in part by 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 and IIS-1900140, and the Princeton SEAS Innovation Award. H. V. Poor is supported in part by NSF CCF-1908308, and in part by a Princeton Schmidt Data-X Research Award. C. Cai is supported in part by Gordon Y. S. Wu Fellowships in Engineering.

A Proofs of master theorems

A.1 Proof of Theorem 4

Given that $u_{\ell}$ is an eigenvector of $M$, one has $Mu_{\ell} = \lambda_{\ell} u_{\ell}$, which together with the decomposition (5.3) and the condition $u_{\ell,\perp} = q^\perp (q^\perp)^T u_{\ell,\perp}$ gives

$$M(q \cos \theta + u_{\ell,\perp} \sin \theta) = \lambda_{\ell} (q \cos \theta + u_{\ell,\perp} \sin \theta)$$

$$\iff \quad Mq \cos \theta + Mq^\perp (q^\perp)^T u_{\ell,\perp} \sin \theta = \lambda_{\ell} q \cos \theta + \lambda_{\ell} u_{\ell,\perp} \sin \theta. \quad (A.1)$$

Left-multiplying both sides of this equation by $q^T$ (resp. $(q^\perp)^T$) and using the assumptions of $u_{\ell,\perp}$ (namely, $q^T u_{\ell,\perp} = 0$ and $q^\perp (q^\perp)^T u_{\ell,\perp} = u_{\ell,\perp}$) give

$$q^T Mq \cos \theta + q^T Mq^\perp (q^\perp)^T u_{\ell,\perp} \sin \theta = \lambda_{\ell} q^T \cos \theta, \quad (A.2a)$$

$$(q^\perp)^T Mq \cos \theta + (q^\perp)^T Mq^\perp (q^\perp)^T u_{\ell,\perp} \sin \theta = \lambda_{\ell} (q^\perp)^T u_{\ell,\perp} \sin \theta. \quad (A.2b)$$

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

$$\lambda_{\ell} I_{n-1} - (q^\perp)^T Mq^\perp (q^\perp)^T u_{\ell,\perp} \sin \theta = (q^\perp)^T Mq \cos \theta. \quad (A.3)$$
Given the assumption that $\lambda_l I_{n-1} - (q^\perp)^T M q^\perp$ is invertible and the fact that $\|(q^\perp)^T u_{l,\perp}\|_2 = \|u_{l,\perp}\|_2 = 1$, 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 $(\lambda_l I_{n-1} - (q^\perp)^T M q^\perp)(q^\perp)^T u_{l,\perp} \neq 0$ 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 $q = u_l$ and $(q^\perp)^T M q = 0$ (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

$$
(q^\perp)^T u_{l,\perp} = \frac{\cos \theta}{\sin \theta}(\lambda_l I_{n-1} - (q^\perp)^T M q^\perp)^{-1}(q^\perp)^T M q.
$$

(A.4)

This together with the unit norm constraint of $u_{l,\perp}$ and $u_{l,\perp} = q^\perp(q^\perp)^T u_{l,\perp}$ implies that

$$
u_{l,\perp} = \pm \frac{q^\perp(\lambda_l I_{n-1} - (q^\perp)^T M q^\perp)^{-1}(q^\perp)^T M q}{\|q^\perp(\lambda_l I_{n-1} - (q^\perp)^T M q^\perp)^{-1}(q^\perp)^T M q\|_2}
$$

(A.5)

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

$$(\lambda_l - q^T M q) \cos \theta = q^T M q^\perp(q^\perp)^T u_{l,\perp} \sin \theta
$$

$$
= q^\perp M q^\perp(\lambda_l I_{n-1} - (q^\perp)^T M q^\perp)^{-1}(q^\perp)^T M q \cos \theta,
$$

$$
\implies \lambda_l - q^T M q = q^\perp M q^\perp(\lambda_l I_{n-1} - (q^\perp)^T M q^\perp)^{-1}(q^\perp)^T M q,
$$

(A.6)

thus establishing the claim (5.4b).

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

$$
\frac{\sin \theta}{\cos \theta} = \frac{\lambda_l - q^T M q}{q^\perp M q^\perp(q^\perp)^T u_{l,\perp}}.
$$

(A.7)

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

$$
\cos^2 \theta = \frac{1}{1 + \frac{|\lambda_l - q^T M q|^2}{\|q^\perp M q^\perp(q^\perp)^T u_{l,\perp}\|^2}}
$$

(i)

$$
= \frac{1}{1 + \frac{|\lambda_l - q^T M q|^2}{\|q^\perp M q^\perp(\lambda_l I_{n-1} - (q^\perp)^T M q^\perp)^{-1}(q^\perp)^T M q\|_2}^2}
$$

(ii)

$$
= \frac{1}{1 + \|q^\perp(\lambda_l I_{n-1} - (q^\perp)^T M q^\perp)^{-1}(q^\perp)^T M q\|_2^2}
$$

(iii)

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

### A.2 Proof of Theorem 5

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

$$
\mathbf{M u}_{l,\parallel} \cos \theta + \mathbf{M u}_{l,\perp} \sin \theta = \lambda_l \mathbf{u}_{l,\parallel} \cos \theta + \lambda_l \mathbf{u}_{l,\perp} \sin \theta,
$$

(A.8)

which together with the conditions $u_{t,\parallel} = QQ^T u_{l,\parallel}$ and $u_{l,\perp} = Q^\perp(Q^\perp)^T u_{l,\perp}$ implies that

$$
\mathbf{M QQ}^T u_{t,\parallel} \cos \theta + \mathbf{M Q}^\perp(Q^\perp)^T u_{l,\perp} \sin \theta = \lambda_t \mathbf{u}_{t,\parallel} \cos \theta + \lambda_l \mathbf{u}_{t,\perp} \sin \theta.
$$

(A.9)

Left-multiplying both sides of this relation by $Q^T$ (resp. $(Q^\perp)^T$) gives

$$
Q^T \mathbf{M QQ}^T u_{t,\parallel} \cos \theta + Q^T \mathbf{M Q}^\perp(Q^\perp)^T u_{l,\perp} \sin \theta = \lambda_t Q^T \mathbf{u}_{t,\parallel} \cos \theta,
$$

(A.10a)
Thus indicating that
\[ \lambda I_k - Q^\top M \lambda Q \top u_{l,\perp} \cos \theta = Q^\perp MQ^\perp (Q^\perp)\top u_{l,\perp} \sin \theta, \]

which follows from (B.2), we see that the triple
\[ (\lambda I_{n-k} - (Q^\perp)^\top MQ^\perp) (Q^\perp)^\top u_{l,\perp} \sin \theta = (Q^\perp)^\top MQQ^\top u_{l,\perp} \cos \theta, \]

where the last identity holds since \( QQ^\top u_{l,\perp} = u_{l,\perp} \). These two relations taken together demonstrate that
\[ (\lambda I_k - Q^\top M) Q^\top u_{l,\perp} \cos \theta = Q^\perp MQ^\perp ((Q^\perp)^\top u_{l,\perp} \sin \theta) \]
\[ = Q^\perp MQ^\perp (\lambda I_{n-k} - (Q^\perp)^\top MQ^\perp)^{-1} (Q^\perp)^\top M u_{l,\perp} \cos \theta. \]

In addition, in view of the invertibility of \( \lambda I_{n-k} - (Q^\perp)^\top MQ^\perp \) (due to the assumption) and \( \| (Q^\perp)^\top u_{l,\perp} \|_2 = \| u_{l,\perp} \|_2 = 1 \), 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 6

#### B.1 Proof of Lemma 1

For notational convenience, divide the matrix \( H \) as follows
\[ H = \begin{bmatrix} H_{ul} & H_{ur} \\ H_{ur} & H_{lr} \end{bmatrix}, \quad H_{ul} \in \mathbb{R}^{r \times r}, \quad H_{ur} \in \mathbb{R}^{r \times (n-r)}, \quad H_{lr} \in \mathbb{R}^{(n-r) \times (n-r)}. \]

In view of the rotational invariance of a symmetric Gaussian matrix, we know that \( RHR^\top \) has the same distribution as \( H \) for any fixed orthonormal matrix \( R \in \mathbb{R}^{n \times n} \) obeying \( RR^\top = I_n \). As a result, it is easily seen that the triple
\[ \left( U^\top H U^*, (U^\perp)^\top H U^*, U^\top H U^* \right) \overset{d}{=} (H_{ul}, H_{lr}, H_{ur}), \]

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

**Controlling \( \| U^\top H U^* \| \).** Apply the standard Gaussian concentration inequalities [Vershynin, 2012] and (B.2) to conclude that with probability at least \( 1 - O(n^{-10}) \),
\[ \| U^\top H U^* \| = \| H_{ul} \| \lesssim \sigma(\sqrt{r} + \sqrt{\log n}). \]

**Controlling \( \| G(\lambda) - G^\perp(\lambda) \| \).** Consider any fixed \( \lambda \) obeying \( 2|\lambda_i^*|/3 \leq |\lambda| \leq 4|\lambda_i^*|/3 \). Recalling the expression of \( G(\lambda) \) in (5.14b), we have
\[ \| G(\lambda) \| = \| U^\top H U^* \| (\lambda I_n - U^\perp (U^\perp)^\top H U^* (U^\perp)^\top) U^\perp (U^\perp)^\top H U^* \]
\[ = \| U^\top H U^* \| (\lambda I_{n-r} - (U^\perp)^\top H U^* (U^\perp)^\top) U^\perp (U^\perp)^\top H U^* \|. \]

Combining this with the fact (B.2), we see that \( \| G(\lambda) \| \) has the same distribution as \( \| H_{ur} (\lambda I_{n-r} - H_{lr})^{-1} H_{ur}^\top \|. \) Repeating the same argument also indicates that \( \| G(\lambda) - G^\perp(\lambda) \| \) has the same distribution as
\[ \| H_{ur} (\lambda I_{n-r} - H_{lr})^{-1} H_{ur}^\top - E[H_{ur} (\lambda I_{n-r} - H_{lr})^{-1} H_{ur}^\top | H_{lr}] \|. \]

This allows us to turn attention to \( H_{ur} (\lambda I_{n-r} - H_{lr})^{-1} H_{ur}^\top \).
As a key observation, $H_{ur}$ and $H_{lr}$ are statistically independent, thus enabling convenient decoupling of the randomness. Let $\gamma_1 \geq \cdots \geq \gamma_{n-r}$ represent the eigenvalues of $H_{lr}$. Denote by $\{h_i\}_{i=1}^{n-r}$ the columns of $H_{ur}$, i.e. $H_{ur} = [h_1, \cdots, h_{n-r}]$, which are independent of $H_0$ and $\{\gamma_i\}$. Invoking the rotational invariance of Gaussian random matrices once again, we see that

$$H_{ur}(\lambda I_{n-r} - H_{lr})^{-1}H_{ur}^\top = \sum_{i=1}^{n-r} \frac{1}{\lambda - \gamma_i} h_i h_i^\top,$$

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

$$\left\| \sum_i H_{ur}(\lambda I_{n-r} - H_{lr})^{-1}H_{ur}^\top - \mathbb{E}\left[ \sum_i H_{ur}(\lambda I_{n-r} - H_{lr})^{-1}H_{ur}^\top | H_{lr} \right] \right\| \\
\lesssim \frac{\sigma^2}{\min_i |\lambda - \gamma_i|} (\sqrt{rn \log n + r \log n})$$

$$\lesssim \frac{\sigma^2}{|\lambda_i^*|} (\sqrt{rn \log n + r \log n})$$

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

$$\|G(\lambda) - G^{\perp}(\lambda)\| \lesssim \frac{\sigma^2}{\lambda_{\min}^*} (\sqrt{rn \log n + r \log n}) \leq \frac{\sigma^2}{\lambda_{\min}^*} (\sqrt{rn \log n + r \log n})$$

for a given $\lambda$.

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

$$|\mathcal{N}_{\epsilon_0}| \lesssim \lambda_i^*/\epsilon_0 \approx n;$$

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

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

$$\|G(\hat{\lambda}) - G^{\perp}(\hat{\lambda})\| \lesssim \frac{\sigma^2}{\lambda_{\min}^*} (\sqrt{rn \log n + r \log n}), \quad \forall \hat{\lambda} \in \mathcal{N}. \quad (B.8)$$

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

$$\|G(\lambda) - G(\hat{\lambda})\| \leq \|H_{ur}(\lambda I_{n-r} - H_{lr})^{-1}H_{ur}^\top - H_{ur}(\hat{\lambda} I_{n-r} - H_{lr})^{-1}H_{ur}^\top\| \\
\leq \|H_{ur}\|^2 \cdot \left\| (\lambda I_{n-r} - H_{lr})^{-1} - (\hat{\lambda} I_{n-r} - H_{lr})^{-1} \right\| \\
\leq \|H_{ur}\|^2 \max_i \left| \frac{1}{\lambda - \gamma_i} - \frac{1}{\lambda - \gamma_i} \right| = \|H_{ur}\|^2 \max_i \left| \frac{\lambda - \hat{\lambda}}{(\lambda - \gamma_i)(\lambda - \gamma_i)} \right| \\
\lesssim \frac{\sigma^2 n \cdot \max_i |\lambda - \hat{\lambda}|}{\lambda_i^2} \leq \sigma^2 n \cdot \frac{\epsilon_0}{\lambda_i^2} \\
\lesssim \frac{\sigma^2}{\lambda_{\min}^*} \quad (B.9)$$

holds with probability $1 - O(n^{-11})$. Here, the penultimate line has made use of the Gaussian concentration bound $\|H_{ur}\| \lesssim \sigma \sqrt{n}$, whereas the last inequality results from (B.7).
Combining the above two facts together, we arrive at

\[
\sup_{\lambda: |\lambda| \in [2\lambda^*_1/3, 4|\lambda^*_1|/3]} \|G(\lambda) - G^\perp(\lambda)\|
\]

\[
= \sup_{\lambda: |\lambda| \in [2\lambda^*_1/3, 4|\lambda^*_1|/3]} \|G(\lambda) - G(\hat{\lambda}) + G^\perp(\lambda) - G^\perp(\hat{\lambda})\|
\]

\[
\leq \sup_{\lambda: |\lambda| \in [2\lambda^*_1/3, 4|\lambda^*_1|/3]} \|G(\lambda) - G(\hat{\lambda})\| + \sup_{\lambda: \hat{\lambda} \in \Lambda_{\alpha}} \|G(\hat{\lambda}) - G^\perp(\hat{\lambda})\|
\]

\[
\leq \frac{\sigma^2}{\lambda_{\min}^2} \left( \sqrt{\tau n \log n + r \log n} \right).
\]

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

\[
\|G^\perp(\lambda) - G^\perp(\hat{\lambda})\| = \|E[G(\lambda) | H_r] - E[G(\hat{\lambda}) | H_r]\| \leq E[\|G(\lambda) - G(\hat{\lambda})\| | H_r] \leq \frac{\sigma^2}{\lambda_{\min}^2},
\]

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

\[
G^\perp(\lambda) = E \left[ U^*^\top HU^*^\perp (\lambda I_n - (U^*^\perp)^\top HU^*^\perp)^{-1} (U^*^\perp)^\top HU^*^\perp (U^*^\perp)^\top \right]_{\lambda: A = A}
\]

\[
= \sigma^2 tr \left[ (\lambda I_n - (U^*^\perp)^\top HU^*^\perp)^{-1} \right] I_r,
\]

which holds since the matrix \( A \) obeys \( A \overset{d}{=} H_{\text{ur}}(\lambda I_n - H_{\text{ur}})^{-1} H_{\text{ur}}^\top \), 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

\[
|\lambda_l - \lambda^*_l - \gamma(\lambda_i)| \leq \varepsilon_{\text{MD}}, \quad \text{for some } 1 \leq i \leq r
\]

\[
\text{or } |\lambda_l| \leq \varepsilon_{\text{MD}}
\]

for any \( 1 \leq l \leq r \). To see this, we recall that for any symmetric matrix \( A \), one has

\[
\min_1 \lambda_i(A) = \sqrt{\lambda_{\min}(A^2)} = \min_{x \in \mathbb{S}^{n-1}} x^\top A^2 x = \min_{x \in \mathbb{S}^{n-1}} \|Ax\|_2,
\]

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

\[
\min_{1 \leq l \leq n} |\lambda_l - \lambda_i(M_{\lambda_i})| = \min_{1 \leq l \leq n} |\lambda_i - \lambda_i(M_{\lambda_i})| \leq \|(\lambda I_n - M_{\lambda_i}) u_{1,l}\|_2 \leq \varepsilon_{\text{MD}}.
\]

This immediately establishes (B.11), since the set of eigenvalues of \( M_{\lambda_i} \) is \( \{ \lambda^*_i + \gamma(\lambda_i) \mid 1 \leq i \leq r \} \cup \{0\} \) (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

\[
M(t) := M^* + tH,
\]

\[
\gamma(\lambda, t) := t^2 \sigma^2 tr \left( (\lambda I_n - t(U^*^\perp)^\top HU^*^\perp)^{-1} \right).
\]
We denote by \( \\{\lambda_{i,t}\}_{i=1}^{n} \) the eigenvalues of \( M(t) \) obeying \( |\lambda_{1,t}| \geq \cdots \geq |\lambda_{n,t}| \); in other words, \( \lambda_{1,t}, \ldots, \lambda_{r,t} \) correspond to the \( r \) eigenvalues of \( M(t) \) with the largest magnitudes. Armed with this notation, we clearly have

\[
\lambda_{i,t} = \lambda_i, \quad 1 \leq i \leq r.
\]

The subsequent analysis consists of three steps.

- First, we establish the correspondence between \( \{\lambda_{i,t} \mid 1 \leq i \leq r\} \) and \( \{\lambda_{i,t}^* \mid 1 \leq i \leq r\} \) 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

\[
\sup_{t \in [1/\sqrt{n}, 1]} |\lambda_{i,t} - \gamma(\lambda_{i,t}, t) - \lambda_i^*| \leq \mathcal{E}_t,
\]

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

\[
\lambda_{i,t} - \gamma(\lambda_{i,t}, t) \in \bigcup_{i=1}^{r} B_{\mathcal{E}_t}(\lambda_i^*),
\]

where \( B_{\mathcal{E}_t}(\lambda) := \{z \mid |z - \lambda| \leq \tau\} \) denotes the ball of radius \( \tau \) centered at \( \lambda \).

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

\[
|\gamma(\lambda_{i,t}, t)| \leq t^2 \sigma^2 \frac{n - r}{|\lambda_{i,t}| - \|tH\|} \leq \frac{1}{n} \cdot \sigma^2 \frac{n}{\lambda_{\text{min}}} = \frac{\sigma^2}{\lambda_{\text{min}}} \leq \sigma/2,
\]

where the last inequality holds due to the assumption \( \sigma \sqrt{n} \lesssim \lambda_{\text{min}}^* \). As a result,

\[
|\lambda_{i,t} - \gamma(\lambda_{i,t}, t) - \lambda_i^*| \leq |\lambda_{i,t} - \lambda_i^*| + |\gamma(\lambda_{i,t}, t)| \leq \sigma \lesssim \mathcal{E}_1/\sqrt{n}
\]

\[
\implies \lambda_{i,t} - \gamma(\lambda_{i,t}, t) \in B(\lambda_i^*, \mathcal{E}_1/\sqrt{n}), \quad 0 \leq t \leq 1/\sqrt{n}.
\]

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

\[
B_{\mathcal{E}_t}(\lambda_i^*) \cap \{\cup_{k \neq l} B_{\mathcal{E}_t}(\lambda_k^*)\} = \emptyset \quad \text{and} \quad B_{\mathcal{E}_t}(\lambda_i^*) \cap \{\cup_{k \neq l} B_{\mathcal{E}_t}(\lambda_k^*)\} = \emptyset.
\]

In other words, \( B_{\mathcal{E}_t}(\lambda_i^*) \) remains an isolated region within the set \( \cup_{i=1}^{r} B_{\mathcal{E}_t}(\lambda_i^*) \) when we increase \( t \) from \( 1/\sqrt{n} \) to 1. This together with the above two facts (namely, the continuity of \( \lambda_{i,t} - \gamma(\lambda_{i,t}, t) \) in \( t \) and \( \mathcal{B}_1(\lambda_i^*) \)) requires that

\[
\lambda_{i,t} - \gamma(\lambda_{i,t}, t) \in B_{\mathcal{E}_t}(\lambda_i^*), \quad 1/\sqrt{n} \leq t \leq 1,
\]

provided that \( \min_{k \neq l} |\lambda_i^* - \lambda_k^*| > 2\mathcal{E}_1 \).

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

\[
|\lambda_i - \gamma(\lambda_i) - \lambda_i^*| \leq \mathcal{E}_{MD}, \quad 1 \leq i \leq r.
\]
B.3 Proof of Lemma 12

Fix an arbitrary $1 \leq t \leq r$. 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 $\|G(\lambda, t) - G^\perp(\lambda, t)\|$, namely,

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

holds simultaneously for all $1/\sqrt{n} \leq t \leq 1$ and $\lambda$ with $|\lambda| \in [2|\lambda_*^1|/3, 4|\lambda_*^1|/3]$. 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 $\mathcal{N}_\epsilon$ be an $\epsilon$-net for $[1/\sqrt{n}, 1]$ such that (1) it has cardinality $|\mathcal{N}_\epsilon| \lesssim \sqrt{n}$; (2) for any $t \in [1/\sqrt{n}, 1]$, there exists some point $\tilde{t} \in \mathcal{N}_\epsilon$ obeying $|\tilde{t} - t| \leq \epsilon_1$.

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

$$\sup_{\lambda: |\lambda| \in [2|\lambda_*^1|/3, 4|\lambda_*^1|/3]} \|G(\lambda, \tilde{t}) - G^\perp(\lambda, \tilde{t})\| \lesssim \frac{t \sigma^2}{\lambda_{\min}} (\sqrt{rn \log n + r \log n}) \leq \frac{t \sigma^2}{\lambda_{\min}} \sqrt{rn \log n}$$

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

- For any $t \in [1/\sqrt{n}, 1]$, let $\tilde{t} \in \mathcal{N}_\epsilon$ be a point obeying $|\tilde{t} - t| \leq \epsilon_1$. Recognizing that $G(\lambda, \tilde{t}) - G(\lambda, t) \overset{d}{=} t^2 H_{\text{ur}}(\lambda I_{n-r} - t H_{t^i})^{-1} H_{t^i}^\top - t^2 H_{\text{ur}}(\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} H_{t^i}^\top$, one can bound

$$\|G(\lambda, \tilde{t}) - G(\lambda, t)\| \leq \|t^2 H_{\text{ur}}(\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} H_{t^i}^\top - t^2 H_{\text{ur}}(\lambda I_{n-r} - t H_{t^i})^{-1} H_{t^i}^\top\|$$

$$\leq \|t^2 H_{\text{ur}}(\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} H_{t^i}^\top - t^2 H_{\text{ur}}(\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} H_{t^i}^\top\| + \|t^2 H_{\text{ur}}(\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} H_{t^i}^\top - t^2 H_{\text{ur}}(\lambda I_{n-r} - t H_{t^i})^{-1} H_{t^i}^\top\|$$

$$\leq \|t - \tilde{t}\| \cdot t + \tilde{t} \cdot \|H_{\text{ur}}^\top\| \|\lambda I_{n-r} - \tilde{t} H_{t^i}\|^{-1} + t^2 \cdot \|H_{\text{ur}}^\top\| \|\lambda I_{n-r} - t H_{t^i}\|^{-1} - (\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} (\lambda I_{n-r} - t H_{t^i})^{-1} \|H_{\text{ur}}^\top\| \|\lambda I_{n-r} - \tilde{t} H_{t^i}\|^{-1} - (\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} \|.$$

Recalling the notation that $\gamma_1 \geq \cdots \geq \gamma_{n-r}$ represent the eigenvalues of $H_{t^i}$, we have

$$\left\| (\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} \right\| = \max_i \left| \frac{1}{\lambda - \tilde{t} \gamma_i} \right| \lesssim \frac{1}{\lambda_{\min}}$$

and

$$\left\| (\lambda I_{n-r} - t H_{t^i})^{-1} - (\lambda I_{n-r} - \tilde{t} H_{t^i})^{-1} \right\| = \max_i \left| \frac{1}{\lambda - \tilde{t} \gamma_i} - \frac{1}{\lambda - t \gamma_i} \right| = \max_i \left| \frac{(t - \tilde{t}) \gamma_i}{(\lambda - \tilde{t} \gamma_i)(\lambda - t \gamma_i)} \right|$$

$$\lesssim \frac{|t - \tilde{t}|}{\lambda_{\min}} \leq \frac{\epsilon_1}{\lambda_{\min}},$$

where we have used the bounds $2|\lambda_*^1|/3 \leq |\lambda| \leq 4|\lambda_*^1|/3$, $|\gamma_i| \leq \lambda_{\min}^* / 3$ and $\tilde{t} \leq t + \epsilon_1 \leq 1$. Combining these with the high-probability bound $\|H_{\text{ur}}\| \lesssim \sigma \sqrt{n}$, we arrive at

$$\|G(\lambda, \tilde{t}) - G(\lambda, t)\| \lesssim \epsilon_1 \cdot t \cdot \sigma^2 n \cdot \frac{1}{\lambda_{\min}^*} + t^2 \cdot \sigma^2 n \cdot \frac{\epsilon_1}{\lambda_{\min}^*} \lesssim \frac{t \sigma^2}{\lambda_{\min}^*} \sqrt{rn \log n},$$

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

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

$$\|G(\lambda, t) - G^\perp(\lambda, t)\| \lesssim \|G(\lambda, \tilde{t}) - G^\perp(\lambda, \tilde{t})\| + \|G(\lambda, \tilde{t}) - G(\lambda, t)\| + \|G^\perp(\lambda, \tilde{t}) - G^\perp(\lambda, t)\|$$

$$\lesssim \frac{t \sigma^2}{\lambda_{\min}^*} \sqrt{rn \log n} + \frac{t \sigma^2}{\lambda_{\min}^*} \sqrt{rn} \lesssim \frac{t \sigma^2}{\lambda_{\min}^*} \sqrt{rn \log n}$$

holds simultaneously for all $\lambda$ with $|\lambda| \in [2|\lambda_*^1|/3, 4|\lambda_*^1|/3]$ and all $t \in [1/\sqrt{n}, 1]$. Finally, taking a union bound over $1 \leq t \leq r$ 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 $M^{(i)}$, 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

\[
\begin{align*}
|\lambda_k^{(i)} - \gamma(\lambda_k^{(i)}) - \lambda_k^*| &\leq \mathcal{E}_{\text{MD}}, \quad 1 \leq k < l, \quad \text{(C.1a)} \\
|\lambda_k^{(i)} - \gamma(\lambda_k^{(i)}) - \lambda_{k+1}^*| &\leq \mathcal{E}_{\text{MD}}, \quad l \leq k < r, \quad \text{(C.1b)} \\
|\lambda_k^{(i)}| &\leq \|H\| \lesssim \sigma\sqrt{n}, \quad k \geq r, \quad \text{(C.1c)}
\end{align*}
\]

where $\mathcal{E}_{\text{MD}} = C_1 \sigma\sqrt{r} \log n$ 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

\[
\begin{align*}
\lambda_k - \gamma(\lambda_k) &\in \mathcal{B}_{\text{stab}}(\lambda_k^*), \quad 1 \leq k \leq r \\
|\lambda_k| &\leq \|H\| \lesssim \sigma\sqrt{n}, \quad k > r
\end{align*}
\]

as shown in Theorem 6, we are positioned to prove the claim (5.33) as follows.

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

\[
|\lambda_k - \lambda| \geq |\lambda_k - \gamma(\lambda_k)| - |\gamma(\lambda_k)| - |\lambda| \geq |\lambda_k^*| - \mathcal{E}_{\text{MD}} - |\gamma(\lambda_k)| - O(\sigma\sqrt{n}) \tag{C.2}
\]

\[
\|H\| \lesssim \lambda_k^* - |\lambda_k^*| - O(\sigma\sqrt{n}) \lesssim |\lambda_k^*|, \quad \text{(ii)}
\]

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

\[
|\gamma(\lambda_k)| = \left| \sum_i \frac{\sigma^2}{\lambda_i - \lambda_k (U^* H U^*)} \right| \leq \frac{\sigma^2 n}{|\lambda| - \|H\|} \lesssim \frac{\sigma^2 n}{\lambda_{\min}^*}, \quad \text{(C.3)}
\]

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

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

\[
f(\lambda) := \lambda - \gamma(\lambda), \quad \text{(C.4)}
\]

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 $\lambda$ with $|\lambda| \in [\lambda_{\min}^*/3, 2\lambda_{\max}^*/3]$ one has

\[
f'(\lambda) = 1 + \sum_i \frac{\sigma^2}{(\lambda - \lambda_i (U^* H U^*))^2} \geq 1 - \frac{\sigma^2 n}{\frac{1}{3} \lambda_{\min}^* - \|H\|} \geq \frac{1}{2},
\]

\[
f'(\lambda) \leq 1 + \frac{\sigma^2 n}{\|H\|^2} \leq \frac{3}{2}
\]

with the proviso that $\sigma\sqrt{n} \leq c_0 \lambda_{\min}^*$ for some constant $c_0 > 0$ small enough. This means that within the range $|\lambda| \in [2\lambda_{\min}^*/3, 4\lambda_{\max}^*/3]$, 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

\[
\frac{2}{3} \leq \frac{df^{-1}(\tau)}{d\tau} \leq 2 \quad \forall \tau \text{ with } |\tau| \in [\lambda_{\min}^*/2, 3\lambda_{\max}^*/2].
\]

(5.5)
In view of (5.23) and the condition that $\lambda - \gamma(\lambda) \in \mathcal{B}_{\text{MD}}(\lambda_k^*)$ for some $1 \leq k \leq r$, one can invoke (C.5) to reach

$$|\lambda_l - \lambda| = \left| f^{-1}(\lambda^*_l + O(\mathcal{E}_{\text{MD}})) - f^{-1}(\lambda_k^* + O(\mathcal{E}_{\text{MD}})) \right|$$

$$\geq \inf_{\tau : |\tau| \leq |\lambda_{\min}/2, 3\lambda_{\max}/2|} \left| \frac{d f^{-1}(\tau)}{d\tau} \right| |\lambda_l^* - \lambda_k^* + O(\mathcal{E}_{\text{MD}})|$$

$$\geq \frac{2}{3} |\lambda_l^* - \lambda_k^* + O(\mathcal{E}_{\text{MD}})| \geq |\lambda_l^* - \lambda_k^*|,$$

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

- 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 6. The idea is to invoke Theorem 5 to analyze the spectrum of $M^{(l)}$. Before proceeding, we introduce several notation tailored to this setting as well as a few simple facts. To begin with, we define

$$M^{(l)} := (u^{(l)}_i)_{\perp}^T M^* u^{(l)}_{i} = (u^{(l)}_{i})_{\perp}^T U^* \Lambda^* U^* u^{(l)}_{i} = U^{(l)} \Lambda^{(l)} U^{(l)} \top,$$

$$H^{(l)} := (u^{(l)}_i)_{\perp}^T Hu^{(l)}_i = (u^{(l)}_{i})_{\perp}^T \sigma^2 \Lambda^* U^* u^{(l)}_{i},$$

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

$$G^{(l)}(\lambda) := \mathbb{E} \left[ G^{(l)}(\lambda) \mid (U^* (l))_{\perp}^T H^{(l)} U^* (l) \right],$$

$$G^{(l)}(\lambda) := \mathbb{E} \left[ G^{(l)}(\lambda) \mid (U^* (l))_{\perp}^T H^{(l)} U^* (l) \right],$$

where $U^* (l)$ is defined in (5.32) and the expectation is taken assuming that $\lambda$ is independent of $H$. By construction, one has $u^{(l)}_i u^{(l)}_{i\perp} = U^* (l)$, and consequently

$$(U^* (l))_{\perp}^T H^{(l)} U^* (l) = (U^* (l))_{\perp}^T (u^{(l)}_i)_{\perp}^T Hu^{(l)}_i U^* (l) = (U^* (l))_{\perp}^T \sigma^2 \Lambda^* U^* U^* (l),$$

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

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

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

$$G^{(l)}(\lambda) = U^* (l)_{\perp}^T \sigma^2 \Lambda^* U^* (l)_{\perp}^T H \sigma^2 \Lambda^* U^* (l)_{\perp}^T (U^* (l))_{\perp}^T (u^{(l)}_{i})_{\perp}^T Hu^{(l)}_i (u^{(l)}_i)_{\perp}^T U^*$$

$$= P^{(l)}_{\perp} U^* (l)_{\perp}^T \sigma^2 \Lambda^* U^* (l)_{\perp}^T (U^* (l))_{\perp}^T (u^{(l)}_{i})_{\perp}^T Hu^{(l)}_i (u^{(l)}_i)_{\perp}^T U^*$$

$$= \mathbb{E} \left[ G^{(l)}(\lambda) \mid (U^* (l))_{\perp}^T \sigma^2 \Lambda^* U^* (l) \right],$$

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

With the above preparation in place, we can repeat the proof of Theorem 6 to obtain

$$\|U^* (l)^T H^{(l)} U^* (l)\| \leq \sigma \left( \sqrt{r} + \sqrt{\log n} \right)$$

$$\sup_{\lambda : |\lambda| \in [2\lambda_{\min}/3, 4\lambda_{\max}/3]} \|G^{(l)}(\lambda) - G^{(l)}(\lambda)\| \leq \frac{\sigma^2}{\lambda_{\min}} \left( \sqrt{r} \log n + r \log n \right)$$

$$G^{(l)}(\lambda) = \gamma(\lambda) P^{(l)}_{\perp} U^* (l)^T \sigma^2 \Lambda^* U^* (l)$$
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 $M^{(l)} + G^{(l)\perp}(\lambda)$ is given by

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

As a result, repeating the same arguments of Theorem 6 (which we omit for brevity) immediately establishes the claim of this lemma.

C.2 Proof of Lemma 4

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

$$
\| (\lambda I_{n-1} - M^{(l)})^{-1}(u^{(l)\perp}_i)^\top Hu^*_i \|_2^2 = \| U^{(l)}(\lambda I_{n-1} - \Lambda^{(l)})^{-1}U^{(l)\top}(u^{(l)\perp}_i)^\top Hu^*_i \|_2^2 
$$

$$
= \sum_{1 \leq k \leq n} \left( \frac{u^{(l)\top}_i (u^{(l)\perp}_i)^\top H u^*_i}{\lambda_i - \lambda_k^{(l)}} \right)^2.
$$

(C.7)

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

$$
U^{(l)\top}(u^{(l)\perp}_i)^\top H u^*_i \sim \mathcal{N}(0, \sigma^2 I_{n-1}).
$$

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

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

$$
\max_{1 \leq i \leq n} \left| Z_i \right| \lesssim \sigma \sqrt{\log n};
$$

(C.8a)

$$
\max_{1 \leq i \leq n} \left| Z_i - \sigma \right| \cdot \max_{1 \leq i \leq n} \left| Z_i + \sigma \right| \lesssim \sigma^2 \log n.
$$

(C.8b)

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

$$
\sum_{1 \leq k < r} \left( \frac{u^{(l)\top}_i (u^{(l)\perp}_i)^\top H u^*_i}{\lambda_i - \lambda_k^{(l)}} \right)^2 \lesssim \frac{\sigma^2 r \log n}{\min_{i \neq l} |\lambda_i^* - \lambda_l^*|^2} = \frac{\sigma^2 r \log n}{(\Delta_i^*)^2}.
$$

(C.9)

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

$$
\left| \sum_{r \leq k < n} \left( \frac{u^{(l)\top}_i (u^{(l)\perp}_i)^\top H u^*_i}{\lambda_i - \lambda_k^{(l)}} \right)^2 - \sum_{r \leq k < n} \frac{\sigma^2}{(\lambda_i - \lambda_k^{(l)})^2} \right| \lesssim \frac{\sigma^2 \sqrt{n \log n}}{\lambda_i^2}.
$$

(C.10)

The proof of this claim is deferred to the end of this section. It then suffices to control the term $\sum_{r \leq k < n} \frac{\sigma^2}{(\lambda_i - \lambda_k^{(l)})^2}$, 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 $1 - O(n^{-10})$,

$$
\left| \sum_{r \leq k \leq n-1} \frac{1}{(\lambda_i - \lambda_k^{(l)})^2} - \sum_{r+1 \leq k \leq n} \frac{1}{(\lambda_i - \lambda_k)^2} \right| \lesssim \frac{1}{\lambda_i^2}.
$$

(C.11)
As a consequence, we have
\[
\left| \sum_{r \leq k \leq n-1} \frac{1}{(\lambda_r - \lambda_k^{(l)})^2} \right| \leq \sum_{r+1 \leq k \leq n} \frac{1}{(\lambda_r - \lambda_k)^2} \lesssim \frac{n}{\lambda^2}. \tag{C.12}
\]

Therefore, combining (C.10) and (C.11) gives
\[
\left| \sum_{r \leq k \leq n-1} \left( \frac{\mathbf{u}_k^{(l)^\top} (\mathbf{u}_k^{i,\perp})^\top H \mathbf{u}_k^*}{\lambda_r - \lambda_k^{(l)}} \right)^2 - \sum_{r+1 \leq k \leq n} \frac{\sigma^2}{(\lambda_r - \lambda_k)^2} \right| \lesssim \frac{\sigma^2}{\lambda^2} \sqrt{n \log n} + \frac{\sigma^2 \log n}{\lambda^2} \lesssim \frac{\sigma^2}{\lambda^2} \sqrt{n \log n}.
\]

Inserting (C.9) and (C.13) into (C.7), we arrive at the advertised bound:
\[
\| (\lambda I_{n-1} - M^{(l)})^{-1} (\mathbf{u}_k^{i,\perp})^\top H \mathbf{u}_k^* \|^2_2 = \sum_{r+1 \leq k \leq n} \frac{\sigma^2}{(\lambda_r - \lambda_k)^2} + O \left( \frac{\sigma^2 r \log n}{(\Delta r)^2} + \frac{\sigma^2 \log n}{\lambda^2} \right) \lesssim \frac{\sigma^2 n}{\lambda^2} + \frac{\sigma^2 \log n}{(\Delta r)^2} \ll 1,
\]

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

**Proof of the claim (C.10).** Note that \( \mathbf{u}_k^{(l)^\top} (\mathbf{u}_k^{i,\perp})^\top H \mathbf{u}_k^* \) is independent of \( \Lambda^{(l)} \) but depends on \( \Lambda \). Therefore, we shall use the epsilon-net argument (i.e., Lemma 20) to bound it. Before proceeding, observe that from (3.33), (C.3) and the condition \( \sigma \sqrt{n} \ll \lambda_{\text{min}}^* \), the following holds for any \( \lambda \) obeying \( \lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\text{m0}}} (\lambda^*_r) \):
\[
|\lambda - \lambda^{(l)}_k| \geq |\lambda_r - \lambda^{(l)}_k| - |\lambda - \lambda_r - \gamma(\lambda_r)| - |\gamma(\lambda)| \lesssim |\lambda^*_r|, \quad k \geq r. \tag{C.14}
\]

Now we begin to check the conditions of Lemma 20. Since \( |f(x) - f(y)| \leq \sup_x |f'(x)||x - y| \), the following holds with probability at least \( 1 - O(n^{-20}) \) for all \( \lambda \) with \( \lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\text{m0}}} (\lambda^*_r) \):
\[
- \frac{d}{d\lambda} \sum_{r \leq k \leq n} \frac{(\mathbf{u}_k^{(l)^\top} (\mathbf{u}_k^{i,\perp})^\top H \mathbf{u}_k^*)^2 - \sigma^2}{(\lambda - \lambda_k^{(l)})^2} \leftrightharpoons \sum_{r \leq k \leq n} \frac{(\mathbf{u}_k^{(l)^\top} (\mathbf{u}_k^{i,\perp})^\top H \mathbf{u}_k^*)^2 - \sigma^2}{(\lambda - \lambda_k^{(l)})^3} \lesssim \sigma^2 n \log n \frac{1}{\lambda^*_r},
\]

where we use (C.8b) and (C.14). In addition, for any fixed \( \lambda \) obeying \( \lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\text{m0}}} (\lambda^*_r) \), one has
\[
\max_{r \leq k \leq n} \frac{1}{(\lambda - \lambda_k^{(l)})^2} \| (\mathbf{u}_k^{(l)^\top} (\mathbf{u}_k^{i,\perp})^\top H \mathbf{u}_k^*)^2 - \sigma^2 \|_{\psi_1} \lesssim \frac{\sigma^2}{\lambda^2}, \quad \lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\text{m0}}} (\lambda^*_r); \quad \sum_{r \leq k \leq n} \frac{1}{(\lambda - \lambda_k^{(l)})^4} \mathbb{E} \left[ (\mathbf{u}_k^{(l)^\top} (\mathbf{u}_k^{i,\perp})^\top H \mathbf{u}_k^*)^2 - \sigma^2 \right]^2 \lesssim \frac{\sigma^4 n}{\lambda^2}, \quad \lambda - \gamma(\lambda) \in \mathcal{B}_{\mathcal{E}_{\text{m0}}} (\lambda^*_r).
where $\|\cdot\|_{\psi_2}$ denote the sub-exponential norm. We can then apply the matrix Bernstein inequality [Koltchinskii, 2011, Corollary 2.1] to find: with probability exceeding $1 - O(n^{-20})$, 

$$\left| \sum_{r \leq k \leq n} \frac{(u_k^{(l)})^\top (u_k^{(l)})^\top H u_k^{(l)} - \sigma^2}{(\lambda - \lambda_k^{(l)})^2} \right| \lesssim L \log^2 n + \sqrt{V \log n} \lesssim \frac{\sigma^2 \sqrt{n \log n}}{\lambda_k^{(l)}}.$$ 

Recognizing the fact that $\{\lambda: \lambda - \gamma(\lambda) \in B_{\expo}(\lambda_i^*)\} \subset [\lambda_i^* - |\lambda_i^*|/3, \lambda_i^* + |\lambda_i^*|/3]$, 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 $|\lambda_l - \lambda_k^{(l)}| \gtrsim |\lambda_l^*|$ for all $k \geq r$, thus leading to

$$\left| \sum_{r \leq k \leq n-1} \frac{1}{(\lambda_l - \lambda_k^{(l)})^2} \right| \lesssim \frac{n}{\lambda_l^{*2}}.$$

In addition, the upper bound for $\sum_{r \leq k \leq n-1} 1/(\lambda_l - \lambda_k)^2$ 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 $M^{(l)} = (u_l^{(l)})^\top M u_l^{(l)}$ and $M$.

Without loss of generality, assume that $\lambda_l^* > 0$, and that there are $m$ (resp. $r - m$) eigenvalues of $M^*$ larger (resp. smaller) than 0. By Weyl’s inequality (similar to (5.10) in the proof of Theorem 6), it is easily seen that there are $m$ eigenvalues of $M$ larger than $c\sigma \sqrt{n}$ and that there are $r - m$ eigenvalues of $M$ 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 $M$ satisfying $|\lambda_1| \geq \cdots \geq |\lambda_n|$, we further denote by $\{\phi_k\}_{k=1}^n$ the eigenvalues of $M$ so that $\phi_1 \geq \cdots \geq \phi_n$. Consider the set of eigenvalues of $M$ with magnitudes upper bounded by $c\sigma \sqrt{n}$. We have the following relation:

$$\sum_{k: r+1 \leq k \leq n} \frac{1}{(\lambda_l - \lambda_k)^2} = \sum_{k: m+1 \leq k \leq n-r+m} \frac{1}{(\lambda_l - \phi_k)^2}.$$  

(C.15)

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

$$\sum_{k: r \leq k \leq n-1} \frac{1}{(\lambda_l - \lambda_k^{(l)})^2} = \sum_{k: m \leq k \leq n-r+m-1} \frac{1}{(\lambda_l - \phi_k^{(l)})^2},$$  

(C.16)

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

$$\left| \sum_{k: m \leq k \leq n-r+m-1} \frac{1}{(\lambda_l - \phi_k^{(l)})^2} - \sum_{k: m+1 \leq k \leq n-r+m} \frac{1}{(\lambda_l - \phi_k)^2} \right| \lesssim \frac{1}{\lambda_l^{*2}}.$$  

(C.17)

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

$$\phi_{k+1}^{(l)} \leq \phi_k^{(l)} \leq \phi_k \quad 1 \leq k < n.$$  

(C.18)

By the assumption $\lambda_l^* > 0$, one has $\lambda_l \geq \lambda_l^* - \|H\| \gtrsim \lambda_l^*$ and thus $\phi_k \leq \lambda_l$ for all $k \geq m$. Consequently, we know from (C.18) that for all $k \geq m$,

$$\phi_{k+1} \leq \phi_k \leq \lambda_l,$$

which further implies that

$$\frac{1}{(\lambda_l - \phi_{k+1})^2} \leq \frac{1}{(\lambda_l - \phi_k^{(l)})^2} \leq \frac{1}{(\lambda_l - \phi_k)^2}, \quad k \geq m.$$  

This enables us to bound

$$\sum_{k: m \leq k \leq n-r+m-1} \frac{1}{(\lambda_l - \phi_k^{(l)})^2} \leq \frac{1}{(\lambda_l - \phi_k)^2} + \sum_{k: m+1 \leq k \leq n-r+m-1} \frac{1}{(\lambda_l - \phi_k)^2}.$$

This leads to

$$\sum_{k: m \leq k \leq n-r+m-1} \frac{1}{(\lambda_l - \phi_k^{(l)})^2} \leq \frac{1}{(\lambda_l - \phi_k)^2} + \sum_{k: m+1 \leq k \leq n-r+m-1} \frac{1}{(\lambda_l - \phi_k)^2}.$$  

42
Consequently, we conclude that

\[
\sum_{k:m \leq k \leq n-r+m-1} \frac{1}{(\lambda_l - \phi(l)^k)^2} \geq \sum_{k:m \leq k \leq n-r+m-1} \frac{1}{(\lambda_l - \phi(l)+1)^2} = \sum_{k:m+1 \leq k \leq n-r+m} \frac{1}{(\lambda_l - \phi(l)^k)^2}.
\]

where the last relation holds since \(|\phi(l)^m| \leq \|H\|\) and hence \(\lambda_l - \phi(l)^k = \lambda_l - \lambda^*_l\).

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

### C.3 Proof of Lemma 5

Since \((u^*_l)^\top Hu^*_l \sim \mathcal{N}(0, \sigma^2 I_{n-1})\) is a Gaussian random vector independent from \(M^{(l)}\) but dependent on \(\lambda_l\), our proof strategy is to apply Lemma 20.

To this end, recall the definition

\[
u_k^{(l)} := (u^{(l)}_k)^\top u^*_k
\]

and the eigen-decomposition of

\[M^{(l)} = U^{(l)} \Lambda^{(l)} U^{(l)\top} = \sum_{1 \leq i \leq n} \lambda_i^{(l)} u_i^{(l)} (u_i^{(l)})^\top.
\]

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

\[
V := \sup_{\lambda - \gamma(\lambda) \in \mathcal{B}_{\Delta^*}(\lambda^*_l)} \left\| \sum_{k:k \neq l} a^\top u_k^{(l)} u_k^{(l)^\top} (\lambda I - M^{(l)})^{-1} \right\|_2 \lesssim \sum_{k:k \neq l} \left| \frac{a^\top u_k^{(l)}}{|\lambda^*_l - \lambda^*_k|} \right| + \frac{1}{|\lambda^*_l|}.
\]

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

\[
\left| (u^{(l)}_i)^\top Hu_i^* \cdot \sum_{k:k \neq l} a^\top u_k^{(l)} u_k^{(l)^\top} (\lambda I - M^{(l)})^{-1} \right| \lesssim \sigma \sqrt{\log \left( \frac{n \kappa \lambda_{\max}}{\Delta^*_l} \right)} \cdot V.
\]

In addition, we collect a basic fact regarding the derivatives of matrices: for any invertible matrix \(A\),

\[
\frac{dA^{-1}}{dx} = -A^{-1} \frac{dA}{dx} A^{-1}.
\]

With this identity in mind, one can derive: with probability at least \(1 - O(n^{-10})\), for all \(\lambda\) with \(\lambda - \gamma(\lambda) \in \mathcal{B}_{\Delta^*}(\lambda^*_l)\),

\[
\frac{d}{d\lambda} \sum_{k:k \neq l} a^\top u_k^{(l)*} u_k^{(l)} (\lambda I - M^{(l)})^{-1} (u^{(l)}_i)^\top Hu_i^* \right| 
\]

\[
= \left| \sum_{k:k \neq l} a^\top u_k^{(l)*} u_k^{(l)^\top} (\lambda I - M^{(l)})^{-2} (u^{(l)}_i)^\top Hu_i^* \right| 
\]

\[
\leq n \cdot \max_{1 \leq i < n} \frac{1}{(\lambda - \lambda^{(l)}_i)^2} \cdot \left| \sum_{k:k \neq l} a^\top u_k^{(l)*} u_k^{(l)^\top} \right|_2 \cdot \left| (u^{(l)}_i)^\top Hu_i^* \right|_2 
\]

\[
\lesssim n \cdot \frac{\Delta^*_l \lambda^*_l}{\lambda^*_l} \sum_{k:k \neq l} \left| a^\top u_k^{(l)} \right| \left| u_k^{(l)^\top} u_i^{(l)^\top} \right|_2 \cdot \sigma \sqrt{n \log n}
\]

\[
\lesssim n \cdot \frac{\Delta^*_l \lambda^*_l}{\lambda^*_l} \sum_{k:k \neq l} \left| a^\top u_k^{(l)} \right| \left| u_k^{(l)^\top} u_i^{(l)^\top} \right|_2 \cdot \sigma \sqrt{n \log n}
\]
Here, (i) arises from (5.33) in Lemma 8 and the standard Gaussian concentration inequality; (ii) holds since $\|u_k^* u_i^*(l)^T\|_2 \leq \|u_i^*\|_2 \|u_i^*(l)^T\| \leq 1$; (iii) is due to $\max_{i \neq j} |\lambda_i^* - \lambda_j^*| \leq \lambda_{\max}^*$; (iv) uses the definition of $V$ in (C.19). Moreover, it is easy to see that $\{\lambda; \lambda - \gamma(\lambda) \in B_{\log(n)}(\lambda^*_i)\} \subset [\lambda_i^* - \lambda^*_j/3, \lambda_j^* + |\lambda_i^*|/3]$. As a consequence, we invoke Lemma 20 and the union bound to find: with probability at least $1 - O(n^{-10})$:

$$\sum_{k: k \neq l} a_k^* u_i^*(l)^T (\lambda_i I - M^{(l)})^{-1} (u_i^*)^T H u_l^* \leq \sigma \sqrt{r \log \left( \frac{nk\lambda_{\max}^*}{\Delta^2} \right)} \cdot V$$

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 $\lambda$ such that $\lambda - \gamma(\lambda) \in B_{\log(n)}(\lambda^*_i)$, one can express

$$\left\| \sum_{k: k \neq l} a_k^* u_i^*(l)^T (\lambda_i I - M^{(l)})^{-1} \right\|_2 = \left\| \sum_{k: k \neq l} a_k^* u_i^*(l)^T U^{(l)} (\lambda_i I - M^{(l)})^{-1} U^{(l)^T} \right\|_2 = \left\| \sum_{k: k \neq l} a_k^* u_i^*(l)^T U^{(l)} \right\|_2 = \left\| \sum_{1 \leq i < n} \left( \frac{1}{\lambda_i - \lambda_i^{(l)}} \sum_{k: k \neq l} a_k^* u_i^*(l)^T u_i^* \right)^2 \right\|_2.$$  

(C.20)

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 $|\lambda - \lambda_i^{(l)}| \geq |\lambda_i^*|$ for all $i \geq r$. This in turn yields

$$\left\| \sum_{r \leq i \leq n-1} \left( \frac{1}{\lambda - \lambda_i^{(l)}} \sum_{k: k \neq l} a_k^* u_i^*(l)^T u_i^* \right)^2 \right\|_2 \leq \left\| \sum_{r \leq i \leq n-1} \left( \sum_{k: k \neq l} a_k^* u_i^*(l)^T u_i^* \right)^2 \right\|_2 \leq \frac{1}{|\lambda_i^*|} \left\| \sum_{k: k \neq l} a_k^* u_i^*(l)^T U^{(l)} \right\|_2 \leq \frac{1}{|\lambda_i^*|} \left\| a \right\|_2 \cdot \left\| \sum_{k: k \neq l} u_i^* u_k^*(l)^T \right\| \cdot \|U^{(l)}\| = \frac{1}{|\lambda_i^*|}.$$  

(C.21)

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

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

$$\left\| \sum_{1 \leq i < r} \left( \frac{1}{\lambda - \lambda_i^{(l)}} \sum_{k: k \neq l} a_k^* u_i^*(l)^T u_i^* \right)^2 \right\|_2 = \left\| \sum_{1 \leq i < r} \left( \frac{1}{\lambda - \lambda_i^{(l)}} \sum_{k: k \neq l} a_k^* u_i^*(l)^T u_i^* \right) u_i^{(l)^2} \right\|_2.$$
\[
\begin{align*}
&= \left\| \sum_{k:k \neq l} a^\top u_k^* \sum_{1 \leq i \leq r} \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} u_i(l) \right\|_2 \\
&\leq \sum_{k:k \neq l} \|a^\top u_k^*\| \left\| \sum_{1 \leq i \leq r} \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} u_i(l) \right\|_2 \\
&= \sum_{k:k \neq l} \|a^\top u_k^*\| \sqrt{\sum_{1 \leq i \leq r} \left( \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} \right)^2}. \quad (C.22)
\end{align*}
\]

The preceding inequality motivates us to control the quantity \(\sum_{1 \leq i \leq r} \left( \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} \right)^2\). Towards this, let us decompose it as follows

\[
\sum_{1 \leq i \leq r} \left( \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} \right)^2 = \sum_{i \in A_1} \left( \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} \right)^2 + \sum_{i \in A_2} \left( \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} \right)^2.
\]

Here, the sets \(A_1\) and \(A_2\) are defined respectively by

\[
A_1 := \{1 \leq i < r \mid \lambda_i^{(l)} - \gamma(\lambda_i^{(l)}) \in \mathcal{B}_{\varepsilon_k}(\lambda_k^*)\},
\]

\[
A_2 := \{1 \leq i < r \mid \lambda_i^{(l)} - \gamma(\lambda_i^{(l)}) \notin \mathcal{B}_{\varepsilon_k}(\lambda_k^*)\},
\]

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

- For each \(i \in A_1\), we claim that

\[
|\lambda - \lambda_i^{(l)}| \geq |\lambda - f^{-1}(\lambda_k^*)| - |f^{-1}(\lambda_k^*) - \lambda_i^{(l)}| \geq \frac{1}{2} |\lambda - f^{-1}(\lambda_k^*)| \gtrsim |\lambda_i^* - \lambda_k^*|. \quad (C.23)
\]

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

\[
|\lambda - f^{-1}(\lambda_k^*)| \geq \frac{1}{2} |f(\lambda) - f(f^{-1}(\lambda_k^*))| = \frac{1}{2} |(\lambda - \gamma(\lambda)) - \lambda_k^*| \\
\geq \frac{1}{2} |\lambda_i^* - \lambda_k^*| - \frac{1}{2} |\lambda - \gamma(\lambda) - \lambda_i^*| \\
\geq \frac{1}{2} |\lambda_i^* - \lambda_k^*| - \frac{1}{2} \mathcal{E}_{\text{MD}} \\
\gtrsim |\lambda_i^* - \lambda_k^*|.
\]

In a similar manner, we can also derive

\[
|f^{-1}(\lambda_k^*) - \lambda_i^{(l)}| \leq 2 |f(f^{-1}(\lambda_k^*)) - f(\lambda_i^{(l)})| = 2 |\lambda_k^* - (\lambda_i^{(l)} - \gamma(\lambda_i^{(l)}))|.
\]

Therefore, for any \(\lambda_i^{(l)}\) such that \(\lambda_i^{(l)} - \gamma(\lambda_i^{(l)}) \in \mathcal{B}_{\varepsilon_k}(\lambda_k^*)\), one has

\[
|f^{-1}(\lambda_k^*) - \lambda_i^{(l)}| \leq 2 \mathcal{E}_k + 2 \mathcal{E}_{\text{MD}} \leq 2c |\lambda_i^* - \lambda_k^*| \leq \frac{1}{2} |\lambda_i - f^{-1}(\lambda_k^*)|.
\]

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

\[
\sum_{i \in A_1} \left( \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} \right)^2 \lesssim \frac{1}{(\lambda_i^* - \lambda_k^*)^2} \sum_{i \in A_1} \left( \frac{u_k^{*\top}(l) u_i(l)}{\lambda - \lambda_i(l)} \right)^2 \leq \frac{1}{(\lambda_i^* - \lambda_k^*)^2}.
\]
Recall the definitions of $C.4$ Proof of Lemma 6

exceeding $\|\lambda I - \lambda^*\|_2 \leq \varepsilon_{MD} = \sigma \sqrt{r} \log n$.

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

$$\|u^{(l)\top}_k u^{(l)}_{i,j}\| \leq \|u^{(l)\top}_k u^{(l)}_{i,j}\|,$$

and therefore,

$$\|\lambda^{(l)} I_{r-1} - \lambda^* I_{r-1} I^{(l)} u^{(l)}_{i,j}\|_2 \geq |\lambda^{(l)} I_{r-1} - \lambda^* I_{r-1} I^{(l)} u^{(l)}_{i,j}| \geq |\lambda^* I_{r-1} - \lambda^* I_{r-1} I^{(l)} u^{(l)}_{i,j}|.$$This in turn allows us to derive that

$$\left(\frac{u^{(l)\top}_k u^{(l)}_{i,j}}{\lambda - \lambda^{(l)} I_{r-1}}\right)^2 \leq \left(\frac{u^{(l)\top}_k u^{(l)}_{i,j}}{\lambda - \lambda^{(l)} I_{r-1}}\right)^2 \leq \left(\frac{u^{(l)\top}_k u^{(l)}_{i,j}}{(\Delta^*)^2}\right)^2 \leq \frac{\sigma^2 r \log^2 n}{(\lambda^* - \lambda^*)^2(\Delta^*)^2}.$$Putting the above two sums together reveals that

$$\sum_{1 \leq i < r} \left(\frac{1}{\lambda - \lambda^{(l)} I_{r-1}} \sum_{k:k \neq l} a^{\top} u^{*\perp}_k u^{*\perp}_i \right)^2 \leq \sum_{k:k \neq l} \left|a^{\top} u^{*\perp}_k \right|^2 \sum_{1 \leq i < r} \left(\frac{u^{(l)\top}_k u^{(l)}_{i,j}}{\lambda - \lambda^{(l)} I_{r-1}}\right)^2 \leq \sum_{k:k \neq l} \left|a^{\top} u^{*\perp}_k \right|^2 \left(\frac{\sigma^2 r \log^2 n}{(\lambda^* - \lambda^*)^2(\Delta^*)^2}\right)^2 \leq \sum_{k:k \neq l} \left|\frac{a^{\top} u^{*\perp}_k}{\lambda^* - \lambda^*}\right|.$$Substituting the above two partial sums (C.21) and (C.24) into (C.20), we conclude that: with probability exceeding $1 - O(n^{-10}),$

$$\left\|\sum_{k:k \neq l} a^{\top} u^{*\perp}_k u^{*\perp}_i (\lambda I_{n-1} - M^{(l)})^{-1}\right\|_2 \leq \sum_{k:k \neq l} \left|\frac{a^{\top} u^{*\perp}_k}{\lambda^* - \lambda^*}\right|$$holds for any $\lambda$ such that $\lambda - \gamma(\lambda) \in \mathcal{B}_{\varepsilon_{MD}}(\lambda^*)$, as claimed in (C.19).

### C.4 Proof of Lemma 6

Recall the definitions of $u^{\perp}_{i}$, $u_{i,\perp}$ and $U^{*\perp}$ in (5.13), (5.24) and (5.32), respectively. Let us rewrite

$$\langle P_{U^{\perp}} a, P_{U^{\perp}} u_i \rangle \stackrel{(i)}{=} \langle (U^{*\perp})^\top a, (U^{*\perp})^\top u_i \rangle = \langle (U^{*\perp})^\top a, (U^{*\perp})^\top (u_{i}^* u_{i}^\top u_i + P_{u_{i}^*} u_i) \rangle \stackrel{(ii)}{=} \langle (U^{*\perp})^\top a, (U^{*\perp})^\top P_{u_{i}^*} u_i \rangle.$$
\[(iii) \left| \langle (U^* \perp)^T a, (U^* \perp)^T u_{l, \perp} \rangle \right| \cdot \left\| P_{u_{l, \perp}^*} u_{l} \right\|_2 \]

\[(iv) \left| \langle (u_{l, \perp}^+ U^* \perp)^T a, (u_{l, \perp}^+ U^* \perp)^T u_{l, \perp} \rangle \right| \cdot \left\| P_{u_{l, \perp}^*} u_{l} \right\|_2 \]

\[= \left| \langle U^* \perp (u_{l, \perp}^+)^T a, (U^* \perp)^T u_{l, \perp} \rangle \right| \cdot \left\| P_{u_{l, \perp}^*} u_{l} \right\|_2 \]

\[(v) \left| \langle P_{U^* \perp} ((u_{l, \perp}^+)^T a), P_{U^* \perp} ((u_{l, \perp}^+)^T \tilde{u}_{l, \perp}) \rangle \right| \cdot \left\| P_{u_{l, \perp}^*} u_{l} \right\|_2 \]

\[= \frac{1}{\left\| \tilde{u}_{l, \perp} \right\|_2} \left| \langle P_{U^* \perp} ((u_{l, \perp}^+)^T a), P_{U^* \perp} ((u_{l, \perp}^+)^T \tilde{u}_{l, \perp}) \rangle \right| \cdot \left\| P_{u_{l, \perp}^*} u_{l} \right\|_2 \]

(C.25)

Here, (i) follows since \((U^* \perp)^T U^* \perp = I_{n-r}\); (ii) holds true since \((U^* \perp)^T u_{l, \perp}^* = 0\); (iii) holds due to the definition \(u_{l, \perp} := (P_{u_{l, \perp}^*} u_{l})/\|P_{u_{l, \perp}^*} u_{l}\|_2\); (iv) results from the fact \(u_{l, \perp}^+ U^* \perp = U^* \perp\); (v) holds true since \((U^* \perp)^T U^* \perp = I_{n-r}\); (vi) arises from (5.4e) in Theorem 4, where we denote (i.e., \(u_{l, \perp}\) is the normalized version of \(\tilde{u}_{l, \perp}\))

\[\tilde{u}_{l, \perp} := u_{l, \perp} (\lambda I_{n-1} - (u_{l, \perp}^+)^T M u_{l, \perp})^{-1} (u_{l, \perp}^+)^T M u_{l, \perp}^*.\]

(C.26)

Our proof strategy is to show that \(P_{U^* \perp} \) is a random vector uniformly distributed in the unit sphere of the subspace \(U^* \perp\). 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 \(1 - O(n^{-10})\),

\[\left| \langle P_{U^* \perp} ((u_{l, \perp}^+)^T a), P_{U^* \perp} ((u_{l, \perp}^+)^T \tilde{u}_{l, \perp}) \rangle \right| \leq \sqrt{\frac{\log n}{n}} \left\| P_{U^* \perp} ((u_{l, \perp}^+)^T a) \right\|_2 \left\| P_{U^* \perp} ((u_{l, \perp}^+)^T \tilde{u}_{l, \perp}) \right\|_2 \]

\[\leq \sqrt{\frac{\log n}{n}} \left\| P_{U^* \perp} a \right\|_2 \left\| P_{U^* \perp} \tilde{u}_{l, \perp} \right\|_2,\]

where we use \(u_{l, \perp}^+ U^* \perp = U^* \perp\) 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:

\[\left| \langle P_{U^* \perp} a, P_{U^* \perp} u_{l} \rangle \right| \leq \sqrt{\frac{\log n}{n}} \left\| P_{U^* \perp} a \right\|_2 \left\| P_{U^* \perp} u_{l} \right\|_2.\]

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

\[R = P_{U^* \perp} + U^* \perp Q \langle U^* \perp \rangle^T \in \mathbb{R}^{(n-1) \times (n-1)};\]

the matrix \(R\) rotates vectors in the subspace spanned by \(U^* \perp\) according to \(Q\), while preserving the part in the subspace spanned by \(U^* \perp\). We make note of two important “rotational invariance” properties as follows.

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

\[R(u_{l, \perp}^+)^T H u_{l, \perp}^* = U^* \perp (U^* \perp)^T (u_{l, \perp}^+)^T H u_{l, \perp}^* + U^* \perp Q \langle U^* \perp \rangle^T (u_{l, \perp}^+)^T H u_{l, \perp}^*\]

\[= (u_{l, \perp}^+)^T U^* \perp U^* \perp Q \langle U^* \perp \rangle^T H u_{l, \perp}^*\]

\[= (u_{l, \perp}^+)^T (I_n - u_{l, \perp}^+ u_{l, \perp}^T) H u_{l, \perp}^*\]

\[= (u_{l, \perp}^+)^T H u_{l, \perp}^*.\]

Here, the second line arises from the definitions of \(U^* \perp\) and \(U^* \perp\) in (5.32); the third line follows because \(Q \langle U^* \perp \rangle^T H u_{l, \perp}^* \equiv (U^* \perp)^T H u_{l, \perp}^*\); the last line holds due to the fact \((u_{l, \perp}^+)^T u_{l, \perp}^* = 0\).
Using the statistical independence between these two parts, we reach

\[ R(u_i^\perp)^T H u_i^\perp R = (u_i^\perp)^T (U_i^\perp U_i^\perp + U_i^\perp Q (U_i^\perp)^T) H (U_i^\perp U_i^\perp + U_i^\perp Q (U_i^\perp)^T) u_i^\perp \]

\[ = (u_i^\perp)^T (U_i^* U_i^* + U_i^* Q (U_i^*)^T) H (U_i^* U_i^* + U_i^* Q (U_i^*)^T) u_i^\perp \]

\[ \overset{d}{=} (u_i^\perp)^T H u_i^\perp, \]

where the second line comes from \((U_i^* U_i^* - U_i^\perp U_i^\perp) u_i^\perp = u_i^\perp u_i^\perp u_i^\perp = 0\), and the last line holds since \(U_i^* U_i^* + U_i^* Q (U_i^*)^T\) is a rotation matrix.

Using the statistical independence between these two parts, we reach

\[ R(u_i^\perp)^T \tilde{u}_{i,\perp} = R(\lambda_i I_{n-1} - (u_i^\perp)^T M u_i^\perp)^{-1} (u_i^\perp)^T M u_i^* = (\lambda_i I_{n-1} - R(u_i^\perp)^T M u_i^\perp R)^{-1} R(u_i^\perp)^T M u_i^* \]

\[ \overset{d}{=} (\lambda_i I_{n-1} - (u_i^\perp)^T H u_i^\perp)^{-1} (u_i^\perp)^T H u_i^* \]

\[ = (\lambda_i I_{n-1} - (u_i^\perp)^T M u_i^\perp)^{-1} (u_i^\perp)^T M u_i^* \]

\[ = (u_i^\perp)^T \tilde{u}_{i,\perp}, \]

where the last step replies on the definition of \(\tilde{u}_{i,\perp}\) in (C.26) and the fact \((u_i^\perp)^T u_i^\perp = I_{n-1}\). This enables us to conclude that \(\| P_{u_i^\perp} (\tilde{u}_i^\perp)^T \tilde{u}_{i,\perp} \|_2 \) is uniformly distributed in the unit sphere spanned by \(U_i^*)^\perp\).

\section{D Proof of auxiliary lemmas in the analysis for Theorem 2}

\subsection{D.1 Proof of Lemma 7}

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

\[ \| A \| = \| [U^*, U^*]^T A [U^*, U^*]^\perp \| \]

\[ \overset{\text{D.1}}{=} \| U^* A U^* + U^*^T A U^*^\perp \| \]

\[ \leq \| U^* A U^* \| + \| U^*^T A U^*^\perp \| + \| (U^*)^\perp A U^* \| + \| (U^*)^\perp A U^*^\perp \|, \]

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

\[ \| \frac{1}{n} S S^T - \Sigma \| \leq \| U^* \left( \frac{1}{n} S S^T - \Sigma \right) U^* \| + \| (U^*^\perp)^T \left( \frac{1}{n} S S^T - \Sigma \right) U^* \| \]

\[ = \| \frac{1}{n} S \| S \| - \Lambda \| + 2 \| \frac{1}{n} S \| S \| - \sigma^2 I_{p-r} \| \]

where we remind the readers of the notation \(S := U^* S, S^\perp := (U^*^\perp)^T S\) and \(\Lambda := U^* \Sigma U^*\) introduced in (5.46).

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

\[ \| G \| \lesssim \sqrt{p} + \sqrt{n}, \]

\[ \| \frac{1}{n} G G^T - I_p \| \lesssim \sqrt{\frac{p}{n}} + \frac{p}{n} + \sqrt{\frac{\log n}{n}}, \]

\[ \| \frac{1}{n} G^T G - \frac{p}{n} I_n \| \lesssim 1 + \sqrt{\frac{p}{n}}. \]
With these bounds in place, we can start to bound the spectral norms of the quantities in (D.1). Note that
the columns of $S_{\parallel}$ (resp. $S_{\perp}$) are i.i.d. zero-mean Gaussian random vectors with covariance $\Lambda$ (resp. $\sigma^2 I_{p-r}$). Since we can rewrite $S_{\parallel} = \Lambda^{1/2} Z$ with $Z \in \mathbb{R}^{r \times n}$ 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 $1 - O(n^{-10})$,

$$
\| \frac{1}{n} S_{\perp} S_{\perp}^\top - \sigma^2 I_{p-r} \| \lesssim \sigma^2 \left( \sqrt{\frac{P}{n}} + \frac{P}{n} + \sqrt{\frac{\log n}{n}} \right) \quad \text{and} \quad (D.4)
$$

$$
\| \frac{1}{n} S_{\parallel} S_{\parallel}^\top - \Lambda \| = \left\| \frac{1}{n} \Lambda^{1/2} Z Z^\top \Lambda^{1/2} - \Lambda \right\| \lesssim \| \Lambda \| \left\| \frac{1}{n} Z Z^\top - I_r \right\| \lesssim (\lambda_{\max}^* + \sigma^2) \left( \sqrt{\frac{r}{n}} + \frac{r}{n} + \sqrt{\frac{\log n}{n}} \right)
$$

$$
\lesssim (\lambda_{\max}^* + \sigma^2) \sqrt{\frac{r \log n}{n}}, \quad (D.5)
$$

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

$$
\left\| \frac{1}{n} S_{\perp} S_{\parallel}^\top \right\| = \frac{1}{n} \| S_{\perp} Z Z^\top \| \| \Lambda^{1/2} \| \lesssim \frac{1}{n} \| S_{\perp} Z Z^\top \| \| \Lambda^{1/2} \| \lesssim \sigma \left( \sqrt{\frac{pn \log n}{n}} + \sqrt{\frac{pr \log n}{n}} \right) \cdot \sqrt{\lambda_{\max}^* + \sigma^2}
$$

$$
\lesssim \sqrt{(\lambda_{\max}^* + \sigma^2) \sigma^2 \frac{P}{n} \log n}, \quad (D.6)
$$

where the last step holds since $n \geq r$. 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

$$
\Sigma_{l, \perp} := \frac{1}{n} \mathbb{E} \left[ S_{l, \perp} S_{l, \perp}^\top \right] \quad (D.7)
$$

$$
= \frac{1}{n} \left( u_{l, \perp}^\top \right) \mathbb{E} \left[ S S^\top \right] u_{l, \perp} = \left( u_{l, \perp}^\top \right) \Sigma u_{l, \perp}^\top
$$

$$
= \left( u_{l, \perp}^\top \right) U^* \Lambda^* U^\top u_{l, \perp}^\top + \sigma^2 I_{p-1}
$$

$$
= U^{(l)} \Lambda^{(l)} U^{(l)} \top + \sigma^2 I_{p-1},
$$

where $U^{(l)}$ and $\Lambda^{(l)}$ have been defined in (5.72). Our proof strategy is to invoke Theorem 5 by treating $\frac{1}{n} S_{l, \perp} S_{l, \perp}^\top$ (resp. $U^{(l)}$) as $M$ (resp. $Q$).

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

$$
K^{(l)}(\lambda) := U^{(l)} \top \frac{1}{n} S_{l, \perp} S_{l, \perp}^\top U^{(l)} \parallel \left( \lambda I_{p-r} - U^{(l)} \perp \right) \frac{1}{n} S_{l, \perp} S_{l, \perp}^\top U^{(l)} \perp \left( U^{(l)} \perp \right) \top \frac{1}{n} S_{l, \perp} S_{l, \perp}^\top U^{(l)} \perp
$$

$$
K^{(l)}(\lambda) := \mathbb{E} \left[ G^{(l)}(\lambda) \left| (U^{(l)} \perp) \top S_{l, \perp} S_{l, \perp}^\top U^{(l)} \perp \right. \right].
$$

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

$$
u_{l, \perp}^\top U^{(l)} \perp = U^{\perp},
$$

$$
(U^{(l)} \perp) \top S_{l, \perp} = (U^{(l)} \perp) \top (u_{l, \perp} \perp)^\top S = (U^{\perp}) \top S = S_{\perp}.
$$

Straightforward calculation allows us to simplify the above expressions as follows

$$
K^{(l)}(\lambda) := \frac{1}{n} P^{(l)} \top S_{\parallel} \frac{1}{n} S_{\perp} \left( \lambda I_{p-r} - \frac{1}{n} S_{\perp} S_{\perp}^\top \right)^{-1} S_{\perp} \top S_{\parallel} P^{(l)} = P^{(l)} \top K(\lambda) P^{(l)},
$$

49
One can then adopt a similar argument as in the proof of Theorem 7 to demonstrate that: with probability at least $1 - O(n^{-10})$, 

$$
\norm{\left( \gamma_i^{(l)} I_{r-1} - (1 + \gamma_i^{(l)}) (\Lambda^{(l)} + \sigma^2 I_{r-1}) \right) U^{(l)\top} u_{i\|}}_2 \\
\leq \frac{1}{n} \norm{P^{(l)} S_{\|} S^{\top} P^{(l)} - \Lambda^{(l)} - \sigma^2 I_{r-1}} + \sup_{\lambda, \gamma \in [2\lambda I/3, 4\lambda I/3]} \norm{(K(\gamma_i^{(l)}) - K(\gamma_i^{(l)})) U^{(l)\top} u_{i\|}}_2 \\
\leq \frac{1}{n} \norm{S_{\|} - \Lambda^* - \sigma^2 I} + \sup_{\lambda, \gamma \in [2\lambda I/3, 4\lambda I/3]} \norm{K(\gamma_i^{(l)}) - K(\gamma_i^{(l)})}_2 \lesssim \varepsilon_{\text{PCA}}.
$$

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

$$
\left| \frac{1}{1 + \gamma_i^{(l)}} - \lambda_k^2 - \sigma^2 \right| \lesssim \varepsilon_{\text{PCA}}.
$$

• 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 $(U^{(l)\top}) S_{\perp} = S_{\perp}$, we can invoke Lemma 21 to see that

$$
\gamma_i^{(l)} \leq \lambda_1 \left( \frac{1}{n} S_{\perp} S_{\perp}^{\top} \right).
$$

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

$$
\left| \lambda_i \left( \frac{1}{n} S_{\perp} S_{\perp}^{\top} \right) - \sigma^2 \frac{n \vee (p - r)}{n} \right| \lesssim \sqrt{\frac{p + \log n}{n}}, \quad 1 \leq i \leq (p - r) \wedge n,
$$

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

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

for all $r \leq i \leq n \wedge (p - 1)$.

– Next, we move on to consider the lower bound. Observe that the matrix $\Sigma_{\ell \perp}$ defined in (D.7) satisfies the following properties: (i) $\Sigma_{\ell \perp} \succeq \sigma^2 I_{p-1}$; (ii) $\Sigma_{\ell \perp}^{-1/2} S_{\ell \perp}$ 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 $1 \leq i \leq (p - 1) \wedge n$ as follows

$$
\gamma_i^{(l)} = \lambda_i \left( \frac{1}{n} S_{\ell \perp} S_{\ell \perp}^{\top} \right) \geq \lambda_i \left( \frac{1}{n} S_{\perp} S_{\perp}^{\top} \right)
\\
= \lambda_i \left( \frac{1}{n} (\Sigma_{\ell \perp}^{-1/2} S_{\ell \perp})^{\top} \Sigma_{\ell \perp} \Sigma_{\ell \perp}^{-1/2} S_{\ell \perp} \right)
\\
\geq \sigma^2 \lambda_i \left( \frac{1}{n} (\Sigma_{\ell \perp}^{-1/2} S_{\ell \perp})^{\top} \Sigma_{\ell \perp}^{-1/2} S_{\ell \perp} \right)
\\
\geq \sigma^2 \lambda_i \left( \frac{1}{n} (\Sigma_{\ell \perp}^{-1/2} S_{\ell \perp})^{\top} (\Sigma_{\ell \perp}^{-1/2} S_{\ell \perp}) \right),
$$

50
Before continuing, we introduce several useful notation that will be used throughout. Moreover, by virtue of the rotational invariance of i.i.d. Gaussian random matrices, it is readily seen that

\[
\mathbf{V}_1 = \mathbf{V}_2 = \cdots = \mathbf{V}_r,
\]

In addition, we note that the vector

\[
\mathbf{v}_i = \frac{1}{\sqrt{n}} \mathbf{S}_{i,\perp} \mathbf{u}_i,
\]

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

\[
\mathbf{v}_i^\top \mathbf{s}_{i,\perp} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \lambda_i^* + \sigma^2), \quad 1 \leq i < p. \tag{D.8}
\]

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

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

\[
\left\| \frac{1}{n} \mathbf{S}_{i,\perp} \right\|_2^2 = \frac{1}{n} \left\| \mathbf{S}_{i,\perp} \right\|_2^2 \leq \frac{1}{n} \left\| \frac{1}{n} \mathbf{S}_{i,\perp} \right\|_2^2 \leq \frac{1}{n} \left\| \mathbf{S}_{i,\perp} \right\|_2^2.
\]

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
\[ \gamma_i(t) \leq \lambda_{\text{max}}^* + \sigma^2 \quad \text{and} \quad (\gamma_i(t) - \lambda_i)^2 \geq \min_{i:j \neq i} (\lambda_j^* - \lambda_i^*)^2 \]
for all $1 \leq i < r$. In addition, recall that $\mathbf{v}_i^{(t)\top} \mathbf{s}_{i\|} (1 \leq i < p)$ are i.i.d. zero-mean Gaussian random variables with variance $\lambda_i^* + \sigma^2$ (see (D.8)). Invoking standard Gaussian inequalities shows that with probability at least $1 - O(n^{-10})$,
\[ \max_{1 \leq i < r} (\mathbf{v}_i^{(t)\top} \mathbf{s}_{i\|})^2 \lesssim (\lambda_i^* + \sigma^2) \log n. \]  
As a result, we obtain
\[ \sum_{1 \leq i < r} \frac{\gamma_i(t)(\mathbf{v}_i^{(t)\top} \mathbf{s}_{i\|})^2}{(\lambda_i - \gamma_i(t))^2} \lesssim \frac{(\lambda_{\text{max}}^* + \sigma^2)(\lambda_i^* + \sigma^2) r \log n}{(\Delta_i)^2} \]  
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 $\mathbf{v}_i^{(t)\top} \mathbf{s}_{i\|} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \lambda_i^* + \sigma^2)$. This suggests that we decompose
\[ \sum_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i(t)(\mathbf{v}_i^{(t)\top} \mathbf{s}_{i\|})^2}{(\lambda_i - \gamma_i(t))^2} = \sum_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i(t)(\lambda_i^* + \sigma^2)}{(\lambda_i - \gamma_i(t))^2} + \sum_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i(t)((\mathbf{v}_i^{(t)\top} \mathbf{s}_{i\|})^2 - (\lambda_i^* + \sigma^2))}{(\lambda_i - \gamma_i(t))^2}, \]
and control $\alpha_1$ as well as $\alpha_2$ individually.

To begin with, let us consider $\alpha_1$, which requires estimating $\lambda_i^* + \sigma^2$ and $\sum_{r \leq i \leq n \wedge (p-1)} \gamma_i(t)/(\lambda_i - \gamma_i(t))^2$. This task is accomplished in Lemma 15 and Lemma 16 stated below.

**Lemma 15.** *Instate the assumptions of Theorem 2. With probability at least $1 - O(n^{-10})$, we have*
\[ \left| \sum_{t \geq r} \frac{\gamma_i(t)}{(\lambda_i - \gamma_i(t))^2} - \sum_{t > r} \frac{\lambda_i}{(\lambda_i - \lambda_i)^2} \right| \lesssim \frac{\sigma^2}{\lambda_i^2} \left( 1 + \frac{p}{n} \right). \]  
*Proof. See Appendix D.3.1.*

**Lemma 16.** *Instate the assumptions of Theorem 2. With probability at least $1 - O(n^{-10})$, one has*
\[ \left| \frac{\lambda_i}{1 + \frac{1}{n} \sum_{t > r} \frac{\lambda_i}{\lambda_i - \lambda_i^*}} - (\lambda_i^* + \sigma^2) \right| \lesssim \frac{(\lambda_{\text{max}}^* + \sigma^2) r \log n}{n}. \]  
*Proof. See Appendix D.3.2.*
With these two lemmas in place, we are ready to control \( \alpha_1 \). According to (D.14), we have

\[
\alpha_1 \lesssim \frac{(\lambda_1^* + \sigma^2)\sigma^2 p}{\lambda_1^*}. \tag{D.17}
\]

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

\[
|\alpha_1 - c_l \cdot n| = \left| (\lambda_1^* + \sigma^2) \sum_{i \geq r} \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} - c_l \cdot n \right|
\]

\[
= \left| (\lambda_1^* + \sigma^2) \left( \sum_{i > r} \frac{\lambda_i}{(\lambda_i - \lambda_i - \lambda_i)^2} + \epsilon_1 \right) - (\lambda_1^* + \sigma^2 + \epsilon_2) \sum_{i > r} \frac{\lambda_i}{(\lambda_i - \lambda_i)^2} \right|
\]

\[
\leq (\lambda_1^* + \sigma^2) |\epsilon_1| + \frac{\sigma^2 p}{\lambda_1^*} |\epsilon_2|
\]

\[
\leq (\lambda_1^* + \sigma^2) \cdot \frac{\sigma^2}{\lambda_1^*} (1 + \frac{p}{n}) + \frac{\sigma^2 p}{\lambda_1^*} \cdot (\lambda_{\max}^* + \sigma^2) \sqrt{\frac{r \log n}{n}}
\]

\[
\approx (\lambda_1^* + \sigma^2) \frac{\sigma^2}{\lambda_1^*} + (\lambda_{\max}^* + \sigma^2) \frac{\sigma^2 p}{\lambda_1^*} \sqrt{\frac{r \log n}{n}}. \tag{D.18}
\]

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 \( \alpha_2 \). Observe that \( \{(\mathbf{v}_i^{(l)}\mathbf{S}_i^{(l)T})^2 - (\lambda_i^* + \sigma^2)\}_{i \geq r} \) is a sequence of zero-mean sub-exponential random variables, which is independent of \( \Gamma^{(l)} \) but depends on \( \lambda_i \). Hence, we shall apply the epsilon-net argument (cf. Lemma 20) to bound \( \alpha_2 \). To do so, let us first verify the conditions required therein. With probability exceeding \( 1 - O(n^{-20}) \), one has

\[
\left| \frac{d}{d\lambda} \sum_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^3} (\mathbf{v}_i^{(l)}\mathbf{S}_i^{(l)T})^2 - (\lambda_i^* + \sigma^2) \right|
\]

\[
\leq (p \wedge n) \cdot \max_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^3} \cdot \max_{r \leq i \leq n \wedge (p-1)} \left| (\mathbf{v}_i^{(l)}\mathbf{S}_i^{(l)T})^2 - (\lambda_i^* + \sigma^2) \right|
\]

\[
\lesssim (p \wedge n) \cdot \frac{\sigma^2 (p \vee n)}{\lambda_i^* n} \cdot (\lambda_i^* + \sigma^2) \log n
\]

\[
= \frac{(\lambda_i^* + \sigma^2)\sigma^2 p \log n}{\lambda_i^*}.
\]

for all \( \lambda \) with \( \lambda/(1 + \beta(\lambda)) \in \mathcal{B}_{EPCA}(\lambda_i^* + \sigma^2) \). In addition, it is seen that

\[
\left\| (\mathbf{v}_i^{(l)}\mathbf{S}_i^{(l)T})^2 - (\lambda_i^* + \sigma^2) \right\|_{\psi_1} \lesssim \lambda_i^* + \sigma^2,
\]

\[
\mathbb{E} \left( (\mathbf{v}_i^{(l)}\mathbf{S}_i^{(l)T})^2 - (\lambda_i^* + \sigma^2) \right)^2 \lesssim (\lambda_i^* + \sigma^2)^2,
\]

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 \( \lambda \) obeying \( \lambda/(1 + \beta(\lambda)) \in \mathcal{B}_{EPCA}(\lambda_i^* + \sigma^2) \), one has

\[
\left| \sum_{r \leq i \leq n \wedge (p-1)} \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} (\mathbf{v}_i^{(l)}\mathbf{S}_i^{(l)T})^2 - (\lambda_i^* + \sigma^2) \right|
\]
Instate the assumptions of Theorem 2. Suppose that $n < p$. To remedy the issue, we provide a more precise estimate for terms (D.12) in the following lemma.

Proof. See Appendix D.3.3.

\[
\sum_{i \geq r} \frac{\gamma_i^{(l)} (v_i^{(l)} \| s_{l,i})^2}{n (\lambda_i - \gamma_i^{(l)})^2} \leq \frac{\sigma^2 p}{\lambda_i - \sigma^2 p/n} + \frac{\lambda_i}{\lambda_i^*} \frac{1}{n} \sum_{i > r} \frac{\lambda_i}{\lambda_i - \lambda_i^*} \sum_{r < i \leq n} \frac{\lambda_i - \sigma^2 p/n}{(\lambda_i - \gamma_i^{(l)})^2}
\]

\[
+ O\left(\frac{\sigma^2 pr \log n}{\min_{i \neq l} |\lambda_i^* - \lambda_i^*| n} + \frac{\sigma^2 \kappa \sqrt{pr \log n}}{\lambda_i^* n}\right).
\]

\[
\sum_{i \geq r} \frac{\gamma_i^{(l)} (v_i^{(l)} \| s_{l,i})^2}{n (\lambda_i - \gamma_i^{(l)})^2} \leq c_l \leq \frac{\sigma^2 p \log n}{\min_{i \neq l} |\lambda_i^* - \lambda_i^*| n} + \frac{\sigma^2 \kappa \sqrt{pr \log n}}{\lambda_i^* n}.
\]

\[
\sum_{i \geq r} \frac{\gamma_i^{(l)} (v_i^{(l)} \| s_{l,i})^2}{n (\lambda_i - \gamma_i^{(l)})^2} - c_l \leq \frac{\sigma^2 p \log n}{\min_{i \neq l} |\lambda_i^* - \lambda_i^*| n} + \frac{\sigma^2 \kappa \sqrt{pr \log n}}{\lambda_i^* n}.
\]
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):

\[
\bigg\| \left( \lambda_i I_{p-1} - \frac{1}{n} S_{l,\perp} S_{l,\perp}^\top \right) \bigg\|_2 \leq \frac{1}{n} S_{l,\perp} S_{l,\perp}^\top \bigg\|_2 \leq \frac{(\lambda_{\text{max}}^* + \sigma^2)(\lambda_i^* + \sigma^2) r \log n}{\min_{i: \not= l} |\lambda_i^* - \lambda_i^*|^2 n} + \frac{(\lambda_i^* + \sigma^2) \sigma^2 p \log n}{\lambda_i^* n} \ll 1,
\]

where the last step holds due to the assumptions (3.11a) and (3.11b).

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

\[
\bigg\| \left( \lambda_i I_{p-1} - \frac{1}{n} S_{l,\perp} S_{l,\perp}^\top \right) \bigg\|_2 \leq \frac{(\lambda_{\text{max}}^* + \sigma^2)(\lambda_i^* + \sigma^2) r \log n}{\min_{i: \not= l} |\lambda_i^* - \lambda_i^*|^2 n} + \frac{\sigma^2 p \log n}{\lambda_i^* n} \ll 1,
\]

As for the case with \( n < p \), substituting (D.11) and (D.23) into (D.9) yields

\[
\bigg\| \left( \lambda_i I_{p-1} - \frac{1}{n} S_{l,\perp} S_{l,\perp}^\top \right) \bigg\|_2 \leq \frac{(\lambda_{\text{max}}^* + \sigma^2)(\lambda_i^* + \sigma^2) r \log n}{\min_{i: \not= l} |\lambda_i^* - \lambda_i^*|^2 n} + \frac{\sigma^2 p \log n}{\lambda_i^* n} \ll 1,
\]

where in the last line we use the conditions \( \min_{i: \not= l} |\lambda_i^* - \lambda_i^*| \leq \lambda_{\text{max}}^* \) and \( \sigma^2 p/n \ll \lambda_i^* \) (according to the noise assumption (3.11a)).

D.3.1 Proof of Lemma 15

To begin with, let us consider (D.13). By construction, we have \( S_{l,\perp} S_{l,\perp}^\top = (u^*_{l,\perp})^T S S^T u^*_{l,\perp} \), and it follows from Lemma 21 that \( \lambda_i \leq \gamma_i^{(l)} \leq \lambda_i^* \) for all \( 1 \leq i < p \). Simple calculation yields

\[
\frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} = \frac{\lambda_i}{(\lambda_i - \lambda_i^*)^2} = \frac{(\gamma_i^{(l)} - \lambda_i)(\lambda_i^* - \lambda_i)}{(\lambda_i - \gamma_i^{(l)})^2(\lambda_i - \lambda_i^*)^2},
\]

and consequently

\[
\frac{\lambda_i}{(\lambda_i - \lambda_i^*)^2} \leq \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} \leq \frac{\lambda_i}{(\lambda_i - \lambda_i^*)^2} \leq \frac{\lambda_i}{(\lambda_i - \lambda_i^*)^2}, \quad i \geq r.
\]

We can then invoke a similar argument used in the proof of Lemma 14 to bound

\[
\sum_{r < i \leq n} \frac{\lambda_i}{(\lambda_i - \lambda_i^*)^2} \leq \sum_{r < i \leq n} \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} \leq \sum_{r < i \leq n} \frac{\lambda_i}{(\lambda_i - \lambda_i^*)^2} + \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2}.
\]

Hence, the conclusion immediately follows since \( \gamma_i^{(l)}/(\lambda_i - \gamma_i^{(l)})^2 \ll \sigma^2(1 + p/n)\lambda_i^* \) 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) \( |\lambda_i - \gamma_i^{(l)}| \geq \lambda_i^* \) for all \( i \geq r \); (ii) \( |\gamma_i^{(l)}| \leq \sigma^2 (p \vee n)/n \) for all \( i \leq n \wedge (p - 1) \); (iii) \( \gamma_i^{(l)} = 0 \) for \( n \wedge (p - 1) < i < p \). Therefore, we can upper bound

\[
\left| \sum_{i \geq r} \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} \right| \leq (p \wedge n) \frac{\sigma^2 (p \vee n)}{\lambda_i^* n} \leq \frac{\sigma^2 p}{\lambda_i^* n},
\]

and the upper bound for \( \sum_{i > r} \lambda_i/(\lambda_i - \lambda_i^*)^2 \) 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

$$\beta(\lambda_l) = \frac{1}{n} \sum_{1 \leq i \leq p-r} \frac{\lambda_i}{\lambda_l - \lambda_i} \frac{\lambda_i}{\lambda_i^n S_{S^T}}.$$  

From Lemma 21, we know that $\lambda_{i+r} \leq \lambda_i \left( \frac{1}{n} S_{S^T} \right) \leq \lambda_i$ for each $1 \leq i \leq p - r$, and thus

$$\frac{\lambda_{i+r}}{\lambda_l - \lambda_{i+r}} \leq \frac{\lambda_i}{\lambda_l - \lambda_i} \frac{\lambda_i}{\lambda_i^n S_{S^T}} \leq \frac{\lambda_i}{\lambda_l - \lambda_i}, \quad i > r.$$  

Hence, we have

$$0 \leq \beta(\lambda_l) - \frac{1}{n} \sum_{r < i \leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} \leq 1 \sum_{1 \leq i \leq r} \frac{\lambda_i}{\lambda_l - \lambda_i} \frac{\lambda_i}{\lambda_i^n S_{S^T}}.$$  

(D.24)

As shown in ((5.44)), the eigenvalue $\lambda_l$ satisfies $\lambda_l / (1 + \beta(\lambda_l)) = \lambda_l^* + \sigma^2 + O(\epsilon_{PCA})$ where $\epsilon_{PCA}$ is defined in (5.59). In particular, we note that for the case with $n < p$, the assumption (3.11a) guarantees that

$$\sigma^2 = o(\lambda_{\text{min}}^*) \quad \text{and} \quad \epsilon_{PCA} := (\lambda_{\text{max}}^* + \sigma^2) \sqrt{\frac{p}{n}} \log n = o(\lambda_{\text{min}}^*).$$  

(D.25)

Combined with (5.61), this also implies that

$$\lambda_l \simeq \lambda_l^*.$$  

(D.26)

With these estimates in place, one can use (D.24) and the high-probability bound $\left\| \frac{1}{n} S_{S^T} \right\| \lesssim \sigma^2 (1 + p/n) \ll \lambda_l^*$ in (D.4) to derive

$$\left| \frac{1}{n} \sum_{r < i \leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} - \beta(\lambda_l) \right| \leq \frac{r}{n} \sum_{1 \leq i \leq r} \left| \frac{\lambda_i}{\lambda_l - \lambda_i} \frac{\lambda_i}{\lambda_i^n S_{S^T}} \right| \lesssim \frac{r}{n} \sigma^2 \left( 1 + \frac{p}{n} \right) = o\left( \frac{r}{n} \right).$$  

(D.27)

where the last step holds due the noise condition (3.11a). Meanwhile, we can combine (5.61) with (D.27) to find

$$\left| \frac{1}{n} \sum_{r < i \leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} \right| \ll 1.$$  

(D.28)

as long as $n \gg r$. Plugging this into ((5.44)) reveals that

$$\lambda_l^* + \sigma^2 = \frac{\lambda_l}{1 + \beta(\lambda_l)} + O(\epsilon_{PCA})$$

$$\overset{(i)}{=} \frac{\lambda_l}{1 + \frac{1}{n} \sum_{r < i \leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} + o\left( \frac{r}{n} \right) + O(\epsilon_{PCA})}$$

$$\overset{(ii)}{=} \frac{\lambda_l}{1 + \frac{1}{n} \sum_{r < i \leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} + o\left( \frac{\lambda_l^* r}{n} \right) + O(\epsilon_{PCA})}$$

$$\overset{(iii)}{=} \frac{\lambda_l}{1 + \frac{1}{n} \sum_{r < i \leq p} \frac{\lambda_i}{\lambda_l - \lambda_i} + O(\epsilon_{PCA})},$$  

(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 \( v_i^{(l)^\top} s_i^{(l)^\top} \overset{i.i.d.}{\sim} N(0, \lambda_i^* + \sigma^2) \) (see (D.8)), we are motivated to first decompose

\[
\sum_{r \leq i \leq n} \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} = \sum_{r \leq i \leq n} \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} \left( v_i^{(l)^\top} s_i^{(l)^\top} \right)^2 \\
+ \sum_{r \leq i \leq n} \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} (\lambda_i^* + \sigma^2) \tag{D.30}
\]

In what follows, we shall control \( \alpha_1, \alpha_2 \) and \( \alpha_3 \) separately in a reverse order.

Controlling \( \alpha_3 \). We intend to apply Lemma 20 in Section F to control \( \alpha_3 \). Before proceeding, we pause to make a few observations. It is straightforward to compute that \( f'(x) = \frac{1}{(\lambda - x)^2} \) and \( g'(x) = \frac{\lambda + (k-1)x}{(\lambda - x)^3} \) for the function \( f(x) := \frac{1}{(\lambda - x)^2} \) and \( g(x) := \frac{x}{(\lambda - x)^3} \). Since \(|f(x) - f(y)| \leq \{\sup_x |f'(z)|\} |x - y| \) for any function \( f(\cdot) \), one can demonstrate that: for all \( \lambda \) satisfying \( \lambda / (1 + \beta(\lambda)) \in \mathcal{B}_{\varepsilon_A}(\lambda_i^* + \sigma^2) \), we claim that the following holds.

\[
\max_{1 \leq i \leq n} \left| \frac{1}{\lambda - \gamma_i^{(l)}} - \frac{1}{\lambda - \sigma^2 p/n} \right| \lesssim \max_{1 \leq i \leq n} \left| \gamma_i^{(l)} - \sigma^2 p/n \right| \qquad \lesssim \frac{\sigma^2}{\lambda_i^2} \sqrt{p/n}, \tag{D.31a}
\]

\[
\max_{1 \leq i \leq n} \left| \frac{1}{(\lambda - \gamma_i^{(l)})^2} - \frac{1}{(\lambda - \sigma^2 p/n)^2} \right| \lesssim \max_{1 \leq i \leq n} \left| \gamma_i^{(l)} - \sigma^2 p/n \right| \qquad \lesssim \frac{\sigma^2}{\lambda_i^2} \sqrt{p/n}, \tag{D.31b}
\]

\[
\max_{1 \leq i \leq n} \left| \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda - \sigma^2 p/n)^2} \right| \lesssim \max_{1 \leq i \leq n} \left| \gamma_i^{(l)} - \sigma^2 p/n \right| \qquad \lesssim \frac{\lambda_i^*}{\lambda_i^2} \cdot \frac{\sigma^2}{\lambda_i^2} \sqrt{p/n}, \tag{D.31c}
\]

and

\[
\max_{1 \leq i \leq n} \left| \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^3} - \frac{\sigma^2 p/n}{(\lambda - \sigma^2 p/n)^3} \right| \lesssim \max_{1 \leq i \leq n} \left| \gamma_i^{(l)} - \sigma^2 p/n \right| \qquad \lesssim \frac{\lambda_i^*}{\lambda_i^2} \cdot \frac{\sigma^2}{\lambda_i^2} \sqrt{p/n}. \tag{D.31d}
\]

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 $\lambda, \gamma$ satisfying $\lambda/(1 + \beta(\lambda)) \in B_{\mathcal{E}_{\text{PCA}}} (\lambda^*_i + \sigma^2)$ and $|\gamma - \sigma^2 p/n| \lesssim \sigma^2 \sqrt{p/n}$, one has

$$|\lambda - \gamma| \gtrsim \lambda^*_i;$$

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 $\lambda$ satisfying $\lambda/(1 + \beta(\lambda)) \in B_{\mathcal{E}_{\text{PCA}}} (\lambda^*_i + \sigma^2)$,

$$\sum_{r \leq i \leq n} \left( \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda - \sigma^2 p/n)^2} \right) \left( (v_i^{(l)})^T s_i^\top \right)^2 - (\lambda^*_i + \sigma^2) \right| \lesssim n \cdot \max_{1 \leq i \leq n} \left| \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda - \sigma^2 p/n)^2} \right| \cdot \max_{1 \leq i \leq n} \left| (v_i^{(l)})^T s_i^\top \right|^2 - (\lambda^*_i + \sigma^2) \right| \lesssim n \cdot \frac{\sigma^2}{\lambda^*_i^2} \sqrt{\frac{p}{n}} \cdot (\lambda^*_i + \sigma^2)^2 \log n \asymp \frac{\sigma^2 \sqrt{mp \log n}}{\lambda^*_i},$$

where the last line holds due to the upper bound (D.31c), the fact $\sigma^2 \leq \sigma^2 p/n \ll \lambda^*_i$ (from the noise assumption (3.11a)), 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 $\lambda$ satisfying $\lambda/(1 + \beta(\lambda)) \in B_{\mathcal{E}_{\text{PCA}}} (\lambda^*_i + \sigma^2)$, with probability at least $1 - O(n^{-10})$,

$$\sum_{r \leq i \leq n} \left( \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda - \sigma^2 p/n)^2} \right) \left( (v_i^{(l)})^T s_i^\top \right)^2 - (\lambda^*_i + \sigma^2) \right| \lesssim n \cdot \max_{1 \leq i \leq n} \left| \frac{\gamma_i^{(l)}}{(\lambda - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda - \sigma^2 p/n)^2} \right| \cdot \max_{1 \leq i \leq n} \left| (v_i^{(l)})^T s_i^\top \right|^2 - (\lambda^*_i + \sigma^2) \right| \lesssim \frac{\sigma^2}{\lambda^*_i^2} \sqrt{\frac{p}{n}} \cdot (\log^2 n + \sqrt{n \log n}) \asymp \frac{\sigma^2 \sqrt{\log n}}{\lambda^*_i},$$

where the second line arises from the the matrix Bernstein inequality, and the last line follows from (D.31d) and the facts $\lambda^*_i + \sigma^2 \asymp \lambda^*_i$ (given the assumption that $\sigma^2 \leq \sigma^2 p/n \ll \lambda^*_i$). Taking this together with the fact

$$\{ \lambda: \lambda/(1 + \beta(\lambda)) \in B_{\mathcal{E}_{\text{PCA}}} (\lambda^*_i + \sigma^2) \} \subseteq [2\lambda^*_i/3, 4\lambda^*_i/3],$$

can apply Lemma 20 to show that

$$|\alpha_3| \lesssim \frac{\sigma^2}{\lambda^*_i} \sqrt{p \log n} \quad \text{(D.32)}$$

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

**Controlling $\alpha_2$.** With regards to $\alpha_2$, we claim that the following upper bound holds, whose proof is deferred to the end of this section.

$$\sum_{r \leq i \leq n} \left( \frac{\lambda_i}{(\lambda_i - \lambda)^2} - \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} \right) - \sum_{r \leq i \leq n} \left( \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} \right) \lesssim \frac{\sigma^2}{\lambda^*_i} \sqrt{\frac{p}{n}}, \quad \text{(D.33)}$$

Combing this with $\mathcal{E}_{\text{PCA}}$ defined in (5.59), we arrive at

$$\alpha_2 = \sum_{r \leq i \leq n} \left( \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} \right) (\lambda^*_i + \sigma^2).$$

58
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 \( \mathcal{E}_{\text{PCA}} \); (iv) holds true under the condition \( \sigma^2 < \lambda_{\max}^* \) (see (D.25)).

Controlling \( \alpha_1 \). Regarding \( \alpha_1 \), the key step lies in controlling \( \sum_{r\leq i\leq n}(v_i^{(l)}\mathbf{T} s_{\perp})^2 \). Recall that \( U^{(l)} \sqrt{\Gamma^{(l)}} V^{(l)}\mathbf{T} \) is the SVD of \( \frac{1}{\sqrt{n}}F_{\perp} \) with \( V^{(l)} := [v_1^{(l)}, \ldots, v_n^{(l)}] \in \mathbb{R}^{n \times n} \). Towards this, we invoke Theorem 4 to derive the following identity:

\[
\lambda_i^{(l)} = \frac{1}{n} u_i^{(l)\mathbf{T}} SS^T u_i^{(l)} + \frac{1}{n} u_i^{(l)\mathbf{T}} SS^T u_{1\perp}^{(l)} \left( \lambda_i I_{p-1} - \frac{1}{n} (u_{1\perp}^{(l)}\mathbf{T}) SS^T u_{1\perp}^{(l)} \right) - \frac{1}{n} (u_{1\perp}^{(l)})^\mathbf{T} SS^T u_{1\perp}^{(l)}
\]

\[
\lambda_i^{(l)} = \frac{1}{n} ||s_{\perp}||^2 + \frac{1}{n} s_{\perp}^\mathbf{T} \left( \lambda_i I_{p-1} - \frac{1}{n} s_{\perp}^\mathbf{T} S_{\perp}^{(l)} \right) - \frac{1}{n} s_{\perp}^\mathbf{T} S_{\perp}^{(l)}
\]

\[
\lambda_i^{(l)} = \frac{1}{n} \sum_{1\leq i\leq n} (v_i^{(l)}\mathbf{T} s_{\perp})^2 + \frac{1}{n} \sum_{1\leq i\leq n} \frac{\gamma_i^{(l)}}{\lambda_i - \gamma_i^{(l)}} (v_i^{(l)}\mathbf{T} s_{\perp})^2
\]

\[
\lambda_i^{(l)} = \frac{1}{n} \sum_{1\leq i\leq n} \frac{(v_i^{(l)}\mathbf{T} s_{\perp})^2}{\lambda_i - \gamma_i^{(l)}}
\]

where (i) arises from (5.4b); (ii) relies on the definitions of \( s_{\perp} \) and \( s_{\perp} \) in (D.15); (iii) follows since \( \{v_i^{(l)}\}_{1\leq i\leq n} \) forms a set of orthonormal bases in \( \mathbb{R}^n \). Rearranging terms, we are left with

\[
n = \sum_{1\leq i\leq n} \frac{(v_i^{(l)}\mathbf{T} s_{\perp})^2}{\lambda_i - \gamma_i^{(l)}} = \sum_{1\leq i<r} \frac{(v_i^{(l)}\mathbf{T} s_{\perp})^2}{\lambda_i - \gamma_i^{(l)}} + \sum_{r\leq i\leq n} \frac{(v_i^{(l)}\mathbf{T} s_{\perp})^2}{\lambda_i - \gamma_i^{(l)}}
\]

59
As a result, we obtain the following decomposition:

\[
\sum_{r \leq i \leq n} \frac{(v_i^T s_i^T)^2}{\lambda_i - \sigma^2 / n} = n - \sum_{i < r \leq n} \frac{(v_i^T s_i^T)^2}{\lambda_{i+1} - \sigma^2 / n} - \sum_{r \leq i \leq n} \frac{1}{\lambda_i - \gamma_i} \sum_{r \leq i \leq n} \frac{1}{\lambda_i - \gamma_i} - \sum_{r \leq i \leq n} \frac{1}{\lambda_i - \gamma_i} \frac{1}{\lambda_i - \gamma_i} - \sum_{r \leq i \leq n} \frac{1}{\lambda_i - \gamma_i} \frac{1}{\lambda_i - \gamma_i} \frac{1}{\lambda_i - \gamma_i} \frac{1}{\lambda_i - \gamma_i}.
\]

(D.35)

In what follows, we shall control \(\varphi_1, \varphi_2, \varphi_3\) separately.

- We start with \(\varphi_1\). Given that \(v_i^T s_i^T \sim \mathcal{N}(0, \lambda_i^* + \sigma^2)\), we can develop an upper bound as follows: with probability at least \(1 - O(n^{-10})\),

\[
|\varphi_1| \leq \frac{\sum_{i < r < n} (v_i^T s_i^T)^2}{\min_i: i \neq i} \leq \frac{(\lambda_i^* + \sigma^2) r \log n}{\min_i: i \neq i} \leq \frac{\lambda_i^* r \log n}{\min_i: i \neq i}.
\]  

(D.36)

Here, (i) utilizes the condition \(|\lambda_i - \gamma_i| \geq \min_i: i \neq i |\lambda_i^* - \lambda_i^*|\) (in view of Lemma 8), (ii) holds due to (D.10), whereas (iii) arises from the noise assumption \(\sigma^2 \leq \lambda_i^*\).

- As for \(\varphi_2\), recall from Lemma 21 that \(\lambda_{i+1} \leq \gamma_i \leq \lambda_i\) for all \(1 \leq i < p\). This in turn leads to

\[
\sum_{r \leq i \leq n} \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \geq \sum_{r < i \leq n} \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \geq \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i}.
\]

This taken collectively with (D.31a) yields

\[
\left| \sum_{r \leq i \leq n} \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \right| \leq \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}}.
\]  

(D.37)

In addition, it is also seen from (D.31a) that

\[
\left| \sum_{r \leq i \leq n} \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \right| \leq \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}}.
\]  

(D.38)

Therefore, we can obtain

\[
\left| \lambda_i^* + \sigma^2 \right| \sum_{r \leq i \leq n} \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \leq \left| \lambda_i^* + \sigma^2 \right| \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \leq \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right).
\]

\[
\leq \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}} + \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}} + \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}} \leq \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}} + \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}} + \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}}.
\]

Therefore, we can obtain

\[
\left| \lambda_i^* + \sigma^2 \right| \sum_{r \leq i \leq n} \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \leq \left| \lambda_i^* + \sigma^2 \right| \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \leq \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right) \left( \frac{1}{\lambda_i - \gamma_i} - \frac{1}{\lambda_i - \gamma_i} \right).
\]
where the last step uses (D.15), (D.16), (D.37) and (D.38). This reveals that
\[
\varphi_2 = \lambda_i \frac{1}{1 + \frac{1}{n} \sum_{r<i \leq p} \frac{\lambda_r}{\lambda_r - \lambda_i}} \sum_{r<i \leq n} \left( \frac{1}{\lambda_r - \lambda_i} - \frac{1}{\lambda_i - \sigma^2p/n} \right) + O\left( \frac{\sigma^2}{\lambda_i^3} \sqrt{pr \log n} \right). \tag{D.39}
\]

• Turning to \( \varphi_3 \), we shall apply Lemma 20 to bound it. Similar to the analysis above for bounding \( \alpha_3 \), one can check that the following holds with probability at least \( 1 - O(n^{-10}) \): for all \( \lambda \) satisfying \( \lambda/(1 + \beta(\lambda)) \in B_{\epsilon_{\text{PCA}}} (\lambda^*_i + \sigma^2) \),
\[
\frac{d}{d\lambda} \sum_{r \leq i \leq n} \left( \frac{1}{\lambda - \gamma_i} - \frac{1}{\lambda - \sigma^2p/n} \right) \left( (\mathbf{v}_i^{(l)^\top} \mathbf{s}_{l,\|}^\top)^2 - (\lambda_i^* + \sigma^2) \right)
\]
\[
= \left| \sum_{r \leq i \leq n} \left( \frac{1}{\lambda - \gamma_i} - \frac{1}{\lambda - \sigma^2p/n} \right) \left( (\mathbf{v}_i^{(l)^\top} \mathbf{s}_{l,\|}^\top)^2 - (\lambda_i^* + \sigma^2) \right) \right|
\]
\[
\leq n \cdot \max_{r \leq i \leq n} \left| \frac{1}{\lambda - \gamma_i} - \frac{1}{\lambda - \sigma^2p/n} \right| \cdot \max_{r \leq i \leq n} \left| (\mathbf{v}_i^{(l)^\top} \mathbf{s}_{l,\|}^\top)^2 - (\lambda_i^* + \sigma^2) \right|
\]
\[
\leq n \frac{\sigma^2}{\lambda_i^3} \sqrt{\frac{p}{n}} (\lambda_i^* + \sigma^2) \log n \approx \frac{\sigma^2}{\lambda_i^2} \sqrt{\log n},
\]
where the last line comes from (D.31b). In addition, for any fixed \( \lambda \) such that \( \lambda/(1 + \beta(\lambda)) \in B_{\epsilon_{\text{PCA}}} (\lambda_i^* + \sigma^2) \), we can use the matrix Bernstein inequality [Koltchinskii, 2011, Corollary 2.1] to demonstrate that: with probability at least \( 1 - O(n^{-10}) \),
\[
\left| \frac{1}{\lambda - \gamma_i} - \frac{1}{\lambda - \sigma^2p/n} \right| \left( (\mathbf{v}_i^{(l)^\top} \mathbf{s}_{l,\|}^\top)^2 - (\lambda_i^* + \sigma^2) \right)
\]
\[
\leq \max_{r \leq i \leq n} \left| \frac{1}{\lambda - \gamma_i} - \frac{1}{\lambda - \sigma^2p/n} \right| \cdot (\lambda_i^* + \sigma^2)(\log^2 n + \sqrt{n \log n})
\]
\[
\leq \frac{\sigma^2}{\lambda_i^3} \sqrt{\log n},
\]
where the last line comes from (D.31a). With these in place, we invoke Lemma 20 to conclude that
\[
|\varphi_3| \lesssim \frac{\sigma^2}{\lambda_i^3} \sqrt{p \log n} \tag{D.40}
\]
with probability at least \( 1 - O(n^{-10}) \).

• Substituting (D.36), (D.39) and (D.40) into (D.35) reveals that: with probability exceeding \( 1 - O(n^{-10}) \),
\[
\sum_{r \leq i \leq n} (\mathbf{v}_i^{(l)^\top} \mathbf{s}_{l,\|}^\top)^2 = n \frac{\lambda_i}{1 + \frac{1}{n} \sum_{r<i \leq n} \frac{\lambda_r}{\lambda_r - \lambda_i}} \sum_{r<i \leq n} \left( \frac{1}{\lambda_r - \lambda_i} - \frac{1}{\lambda_i - \sigma^2p/n} \right)
\]
\[
+ O\left( \frac{\sigma^2}{\lambda_i^3} \sqrt{pr \log n} \right) + \frac{\lambda_i^* \log n}{\min_{i: i \neq l} |\lambda_i^* - \lambda_i^*|}, \tag{D.41}
\]
As a consequence, we arrive at
\[
\alpha_1 = \frac{\sigma^2p/n}{\lambda_i - \sigma^2p/n} \sum_{r \leq i \leq n} (\mathbf{v}_i^{(l)^\top} \mathbf{s}_{l,\|}^\top)^2
\]
\[
= \frac{\sigma^2p/n}{\lambda_i - \sigma^2p/n} \left( n - \frac{1}{1 + \frac{1}{n} \sum_{r<i \leq n} \frac{\lambda_r}{\lambda_r - \lambda_i}} \sum_{r<i \leq n} \left( \frac{1}{\lambda_r - \lambda_i} - \frac{1}{\lambda_i - \sigma^2p/n} \right) \right)
\]
\[
+ o\left( \frac{\sigma^2}{\lambda_i^3} \sqrt{pr \log n} \right) + O\left( \frac{\sigma^2pr \log n}{\min_{i: i \neq l} |\lambda_i^* - \lambda_i^*| n} \right), \tag{D.42}
\]
where we have made use of the bound \( \lambda_i - \sigma^2p/n \gtrsim \lambda_i^* \) and \( \sigma^2p/n = o(\lambda_i^*) \).
Combining the bounds on $\alpha_1$, $\alpha_2$ and $\alpha_3$. Putting (D.30), (D.32), (D.34) and (D.42) together, we conclude

\[
\sum_{r \leq i < n} \frac{\gamma_i^{(l)} (v_i^{(l)} s_i)^T}{(\lambda_i - \gamma_i^{(l)})^2} \leq \alpha_1 + \alpha_2 + \alpha_3
\]

\[
= \frac{\sigma^2 p/n}{\lambda_i - \sigma^2 p/n} \left( n - \frac{1}{1 + \frac{1}{p} \sum_{r < i \leq n} \frac{\lambda_i}{\lambda_i - \lambda_r} \sum_{r < i \leq n} \left( \frac{1}{\lambda_i - \lambda_r} - \frac{1}{\lambda_i - \sigma^2 p/n} \right) \right)
\]

\[
+ \frac{\lambda_i}{1 + \frac{1}{p} \sum_{r < i \leq n} \frac{\lambda_i}{\lambda_i - \lambda_r} \sum_{r < i \leq n} \left( \frac{1}{\lambda_i - \lambda_r} - \frac{1}{\lambda_i - \sigma^2 p/n} \right) + O\left( \frac{\sigma^2 p}{\lambda_i^2} \sqrt{p} \log n + \frac{\sigma^2 p r \log n}{\min_{i \neq l} |\lambda_i^* - \lambda_l^*| n} \right)
\]

\[
= \frac{\sigma^2 p}{\lambda_i - \sigma^2 p/n} + \frac{\lambda_i}{1 + \frac{1}{p} \sum_{r < i \leq n} \frac{\lambda_i}{\lambda_i - \lambda_r} \sum_{r < i \leq n} \left( \frac{1}{\lambda_i - \lambda_r} - \frac{1}{\lambda_i - \sigma^2 p/n} \right) + O\left( \frac{\sigma^2 p}{\lambda_i^2} \sqrt{p} \log n + \frac{\sigma^2 p r \log n}{\min_{i \neq l} |\lambda_i^* - \lambda_l^*| n} \right)
\]

as claimed.

**Proof of the inequality (D.33).** Given that $\lambda_{i+1} \leq \gamma_i^{(l)} \leq \lambda_i$ for all $1 \leq i < p$, we can bound

\[
\sum_{r < i \leq n} \frac{\lambda_i}{(\lambda_i - \lambda_r)^2} + \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} \leq \sum_{r < i \leq n} \frac{(\lambda_i^2 - \gamma_i^{(l)}^2)}{(\lambda_i - \gamma_i^{(l)})^2} \leq \sum_{r < i \leq n} \frac{\lambda_i}{(\lambda_i - \lambda_r)^2} + \frac{\gamma_r^{(l)}}{(\lambda_r - \gamma_r^{(l)})^2}.
\]

By subtracting $\sum_{r < i \leq n} \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2}$ from both sides and rearranging terms, we have

\[
\left| \sum_{r < i \leq n} \left( \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} \right) - \sum_{r < i \leq n} \left( \frac{\lambda_i}{(\lambda_i - \lambda_r)^2} - \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} \right) \right|
\]

\[
\leq \left| \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} \right| + \left| \frac{\gamma_r^{(l)}}{(\lambda_r - \gamma_r^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_r - \sigma^2 p/n)^2} \right|
\]

In view of the basic property $|f(x) - f(y)| \leq \{\sup_{z} |f'(z)|\} |x - y|$, we can upper bound

\[
\left| \frac{\gamma_i^{(l)}}{(\lambda_i - \gamma_i^{(l)})^2} - \frac{\sigma^2 p/n}{(\lambda_i - \sigma^2 p/n)^2} \right| \leq \max_{\gamma \in \gamma_i^{(l)} - \sigma^2 p/n \leq \lambda_i} \left| \frac{\lambda_i + \gamma}{(\lambda_i - \gamma)^2} \cdot |\gamma_i^{(l)} - \sigma^2 p/n| \right|
\]

\[
\leq \frac{\lambda_i^2}{\lambda_i^4} \sigma^2 \sqrt{\frac{p}{n}} = \frac{\sigma^2}{\lambda_i^2} \sqrt{\frac{p}{n}}
\]

for any $r \leq i \leq n$. Here, the last line holds because (i) $|\lambda_i - \lambda_r| \leq \sigma^2 (p/n + \sqrt{p/n}) \leq \lambda_r^*$ holds due to the assumption (3.11a), and hence $|\lambda_i - \gamma| \geq \lambda_r^*$ and $\lambda_i + \gamma \leq \lambda_i^*$; (ii) $|\gamma_i^{(l)} - \sigma^2 p/n| \leq \sigma^2 \sqrt{p/n}$ 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 $1 - O(n^{-20})$:

\[
V := \frac{1}{n} \sup_{\lambda_1, \ldots, \lambda_n \in \mathbb{E} \cap (\lambda_1^* + \sigma^2)} \left\| \sum_{k: k \neq l} a_k^* u_k^{(l)^T} (\lambda I_p - \frac{1}{n} S_{l, l} S_{l, l}^T)^{-1} \frac{1}{n} S_{l, l} \right\|_2
\]

62
\[ \sum_{k:k \neq l} |a_k^t u_k^\perp| \sqrt{\frac{\lambda_i^* - \lambda_k^*}{\lambda_i^* - \lambda_k^*}} \sqrt{\frac{(\lambda_{\max}^* + \sigma^2)(k^2 + r)}{n}}. \]  

(D.43)

Consequently, we can apply the Gaussian concentration inequality to show that: for any fixed \( \lambda/(1 + \beta(\lambda)) \in \mathcal{B}_{\varepsilon_{PCA}}(\lambda_i^* + \sigma^2) \), one has

\[
\left| \sum_{k:k \neq l} a_k^t u_k^\perp u_k^{(t)T} \left( \lambda I_{p-1} - \frac{1}{n} S_{l,\perp} S_{l,\perp}^T \right)^{-1} \frac{1}{n} S_{l,\perp} \cdot s_{l,\perp} \right| \lesssim \sqrt{\frac{(\lambda_i^* + \sigma^2)k}{n}} \log \left( \frac{n \kappa \lambda_{\max}^*}{\Delta_i^*} \right) \cdot V
\]

with probability exceeding \( 1 - O(\kappa^{-10}(\lambda_{\max}^*/\Delta_i^*)^{-20}n^{-20}) \).

In addition, one can derive: for all \( \lambda \) such that \( \lambda/(1 + \beta(\lambda)) \in \mathcal{B}_{\varepsilon_{PCA}}(\lambda_i^* + \sigma^2) \),

\[
\left| \sum_{k:k \neq l} a_k^t u_k^\perp u_k^{T} \left( \lambda I_{p-1} - \frac{1}{n} S_{l,\perp} S_{l,\perp}^T \right)^{-1} \frac{1}{n} S_{l,\perp} \cdot s_{l,\perp} \right| \lesssim \sqrt{\frac{(\lambda_i^* + \sigma^2)k}{n}} \log \left( \frac{n \kappa \lambda_{\max}^*}{\Delta_i^*} \right) \cdot V
\]

holds with probability at least \( 1 - O(n^{-20}) \). Here, (i) uses Lemma 8 and the high-probability fact that \( \|s_{l,\perp}\|_2 \lesssim (\lambda_i^* + \sigma^2)\sqrt{n \log n} \), (ii) holds since \( \|u_k^\perp u_i^\perp\|_2 \leq \|u_k^\perp\|_2 \cdot \|u_i^\perp\|_2 \leq 1 \), 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

\[
\left\| \sum_{k:k \neq l} a_k^t u_k^\perp u_k^{(t)T} \left( \lambda I_{p-1} - \frac{1}{n} S_{l,\perp} S_{l,\perp}^T \right)^{-1} \frac{1}{n} S_{l,\perp} \cdot s_{l,\perp} \right\|_2 \lesssim \sqrt{\frac{(\lambda_i^* + \sigma^2)k}{n}} \log \left( \frac{n \kappa \lambda_{\max}^*}{\Delta_i^*} \right) \cdot V
\]

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 \( \frac{1}{\sqrt{n}} S_{l,\perp} = U^{(l)} \sqrt{\Gamma^{(l)}} V^{(l)T} \). Similar to (C.20), any \( \lambda \) such that \( \lambda/(1 + \beta(\lambda)) \in \mathcal{B}_{\varepsilon_{PCA}}(\lambda_i^* + \sigma^2) \), one can rewrite

\[
\left\| \sum_{k:k \neq l} a_k^t u_k^\perp u_k^{(t)T} \left( \lambda I_{p-1} - \frac{1}{n} S_{l,\perp} S_{l,\perp}^T \right)^{-1} \frac{1}{\sqrt{n}} S_{l,\perp} \right\|_2
\]

\[
= \left\| \sum_{k:k \neq l} a_k^t u_k^\perp u_k^{(t)T} U^{(l)} (\lambda I_{p-1} - \Gamma^{(l)})^{-1} U^{(l)T} \sqrt{\Gamma^{(l)}} V^{(l)T} \right\|_2
\]

63
\[
\sum_{k,k \neq l} a_k^T u_k^* u_k^*(l)^T U(l) (\lambda_{l,p-1} - \Gamma(l))^{-1} \sqrt{\Gamma(l)}
\]
\[
= \sqrt{\sum_{i \leq n \wedge (p-1)} \frac{\gamma_i(l)}{(\lambda_i - \gamma_i(l))^2} \left( \sum_{k,k \neq l} a_k^T u_k^* u_k^*(l)^T u_i(l)^2 \right)}.
\]

• With regards to the sum over the range \(i \geq r\), it is seen from Lemma 8 and the assumption (3.11a) that for all \(i \geq r\), \(\gamma_i(l) \leq \sigma^2(1 + p/n) \leq \lambda^*_\max + \sigma^2\) and \(|\lambda_i - \gamma_i(l)| \geq \lambda_i^* \geq |\lambda_i^* - \lambda_k^*|/\kappa\) for any \(k \neq i\). This enables us to derive

\[
\sum_{r \leq n \wedge (p-1)} \frac{\gamma_i(l)}{(\lambda_i - \gamma_i(l))^2} \left( \sum_{k,k \neq l} a_k^T u_k^* u_k^*(l)^T u_i(l)^2 \right)
\]
\[
\leq \sqrt{\left(\lambda^*_\max + \sigma^2\right) \sum_{r \leq n \wedge (p-1)} \left( \sum_{k,k \neq l} \frac{\kappa}{\lambda_k^* - \lambda_i^*} a_k^T u_k^* u_k^*(l)^T u_i(l)^2 \right)}
\]
\[
\leq \sqrt{\lambda^*_\max + \sigma^2} \left\| \sum_{k,k \neq l} \frac{\kappa}{\lambda_k^* - \lambda_i^*} a_k^T u_k^* u_k^*(l)^T U(l) \right\|
\]
\[
\leq \sqrt{\lambda^*_\max + \sigma^2} \kappa_2 \left( \sum_{k,k \neq l} \frac{a_k^T u_k^*}{\lambda_k^* - \lambda_i^*} \left\| u_k^*(l)^T U(l) \right\|_2 \right)
\]
\[
\leq \sqrt{\lambda^*_\max + \sigma^2} \kappa_2 \left( \sum_{k,k \neq l} \frac{a_k^T u_k^*}{\lambda_k^* - \lambda_i^*} \right)
\]

where the last line holds since \(\left\| u_k^*(l)^T U(l) \right\|_2 \leq \left\| u_k^*(l) \right\|_2 \left\| U(l) \right\| \leq 1\).

• Turning to the sum over the range \(i < r\), we can control

\[
\sum_{1 \leq i \leq r} \frac{\gamma_i(l)}{(\lambda_i - \gamma_i(l))^2} \left( \sum_{k,k \neq l} a_k^T u_k^* u_k^*(l)^T u_i(l)^2 \right)
\]
\[
= \sum_{1 \leq i \leq r} \frac{(\gamma_i(l))^{1/2}}{\lambda_i - \gamma_i(l)} \sum_{k,k \neq l} a_k^T u_k^* u_k^*(l)^T u_i(l)^2 \left\| u_i(l)^2 \right\|_2
\]
\[
= \sum_{k,k \neq l} \left\| a_k^T u_k^* \right\| \sum_{1 \leq i \leq r} \frac{(\gamma_i(l))^{1/2} u_k^*(l)^T u_i(l)^2}{\lambda_i - \gamma_i(l)} \left\| u_i(l)^2 \right\|_2
\]
\[
= \sum_{k,k \neq l} \left\| a_k^T u_k^* \right\| \sum_{1 \leq i \leq r} \frac{(\gamma_i(l))^{1/2} u_k^*(l)^T u_i(l)^2}{\lambda_i - \gamma_i(l)} \left\| u_i(l)^2 \right\|_2.
\]

This leads us to control \(\sum_{1 \leq i \leq r-1} \frac{\gamma_i(l)}{(\lambda_i - \gamma_i(l))^2} (u_k^*(l)^T u_i(l)^2) / (\lambda_i - \gamma_i(l))^2\) for each \(k \neq l\), which can be decomposed as follows

\[
\sum_{1 \leq i \leq r-1} \frac{\gamma_i(l)}{(\lambda_i - \gamma_i(l))^2} (u_k^*(l)^T u_i(l)^2) / (\lambda_i - \gamma_i(l))^2 = \sum_{i \in C_1} \frac{\gamma_i(l)}{(\lambda_i - \gamma_i(l))^2} (u_k^*(l)^T u_i(l)^2) / (\lambda_i - \gamma_i(l))^2 + \sum_{i \in C_2} \frac{\gamma_i(l)}{(\lambda_i - \gamma_i(l))^2} (u_k^*(l)^T u_i(l)^2) / (\lambda_i - \gamma_i(l))^2.
\]

Here, the sets \(C_1\) and \(C_2\) are defined respectively as follows

\[
C_1 := \{1 \leq i \leq r \mid \gamma_i(l) / (1 + \gamma_i(l)) \in \mathcal{B}_{\varepsilon_k}(\lambda_i^*)\},
\]
\[
C_2 := \{1 \leq i \leq r \mid \gamma_i(l) / (1 + \gamma_i(l)) \in \mathcal{B}_{\varepsilon_k}(\lambda_i^*)\},
\]

where we take \(\varepsilon_k := c |\lambda_i^* - \lambda_k^*|\) for some sufficiently small constant \(c > 0\). In the sequel, we shall control the above two sums separately.
With respect to the sum over $C_1$, one can apply a similar argument for (C.23) to show $|\lambda_i - \gamma_i^{(l)}| \gtrsim |\lambda_i^* - \lambda_k^*|$ for $i \in C_1$. This enables us to bound

$$\sum_{i \in C_1} \gamma_i^{(l)} \left( \frac{\lambda_k^{* \top} u_i^{(l)}}{\lambda_i - \gamma_i^{(l)}} \right)^2 \leq \frac{\lambda_i^{* \max} + \sigma^2}{|\lambda_i^* - \lambda_k^*|^2} \sum_{i \in C_1} \left( \frac{\lambda_k^{* \top} u_i^{(l)}}{\lambda_i - \gamma_i^{(l)}} \right)^2 \leq \frac{\lambda_i^{* \max} + \sigma^2}{|\lambda_i^* - \lambda_k^*|^2} \|U_i^{(l)}\|^2 \leq \lambda_i^{* \max} + \sigma^2.
$$

Next, we move on to look at the sum over $C_2$. According to Lemma 8, we have

$$E_{\text{PCA}} \gtrsim \left\| (\gamma_i^{(l)} I_{r-1} - (1 + \beta(\gamma_i^{(l)}))\Lambda_i^{(l)}) U_i^{(l)} \right\|_2 \geq |\gamma_i^{(l)} - (1 + \beta(\gamma_i^{(l)}))\lambda_k^{(l)}| \cdot |u_k^{(l)}\|_2 \gtrsim E_k \cdot |u_k^{(l)}\|_2 \geq |\lambda_k^{*} - \lambda_k^*| \cdot |u_k^{(l)}\|_2,
$$

where we use the fact that $|u_k^{(l)}\|_2 \leq |u_k^{(l)}\|_2$ and $|\gamma_i^{(l)} - (1 + \beta(\gamma_i^{(l)}))\lambda_k^{(l)}| \gtrsim E_k$ for all $i \in C_2$. Therefore, we arrive at the upper bound

$$\frac{\gamma_i^{(l)} (u_k^{(l)}\top u_i^{(l)})^2}{(\lambda_i - \gamma_i^{(l)})^2} \lesssim \frac{(\lambda_i^{* \max} + \sigma^2)^2 E_{\text{PCA}}}{|\lambda_i^* - \lambda_k^*|^2 \min_{i \neq i} |\lambda_i^* - \lambda_k^*|^2} \lesssim \frac{\lambda_i^{* \max} + \sigma^2}{|\lambda_i^* - \lambda_k^*|^2}, \quad i \in C_2,
$$

where we invoke the condition $\min_{i \neq i} |\lambda_i^* - \lambda_k^*| \gtrsim E_{\text{PCA}}$. Taking these two bounds collectively, we reach

$$\sum_{1 \leq i < r} \gamma_i^{(l)} \left( \frac{\lambda_k^{* \top} u_i^{(l)}}{\lambda_i - \gamma_i^{(l)}} \right)^2 \lesssim \frac{\lambda_i^{* \max} + \sigma^2}{|\lambda_i^* - \lambda_k^*|^2} + \frac{(\lambda_i^{* \max} + \sigma^2)^2}{|\lambda_i^* - \lambda_k^*|^2} \geq \frac{(\lambda_i^{* \max} + \sigma^2)^2}{|\lambda_i^* - \lambda_k^*|^2} r,
$$

and hence

$$\sqrt{\sum_{1 \leq i < r} \gamma_i^{(l)} \left( \frac{\lambda_k^{* \top} u_i^{(l)}}{\lambda_i - \gamma_i^{(l)}} \right)^2} \lesssim \sqrt{(\lambda_i^{* \max} + \sigma^2)^2} \sum_{k \neq k} |a^\top u_k^*| \prod \left| \frac{a^\top u_k^*}{|\lambda_i^* - \lambda_k^*|^2} \right|.
$$

Combining the preceding two bounds, we finish the proof for (D.43).

### E Proof for minimax lower bounds (Theorem 3)

Fix an arbitrary $1 \leq l \leq r$ and an arbitrary $k \neq l$ and $1 \leq k \leq r$. In what follows, we intend to prove the following two claims:

\[ \inf_{u_{a,l}, \Sigma \in M_1} \sup_{\Sigma \in M_1} \mathbb{E} \left[ \min \left| u_{a,l}^\top u_{l}(\Sigma) \right| \right] \gtrsim \frac{(\lambda_k^* + \sigma^2)(\lambda_i^* + \sigma^2)}{|\lambda_i^* - \lambda_k^*|^2 n} |a^\top u_i^*| + \frac{\sqrt{(\lambda_k^* + \sigma^2)(\lambda_i^* + \sigma^2)}}{|\lambda_i^* - \lambda_k^*|} |a^\top u_k^*|,
\]

(E.1)

\[ \inf_{u_{a,l}, \Sigma \in M_2} \sup_{\Sigma \in M_2} \mathbb{E} \left[ \min \left| u_{a,l}^\top u_{l}(\Sigma) \right| \right] \gtrsim \frac{(\lambda_i^* + \sigma^2)^2}{\lambda_i^* n} \|P_{U^\perp} a\|_2,
\]

(E.2)

where the infimum is over all estimators, and $M_1(\Sigma^*)$ and $M_2(\Sigma^*)$ are defined right before the statement of Theorem 3. It is self-evident that Theorem 3 follows from these two claims by taking the maximum over all $k \neq l$. 

65
E.1 Proof of the lower bound (E.1)

Step 1: constructing a collection of hypotheses. Let us consider the following two hypotheses:

\[
\mathcal{H}_0 : s_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma^*) + \sigma I_p, \quad 1 \leq i \leq n; \\
\mathcal{H}_k : s_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma_k + \sigma^2 I_p), \quad 1 \leq i \leq n.
\]

Here, the covariance matrix \(\Sigma_k\) is defined as follows:

\[
\Sigma_k := \lambda_1^* u_1 u_1^T + \lambda_k^* u_k u_k^T + \sum_{i \neq k, l, 1 \leq i \leq r} \lambda_i^* u_i u_i^T.
\]

where \(u_i\) and \(u_k\) are defined as

\[
[u_i, u_k] := [u_i^*, u_k^*] \begin{bmatrix} \cos \theta_n & -\sin \theta_n \\ \sin \theta_n & \cos \theta_n \end{bmatrix}
\]

for some \(\theta_n \in [-\pi/2, \pi/2]\) to be specified later. Straightforward calculation yields

\[
u_i u_i^T + u_k u_k^T = u_i^* u_i^T + u_k^* u_k^T.
\]

This identity further leads to

\[
\Sigma_k - \Sigma^* = \lambda_1^* u_1 u_1^T + \lambda_k^* u_k u_k^T - (\lambda_1^* u_1^* u_1^T + \lambda_k^* u_k^* u_k^T)
\]

\[
= \lambda_1^* (u_1 u_1^T + u_k u_k^T) + (\lambda_k^* - \lambda_1^*) u_k u_k^T - (\lambda_1^* u_1^* u_1^T + u_k^* u_k^T) + (\lambda_k^* - \lambda_1^*) u_k^* u_k^T)
\]

\[
= (\lambda_k^* - \lambda_1^*) (u_k u_k^T - u_k^* u_k^T).
\]

In addition, it is also seen that

\[
\|\Sigma_k - \Sigma^*\|_F = |\lambda_k^* - \lambda_1^*| \cdot \|u_k u_k^T - u_k^* u_k^T\|_F \leq |\lambda_k^* - \lambda_1^*| \cdot (\|u_k (u_k - u_k^*)^T\|_F + \|(u_k - u_k^*) u_k^T\|_F)
\]

\[
= |\lambda_k^* - \lambda_1^*| \cdot (\|u_k\|_2 \|u_k - u_k^*\|_2 + \|u_k - u_k^*\|_2 \|u_k\|_2)
\]

\[
= 2 |\lambda_k^* - \lambda_1^*| \cdot \|u_k - u_k^*\|_2
\]

\[
\leq 2 |\lambda_k^* - \lambda_1^*| \cdot (\|u_k\|_2 + \|u_k - u_k^*\|_2)
\]

\[
\leq 2 |\lambda_k^* - \lambda_1^*| \cdot (\|u_k\|_2 + 2 \|u_k - u_k^*\|_2)
\]

\[
\leq 4 |\lambda_k^* - \lambda_1^*| \cdot |\theta_n|.
\]

where (i) arises from the definition of \(u_i\) in (E.3); (ii) holds since \(\sin \theta \leq |\theta|\).

In what follows, we denote by \(P_0^H\) and \(P_k^H\) the distribution of \(S\) under the hypothesis \(\mathcal{H}_0\) and \(\mathcal{H}_k\), respectively, and let \(P_0^k\) and \(P_k^k\) denote the distribution of \(s_i\) (i-th column of \(S\)) 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]

\[
\text{KL}(\mathcal{N}(0, \Sigma_1) \mid \mid \mathcal{N}(0, \Sigma_0)) = \frac{1}{2} \left( \text{tr}(\Sigma_0^{-1} \Sigma_1) - p + \log \frac{\Sigma_0}{\Sigma_1} \right).
\]

Since the KL divergence is additive over independent distributions [Tsybakov, 2009], one has

\[
\text{KL}(P_k^k \mid \mid P_0^k) = \sum_{i=1}^n \text{KL}(P_k^k \mid \mid P_0^0) = \frac{1}{2} \sum_{i=1}^n \left( \text{tr}(\Sigma_k + \sigma^2 I_p)^{-1} \Sigma_k + \sigma^2 I_p) - p \right).
\]

This suggests that we need to compute \(\text{tr}(\Sigma_k + \sigma^2 I_p)^{-1} \Sigma_k + \sigma^2 I_p)\). By construction in (E.3), we know that \(u_i\) and \(u_k\) span the same subspace as \(u_i^*\) and \(u_k^*\), and are orthogonal to \(\{u_i^*\}_{i \neq k, l}\). Denote by
\( U^* \perp \in \mathbb{R}^{p \times (p-r)} \) the matrix whose columns form an orthonormal basis of the complement to the subspace spanned by \( U^* \). One can then derive

\[
(\Sigma^* + \sigma^2 I_p)^{-1}(\Sigma_k + \sigma^2 I_p) = \left( \sum_{1 \leq i \leq r} \frac{1}{\lambda_i^* + \sigma^2} u_i^* u_i^{*T} + \frac{1}{\sigma^2} U^{*\perp}(U^{*\perp})^T \right) 
\times \left( (\lambda_i^* + \sigma^2) u_i^* u_i^{*T} + \sum_{i \neq k, l, 1 \leq i \leq r} (\lambda_i^* + \sigma^2) u_i^* u_i^{*T} + \sigma^2 U^{*\perp}(U^{*\perp})^T \right)
\]

As a result, we find

\[
\text{tr}((\Sigma^* + \sigma^2 I_p)^{-1}(\Sigma_k + \sigma^2 I_p)) \]

\[
\overset{(i)}{=} \text{tr} \left( \left( \frac{1}{\lambda_i^* + \sigma^2} u_i^* u_i^{*T} + \frac{1}{\lambda_k^* + \sigma^2} u_k^* u_k^{*T} \right) \left( (\lambda_i^* + \sigma^2) u_i^* u_i^{*T} + (\lambda_k^* + \sigma^2) u_k^* u_k^{*T} \right) \right) + p - 2
\]

\[
= |u_i^T u_i|^2 + \frac{\lambda_i^* + \sigma^2}{\lambda_i^* + \sigma^2} |u_i^T u_k|^2 + |u_k^T u_k|^2 + |u_k^T u_i|^2 + |u_i^T u_k|^2 + |u_k^T u_i|^2 + |u_i^T u_k|^2 + |u_k^T u_i|^2 + p - 2
\]

\[
\overset{(ii)}{=} \cos^2 \theta_n + \frac{\lambda_i^* + \sigma^2}{\lambda_i^* + \sigma^2} \sin^2 \theta_n + \frac{\lambda_i^* + \sigma^2}{\lambda_k^* + \sigma^2} \sin^2 \theta_n + \cos^2 \theta_n + p - 2
\]

\[
= (\lambda_i^* + \sigma^2)^2 + (\lambda_k^* + \sigma^2)^2 \sin^2 \theta_n - 2 \sin^2 \theta_n + p
\]

\[
= \frac{(\lambda_i^* - \lambda_k^*)^2}{(\lambda_i^* + \sigma^2)(\lambda_k^* + \sigma^2)} \sin^2 \theta_n + p.
\]

Here, (i) holds since \( \text{tr}(u_i^* u_i^{*T}) = 1 \) and \( \text{tr}(U^{*\perp}(U^{*\perp})^T) = \text{tr}((U^{*\perp})^T U^{*\perp}) = \text{tr}(I_{p-r}) = p - r \); (ii) follows from the following observations:

\[
u_i^T u_i = u_i^* u_i \cos \theta_n + u_i^* u_k \sin \theta_n = \cos \theta_n;
\]

\[
u_i^T u_k = u_i^* u_k \cos \theta_n + u_k^* u_k \sin \theta_n = \sin \theta_n;
\]

\[
u_i^T u_k = -u_i^* u_k \sin \theta_n + u_i^* u_i \cos \theta_n = -\sin \theta_n;
\]

\[
u_k^T u_i = -u_k^* u_i \sin \theta_n + u_k^* u_k \cos \theta_n = \cos \theta_n;
\]

where we have used the construction (E.3) and the fact that \( u_i^* u_k^* = 0 \). Therefore, combining the above identities allows us to conclude that

\[
\text{KL}(p^k \parallel p^0) = \frac{n(\lambda_i^* - \lambda_k^*)^2}{2(\lambda_i^* + \sigma^2)(\lambda_k^* + \sigma^2)} \sin^2 \theta_n.
\]

**Step 3: invoking Fano’s inequality.** Suppose that we choose \( \theta_n \)

\[
|\theta_n| = c_n \sqrt{\frac{(\lambda_i^* + \sigma^2)(\lambda_k^* + \sigma^2)}{(\lambda_i^* - \lambda_k^*)^2 n}}
\]

where \( c_n \approx 1 \) 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

\[
\|\Sigma_k - \Sigma^*\|_F \leq \sqrt{\frac{(\lambda_i^* + \sigma^2)(\lambda_k^* + \sigma^2)}{n}}.
\]
In other words, \( \Sigma_k \in \mathcal{M}_1(\Sigma^*) \). Moreover, plugging the value (E.7) of \( \theta_n \) into (E.6) and using the facts \( |\sin \theta| \leq |\theta| \) as well as \( \max_n c_n = 1/4 \) yields

\[
\text{KL}(\mathbb{P}_k \parallel \mathbb{P}^0) \leq 1/16.
\]

It then follows from Fano’s inequality [Tsybakov, 2009, Theorem 2] that

\[
p_{c,k} := \inf_{\psi} \max \left\{ \mathbb{P}\{\psi \text{ rejects } \mathcal{H}_0 \mid \mathcal{H}_0\}, \mathbb{P}\{\psi \text{ rejects } \mathcal{H}_k \mid \mathcal{H}_k\} \right\} \geq 1/5,
\]

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

\[
\inf_{u_n} \sup_{\Sigma \in \mathcal{M}_1(\Sigma^*)} \mathbb{E}\left[ \min |u_{a,l} \pm a^\top u_l(\Sigma)| \right] \geq p_{c,k} \min |a^\top u_l \pm a^\top u_l^*| \geq \min |a^\top u_l \pm a^\top u_l^*|.
\]

Observe that once we prove

\[
\min |a^\top u_l \pm a^\top u_l^*| \geq \frac{1}{8\pi^2} (\theta_n^2 \cdot |a^\top u_l^*| + |\theta_n| \cdot |a^\top u_l^*|), \tag{E.8}
\]

then (E.7) would immediately lead to the advertised bound

\[
\inf_{u_n} \sup_{\Sigma \in \mathcal{M}_1(\Sigma^*)} \mathbb{E}\left[ \min |u_{a,l} \pm a^\top u_l(\Sigma)| \right] \geq c_n \frac{(\lambda_n^* + \sigma^2)(\lambda_n^* + \sigma^2)|a^\top u_l^*|}{(\lambda_n^* - \lambda_k^*)^2 n} + c_n \frac{\sqrt{(\lambda_n^* + \sigma^2)(\lambda_n^* + \sigma^2)|a^\top u_k^*|}}{|\lambda_n^* - \lambda_k^*| \sqrt{n}} \geq \frac{(\lambda_n^* + \sigma^2)(\lambda_n^* + \sigma^2)|a^\top u_l^*|}{(\lambda_n^* - \lambda_k^*)^2 n} + \frac{\sqrt{(\lambda_n^* + \sigma^2)(\lambda_n^* + \sigma^2)|a^\top u_k^*|}}{|\lambda_n^* - \lambda_k^*| \sqrt{n}}
\]

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

\[
|a^\top u_l^* - a^\top u_l| = |a^\top u_l^* - a^\top (u_l^* \cos \theta_n + u_k^* \sin \theta_n)| = |a^\top u_l^*(1 - \cos \theta_n) - a^\top u_k^* \sin \theta_n| = |2a^\top u_l^* \cos^2(\theta_n/2) - a^\top u_k^* \sin \theta_n| = \frac{(i)}{2}(2a^\top u_l^* \sin^2(\theta_n/2) + |a^\top u_k^* \sin \theta_n|) \geq \frac{2}{\pi^2} (\theta_n^2 \cdot |a^\top u_l^*| + |\theta_n| \cdot |a^\top u_k^*|) \tag{E.9}
\]

where (i) holds true as long as we choose \( \text{sign}(\theta_n) = -\text{sign}(a^\top u_l^*/a^\top u_k^*) \); (ii) relies on the fact \( |\sin \theta_2| \geq \frac{2}{\pi} |\theta_2| \) for \( \theta_2 \in [-\pi, \pi] \). In addition, we can derive

\[
|a^\top u_l^* + a^\top u_l| = |a^\top u_l^* + a^\top (u_l^* \cos \theta_n + u_k^* \sin \theta_n)| = |a^\top u_l^*(1 + \cos \theta_n) + a^\top u_k^* \sin \theta_n| = |2a^\top u_l^* \cos^2(\theta_n/2) + a^\top u_k^* \sin(\theta_n/2) \cos(\theta_n/2)| = \cos(\theta_n/2)|2a^\top u_l^* \cos(\theta_n/2) + a^\top u_k^* \sin(\theta_n/2)| \geq \frac{1}{2} |a^\top u_l^* \cos(\theta_n/2) + a^\top u_k^* \sin(\theta_n/2)| \geq \frac{1}{2} \frac{1}{4} |a^\top u_l^* \sin(\theta_n/2 + \omega_k)| \tag{E.10}
\]

where (i) holds due to \( |\theta_n| \leq 1/4 \) by the choice of \( c_n \) in (E.7) and the sample size condition (3.16); \( \omega_k \in [-\frac{\pi}{2}, \frac{\pi}{2}] \) in (ii) is defined such that \( \tan \omega_k = a^\top u_l^*/a^\top u_k^* \). In particular, recall that the sign of \( \theta_n \) is chosen such
that \( \text{sign}(\theta_n) = -\text{sign}(\mathbf{a}^T \mathbf{u}_i^* / \mathbf{a}^T \mathbf{u}_i) \), one has \( \text{sign}(\theta_n) = -\text{sign}(\omega_k) \). Next, our goal is to show if we choose \( c_n \in \{1/64, 1/16, 1/4\} \) of \( \theta_n \) in (E.7) suitably, one has
\[
\sin |\theta_n/2 + \omega_k| \geq \frac{1}{2\pi} |\theta_n|.
\] (E.11)

To this end, for each \( n \), we choose \( c_n \) of \( \theta_n \) 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 \( \theta_n \) and \( \omega_k \) are different. Combined with the inequality \( |\sin \theta| \geq \frac{\theta}{2\theta} \) for \( \theta \in [-\pi, \pi] \), 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 \( \pi/2 \geq |\theta_n/2 + \omega_k| \geq |\omega_k| - |\theta_n|/2 \geq |\omega_k|/2 \). This implies that \( \sin |\theta_n/2 + \omega_k| \geq \sin |\theta_n/2| \geq \frac{1}{2\pi} |\theta_n| \).

- Otherwise, (i.e. \( |\omega_k|/2 < |\theta_n|/2 < 2 |\omega_k| \)), one can adjust \( c_n \) to be either 1/4 or 1/16 (namely, increasing it or decreasing it by 4 times). After doing so, it is easily seen that \( \theta_n \) 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
\[
|\mathbf{a}^T \mathbf{u}_i^* + \mathbf{a}^T \mathbf{u}_i| \geq \frac{1}{8\pi} \left( |\theta_n| \cdot |\mathbf{a}^T \mathbf{u}_i^*| + |\theta_n| \cdot |\mathbf{a}^T \mathbf{u}_i^*| \right) \geq \frac{1}{8\pi^2} \left( \theta_n^2 \cdot |\mathbf{a}^T \mathbf{u}_i^*| + |\theta_n| \cdot |\mathbf{a}^T \mathbf{u}_i^*| \right),
\]
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:

\[
\mathcal{H}_0 : \mathbf{s}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma^* + \sigma^2 I_p), \quad 1 \leq i \leq n;
\]
\[
\mathcal{H}_1 : \mathbf{s}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma + \sigma^2 I_p), \quad 1 \leq i \leq n.
\]

Here, the covariance matrix \( \Sigma \) is defined to be
\[
\Sigma := \lambda_i^* \mathbf{u}_l^T \mathbf{u}_l^* + \sum_{i : i \neq l} \lambda_i^* \mathbf{u}_i^* \mathbf{u}_i^T,
\]
where \( \mathbf{u}_l \) is defined as
\[
\mathbf{u}_l := \frac{\mathbf{u}_i^* + \delta_n \mathbf{a}_\perp}{\|\mathbf{u}_i^* + \delta_n \mathbf{a}_\perp\|_2} = \frac{\mathbf{u}_i^* + \delta_n \mathbf{a}_\perp}{\sqrt{1 + \delta_n^2}} \quad \text{with} \quad \mathbf{a}_\perp := \frac{P_{U^\perp} \mathbf{a}}{\|P_{U^\perp} \mathbf{a}\|_2}
\]
for some \( 0 < \delta_n < 1 \) to be specified later. We note that \( \mathbf{u}_i^T \mathbf{a}_\perp = 0 \) for all \( 1 \leq i \leq r \) and \( \mathbf{u}_i^T \mathbf{u}_l = 0 \) for all \( i \neq l \). As can be straightforwardly verified, one has
\[
\|\mathbf{u} - \mathbf{u}_l\|_2 \leq \left( 1 - \frac{1}{\sqrt{1 + \delta_n^2}} \right) \|\mathbf{u}_l\|_2 + \frac{\delta_n}{\sqrt{1 + \delta_n^2}} \|\mathbf{a}_\perp\|_2 = \frac{\sqrt{1 + \delta_n^2} - 1 + \delta_n}{\sqrt{1 + \delta_n^2}} \leq 2\delta_n,
\] (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 \( \mathbf{S} \) 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 \( \mathbf{s}_i \) (i-th column of \( \mathbf{S} \)) under \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \), respectively.
Step 2: bounding the KL divergence between hypotheses. Let us define vector $\hat{u}_l$ as

$$\hat{u}_l := \frac{u_l^* - \frac{1}{\sqrt{\sigma}} a_{\perp}}{\|u_l^* - \frac{1}{\sqrt{\sigma}} a_{\perp}\|_2} = \frac{u_l^* - \frac{1}{\sqrt{\sigma}} a_{\perp}}{\sqrt{1 + \frac{1}{\delta_n^2}}}.$$

where the last step holds since $u_l^*$ is orthogonal to $a_{\perp}$. Note that $\hat{u}_l$ is a unit vector orthogonal to the subspace spanned by $u_l$ and $\{u_i^*\}_{i \neq l}$, namely, $\hat{u}_l u_l = 0$ and $u_l^T \hat{u}_l = 0$ for all $i \neq l$. Similar to the proof for the claim (E.1), one can derive

$$\text{tr}((\Sigma^* + \sigma^2 I_p)^{-1}(\Sigma + \sigma^2 I_p)) = \text{tr}\left(\frac{1}{\lambda_l^* + \sigma^2} u_l^* u_l^T + \frac{1}{\sigma^2} a_{\perp} a_{\perp}^T\right)((\lambda_l^* + \sigma^2) u_l^* u_l^T + \sigma^2 a_{\perp} a_{\perp}^T)\right) + p - 2$$

$$= (\bar{u}_l^* u_l^*)^2 + \frac{\sigma^2}{\lambda_l^* + \sigma^2} (\bar{u}_l^* u_l^*)^2 + \frac{\lambda_l^* + \sigma^2}{\sigma^2} (\bar{u}_l^* a_{\perp})^2 + (\hat{u}_l^* a_{\perp})^2 + p - 2$$

$$= \frac{1}{1 + \delta_n^2} \left(1 + \frac{\lambda_l^* + \sigma^2}{\sigma^2} \delta_n^2\right) + \frac{1}{1 + \delta_n^2} \left(\frac{\sigma^2}{\lambda_l^* + \sigma^2} \delta_n^4 + 1\right) + p - 2$$

$$= \frac{\lambda_l^*^2}{(\lambda_l^* + \sigma^2) \sigma^2} \frac{1}{1 + \delta_n^2} + p,$$

where the second step is due to $u_l^T a_{\perp} = 0$ and the third line follows from the following facts:

$$\bar{u}_l^* u_l^* = \frac{u_l^* u_l^* + \delta_n a_{\perp} a_{\perp}^T}{\|u_l^* + \delta_n a_{\perp}\|_2} = \frac{1}{\sqrt{1 + \delta_n^2}};$$

$$\bar{u}_l^* a_{\perp} = \frac{u_l^* a_{\perp} + \delta_n a_{\perp} a_{\perp}^T}{\|u_l^* + \delta_n a_{\perp}\|_2} = \frac{\delta_n}{\sqrt{1 + \delta_n^2}};$$

$$\hat{u}_l^* u_l^* = \frac{u_l^* u_l^* - \frac{1}{\sqrt{\sigma}} a_{\perp} a_{\perp}^T}{\|u_l^* - \frac{1}{\sqrt{\sigma}} a_{\perp}\|_2} = \frac{\delta_n}{\sqrt{1 + \delta_n^2}};$$

$$\hat{u}_l^* a_{\perp} = \frac{u_l^* a_{\perp} - \frac{1}{\sqrt{\sigma}} a_{\perp} a_{\perp}^T}{\|u_l^* - \frac{1}{\sqrt{\sigma}} a_{\perp}\|_2} = \frac{1}{\sqrt{1 + \delta_n^2}}.$$

As a consequence, we can upper bound the KL divergence as follows

$$\text{KL}(P^1 \parallel P^0) = \sum_{i=1}^{n} \text{KL}(P^1_i \parallel P^0_i) = \frac{1}{2} \sum_{i=1}^{n} \text{tr}((\Sigma^* + \sigma^2 I_p)^{-1}(\Sigma + \sigma^2 I_p)) - p$$

$$= \frac{n \lambda_l^*}{2(\lambda_l^* + \sigma^2) \sigma^2} \frac{\delta_n^2}{1 + \delta_n^2} \leq \frac{\delta_n^2 n \lambda_l^*}{2(\lambda_l^* + \sigma^2) \sigma^2}.$$

Step 3: invoking Fano’s inequality. From the preceding upper bound on the KL divergence, it is easy to see that $\text{KL}(P^1 \parallel P^0) \leq 1/16$ if we choose

$$\delta_n = c_n \sqrt{\frac{(\lambda_l^* + \sigma^2) \sigma^2}{\lambda_l^*^2 n}} \leq 1,$$

where $c_n \simeq 1$ obeys $c_n \in \{1/64, 1/16, 1/4\}$ and the last step holds due to the assumption (3.16). It follows from Fano’s inequality [Tsybakov, 2009, Theorem 2] that

$$p_\epsilon := \inf_{\psi} \max \{P(\psi \text{ rejects } H_0 \mid H_0), P(\psi \text{ rejects } H_1 \mid H_1)\} \geq 1/5,$$

where the infimum is taken over all tests. Further, we know from (E.12), (E.13) and $c_n \leq 1/4$ that

$$\|\bar{u} - u_l^*\|_2 \leq \sqrt{\frac{(\lambda_l^* + \sigma^2) \sigma^2}{\lambda_l^*^2 n}},$$
namely, $\Sigma_1 \in \mathcal{M}_2(\Sigma^*)$.

Next, let us continue to control $\min |a^\top \tilde{u}_l \pm a^\top u^*_i|$. Our goal is to show
\[
\min |a^\top \tilde{u}_l \pm a^\top u^*_i| \geq \delta_n \|P_{U^\perp} a\|_2,
\]
and we shall use the same argument as for (E.8) to prove it. Towards this, let us first consider $|a^\top \tilde{u}_l - a^\top u^*_i|$. By construction, one can derive
\[
a^\top \tilde{u}_l - a^\top u^*_i = a^\top u^*_i + \delta_n a^\top a_{\perp} \quad \frac{1}{\|P_{U^\perp}\|_2} \frac{1}{\|P_{U^\perp}\|_2} - a^\top u^*_i \leq \frac{\delta_n}{\sqrt{1 + \delta^2_n}} \|P_{U^\perp} a\|_2 - \left(1 - \frac{1}{\sqrt{1 + \delta^2_n}}\right) a^\top u^*_i \tag{E.14}
\]
where the last step holds because
\[
a^\top a_{\perp} = a^\top P_{U^\perp} a / \|P_{U^\perp} a\|_2 = (P_{U^\perp} a)^\top P_{U^\perp} a / \|P_{U^\perp} a\|_2 = \|P_{U^\perp} a\|_2.
\]

Moreover, it is straightforward to verify that
\[
\frac{1}{4} \delta^2_n \leq 1 - \frac{1}{\sqrt{1 + \delta^2_n}} \leq \sqrt{1 + \delta^2_n} - 1 \leq \frac{1}{2} \delta^2_n \tag{E.15}
\]
for $0 < \delta_n < 1$. With these basic facts in place, let us first choose the pre-factor $c_n \approx 1$ 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 $|\eta_1| \geq 2 |\eta_2|$, then one has
  \[
  |a^\top \tilde{u}_l - a^\top u^*_i| \geq |\eta_1| - |\eta_2| \geq \frac{|\eta_1|}{2} \geq \frac{\delta_n}{4} \|P_{U^\perp} a\|_2,
  \]
  where we have used the fact that $\delta_n \leq 1$.

- If $|\eta_1| \leq |\eta_2|/2$, then we know that
  \[
  |a^\top \tilde{u}_l - a^\top u^*_i| \geq |\eta_2| - |\eta_1| \geq |\eta_1| \geq \frac{\delta_n}{2} \|P_{U^\perp} a\|_2
  \]
as long as $\delta_n \leq 1$.

- 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 $\eta_1$ and $\eta_2$ satisfy $|\eta_1| \leq |\eta_2|/2$, thus belonging to the second case discussed above and hence $|a^\top \tilde{u}_l - a^\top u^*_i| \geq \delta_n \|P_{U^\perp} a\|_2$. 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
\[
|a^\top \tilde{u}_l - a^\top u^*_i| \geq \frac{\delta_n}{4} \|P_{U^\perp} a\|_2.
\]

Similarly, we can also derive
\[
|a^\top \tilde{u}_l + a^\top u^*_i| = \left| \frac{a^\top u^*_i + \delta_n a^\top a_{\perp}}{\sqrt{1 + \delta^2_n}} + a^\top u^*_i \right| = \left| \frac{\delta_n}{\sqrt{1 + \delta^2_n}} \|P_{U^\perp} a\|_2 + \left(1 + \frac{1}{\sqrt{1 + \delta^2_n}}\right) a^\top u^*_i \right|
\]
\[
\geq \left| \frac{\delta_n}{\sqrt{1 + \delta^2_n}} \|P_{U^\perp} a\|_2 - \left(1 + \frac{1}{\sqrt{1 + \delta^2_n}}\right) a^\top u^*_i \right|
\]
\[
\geq \frac{\delta_n}{\sqrt{1 + \delta^2_n}} \|P_{U^\perp} a\|_2
\]

71
Taking these two relations collectively yields the advertised bound:
\[
\min \left| \mathbf{a}^\top \mathbf{u}_i \pm \mathbf{a}^\top \mathbf{u}^*_i \right| \gtrsim \delta_n \left\| \mathbf{P}_{\mathcal{U}^*} \mathbf{a} \right\|_2.
\]

As a consequence, one can readily apply the standard reduction scheme in [Tsybakov, 2009, Chapter 2.2] again to arrive that
\[
\inf \sup \mathbb{E} \left[ \min \left| u_{a,i} \pm \mathbf{a}^\top \mathbf{u}_i (\Sigma) \right| \right] \gtrsim_p \min \left| \mathbf{a}^\top \mathbf{\tilde{u}}_i \pm \mathbf{a}^\top \mathbf{u}^*_i \right| \gtrsim \min \left| \mathbf{a}^\top \mathbf{\tilde{u}}_i \pm \mathbf{a}^\top \mathbf{u}^*_i \right|
\]
\[
\gtrsim c_n \sqrt{\frac{(\lambda_i^* + \sigma^2)^2}{\lambda_i^2 n}} \left\| \mathbf{P}_{\mathcal{U}^*} \mathbf{a} \right\|_2 \gtrsim \sqrt{\frac{(\lambda_i^* + \sigma^2)^2}{\lambda_i^2 n}} \left\| \mathbf{P}_{\mathcal{U}^*} \mathbf{a} \right\|_2
\]
where the last step holds since \( \min_n c_n = 1/64 \).

F \hspace{1em} \text{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 \( \{ \mathbf{h}_i \}_{i=1}^n \) be a sequence of independent zero-mean Gaussian random vectors in \( \mathbb{R}^r \) with covariance matrix \( \sigma^2 \mathbf{I}_r \), and let \( \mathbf{a} = [a_i]_{1 \leq i \leq n} \in \mathbb{R}^n \) be a fixed vector. Then with probability at least \( 1 - O \left( n^{-10} \right) \), one has
\[
\left\| \sum_{1 \leq i \leq n} a_i (\mathbf{h}_i \mathbf{h}_i^\top - \sigma^2 \mathbf{I}_r) \right\| \leq C_1 \sigma^2 \left( \left\| \mathbf{a} \right\|_2 \sqrt{r \log n} + \left\| \mathbf{a} \right\|_\infty (r \log n + \log^2 n) \right)
\]
\[
\leq C_2 \sigma^2 \left\| \mathbf{a} \right\|_\infty \left( \sqrt{r n \log n} + r \log n \right) \quad \text{(F.1)}
\]
for some sufficiently large constants \( C_1, C_2 > 0 \). Here, \( \left\| \mathbf{a} \right\|_\infty := \max_{1 \leq i \leq n} |a_i| \).

**Lemma 19.** Let \( \{ \mathbf{h}_i \}_{i=1}^n \) and \( \{ \mathbf{g}_i \}_{i=1}^n \) 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:
\[
\left\| \sum_{1 \leq i \leq n} \mathbf{h}_i \mathbf{g}_i^\top \right\| \leq C_3 \left( \sqrt{p n \log n} + \sqrt{pr \log n} \right)
\]
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 \( 1 - \delta_1 \),
\[
\left| \sum_{1 \leq i \leq n} f_i(X_i, \lambda) \right| \leq \frac{\varepsilon}{2};
\]
2. with probability at least \( 1 - \delta_2 \),
\[
\sup_{\lambda \in \mathcal{I}} \left| \frac{d}{d \lambda} \sum_{1 \leq i \leq n} f_i(X_i, \lambda) \right| \leq L.
\]

Then with probability exceeding \( 1 - \frac{8 |\mathcal{I}|}{\varepsilon} \delta_1 - \delta_2 \), one has
\[
\sup_{\lambda \in \mathcal{I}} \left| \sum_{1 \leq i \leq n} f_i(X_i, \lambda) \right| \leq \varepsilon.
\]
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 $M$ be a symmetric matrix in $\mathbb{R}^{n \times n}$ and $U$ be an orthonormal matrix in $\mathbb{R}^{n \times r}$ satisfying $U^T U = I_r$. Then one has

$$
\lambda_{n-r+i}(M) \leq \lambda_i(U^T M U) \leq \lambda_i(M), \quad 1 \leq i \leq r,
$$

where $\lambda_i(A)$ denote the $i$-th largest eigenvalue of matrix $A$."

### F.1 Proof of Lemma 18

As can be easily seen, the second inequality in (F.1) follows immediately from the elementary bound $\|a\|_2 \leq \|a\|_\infty \sqrt{n}$. Hence, the proof boils down to justifying the first inequality in (F.1).

We shall invoke the truncated matrix Bernstein inequality [Hopkins et al., 2016, Proposition A.7] to control the spectral norm of $\sum_i a_i (h_i h_i^T - \sigma^2 I_r)$, 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 $\sum_i a_i (h_i h_i^T - \sigma^2 I_r)$, (2) a high-probability upper bound $L$ on $\|a_i (h_i h_i^T - \sigma^2 I_r)\|$, (3) the expectation of the truncated summand $\max_i \|a_i (h_i h_i^T - \sigma^2 I_r)\| \leq L)$. We shall look at each of them separately.

1. Straightforward computation gives

$$
\Sigma := \sum_{i=1}^n a_i^2 E[(h_i h_i^T - \sigma^2 I_r)^2] = (r+1)\sigma^4 \sum_{i=1}^n a_i^2 I_r = (r+1)\sigma^4 \|a\|^2 \|I_r\|.
$$

2. We now turn to bounding the spectral norm of each summand $a_i (h_i h_i^T - \sigma^2 I_r)$, which clearly satisfies

$$
\|a_i (h_i h_i^T - \sigma^2 I_r)\| \leq |a_i| \cdot (\|h_i\|_2^2 + \sigma^2).
$$

By virtue of the Gaussian concentration inequality [Hsu et al., 2012, Proposition 1.1], we obtain

$$
P\{\|h_i\|_2^2 - \sigma^2 r \geq t\} \leq \exp\left(-\frac{1}{16} \min\left\{ \frac{t^2}{r \sigma^2}, \frac{t}{\sigma^2} \right\}\right).
$$

In particular, this implies that with probability at least $1 - O(n^{-20})$, one has

$$
\|h_i\|_2^2 \lesssim \sigma^2 (r + \log n).
$$

In what follows, we shall set

$$
L := C\sigma^2 (r + \log n)
$$

for some sufficiently large constant $C > 0$.

3. We then look the truncated mean. To this end, we observe that

$$
E[\|h_i\|_2^2 1\{\|h_i\|_2^2 \geq L\}] \leq LP\{\|h_i\|_2^2 \geq L\} + \int_{L}^{\infty} P\{\|h_i\|_2^2 \geq t\} \ dt 
\leq O(n^{-20})L + \int_{L}^{\infty} P\{\|h_i\|_2^2 \geq t\} \ dt.
$$

For any $t \geq L/2$, it is seen that min $\{t^2/(r \sigma^2), t/\sigma^2\} \geq t/\sigma^2$, and hence

$$
\int_{L/2}^{\infty} P\{\|h_i\|_2^2 \geq t\} \ dt \leq \int_{L/2}^{\infty} P\{\|h_i\|_2^2 - \sigma^2 r \geq t\} \ dt \leq \int_{L/2}^{\infty} \exp\left(-\frac{t}{16\sigma^2}\right) \ dt 
\lesssim \sigma^2 \exp\left(-\frac{C(r + \log n)}{32}\right) \leq \frac{L}{n^2},
$$

provided that $C > 0$ is sufficiently large. As a result, taking this together with $\|h_i h_i^T - \sigma^2 I_r\| \leq \|h_i\|_2^2 + \sigma^2$, we arrive at

$$
R := E[\|h_i h_i^T - \sigma^2 I_r\| 1\{\|h_i h_i^T - \sigma^2 I_r\| \geq L\}] \leq E[\|h_i\|_2^2 + \sigma^2] 1\{\|h_i\|_2^2 + \sigma^2 \geq L\} \lesssim \frac{L}{n^2}.
$$

73
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 − \( O(n^{-11}) \),

\[
\left\| \sum_i \phi_i (h_i h_i^\top - \sigma^2 I_r) \right\| \leq \sqrt{\|\Sigma\| \log n + \|a\|_\infty (nR + L \log n)} \\
\leq \sqrt{\|\Sigma\| \log n + \|a\|_\infty L \log n} \\
\leq \sqrt{n} \sum_{i=1}^n a_i^2 (r + 1) \sigma^4 \log n + \|a\|_\infty \sigma^2 (r \log n + \log^2 n) \\
\leq \sigma^2 \left( \|a\|_2 \sqrt{\log n} + \|a\|_\infty (r \log n + \log^2 n) \right),
\]

where (i) arises from (F.6), and (ii) relies on (F.2) and (F.5). This completes the proof.

### F.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 \( \sum_{i=1}^n h_i g_i^\top \), 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 \( \{h_i\} \) and \( \{g_i\} \), the covariance matrices can be computed as

\[
\Sigma_1 := \sum_{i=1}^n \mathbb{E}[h_i g_i^\top g_i h_i^\top] = \sum_{i=1}^n \mathbb{E}[\|g_i\|_2^2] \mathbb{E}[h_i h_i^\top] = np I_r; \quad (F.7)
\]

\[
\Sigma_2 := \sum_{i=1}^n \mathbb{E}[g_i h_i^\top h_i g_i^\top] = \sum_{i=1}^n \mathbb{E}[\|h_i\|_2^2] \mathbb{E}[g_i g_i^\top] = nr I_r. \quad (F.8)
\]

- As for the spectral norm of each summand \( h_i g_i^\top \), we know from (F.4) that with probability at least 1 − \( O(n^{-20}) \),

\[
\|h_i g_i^\top\| = \|h_i\|_2 \|g_i\|_2 \leq \sqrt{(r + \log n)(p + \log n)} \approx \sqrt{pr} + \sqrt{p \log n + \log n}.
\]

Therefore, this suggests that we define

\[
L := C \left( \sqrt{pr} + \sqrt{p \log n + \log n} \right)
\]

for some sufficiently large constant \( C > 0 \).

- Next, we turn to the truncated mean. Observe that

\[
\mathbb{E}[\|h_i g_i^\top\| \mathbb{1}\{\|h_i g_i^\top\| \geq L\}] \leq LP\{\|h_i g_i^\top\| \geq L\} + \int_L^\infty \mathbb{P}\{\|h_i g_i^\top\| \geq t\} \, dt \\
\leq O(n^{-20}) L + \int_L^\infty \mathbb{P}\{\|h_i g_i^\top\| \geq t\} \, dt \\
\leq O(n^{-20}) L + \int_L^\infty \mathbb{P}\{\|h_i\|_2^2 \geq t\} \, dt + \int_L^\infty \mathbb{P}\{\|g_i\|_2^2 \geq t\} \, dt,
\]

where the last holds arises from the following bound due to the union bound:

\[
\mathbb{P}\{\|h_i g_i^\top\| \geq t\} = \mathbb{P}\{\|h_i\|_2^2 \|g_i\|_2^2 \geq t^2\} \leq \mathbb{P}\{\|h_i\|_2^2 \geq t\} + \mathbb{P}\{\|g_i\|_2^2 \geq t\}.
\]

In addition, since \( \min \{t^2 / r, t\} \geq t \) for all \( t \geq L/2 \geq 2r \), we can use (F.3) to bound

\[
\int_L^\infty \mathbb{P}\{\|h_i\|_2^2 \geq t\} \, dt \leq \int_L^\infty \mathbb{P}\{\|h_i\|_2^2 - r \geq \frac{t}{2}\} \, dt \leq \int_{L/2}^\infty \mathbb{P}\{\|h_i\|_2^2 - r \geq t\} \, dt
\]

74
\[
\leq \int_0^\infty \exp \left( -\frac{1}{16} \min \left\{ \frac{t^2}{r}, t \right\} \right) \, dt \leq \int_0^\infty \exp \left( -\frac{t}{16} \right) \, dt
\]
\[
\lesssim \exp \left( -\frac{C}{32} \sqrt{pr + p \log n + \log^2 n} \right) \lesssim \frac{L}{n^2}.
\]

Clearly, the same bound also holds for \( \int_0^\infty \mathbb{P} \{ \| g_i \|^2 \geq t \} \, dt \). Therefore, combining these estimates yields
\[
R := \mathbb{E} [ \| h_i g_i^\top \| \mathbb{1} \{ \| h_i g_i^\top \| \geq L \} ] \lesssim \frac{L}{n^2}.
\]

(F.10)

With these parameters in place, one can apply the truncated matrix Bernstein inequality \cite{Hopkins2016}, Proposition A.7 to demonstrate that: with probability at least \( 1 - O \left( \frac{n}{n-10} \right) \),
\[
\left\| \sum_{1 \leq i \leq n} h_i g_i^\top \right\| \lesssim \sqrt{\| \Sigma_1 \| + \| \Sigma_2 \| } \log n + nR + L \log n
\]
\[
\quad \overset{(i)}{\lesssim} \sqrt{\| \Sigma_1 \| } \log n + L \log n
\]
\[
\quad \overset{(ii)}{\lesssim} \sqrt{pn \log n + p^2 \log n + \sqrt{p} \log^3 n + \log^2 n}
\]
\[
\quad \approx \sqrt{pn \log n + p^2 \log n}.
\]

Here, (i) uses (F.7), (F.7) and (F.10); (ii) arises from (F.9). The proof is thus complete.

F.3 Proof of Lemma 20

Let \( \mathcal{N} \) be a \( \frac{4L}{2L} \)-covering of \( \mathcal{I} \) with cardinality \( |\mathcal{N}| \leq \frac{4L|\mathcal{I}|}{\varepsilon} \), and let \( \mathcal{E} \) denote an event such that
\[
\sup_{\lambda \in \mathcal{N}} \left| \sum_{1 \leq i \leq n} f_i(X_i, \lambda) \right| \leq \frac{\varepsilon}{2},
\]
\[
\sup_{\lambda \in \mathcal{I}} \left| \frac{d}{d\lambda} \sum_{1 \leq i \leq n} f_i(X_i, \lambda) \right| \leq L,
\]
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 \( \lambda_0 \in \mathcal{N} \) such that \( |\lambda - \lambda_0| \leq \frac{\varepsilon}{2L} \). One can easily check that on the event \( \mathcal{E} \), one has
\[
\sup_{\lambda \in \mathcal{I}} \left| \sum_{1 \leq i \leq n} f_i(X_i, \lambda) \right| = \sup_{\lambda \in \mathcal{I}} \left| \sum_{1 \leq i \leq n} \left( f_i(X_i, \lambda) - f_i(X_i, \lambda_0) + f_i(X_i, \lambda_0) \right) \right|
\]
\[
\leq \sup_{\lambda \in \mathcal{I}} \left| \frac{d}{d\lambda} \sum_{1 \leq i \leq n} f_i(X_i, \lambda) \right| \cdot |\lambda - \lambda_0| + \sup_{\lambda \in \mathcal{N}} \left| \sum_{1 \leq i \leq n} f_i(X_i, \lambda_0) \right|
\]
\[
\leq L \cdot \frac{\varepsilon}{2L} + \sup_{\lambda \in \mathcal{N}} \left| \sum_{1 \leq i \leq n} f_i(X_i, \lambda_0) \right|
\]
\[
\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\]
thus concluding the proof.

References


