Tackling Small Eigen-Gaps: Fine-Grained Eigenvector Estimation and Inference Under Heteroscedastic Noise

Chen Cheng, Yuting Wei, and Yuxin Chen, *Member, IEEE*

***Abstract*— This paper aims to address two fundamental challenges arising in eigenvector estimation and inference for a low-rank matrix from noisy observations: 1) how to estimate an unknown eigenvector when the eigen-gap (i.e. the spacing between the associated eigenvalue and the rest of the spectrum) is particularly small; 2) how to perform estimation and inference on linear functionals of an eigenvector—a sort of "fine-grained" statistical reasoning that goes far beyond the usual ℓ_2 analysis. We investigate how to address these challenges in a setting where the unknown $n \times n$ matrix is symmetric and the additive noise matrix contains independent (and non-symmetric) entries. Based on eigen-decomposition of the asymmetric data matrix, we propose estimation and uncertainty quantification procedures for an unknown eigenvector, which further allow us to reason about linear functionals of an unknown eigenvector. The proposed procedures and the accompanying theory enjoy several important features: 1) distribution-free (i.e. prior knowledge about the noise distributions is not needed); 2) adaptive to heteroscedastic noise; 3) minimax optimal under Gaussian noise. Along the way, we establish valid procedures to construct confidence intervals for the unknown eigenvalues. All this is guaranteed even in the presence of a small eigen-gap (up to $O(\sqrt{n/\mathrm{poly}\log(n)})$ times smaller than the requirement in prior theory), which goes significantly beyond what generic matrix perturbation theory has to offer.**

***Index Terms*— Eigen-gap, linear form of eigenvectors, confidence interval, uncertainty quantification, heteroscedasticity.**

I. Introduction

A VARIETY of science and engineering applications ask for spectral analysis of low-rank matrices in high dimension [1]. Setting the stage, imagine that we are interested in a large low-rank matrix

$$\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} \in \mathbb{R}^{n\times n}, \tag{1}$$

where r ($r \ll n$) represents the rank, and $\lambda_l^\star$ stands for the lth largest eigenvalue of $\boldsymbol{M}^\star$, with $\boldsymbol{u}_l^\star \in \mathbb{R}^n$ the associated eigenvector. Suppose, however, that we do not have access to perfect measurements about the entries of this matrix; rather, the observations we have available, represented by a data matrix $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$, are contaminated by a substantial amount of random noise (reflected by the noise matrix $\boldsymbol{H}$). The aim is to perform reliable estimation and inference on the unseen eigenvectors of $\boldsymbol{M}^\star$ on the basis of noisy data. Motivated by the abundance of applications (e.g. collaborative filtering, harmonic retrieval, sensor network localization, joint shape matching [2]–[6]), research on eigenspace estimation in this context has flourished in the past several years, typically built upon proper exploitation of low-rank structures. We have now been equipped with a rich suite of modern statistical theory that delivers statistical performance guarantees for a number of spectral estimators (e.g. [7]–[16], [16]–[22]); see [23] for a contemporary overview of spectral methods.

A. Motivation and Challenges

Despite a large body of work tackling the above problem, there are several fundamental yet unaddressed challenges that deserve further attention.

1) Stringent Requirements on Eigen-Gaps: A crucial identifiability issue stands out when estimating individual eigenvectors. In general, one cannot possibly disentangle the eigenvectors $\boldsymbol{u}_1^\star, \cdots, \boldsymbol{u}_r^\star$ unless there is sufficient spacing between adjacent eigenvalues. After all, even in the noiseless setting, one can only hope to recover the subspace spanned by $\{\boldsymbol{u}_i^\star\}$, rather than individual eigenvectors, if all non-zero eigenvalues are identical.

In principle, a "sufficient" eigen-gap criterion in the presence of noise is dictated by the noise levels, or more precisely, the signal-to-noise ratios (SNRs). However, generic linear algebra theory typically imposes fairly stringent, and hence pessimistic, eigen-gap requirements for both eigenvector and eigenvalue estimation. More concretely, imagine we wish to estimate the lth eigenvector $\boldsymbol{u}_l^\star$: generic matrix perturbation theory (e.g. the Davis-Kahan $\sin\boldsymbol{\Theta}$ theorem or the Wedin theorem [12], [24], [25]) typically cannot guarantee meaningful

Manuscript received November 22, 2020; revised June 13, 2021; accepted September 7, 2021. Date of publication September 10, 2021; date of current version October 20, 2021. The work of Chen Cheng was supported by the William R. Hewlett Stanford Graduate Fellowship. The work of Yuting Wei was supported in part by NSF under Grant DMS-2015447/2147546 and Grant CCF-2007911. The work of Yuxin Chen was supported in part by Air Force Office of Scientific Research (AFOSR) Young Investigator Program (YIP) under Award FA9550-19-1-0030; in part by Office of Naval Research (ONR) under Grant N00014-19-1-2120; in part by Army Research Office (ARO) under Grant W911NF-20-1-0097 and Grant W911NF-18-1-0303; and in part by NSF under Grant CCF-1907661, Grant IIS-1900140, Grant IIS-2100158, and Grant DMS-2014279. *(Corresponding author: Yuting Wei.)*

Chen Cheng is with the Department of Statistics, Stanford University, Stanford, CA 94305 USA (e-mail: chencheng@stanford.edu).

Yuting Wei is with the Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: ytwei@wharton.upenn.edu).

Yuxin Chen is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: yuxin.chen@princeton.edu).

Communicated by M. Davenport, Associate Editor for Signal Processing.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2021.3111828.

Digital Object Identifier 10.1109/TIT.2021.3111828

estimation of $\boldsymbol{u}_l^\star$ unless

$$\text{(classical theory)} \qquad \underbrace{\min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star|}_{\text{eigen-gap}} > \|\boldsymbol{H}\|. \tag{2}$$

This eigen-gap requirement can be problematic and hard to satisfy as the noise size grows, casting doubts on our ability to perform informative inference on individual eigenvectors.

2) Fine-Grained Estimation and Inference (Beyond ℓ_2 Guarantees): In many applications, it is often the case that the ultimate goal is not to characterize the ℓ_2 or "bulk" behavior (e.g. the mean squared estimation error) of an eigenvector estimator, but rather to reason about the eigenvectors along a few preconceived yet important directions. Take community recovery for instance: the eigenvector of a certain adjacency matrix encodes community membership of a set of users [26]; if we wish to infer the community memberships of a few important users, or the similarities between a few pairs of users, it boils down to assessing the entrywise behavior or certain pairwise linear functional of an eigenvector estimator. Another example is concerned with harmonic retrieval: the leading eigenvector of a properly arranged Toeplitz matrix represents the time-domain response of the sinusoidal signals of interest [27]; thus, retrieving the underlying frequency involves inferring the Fourier coefficients of this eigenvector at some given frequencies. These problems can be formulated as estimation and inference for *linear functionals of an eigenvector*, namely, quantities of the form $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ $(1 \le l \le r)$ with $\boldsymbol{a}$ some prescribed vector. In principle, this task can be viewed as a sort of *fine-grained* statistical analysis, given that it pursues highly "local" and "delicate" information of interest.

Towards estimating $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$, a natural starting point is a "plug-in" estimator, which computes a reasonable estimate $\widehat{\boldsymbol{u}}_l$ of $\boldsymbol{u}_l^\star$ and outputs $\boldsymbol{a}^\top \widehat{\boldsymbol{u}}_l$. There are several critical challenges that we shall bear in mind. To begin with, a dominant fraction of prior theory focuses on ℓ_2 risk analysis of an eigenvector estimator, which is often too coarse to deliver tight uncertainty assessment for the plug-in estimator. In fact, the ℓ_2 risk bounds alone often lead to highly conservative estimates for the risk under consideration, making it hard to assess the performance of the plug-in estimator. To further complicate matters, there is often a severe bias issue surrounding the plug-in estimator. Even when an estimator $\widehat{\boldsymbol{u}}_l$ is nearly unbiased in a strong entrywise sense (meaning that the estimation bias is dominated by the variability in every single coordinate), this property alone does not preclude the possibility of bias accumulation along the direction $\boldsymbol{a}$. Addressing these issues calls for refined risk analysis as well as careful examination into the bias-variance trade-off.

B. A Glimpse of Our Approach and Our Contributions

1) Eigen-Decomposition Meets Statistical Asymmetry: The current paper makes progress in a setting where the noise matrix $\boldsymbol{H}$ consists of independent (but possibly heterogeneous and heteroscedastic) zero-mean components. Our approach is inspired by the findings of [17]. Consider, for example, the case when $\boldsymbol{H}$ is a random and asymmetric matrix and when $\boldsymbol{M}^\star$ is a rank-1 symmetric matrix. The results in [17] reveal that an eigen-decomposition approach applied to $\boldsymbol{M}$ (without symmetrization) achieves appealing statistical accuracy when estimating the leading eigenvalue of $\boldsymbol{M}^\star$. The key enabler is an implicit bias reduction feature when computing vanilla eigen-decomposition of an asymmetric data matrix. While [17] only provides highly partial results when going beyond the rank-1 case, it hints at the potential benefits of eigen-decomposition in super-resolving the spectrum.

2) Our Contributions: The main contributions of this paper are summarized below, all of which are built upon an eigen-decomposition approach applied to the asymmetric data matrix $\boldsymbol{M}$ (without symmetrization). As an important advantage, the proposed procedures and their accompanying theory are distribution-free, meaning that they do not require prior knowledge about the distribution of the noise and thus are fully adaptive to heteroscedasticity of data.

- We demonstrate that the lth eigenvector $\boldsymbol{u}_l^\star$ and eigenvalue $\lambda_l^\star$ can be estimated with near-optimal accuracy even when the eigen-gap (2) is extremely small. More concretely, for various noise distributions our results only require

$$\text{(our theory)} \qquad \min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star| \gtrsim \|\boldsymbol{H}\| \frac{\mathrm{poly}\log(n)}{\sqrt{n}}, \tag{3}$$

  which is about $O(\sqrt{n})$ times less stringent (up to log factor) than generic linear algebra theory (cf. (2)).
- We propose a new estimator for the linear functional $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ (obtained via proper de-biasing of certain plug-in estimators) that achieves minimax-optimal statistical accuracy. In addition, we demonstrate how to construct valid confidence intervals for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.
- Additionally, we demonstrate how to compute valid confidence intervals for the eigenvalues of interest. Interestingly, de-biasing is not needed at all for performing inference on eigenvalues, as the eigen-decomposition approach implicitly alleviates the estimation bias.

Our findings unveil new insights into the capability of spectral methods in a statistical context, going far beyond what conventional matrix perturbation theory has to offer.

C. Notation

For any vector $\boldsymbol{x}$, denote by $\|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_\infty$ the ℓ_2 norm and the ℓ_∞ norm of $\boldsymbol{x}$, respectively. For any matrix $\boldsymbol{M}$, we let $\|\boldsymbol{M}\|$, $\|\boldsymbol{M}\|_{\mathrm{F}}$, $\|\boldsymbol{M}\|_\infty$ and $\|\boldsymbol{M}\|_{2,\infty}$ represent the spectral norm, the Frobenius norm, the entrywise ℓ_∞ norm (i.e. $\|\boldsymbol{M}\|_\infty := \max_{i,j} |M_{ij}|$), and the two-to-infinity norm (i.e. $\|\boldsymbol{M}\|_{2,\infty} := \sup_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{M}\boldsymbol{x}\|_\infty$) of $\boldsymbol{M}$, respectively. The notation $f(n) = O(g(n))$ or $f(n) \lesssim g(n)$ means that there is a universal constant $c > 0$ such that $|f(n)| \le c|g(n)|$, $f(n) \gtrsim g(n)$ means that there is a universal constant $c > 0$ such that $|f(n)| \ge c|g(n)|$, and $f(n) \asymp g(n)$ means that there exist constants $c_1, c_2 > 0$ such that $c_1|g(n)| \le |f(n)| \le c_2|g(n)|$. The notation $f(n) \gg g(n)$ (resp. $f(n) \ll g(n)$) means that there exists a sufficiently large (resp. small) constant $c > 0$ such that $f(n) \ge cg(n)$ (resp. $f(n) \le cg(n)$).

In addition, we denote by $u_{l,j}$, $w_{l,j}$ and $u^\star_{l,j}$ the jth entry of $\boldsymbol{u}_l$, $\boldsymbol{w}_l$ and $\boldsymbol{u}^\star_l$, respectively. Let $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ represent the standard basis vectors in $\mathbb{R}^n$, $\mathbf{1}_n \in \mathbb{R}^n$ the all-one vector, and $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ the identity matrix. We shall also abbreviate the interval $[b-c, b+c]$ to $[b \pm c]$, and denote $\min|b \pm c| = \min\{|b+c|, |b-c|\}$, $\max|b \pm c| = \max\{|b+c|, |b-c|\}$, and $\min\|\boldsymbol{b} \pm \boldsymbol{c}\| = \min\{\|\boldsymbol{b}-\boldsymbol{c}\|, \|\boldsymbol{b}+\boldsymbol{c}\|\}$ for any norm $\|\cdot\|$. Moreover, denote by $\Phi(\cdot)$ the cumulative distribution function (CDF) of a standard Gaussian random variable.

II. Problem Formulation

A. Model

Imagine we seek to estimate an unknown rank-r matrix

$$\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda^\star_l \boldsymbol{u}^\star_l \boldsymbol{u}^{\star\top}_l =: \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{U}^{\star\top} \in \mathbb{R}^{n\times n}, \tag{4}$$

where $\lambda^\star_1 \geq \lambda^\star_2 \geq \cdots \geq \lambda^\star_r$ denote the r nonzero eigenvalues of $\boldsymbol{M}^\star$, and $\boldsymbol{u}^\star_1, \cdots, \boldsymbol{u}^\star_r$ represent the associated (normalized) eigenvectors. Here, for notational convenience we let

$$\begin{aligned} \boldsymbol{U}^\star &:= \left[\boldsymbol{u}^\star_1, \cdots, \boldsymbol{u}^\star_r\right] \in \mathbb{R}^{n\times r}, \\ \boldsymbol{\Sigma}^\star &:= \begin{bmatrix} \lambda^\star_1 & & \\ & \ddots & \\ & & \lambda^\star_r \end{bmatrix} \in \mathbb{R}^{r\times r}. \end{aligned} \tag{5}$$

We shall also define

$$\lambda^\star_{\max} := \max_{1\leq l\leq r} |\lambda^\star_l|, \ \lambda^\star_{\min} := \min_{1\leq l\leq r} |\lambda^\star_l| \text{ and } \kappa := \lambda^\star_{\max} / \lambda^\star_{\min}. \tag{6}$$

What we have observed is a corrupted copy of $\boldsymbol{M}^\star$, namely,

$$\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}, \tag{7}$$

where $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ stands for a random noise matrix. This paper focuses on the family of noise matrices satisfying the following assumptions:

Assumption 1: The noise matrix $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ satisfies the following properties.

1) **Independence and zero mean**. The entries $\{H_{ij}\}_{1\leq i,j\leq n}$ are independent zero-mean random variables; this indicates that the matrices $\boldsymbol{H}$ and $\boldsymbol{M}$ are, in general, *asymmetric*.
2) **Heteroscedasticity and unknown variances**. Let $\sigma^2_{ij} := \mathsf{Var}(H_{ij})$ denote the variance of H_{ij}. To accommodate more realistic scenarios, we allow σ^2_{ij} to vary across entries — commonly referred to as *heteroscedastic noise*. In addition, we do not a priori know $\{\sigma^2_{ij}\}_{1\leq i,j\leq n}$. Throughout this paper, we assume
$$0 \leq \sigma^2_{\min} \leq \sigma^2_{ij} \leq \sigma^2_{\max}, \qquad 1 \leq i,j \leq n. \tag{8}$$
3) **Magnitudes**. Each H_{ij} $(1 \leq i,j \leq n)$ satisfies either of the following conditions:
   (a) $|H_{ij}| \leq B$;
   (b) H_{ij} has a symmetric distribution obeying $\mathbb{P}\{|H_{ij}| > B\} \leq c_{\mathrm{b}} n^{-12}$ for some constant $c_{\mathrm{b}} > 0$, and $\mathbb{E}[H^2_{ij} \mathbb{1}_{\{|H_{ij}|>B\}}] = o(\sigma^2_{ij})$.

Remark 1: Here, the quantities B, $\{\sigma_{i,j}\}$, $\sigma_{\min}$ and $\sigma_{\max}$ may all depend on n.

Remark 2: As we shall see momentarily, the lower bound $\sigma^2_{\min}$ on the noise variance is imposed only for our statistical inference theory (or more precisely, it is imposed in order to ensure the plausibility to estimate the variance of the estimation error in an accurate manner). This lower bound $\sigma^2_{\min}$ is not needed at all in our eigenvector estimation guarantees (e.g., Theorems 1 and 2).

The careful reader would naturally ask when we would have an asymmetric noise matrix $\boldsymbol{H}$. This might happen when, for example, one has collected two independent samples about each entry of $\boldsymbol{M}^\star$ and chooses to arrange the observed data in an asymmetric manner. Moreover, when the noise matrix is a symmetric Gaussian random matrix and possibly contains missing data, [17, Appendix J] points out some simple asymmetrization tricks that allow one to convert a symmetric data matrix $\boldsymbol{M}$ to an asymmetric counterpart with independent components; see Appendix G for an example.

In addition to the above assumptions on the noise, we assume that the unknown matrix $\boldsymbol{M}^\star$ satisfies a certain incoherence condition, as commonly seen in the low-rank matrix estimation literature.

Definition 1 (Incoherence): The matrix $\boldsymbol{M}^\star$ with eigen-decomposition $\boldsymbol{M}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{U}^{\star\top}$ is said to be μ-incoherent if $\|\boldsymbol{U}^\star\|_\infty \leq \sqrt{\mu/n}$.

B. Goal

The primary goal of the current paper is to perform certain "fine-grained" statistical estimation and inference on the unknown eigenvectors $\{\boldsymbol{u}^\star_1, \cdots, \boldsymbol{u}^\star_r\}$. To be more specific, we aim at developing statistically efficient methods to estimate and construct confidence intervals for linear functionals taking the form $\boldsymbol{a}^\top \boldsymbol{u}^\star_l$ $(1 \leq l \leq r)$, where $\boldsymbol{a} \in \mathbb{R}^n$ is a pre-determined fixed vector. Along the way, we shall also demonstrate how to perform estimation and inference on the unknown eigenvalues $\{\lambda^\star_1, \cdots, \lambda^\star_r\}$. Ideally, all these tasks can be accomplished even when the associated eigen-gaps are very small, without requiring prior knowledge about the noise distributions and noise levels.

III. Estimation

This section presents our algorithms and the accompanying theory for estimating an eigenvector $\boldsymbol{u}^\star_l$, a linear functional $\boldsymbol{a}^\top \boldsymbol{u}^\star_l$ of this eigenvector, as well as the associated eigenvalue $\lambda^\star_l$.

Two popular estimation schemes immediately come into mind: (1) performing eigen-decomposition after symmetrizing the data matrix (e.g. replacing $\boldsymbol{M}$ by $\frac{1}{2}(\boldsymbol{M} + \boldsymbol{M}^\top)$); (2) computing singular value decomposition (SVD) of the asymmetric data matrix $\boldsymbol{M}$. Contrastingly, this paper adopts a far less widely used, and in fact far less widely studied, strategy based on eigen-decomposition without symmetrization; namely, we attempt estimation of $\{\boldsymbol{u}^\star_l\}$ and $\{\lambda^\star_l\}$ via the eigenvectors and the eigenvalues of $\boldsymbol{M}$, respectively, despite the asymmetric nature of $\boldsymbol{M}$ in general. While conventional wisdom does not advocate eigen-decomposition of asymmetric

matrices (due to, say, potential numerical instability), our investigation uncovers remarkable advantages of this approach under the statistical context considered in the present paper.

A. Notation: Eigen-Decomposition Without Symmetrization

Before continuing, we introduce several additional notation that shall be adopted throughout. Owing to the asymmetry of $\boldsymbol{M}$, the left and the right eigenvectors of $\boldsymbol{M}$ do not coincide.

Notation 1: Let $\lambda_1, \cdots, \lambda_r$ denote the top-r leading eigenvalues of $\boldsymbol{M}$ (so that $\min_{1\le l\le r}|\lambda_l|$ is larger in magnitude than any other eigenvalue of $\boldsymbol{M}$). Assume that they are sorted by their real parts, namely, $\mathrm{Re}(\lambda_1) \ge \mathrm{Re}(\lambda_2) \ge \cdots \ge \mathrm{Re}(\lambda_r)$, and denote by $\boldsymbol{u}_l$ (resp. $\boldsymbol{w}_l$) the right (resp. left) eigenvector of $\boldsymbol{M}$ associated with λ_l; this means

$$\boldsymbol{M}\boldsymbol{u}_l = \lambda_l \boldsymbol{u}_l \quad \text{and} \quad \boldsymbol{M}^\top \boldsymbol{w}_l = \lambda_l \boldsymbol{w}_l. \tag{9}$$

In addition, if $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$ are both real-valued, then we assume without loss of generality that $\langle \boldsymbol{u}_l, \boldsymbol{w}_l\rangle \ge 0$.

Remark 3: As we shall justify shortly in Theorem 1, even though $\boldsymbol{M}$ is in general asymmetric, the eigenvalue λ_l and the eigenvectors $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$ are, with high probability, real-valued under the assumptions imposed in this paper.

B. Estimation Algorithms

We are now ready to present our procedures for estimating the lth eigenvector of $\boldsymbol{M}^\star$, on the basis of eigen-decomposition of $\boldsymbol{M}$ (see Notation 1).

- **Estimator for $\boldsymbol{u}_l^\star$ (with the aim of achieving low ℓ_2 risk):**

$$\widehat{\boldsymbol{u}}_l := \frac{1}{\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2}\big(\boldsymbol{u}_l + \boldsymbol{w}_l\big); \tag{10}$$

- **Estimator for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ for a preconceived direction $\boldsymbol{a}$:**

$$\widehat{u}_{\boldsymbol{a},l} := \min\left\{\sqrt{\left|\frac{(\boldsymbol{a}^\top \boldsymbol{u}_l)(\boldsymbol{a}^\top \boldsymbol{w}_l)}{\boldsymbol{w}_l^\top \boldsymbol{u}_l}\right|},\ \|\boldsymbol{a}\|_2\right\}. \tag{11}$$

Here, the rationale behind $\widehat{u}_{\boldsymbol{a},l}$ is this: both $\boldsymbol{a}^\top \boldsymbol{u}_l$ and $\boldsymbol{a}^\top \boldsymbol{w}_l$ might systematically *under-estimate* the quantity of interest $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$. As a result, one is advised to first alleviate the bias via proper de-shrinking, which is the role played by the factor $\frac{1}{\boldsymbol{w}_l^\top \boldsymbol{u}_l}$. The choice of this factor is based on in-depth understanding of the behavior of $\boldsymbol{a}^\top \boldsymbol{u}_l$ and $\boldsymbol{a}^\top \boldsymbol{w}_l$, and will be better understood after we delve into technical details. As an important feature, the proposed estimation procedures do not rely on prior knowledge about the noise distributions.

Remark 4: The estimator (11) involves the term $\|\boldsymbol{a}\|_2$ due to the trivial upper bound $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \le \|\boldsymbol{a}\|_2 \|\boldsymbol{u}_l^\star\|_2 = \|\boldsymbol{a}\|_2$.

C. Theoretical Guarantees

We now embark on theoretical development for the proposed estimators. As alluded to previously, whether we can reliably estimate and infer an eigenvector $\boldsymbol{u}_l^\star$ depends largely upon the spacing between the lth eigenvalue and its adjacent eigenvalues. In light of this, we define formally the eigen-gap metric w.r.t. the lth eigenvalue of $\boldsymbol{M}^\star$ as follows

$$\Delta_l^\star := \begin{cases} \min\limits_{1\le k\le r, k\ne l} |\lambda_l^\star - \lambda_k^\star|, & \text{if } r > 1; \\ \infty, & \text{otherwise.} \end{cases} \tag{12}$$

Most of our theoretical guarantees rely on this crucial metric. Moreover, our theory in this section is built upon a set of assumptions on the noise levels:

Assumption 2: Suppose that the noise parameters defined in Assumption 1 satisfy

$$\Delta_l^\star > 2c_4\kappa^2\ r^2\sigma_{\max}\sqrt{\mu \log n} \tag{13a}$$

$$B\log n \le \sigma_{\max}\sqrt{n\log n} \le c_5 \lambda_{\min}^\star/\kappa^3 \tag{13b}$$

for some sufficiently large (resp. small) universal constant $c_4 > 0$ (resp. $c_5 > 0$).

Remark 5: Consider, for example, the case where $r, \kappa, \mu \asymp 1$: the lower bound on $\Delta_l^\star$ in (13a) is $O(\sqrt{n})$ times smaller than the lower bound on $\lambda_{\min}^\star$ in (13b), meaning that the eigen-gap can be considerably smaller than the minimum eigenvalue of the truth.

The following theorem delivers statistical guarantees for the proposed eigenvector estimators, with the proof postponed to Appendix B. We recall the notation $\min \|\boldsymbol{z} \pm \boldsymbol{u}_l^\star\| := \min\{\|\boldsymbol{z} - \boldsymbol{u}_l^\star\|, \|\boldsymbol{z} + \boldsymbol{u}_l^\star\|\}$ for any norm $\|\cdot\|$.

Theorem 1 (Eigenvector Estimation): Suppose that $\mu\kappa^2\ r^4 \lesssim n$, and that Assumptions 1-2 hold. With probability at least $1 - O(n^{-6})$, the eigenvalue λ_l and the associated eigenvectors $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$ (see Notation 1) are all real-valued, and one has the following:

1) (ℓ_2 and ℓ_∞ guarantees)

$$\begin{aligned} &\min \|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_2 \\ &\lesssim \frac{\sigma_{\max}\sqrt{\kappa^6 n\log n}}{\lambda_{\min}^\star} + \frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\mu\kappa^4 r^2 \log n}, \end{aligned} \tag{14a}$$

$$\begin{aligned} &\min \|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_\infty \\ &\lesssim \frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^4 r\log n} + \frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\frac{\mu^2\kappa^4 r^3 \log n}{n}}; \end{aligned} \tag{14b}$$

these hold unchanged if $\boldsymbol{u}_l$ is replaced by either $\boldsymbol{w}_l$ or $\widehat{\boldsymbol{u}}_l$ (cf. (10));

2) (statistical guarantees for linear forms of an eigenvector) for any fixed vector $\boldsymbol{a}$ with $\|\boldsymbol{a}\|_2 = 1$, the proposed estimator (11) obeys[1]

$$\begin{aligned} &\min \left|\widehat{u}_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \\ &\lesssim \frac{\sigma_{\max} r^2\sqrt{\mu\kappa^4\log n}}{\lambda_{\min}^\star} + \frac{\sigma_{\max}^2 \mu r^2\kappa^4 \log n}{(\Delta_l^\star)^2}|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \\ &\quad + \sigma_{\max}\sqrt{\mu\kappa^4 r^3\log n}\max_{k\ne l}\frac{|\boldsymbol{a}^\top \boldsymbol{u}_k^\star|}{|\lambda_l^\star - \lambda_k^\star|}. \end{aligned} \tag{14c}$$

Interestingly, even though we work with eigen-decomposition of asymmetric matrices, the obtained eigenvalue and eigenvector are provably real-valued as long as a fairly mild eigen-gap condition is satisfied. This

[1]If the rank $r = 1$, then the 2nd and the 3rd terms on the right-hand side of (14c) are set to be zero.

presents an important feature that cannot be derived from generic matrix perturbation theory.

To interpret the effectiveness of this theorem, we find it convenient to concentrate on the case with $r, \kappa, \mu \asymp 1$ under i.i.d. Gaussian noise. The implications in this case are summarized below.

- ℓ_2 *and* ℓ_∞ *guarantees.* The ℓ_2 and ℓ_∞ statistical guarantees derived in Theorem 1 read

$$\min \|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_2 \lesssim \frac{\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star} + \frac{\sigma_{\max}\sqrt{\log n}}{\Delta_l^\star} =: \mathcal{E}_2; \tag{15a}$$

$$\min \|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_\infty \lesssim \frac{1}{\sqrt{n}}\mathcal{E}_2. \tag{15b}$$

  Our ℓ_∞ error bound is about $O(\sqrt{n})$ times smaller than the ℓ_2 risk bound, implying that the energy of the estimation error is more or less dispersed across all entries.
- *Improved eigen-gap requirements.* Consistent estimation of $\boldsymbol{u}_l^\star$ — in the sense of $\min \|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_2 = o(\|\boldsymbol{u}_l^\star\|_2)$ and $\min \|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_\infty = o(\|\boldsymbol{u}_l^\star\|_\infty)$ — is guaranteed as long as[2]

$$\|\boldsymbol{H}\| \log n \lesssim \|\boldsymbol{M}^\star\|; \tag{16a}$$

$$\Delta_l^\star \gtrsim \frac{\|\boldsymbol{H}\| \log n}{\sqrt{n}}. \tag{16b}$$

  While the condition (16a) is commonly seen in prior literature (up to some log factor), the eigen-gap requirement (16b) is in stark contrast to classical matrix perturbation theory (e.g. the Davis-Kahan $\sin\boldsymbol{\Theta}$ theorem or the Wedin theorem [23]–[25]). In fact, prior theory typically requires the spacing between $\lambda_l^\star$ and the rest of the eigenvalues to at least exceed

$$\Delta_l^\star \gtrsim \|\boldsymbol{H}\| \qquad \text{(prior theory)}, \tag{17}$$

  which is about $O(\sqrt{n}/\log n)$ times more stringent than our requirement (16b).
- *The influence of eigen-gaps and* $\{\boldsymbol{a}^\top \boldsymbol{u}_i^\star\}$ *upon estimation accuracy.* For any fixed unit vector $\boldsymbol{a}$, our theoretical guarantees for estimating $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ read

$$\begin{aligned} &\min \left|\widehat{u}_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \\ &\lesssim \frac{\sigma_{\max}\sqrt{\log n}}{\lambda_{\min}^\star} + \frac{\sigma_{\max}^2 \log n}{(\Delta_l^\star)^2}|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \\ &\quad + \underbrace{\sigma_{\max}\sqrt{\log n}\max_{k\neq l}\frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star - \lambda_k^\star\right|}}_{\text{“interferers”}}. \end{aligned} \tag{18}$$

  The first term on the right-hand side of (18) is a universal term that is no larger than $O(1/\sqrt{n})$ times the ℓ_2 bound (15a). The other two terms are more complicated, which depend on not only the spacing of the eigenvalues but also the sizes of $\{\boldsymbol{a}^\top \boldsymbol{u}_i^\star\}_{1\le i\le r}$. More concretely, (1) the influence of the target quantity $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ upon the estimation error is captured by the multiplicative factor $\frac{\sigma_{\max}^2}{(\Delta_l^\star)^2}$, which scales inverse quadratically in $\Delta_l^\star$; (2) the linear functionals of other eigenvectors (namely, $\{\boldsymbol{a}^\top \boldsymbol{u}_k^\star\}_{k\neq l}$) essentially behave as “interferers” that might degrade estimation fidelity; in particular, the influence of $\boldsymbol{a}^\top \boldsymbol{u}_k^\star$ ($k \neq l$) upon estimation loss can be understood through the coefficient $\sigma_{\max}/|\lambda_l^\star - \lambda_k^\star|$, which is inversely proportional to the associated eigen-gap $|\lambda_l^\star - \lambda_k^\star|$. Intuitively, if $\boldsymbol{a}^\top \boldsymbol{u}_k^\star$ ($k \neq l$) becomes large, it results in stronger “interference” when estimating $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$; this adverse effect would be easier to mitigate if the gap $|\lambda_l^\star - \lambda_k^\star|$ increases (so that it becomes easier to differentiate the lth and the kth eigenvectors).
- *The rank-1 case.* When $r = 1$, the preceding theory can be significantly simplified as follows

$$\begin{aligned} \min \|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_2 &\lesssim \frac{\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}, \\ \min \|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_\infty &\lesssim \frac{\sigma_{\max}\sqrt{\log n}}{\lambda_{\min}^\star}, \\ \min \left|\widehat{u}_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| &\lesssim \frac{\sigma_{\max}\sqrt{\log n}}{\lambda_{\min}^\star}. \end{aligned}$$

  In particular, the same estimation error bound — which is about $\sqrt{n}$ times smaller than the ℓ_2 loss — holds for any arbitrary direction as specified by $\boldsymbol{a}$. In other words, the estimation error is more or less identical over any pre-determined direction.

Encouragingly, the above performance guarantees are minimax optimal up to some logarithmic factor, as we shall elucidate in Section III-D.

As it turns out, the feasibility of faithful eigenvector estimation is largely dictated by our ability to locate the lth eigenvalue $\lambda_l^\star$ and to disentangle it from the rest of the spectrum, which becomes particularly challenging if the eigen-gap $\Delta_l^\star$ is very small. In light of this, we develop the following eigenvalue perturbation theory that significantly improves upon generic linear algebra theory.

Theorem 2 (Eigenvalue Estimation): Suppose that $\mu\kappa^2 r^4 \lesssim n$, and that Assumptions 1-2 hold. With probability at least $1 - O(n^{-6})$, one has

$$|\lambda_l - \lambda_l^\star| \le c_4 \sigma_{\max}\sqrt{\mu\kappa^2 r^4 \log n}. \tag{19}$$

In words, as long as the eigen-gap $\Delta_l^\star$ exceeds (13a), the vanilla eigen-decomposition approach (without symmetrization) produces an estimate of $\lambda_l^\star$ with an estimation error at most $\Delta_l^\star/2$, meaning that we have managed to locate $\lambda_l^\star$ with a high resolution. As mentioned previously, this eigen-gap requirement can be interpreted as $\Delta_l^\star \gtrsim \|\boldsymbol{H}\|\frac{\log n}{\sqrt{n}}$ in the Gaussian noise case with $r, \mu, \kappa \asymp 1$. For the sake of comparison, we remind the reader that classical linear algebra theory (e.g. Weyl's inequality and the Bauer-Fike inequality) only provides perturbation bounds with fairly low resolution — namely, bounds at best on the order $|\lambda_l - \lambda_l^\star| \le \|\boldsymbol{H}\|$ — which are highly insufficient and in fact suboptimal for our purpose.

It is worth emphasizing that the ability to obtain highly accurate eigenvalue estimates is the key to tackling small eigen-gaps. Given that we can estimate $\lambda_l^\star$ with a precision

[2]Note that when $\{H_{ij}\}$ are i.i.d. Gaussian, we have $\|\boldsymbol{H}\| \asymp \sigma_{\max}\sqrt{n}$ with high probability.

$\sigma_{\max}\sqrt{\mu\kappa^2\ r^4\log n}$, we can distinguish $\lambda_l^\star$ from the rest of the spectrum even when the associated eigen-gap is on the same order. This also helps explain why a large body of prior theory fell short of accommodating small eigen-gaps. As discussed in [17, Section 4.3], the standard eigen-decomposition approach when applied to symmetric matrices suffers from a non-negligible bias issue, which fails to yield a high-precision eigenvalue estimate. While the recent work [28] developed a de-biasing scheme that allows one to cope with small eigen-gaps in the presence of an i.i.d. symmetric Gaussian noise matrix, the current paper covers a remarkably broader class of noise distributions by exploiting the special property of statistical asymmetry.

Finally, as mentioned before, one might be tempted to first symmetrize the data matrix (i.e. replacing $\boldsymbol{M}$ with $\frac{1}{2}(\boldsymbol{M}+\boldsymbol{M}^\top)$) before computing eigen-decomposition. While [17] has discussed the non-negligible bias of this approach in eigenvalue estimation, it remains unclear whether or not this natural approach suffers from a bias issue as well when estimating the eigenvectors, and, more crucially, whether or not we can still expect reliable eigenvector estimation after symmetrizing the data matrix. Addressing these questions is instrumental in understanding whether asymmetry plays an important role or it merely provides theoretical convenience. While a theory for this is yet to be developed, we shall compare these two approaches numerically and demonstrate the gains of our approach in Section V.

D. Minimax Lower Bounds

To assess the tightness of our statistical guarantees, we develop localized minimax lower bounds on eigenvector estimation, aimed at providing in-depth understanding about how the estimation difficulty changes as a function of certain salient parameters of the problem instance. To derive these lower bounds, we consider the non-asymptotic local minimax framework (see, e.g. [29]–[31]), an approach built upon the concept of the hardest local alternatives that finds its roots in [32].

Before stating our main results, let us first define several sets that are properly localized around the truth. Let $\mathbb{S}^n\subseteq\mathbb{R}^n$ represent the set of $n\times n$ symmetric matrices. Denoting by $\lambda_l(\boldsymbol{A})$ the lth largest eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{u}_l(\boldsymbol{A})$ the associated eigenvector of a symmetric matrix $\boldsymbol{A}$, we define

$$\mathcal{M}_0(\boldsymbol{M}^\star):=\Big\{\boldsymbol{A}\in\mathbb{S}^n\mid \mathsf{rank}(\boldsymbol{A})=r,\ \lambda_i(\boldsymbol{A})=\lambda_i^\star\ (1\le i\le r),\ \|\boldsymbol{A}-\boldsymbol{M}^\star\|_{\mathrm{F}}\le\frac{\sigma_{\min}}{2}\Big\} \tag{20a}$$

$$\mathcal{M}_1(\boldsymbol{M}^\star):=\Big\{\boldsymbol{A}\in\mathbb{S}^n\mid \mathsf{rank}(\boldsymbol{A})=r,\ \lambda_i(\boldsymbol{A})=\lambda_i^\star\ (1\le i\le r),\ \|\boldsymbol{u}_l(\boldsymbol{A})-\boldsymbol{u}_l^\star\|_2\le\frac{\sigma_{\min}}{4|\lambda_l^\star|}\Big\} \tag{20b}$$

$$\mathcal{M}_2(\boldsymbol{M}^\star):=\Big\{\boldsymbol{A}\in\mathbb{S}^n\mid \mathsf{rank}(\boldsymbol{A})=r,\ \lambda_i(\boldsymbol{A})=\lambda_i^\star\ (1\le i\le r),\ \|\boldsymbol{u}_l(\boldsymbol{A})-\boldsymbol{u}_l^\star\|_2\le c_4\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|}\Big\} \tag{20c}$$

for some sufficiently large constant $c_4>0$.

Now we are ready to state our results, the proof of which is deferred to Appendix E.

Theorem 3 (Minimax Lower Bounds): Consider any $1\le l\le r\le n/2$. Suppose that $H_{ij}\overset{\text{ind.}}{\sim}\mathcal{N}(0,\sigma_{ij}^2)$ with $\sigma_{ij}^2\ge\sigma_{\min}^2$, and assume that $4\sigma_{\min}\sqrt{n}\le|\lambda_l^\star|$ and $\sigma_{\min}\le\Delta_l^\star$.

1) There exist some constants $c_0,c_1>0$ such that

$$\inf_{\widehat{\boldsymbol{u}}_l}\sup_{\boldsymbol{A}\in\mathcal{M}_0(\boldsymbol{M}^\star)}\mathbb{E}\Big[\min\|\widehat{\boldsymbol{u}}_l\pm\boldsymbol{u}_l(\boldsymbol{A})\|_2\Big]\ge c_0\frac{\sigma_{\min}}{\Delta_l^\star}; \tag{21a}$$

$$\inf_{\widehat{u}_{\boldsymbol{a},l}}\sup_{\boldsymbol{A}\in\mathcal{M}_0(\boldsymbol{M}^\star)}\mathbb{E}\Big[\min\big|\widehat{u}_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l(\boldsymbol{A})\big|\Big]\ge c_1\left\{\big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big|\frac{\sigma_{\min}^2}{\Delta_l^{\star2}}+\sigma_{\min}\max_{k:k\neq l}\frac{\big|\boldsymbol{a}^\top\boldsymbol{u}_k^\star\big|}{|\lambda_l^\star-\lambda_k^\star|}\right\}. \tag{21b}$$

2) There exists some constant $c_2>0$ such that

$$\inf_{\widehat{u}_{\boldsymbol{a},l}}\sup_{\boldsymbol{A}\in\mathcal{M}_1(\boldsymbol{M}^\star)}\mathbb{E}\Big[\min\big|\widehat{u}_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l(\boldsymbol{A})\big|\Big]\ge c_2\frac{\sigma_{\min}\big\|\boldsymbol{a}-(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)\boldsymbol{a}\big\|_2}{|\lambda_l^\star|}. \tag{21c}$$

3) Then there exists some constant $c_3>0$ such that

$$\inf_{\widehat{\boldsymbol{u}}_l}\sup_{\boldsymbol{A}\in\mathcal{M}_2(\boldsymbol{M}^\star)}\mathbb{E}\Big[\min\|\widehat{\boldsymbol{u}}_l\pm\boldsymbol{u}_l(\boldsymbol{A})\|_2\Big]\ge c_3\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|}. \tag{21d}$$

Here, the infimum in (21a) and (21d) is over all eigenvector estimators, while the infimum in (21b) and (21c) is over all estimators for the linear form of the lth eigenvector.

Remark 6: Note that each part of Theorem 3 captures one bottleneck for the estimation task. Given that enlarging the set of matrices under consideration can only lead to increased minimax lower bound, we can take the supremum over the union $\mathcal{M}_0(\boldsymbol{M}^\star)\cup\mathcal{M}_1(\boldsymbol{M}^\star)\cup\mathcal{M}_2(\boldsymbol{M}^\star)$ in order to yield a minimax lower bound that reflects all these bottlenecks at once.

Consider again the case with $r,\kappa,\mu\asymp 1$ under i.i.d. Gaussian noise, and the case when $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\le(1-\epsilon)\|\boldsymbol{a}\|_2$ for an arbitrarily small constant $\epsilon>0$ (so that $\boldsymbol{a}$ is not perfectly aligned with $\boldsymbol{u}_l^\star$). The theoretical guarantees derived for our estimators (see Theorem 1, or (15) and (18)) match the minimax lower bounds in Theorem 3 up to some logarithmic factor, including both ℓ_2 loss and the risk for estimating an arbitrary linear form $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$. In particular, the dependencies of the estimation risk on $\Delta_l^\star$, $|\lambda_l^\star-\lambda_k^\star|$, and $\{\boldsymbol{a}^\top\boldsymbol{u}_k^\star\}$ in Theorem 1 — which might seem complicated at first glance — are all optimal modulo some log factor. All this confirms the effectiveness and optimality of the proposed estimator in fine-grained eigenvector estimation.

IV. Inference and Uncertainty Quantification

This section moves one step further to the task of statistical inference, with the aim of constructing valid and efficient confidence intervals for the linear functionals $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ and the eigenvalues $\lambda_l^\star$ $(1\le l\le r)$. More precisely, for any target

coverage level $0 < 1-\alpha < 1$, we seek to compute intervals such that

$$\mathbb{P}\left\{\boldsymbol{a}^\top \boldsymbol{u}_l^\star \in [c_{\mathrm{lb}}^{\boldsymbol{a}}, c_{\mathrm{ub}}^{\boldsymbol{a}}]\right\} \approx 1-\alpha \quad \text{or} \quad \mathbb{P}\left\{-\boldsymbol{a}^\top \boldsymbol{u}_l^\star \in [c_{\mathrm{lb}}^{\boldsymbol{a}}, c_{\mathrm{ub}}^{\boldsymbol{a}}]\right\} \approx 1-\alpha. \tag{22}$$

and

$$\mathbb{P}\left\{\lambda_l^\star \in [c_{\mathrm{lb}}^{\lambda}, c_{\mathrm{ub}}^{\lambda}]\right\} \approx 1-\alpha \tag{23}$$

Here, we have taken into account the difficulty in distinguishing $\boldsymbol{u}_l^\star$ from $-\boldsymbol{u}_l^\star$ using only the data $\boldsymbol{M}$. As we shall see, providing precise confidence intervals for the linear functional is much more challenging than the estimation task, and our theory requires additional assumptions beyond Assumption 2 in order for our procedures to have exact coverage.

A. Algorithms

1) Confidence Intervals for Linear Forms of Eigenvectors: We start by constructing confidence intervals for the linear form $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$. Towards this, two ingredients are needed: (1) a nearly unbiased estimate of $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$, and (2) a valid length of the interval (which is typically built upon a certain variance estimate). We describe these ingredients as follows.

a) A modified nearly unbiased estimator $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$: While the estimator $\widehat{u}_{\boldsymbol{a},l}$ (cf. (11)) discussed previously enjoys minimax optimal statistical accuracy, we find it more convenient to work with a modified estimator when conducting uncertainty quantification, particularly in the regime when $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is very small.[3] More specifically, we introduce

$$\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}} := \begin{cases} \frac{1}{2}\boldsymbol{a}^\top\big(\boldsymbol{u}_l + \boldsymbol{w}_l\big), & \text{if } \big|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\big| \text{ is "small"}, \\ \widehat{u}_{\boldsymbol{a},l}, & \text{otherwise.} \end{cases} \tag{24}$$

We shall make precise in Algorithm 1 what it means by "small" in a practical and data-dependent manner. As before, the proposed procedures are fully data-driven, without requiring prior knowledge about the noise distributions. A few informal yet important remarks are in order.

- When $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is very "small" in magnitude, both $\boldsymbol{a}^\top \boldsymbol{u}_l$ and $\boldsymbol{a}^\top \boldsymbol{w}_l$ serve as unbiased estimates of $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$, and the averaging operation further reduces the uncertainty. Here, the quantity $\sqrt{\widehat{v}_{a,l}}$ in Algorithm 1 captures the level of the smallest possible uncertainty when estimating $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.
- When $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is not very "small", the procedure is identical to estimator $\widehat{u}_{\boldsymbol{a},l}$ proposed previously.

b) Construction of confidence intervals: As it turns out, the estimation error of the modified estimator $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$ is well approximated by a zero-mean Gaussian random variable with tractable variance $v_{\boldsymbol{a},l}^\star$ (to be formalized shortly). Motivated by this observation, we propose to first obtain an estimate of the variance of $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$ — denoted by $\widehat{v}_{\boldsymbol{a},l}$. The proposed confidence interval for any given coverage level $0 < 1-\alpha < 1$ is then given by

$$\mathsf{CI}_{1-\alpha}^{\boldsymbol{a}} := \left[\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{v}_{\boldsymbol{a},l}}\right], \tag{25}$$

[3]More precisely, in the regime where $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is very small, the uncertainty of $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$ is nearly Gaussian, while that of $\widehat{u}_{\boldsymbol{a},l}$ is non-Gaussian and more complicated to describe.

where we abbreviate $[b \pm c] := [b-c, b+c]$, and $\Phi(\cdot)$ is the CDF of a standard Gaussian.

c) Variance estimates: We shall take a moment to discuss how to obtain the variance estimate $\widehat{v}_{\boldsymbol{a},l}$. As will be seen momentarily, the proposed estimator $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$ admits — modulo some global sign — the following first-order approximation for a broad range of settings (up to global sign):

$$\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}} \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \underbrace{\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}\big(\boldsymbol{H}+\boldsymbol{H}^\top\big)\boldsymbol{u}_l^\star}_{\text{uncertainty term}}, \tag{26}$$

where $\boldsymbol{a}_l^\perp := \boldsymbol{a} - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)\boldsymbol{u}_l^\star$. For a broad family of noise distributions, the uncertainty term is approximately zero-mean Gaussian with variance

$$\begin{aligned} v_{\boldsymbol{a},l}^\star &:= \mathsf{Var}\left[\frac{1}{2\lambda_l^\star}(\boldsymbol{a}_l^\perp)^\top\big(\boldsymbol{H}+\boldsymbol{H}^\top\big)\boldsymbol{u}_l^\star\right] \\ &= \frac{1}{4\lambda_l^{\star 2}}\sum_{1\le i,j\le n}\big(a_{l,i}^\perp u_{l,j}^\star + a_{l,j}^\perp u_{l,i}^\star\big)^2\sigma_{ij}^2, \end{aligned} \tag{27}$$

where $a_{l,j}^\perp$ denotes the jth entry of $\boldsymbol{a}_l^\perp$. At first glance, evaluating this variance quantity precisely requires prior knowledge about the noise level in addition to the truth $(\boldsymbol{u}_l^\star, \lambda_l^\star)$. To enable a model-agnostic and data-driven estimate of $v_{\boldsymbol{a},l}^\star$, we make the observation that $\frac{1}{4\lambda_l^{\star 2}}\sum_{i,j}\big(a_{l,i}^\perp u_{l,j}^\star + a_{l,j}^\perp u_i^\star\big)^2 H_{ij}^2$ is very close to $v_{\boldsymbol{a},l}^\star$ based on the concentration of measure phenomenon. This in turn motivates us to estimate $v_{\boldsymbol{a},l}^\star$ via a plug-in approach: (1) replacing $\boldsymbol{u}_l^\star$ with an estimate $\widehat{\boldsymbol{u}}_l$, (2) replacing $\boldsymbol{a}_l^\perp$ with a plug-in estimate $\widehat{\boldsymbol{a}}_l^\perp$, (3) using λ_l in place of $\lambda_l^\star$, and (4) replacing H_{ij} with an estimate $\widehat{H}_{ij}$, where

$$\widehat{\boldsymbol{u}}_l := \tfrac{1}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}(\boldsymbol{u}_l+\boldsymbol{w}_l) \quad \text{and} \quad \widehat{\boldsymbol{H}} = [\widehat{H}_{ij}]_{1\le i,j\le n} := \boldsymbol{M} - \boldsymbol{M}_{\mathsf{svd},r}, \tag{28}$$

with $\boldsymbol{M}_{\mathsf{svd},r} := \arg\min_{\mathrm{rank}(\boldsymbol{Z})\le r}\|\boldsymbol{M}-\boldsymbol{Z}\|_{\mathrm{F}}$ the best rank-r approximation of $\boldsymbol{M}$. As a byproduct, this variance estimate in turn allows us to specify whether $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is "small" (the case where $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is comparable to or smaller than the typical size of the uncertainty component).

Remark 7: For estimating $\boldsymbol{H}$ (or equivalently, $\boldsymbol{M}^\star$), it has been shown that the SVD-based approach achieves appealing entrywise accuracy (e.g. [9], [33]).

For ease of reference, the proposed procedure is summarized in Algorithm 1. The computational cost of this procedure mostly lies in computing the eigen-decomposition and the SVD of $\boldsymbol{M}$.

2) Confidence Intervals for Eigenvalues: Moving beyond linear forms of eigenvectors, one might also be interested in performing inference on the eigenvalues of interest. As it turns out, this task is simpler than inferring linear functionals of eigenvectors; one can simply estimate $\lambda_l^\star$ via the lth eigenvalue λ_l (see Notation 1) and compute a confidence interval based on the distributional characterization of λ_l. There is absolutely no need for careful de-biasing, since the eigen-decomposition approach automatically exploits the asymmetry structure of noise to suppress bias. Similar to Section IV-A1, our procedure for performing inference on $\lambda_l^\star$ is distribution-free (i.e. it does not require prior knowledge about the noise variance) and adaptive to heteroscedasticity of noise.

Algorithm 1 Constructing a Confidence Interval for the Linear Form $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

1: Compute the lth eigenvalue λ_l of $\boldsymbol{M}$, and the associated right eigenvector $\boldsymbol{u}_l$ and left eigenvector $\boldsymbol{w}_l$ such that $\mathrm{Re}(\boldsymbol{w}_l^\top \boldsymbol{u}_l) \geq 0$ (see Notation 1).
2: Compute the following estimator

$$\widehat{u}_{\boldsymbol{a},l} := \min\left\{\sqrt{\left|\frac{1}{\boldsymbol{w}_l^\top \boldsymbol{u}_l}\left(\boldsymbol{a}^\top \boldsymbol{u}_l\right)\left(\boldsymbol{a}^\top \boldsymbol{w}_l\right)\right|}, \|\boldsymbol{a}\|_2\right\},$$

$$\widehat{u}_{\boldsymbol{a},l}^{\text{modified}} := \begin{cases} \frac{1}{2}\boldsymbol{a}^\top\left(\boldsymbol{u}_l + \boldsymbol{w}_l\right), & \text{if } \left|\boldsymbol{a}^\top \boldsymbol{u}_l\right| \leq c_{\mathrm{b}}\sqrt{\widehat{v}_{\boldsymbol{a},l}}\log^{1.5} n, \\ \widehat{u}_{\boldsymbol{a},l}, & \text{else,} \end{cases} \tag{29}$$

where $c_{\mathrm{b}} > 0$ is some sufficiently large constant. Here, $\widehat{v}_{\boldsymbol{a},l} = \text{ESTIMATE-VARIANCE}\,\widehat{\boldsymbol{a}}_l^\perp, \widehat{\boldsymbol{u}}_l, \lambda_l$ with $\widehat{\boldsymbol{a}}_l^\perp := \boldsymbol{a} - (\boldsymbol{a}^\top \widehat{\boldsymbol{u}}_l)\widehat{\boldsymbol{u}}_l$ and $\widehat{\boldsymbol{u}}_l := \frac{1}{\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2}(\boldsymbol{u}_l + \boldsymbol{w}_l)$.
3: For any prescribed coverage level $1-\alpha$, compute the confidence interval as

$$\mathsf{CI}_{1-\alpha}^{\boldsymbol{a}} := \left[\widehat{u}_{\boldsymbol{a},l}^{\text{modified}} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{v}_{\boldsymbol{a},l}}\right]. \tag{30}$$

1: **function** ESTIMATE-VARIANCE($\boldsymbol{a}, \boldsymbol{u}, \lambda$)
2: Compute $\widehat{\boldsymbol{H}} := \boldsymbol{M} - \boldsymbol{M}_{\mathsf{svd},r}$ with $\boldsymbol{M}_{\mathsf{svd},r} := \arg\min_{\mathsf{rank}(\boldsymbol{Z})\leq r}\|\boldsymbol{M} - \boldsymbol{Z}\|_{\mathrm{F}}$.
3: **return** $v := \frac{1}{4\lambda^2}\sum_{1\leq i,j\leq n}\left(a_i u_j + a_j u_i\right)^2 \widehat{H}_{ij}^2$.

To be more specific, a crucial observation, which we will make precise shortly, is as follows

$$\lambda_l \;\approx\; \lambda_l^\star + \boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star, \tag{31}$$

where the uncertainty term $\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star$ is approximately Gaussian with variance

$$v_{\lambda,l}^\star := \mathsf{Var}\left[\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star\right] = \sum_{1\leq i,j\leq n}\left(u_{l,i}^\star u_{l,j}^\star\right)^2\sigma_{ij}^2. \tag{32}$$

Similar to the estimator $\widehat{v}_{\boldsymbol{a},l}$, we proposee to estimate $v_{\lambda,l}^\star$ via the following estimator

$$\widehat{v}_{\lambda,l} := \sum_{1\leq i,j\leq n}\left(\widehat{u}_{l,i}\widehat{u}_{l,j}\right)^2\widehat{H}_{ij}^2, \tag{33}$$

with $\widehat{\boldsymbol{u}}_l$ and $\widehat{\boldsymbol{H}}$ defined in (28). All this suggests the following confidence interval for $\lambda_l^\star$:

$$\mathsf{CI}_{1-\alpha}^{\lambda} := \left[\lambda_l \pm \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{v}_{\lambda,l}}\right]. \tag{34}$$

See Algorithm 2 for the complete procedure for performing inference on $\lambda_l^\star$.

B. Theoretical Guarantees

We now set out to justify the validity of the proposed confidence intervals. Before proceeding, we need to impose another set of assumptions on the noise levels:

Algorithm 2 Constructing a Confidence Interval for $\lambda_l^\star$

1: Compute the lth eigenvalue λ_l of $\boldsymbol{M}$, and the associated right eigenvector $\boldsymbol{u}_l$ and left eigenvector $\boldsymbol{w}_l$ such that $\boldsymbol{w}_l^\top \boldsymbol{u}_l \geq 0$ (see Notation 1).
2: For any prescribed coverage level $1-\alpha$, compute the confidence interval as

$$\mathsf{CI}_{1-\alpha}^{\lambda} := \left[\lambda_l \pm \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{v}_{\lambda,l}}\right]. \tag{35}$$

Here, $\widehat{v}_{\lambda,l} = \text{ESTIMATE-VARIANCE}\,\widehat{\boldsymbol{u}}_l, \widehat{\boldsymbol{u}}_l, 1$ with $\widehat{\boldsymbol{u}}_l := \frac{1}{\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2}(\boldsymbol{u}_l + \boldsymbol{w}_l)$.

Assumption 3: Suppose that the noise parameters defined in Assumption 1 satisfy

$$\sigma_{\max}\kappa^3\sqrt{\mu^3 r^3}\log^{1.5} n = o(\Delta_l^\star), \tag{36a}$$

$$\sigma_{\max}\mu\kappa^2 r^2\sqrt{n}\log^{1.5} n = o(\lambda_{\min}^\star), \tag{36b}$$

$$\frac{\sigma_{\max}}{\sigma_{\min}} \asymp 1, \quad B = o\left(\sigma_{\min}\sqrt{\frac{n}{\mu\log n}}\right). \tag{36c}$$

Armed with the above assumptions, we are ready to present theoretical guarantees.

Theorem 4 (Validity of Confidence Intervals (Rank-r)): Assume $\boldsymbol{M}^\star$ is rank-r and μ-incoherent with $\mu^4\kappa^8 r^2\log^2 n = o(n)$. Given any integer $1 \leq l \leq r$, under Assumptions 1 and 3, the confidence interval returned by Algorithm 2 obeys

$$\mathbb{P}\left\{\lambda_l^\star \in \mathsf{CI}_{1-\alpha}^{\lambda}\right\} = 1 - \alpha + o(1) \tag{37}$$

uniformly over all $0 < \alpha < 1$. In addition, fix an arbitrarily small constant $0 < \epsilon < 1$, and consider any fixed vector $\boldsymbol{a}$ with $\|\boldsymbol{a}\|_2 = 1$ obeying

$$\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \leq 1 - \epsilon, \quad \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| = o\left(\frac{\Delta_l^{\star 2}}{|\lambda_l^\star|\sigma_{\max}\kappa^4 r^2\mu\log n}\right), \tag{38a}$$

$$\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right| = o\left(\frac{|\lambda_l^\star - \lambda_k^\star|}{|\lambda_l^\star|\sqrt{\mu\kappa^4 r^3\log n}}\right), \qquad \forall k \neq l. \tag{38b}$$

Then the confidence interval returned by Algorithms 1 obeys

$$\begin{aligned} &\mathbb{P}\left\{\boldsymbol{a}^\top \boldsymbol{u}_l^\star \in \mathsf{CI}_{1-\alpha}^{\boldsymbol{a}}\right\} = 1 - \alpha + o(1) \quad \text{or} \\ &\mathbb{P}\left\{-\boldsymbol{a}^\top \boldsymbol{u}_l^\star \in \mathsf{CI}_{1-\alpha}^{\boldsymbol{a}}\right\} = 1 - \alpha + o(1) \end{aligned} \tag{39}$$

uniformly over all $0 < \alpha < 1$.

We immediately make note of a direct consequence of Theorem 4 (by taking $\boldsymbol{a}$ to be the kth standard basis vector $\boldsymbol{e}_k$), which concerns entrywise confidence intervals for $\boldsymbol{u}_l^\star$.

Corollary 1 (Validity of Entrywise Confidence Intervals (Rank-r)): Assume that $\boldsymbol{M}^\star$ is rank-r and μ-incoherent with $\mu^4 = o(n/\log n)$. Suppose that Assumptions 1 and 3 hold and that $\sqrt{\frac{\mu^2\kappa^6 r^6\log n}{n}}\left|\lambda_l^\star\right| = o\left(\Delta_l^\star\right)$. Then the confidence interval constructed in Algorithms 1 with $\boldsymbol{a} = \boldsymbol{e}_k$ satisfies

$$\begin{aligned} &\mathbb{P}\left\{u_{l,k}^\star \in \mathsf{CI}_{1-\alpha}^{\boldsymbol{e}_k}\right\} = 1 - \alpha + o(1) \quad \text{or} \\ &\mathbb{P}\left\{-u_{l,k}^\star \in \mathsf{CI}_{1-\alpha}^{\boldsymbol{e}_k}\right\} = 1 - \alpha + o(1); \end{aligned} \tag{40}$$

this holds true uniformly over all $0 < \alpha < 1$.

The above results confirm the validity of our inferential procedure in high dimension. Our theory applies to a very broad family of noise distributions, without the need of any prior knowledge about detailed noise distributions or noise levels $\{\sigma_{ij}\}$. The results are fully adaptive to heteroscedasticity of noise, making them appealing for practical scenarios. In the sequel, we discuss several important implications of our results in a more quantitative manner. For simplicity of discussion, we shall concentrate on the case where $r, \kappa, \mu \asymp 1$ and consider Gaussian noise with $\sigma_{\max} \asymp \sigma_{\min}$. We shall also assume $\|\boldsymbol{a}\|_2 = 1$ without loss of generality.

- We start with the rank-1 case (i.e. $r = 1$), in which we have $\Delta_l^\star = \infty$ according to the definition (12). In this case, the conditions (38) admit considerable simplification:
$$|\boldsymbol{a}^\top \boldsymbol{u}_1^\star| \le 1 - \epsilon \tag{41}$$
for any small constant $\epsilon > 0$. This means that our theory covers a very wide range of linear functionals $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$; basically, the proposed confidence interval finds theoretical support unless the preconceived direction $\boldsymbol{a}$ is already highly aligned with the truth $\boldsymbol{u}_1^\star$.
- Going beyond the rank-1 case, the eigen-gap condition imposed in Assumption 3 and Corollary 1 reads
$$\Delta_l^\star \gtrsim \sigma_{\max} \mathrm{poly}\log n \asymp \frac{\|\boldsymbol{H}\| \mathrm{poly}\log n}{\sqrt{n}}, \tag{42}$$
which is allowed to be substantially smaller than the spectral norm $\|\boldsymbol{H}\|$ of the noise matrix.
- Under such eigen-gap requirements, we develop an informative distributional theory underlying the proposed estimators $\widehat{u}_{\boldsymbol{a},l}$, provided that $\{\boldsymbol{a}^\top \boldsymbol{u}_k^\star\}_{1\le k\le r}$ are not too large in the sense that
$$|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \lesssim \frac{{\Delta_l^\star}^2}{|\lambda_l^\star| \sigma_{\max} \log^{1.5} n}; \qquad |\boldsymbol{a}^\top \boldsymbol{u}_k^\star| \lesssim \frac{|\lambda_l^\star - \lambda_k^\star|}{|\lambda_l^\star| \log n}, \quad k \ne l. \tag{43}$$
In words, the condition (43) requires that both the target quantity $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ and the "interferers" are dominated by the respective (normalized) eigen-gaps. An important instance that automatically satisfies such conditions has been singled out in Corollary 1, leading to appealing entrywise distributional characterizations and inferential procedures.
- While our theorems focus on asymmetric noise matrices, one can combine them with a simple asymmetrization trick to yield valid confidence intervals when $\boldsymbol{H}$ is a symmetric matrix with homoscedastic Gaussian noise. See Appendix G for detailed discussion.

We note, however, that the additional requirement (43) makes our inference results less general than our estimation guarantees in Section III-C, except for the rank-1 case. In fact, if (43) is violated, then the influence of the eigen-gaps might become the dominant factor in the uncertainty term and needs to be quantified in a precise fashion. Achieving this calls for more refined theoretical analysis, and we leave it to future investigation.

C. Key Ingredients Behind Theorem 4

Next, we single out two key ingredients that shed light on not only the validity of, but also the statistical efficiency of, our inferential procedures. To be specific, Theorem 4 is mainly built upon distributional guarantees developed for the proposed estimator $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$ and the lth eigenvalue λ_l of $\boldsymbol{M}$. In a nutshell, the quantity $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$ (resp. λ_l) is a nearly unbiased estimator of $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ (resp. $\lambda_l^\star$) and is approximately Gaussian. We formalize this distributional theory as follows, whose proof is deferred to Appendix C.

Theorem 5 (Distributional Characterization (Rank-r)): Instate the assumptions of Theorem 4. Let $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$ (cf. Algorithm 1) be the proposed estimate for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$. Then one can write
$$\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}} = \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \sqrt{v_{\boldsymbol{a},l}^\star} W_{\boldsymbol{a},l} + \zeta \quad \text{or} \quad -\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}} = \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \sqrt{v_{\boldsymbol{a},l}^\star} W_{\boldsymbol{a},l} + \zeta \tag{44a}$$
and
$$\lambda_l = \lambda_l^\star + \sqrt{v_{\lambda,l}^\star} W_{\lambda,l} + \xi, \tag{44b}$$
where $v_{\boldsymbol{a},l}^\star$ and $v_{\lambda,l}^\star$ are defined respectively in (27) and (32), and
$$W_{\boldsymbol{a},l} := \frac{(\boldsymbol{a}_l^\perp)^\top \left(\boldsymbol{H} + \boldsymbol{H}^\top\right) \boldsymbol{u}_l^\star}{2\lambda_l^\star \sqrt{v_{\boldsymbol{a},l}^\star}} \quad \text{and} \quad W_{\lambda,l} := \frac{\boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star}{\sqrt{v_{\lambda,l}^\star}} \tag{45}$$
with $\boldsymbol{a}_l^\perp := \boldsymbol{a} - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)\boldsymbol{u}_l^\star$. The residual terms obey $|\zeta| = o\big(\sqrt{v_{\boldsymbol{a},l}^\star}\big)$ and $|\xi| = o\big(\sqrt{v_{\lambda,l}^\star}\big)$ with probability at least $1 - O(n^{-5})$. In addition, one has
$$\sup_{z\in\mathbb{R}} \big|\, \mathbb{P}(W_{\boldsymbol{a},l} \le z) - \Phi(z) \,\big| = o(1) \quad \text{and} \quad \sup_{z\in\mathbb{R}} \big|\, \mathbb{P}(W_{\lambda,l} \le z) - \Phi(z) \,\big| = o(1). \tag{46}$$

Informally, this theorem reveals the tightness of the following first-order approximation (up to global sign)
$$\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}} \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{1}{2\lambda_l^\star} (\boldsymbol{a}_l^\perp)^\top (\boldsymbol{H} + \boldsymbol{H}^\top) \boldsymbol{u}_l^\star \quad \text{and} \quad \lambda_l \approx \lambda_l^\star + \boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star \tag{47}$$
for a wide range of directions $\boldsymbol{a}$. In addition, the first-order approximations are close in distribution to Gaussian random variables. Such distributional characterizations would immediately lead to $(1-\alpha)$-confidence intervals for any α, if an oracle had informed us of the quantities $v_{\boldsymbol{a},l}^\star$ and $v_{\lambda,l}^\star$. Fortunately, the proposed $\widehat{v}_{\boldsymbol{a},l}$ and $\widehat{v}_{\lambda,l}$ (see Algorithms 1 and 2) serve as highly accurate estimates of $v_{\boldsymbol{a},l}^\star$ and $v_{\lambda,l}^\star$, respectively, and can be employed in place of $v_{\boldsymbol{a},l}^\star$ and $v_{\lambda,l}^\star$. This observation is asserted by the following theorem; the proof is postponed to Appendix D.

Theorem 6 (Accuracy of Variance Estimates (Rank-r)): Instate the assumptions of Theorem 4. With probability at least $1 - O(n^{-10})$, the variance estimators proposed in Algorithms 1-2 satisfy
$$\widehat{v}_{\boldsymbol{a},l} = (1 + o(1)) v_{\boldsymbol{a},l}^\star \quad \text{and} \quad \widehat{v}_{\lambda,l} = (1 + o(1)) v_{\lambda,l}^\star. \tag{48}$$

Clearly, putting Theorems 5-6 together immediately establishes Theorem 4.

V. Numerical Experiments

This section consists of numerical experiments in various settings, in order to demonstrate the performance of our estimation and inference procedures and their accompanying theory.

A. Eigen-Decomposition After Symmetrization?

As mentioned in the discussions after Theorem 2, one may consider first symmetrizing the data matrix with $\frac{1}{2}(\boldsymbol{M}+\boldsymbol{M}^\top)$ before computing the eigen-decomposition. It is unclear whether this procedure provides sensible eigenvector estimators in the presence of small eigen-gaps and heteroscedastic noise. This subsection is devoted to understanding the potential sub-optimality of this approach via a simple example. Specifically, consider the rank-2 model where

$$\boldsymbol{M}^\star = \lambda_1^\star \boldsymbol{u}_1^\star \boldsymbol{u}_1^{\star\top} + \lambda_2^\star \boldsymbol{u}_2^\star \boldsymbol{u}_2^{\star\top}, \tag{49}$$

with $\boldsymbol{H}$ satisfying our usual assumptions. To simplify presentation, we define

1. Spectral-asym: eigen-decomposition applied to the observed asymmetric data matrix $\boldsymbol{M}$;
2. Spectral-sym: eigen-decomposition applied to the symmetrized data matrix $\frac{1}{2}\left(\boldsymbol{M}+\boldsymbol{M}^\top\right)$.

a) A Numerical Example With Heteroscedastic Gaussian Noise: We start by looking at an example with

$$\boldsymbol{u}_1^\star = \frac{1}{\sqrt{n}}\mathbf{1}_n \quad \text{and} \quad \boldsymbol{u}_2^\star = \frac{1}{\sqrt{n}}\begin{bmatrix}\mathbf{1}_{n/2}\\ -\mathbf{1}_{n/2}\end{bmatrix}. \tag{50}$$

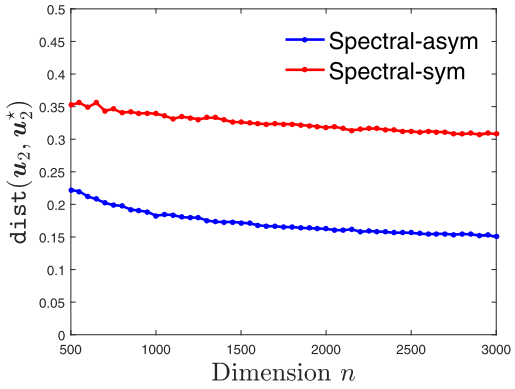
The noise matrix $\boldsymbol{H}$ is assumed to have independent zero-mean Gaussian entries with variance

$$\begin{aligned}\mathsf{Var}(\boldsymbol{H}) &:= \left[\mathsf{Var}(H_{ij})\right]_{1\le i,j\le n}\\ &= \sigma_1^2\begin{bmatrix}\mathbf{1}_{n/2}\mathbf{1}_{n/2}^\top - \frac{1}{2}\boldsymbol{I}_{n/2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} + \sigma_2^2\left(\mathbf{1}_n\mathbf{1}_n^\top - \frac{1}{2}\boldsymbol{I}_n\right),\end{aligned} \tag{51}$$

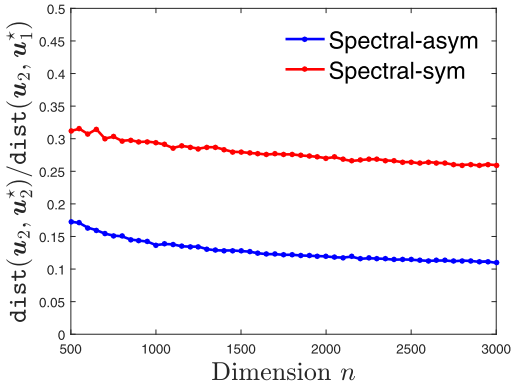
and hence the variance of the symmetrized data satisfies

$$\begin{aligned}\mathsf{Var}\left(\frac{\boldsymbol{H}+\boldsymbol{H}^\top}{2}\right) &:= \left[\mathsf{Var}\left(\tfrac{1}{2}(H_{ij}+H_{ji})\right)\right]_{1\le i,j\le n}\\ &= \frac{\sigma_1^2}{2}\begin{bmatrix}\mathbf{1}_{n/2}\mathbf{1}_{n/2}^\top & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} + \frac{\sigma_2^2}{2}\mathbf{1}_n\mathbf{1}_n^\top.\end{aligned} \tag{52}$$

We plot in Fig. 1 the numerical performance of both Spectral-asym and Spectral-sym in estimating $\boldsymbol{u}_2^\star$. For Spectral-asym, the estimator $\widehat{\boldsymbol{u}}_2$ is constructed as in the expression (10), whereas the second eigenvector of $\frac{1}{2}\left(\boldsymbol{M}+\boldsymbol{M}^\top\right)$ is used for Spectral-sym. As can be seen, Spectral-asym strictly outperforms Spectral-sym in all cases, thus unveiling the real benefits of exploiting asymmetry in eigen-decomposition. The interested reader is deferred to Appendix H for some high-level interpretation of this phenomenon.



(a) ℓ_2 estimation error $\mathsf{dist}(\boldsymbol{u}_2, \boldsymbol{u}_2^\star)$



(b) relative ℓ_2 error $\mathsf{dist}(\boldsymbol{u}_2, \boldsymbol{u}_2^\star)/\mathsf{dist}(\boldsymbol{u}_2, \boldsymbol{u}_1^\star)$

Fig. 1. Numerical performance of Spectral-asym vs. Spectral-sym when $\lambda_1^\star = 1$ and $\lambda_2^\star = 0.95$. Here, we define $\mathsf{dist}(\boldsymbol{u}, \boldsymbol{v}) := \min\{\|\boldsymbol{u}-\boldsymbol{v}\|_2, \|\boldsymbol{u}+\boldsymbol{v}\|_2\}$. (a) plots the ℓ_2 estimation error vs. the dimension n, whereas (b) plots the relative ℓ_2 error vs. n. For each n, the results are averaged over 1000 trials, with $\sigma_1 = 1/\sqrt{n\log n}$ and $\sigma_2 = 0.1/\sqrt{n\log n}$ (see the expression (51)).

B. Estimation

This section is devoted to numerically studying the efficiency of our estimators for linear functionals of the eigenvectors. Given any fixed $\boldsymbol{a}$ with $\|\boldsymbol{a}\|_2 = 1$, our estimator is constructed as in the expression (11). Theorem 1 (in particular, the upper bound (14c)) together with Theorem 3 implies that the estimation error is controlled by the eigen-gap, the true signal strength $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|$, and the magnitude of the "interferers". In the following, we examine qualitatively the effects of these quantities upon the estimation errors through some simple yet representative examples.

a) Settings: Consider the rank-2 case as in Eq. (49) again, where the leading eigenvalue is set to be 1 and the orthonormal pair $\boldsymbol{u}_1^\star$ and $\boldsymbol{u}_2^\star$ are randomly generated. We focus on estimating the linear functionals of the form $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$. In the following, several quantities that are selected to examine how they affect the estimation error $|\widehat{u}_{\boldsymbol{a},2} - \boldsymbol{a}^\top \boldsymbol{u}_2^\star|$ include the ground-truth $|\boldsymbol{a}^\top \boldsymbol{u}_2^\star|$, the interferer $|\boldsymbol{a}^\top \boldsymbol{u}_1^\star|$, and the eigen-gap $\Delta_2^\star = \lambda_1^\star - \lambda_2^\star$. In the following, we consider two scenarios: the case where the observation matrix is generated from a rank-2 underlying matrix $\boldsymbol{M}^\star$ plus heteroscedastic Gaussian noise; and the case where entries are missing at random with sampling rate 0.1 on top of the above-mentioned observation model (more details are given below).

b) Heteroscedastic Gaussian noise: Consider a heteroscedastic Gaussian noise matrix $\boldsymbol{H}$ with independent entries $H_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)$ obeying

$$\mathsf{Var}(\boldsymbol{H}) := \big[\mathsf{Var}(H_{ij})\big]_{1\le i,j\le n} = \begin{bmatrix} \sigma_1^2 \\ (\sigma_1+\delta_\sigma)^2 \\ \vdots \\ (\sigma_1+(n-1)\delta_\sigma)^2 \end{bmatrix} \cdot \mathbf{1}_n^\top, \tag{53}$$

where δ_σ determines the spacing between adjacent standard deviation. The parameters chosen to be

$$\sigma_1 = 0.1/\sqrt{n\log n}, \quad \delta_\sigma = 0.9/((n-1)\sqrt{n\log n}).$$

c) Missing data model: Suppose that we only get to observe a fraction of the entries of $\boldsymbol{M}^\star$; more precisely, each entry of $\boldsymbol{M}^\star$ is observed independently with probability p. In this setting, we can take the observed data matrix via zero-padding and rescaling as follows

$$M_{ij} = \begin{cases} \frac{1}{p}(M_{ij}^\star + \sqrt{p}\widetilde{H}_{ij}), & \text{with probability } p, \\ 0, & \text{otherwise.} \end{cases} \tag{54}$$

This way we guarantee that $\mathbb{E}[\boldsymbol{M}] = \boldsymbol{M}^\star$, and $\widetilde{\boldsymbol{H}} = [\widetilde{H}_{ij}]_{1\le i,j\le n}$ is a heteroscedastic Gaussian noise with variance $\mathsf{Var}(\widetilde{\boldsymbol{H}}) := \big[\mathsf{Var}(\widetilde{H}_{ij})\big]_{1\le i,j\le n}$ equal to (53). In this case, the matrix $\boldsymbol{H} := \boldsymbol{M} - \boldsymbol{M}^\star$ obeys

$$|H_{ij}| \le \left|\frac{1}{p}M_{ij}^\star\right| + \left|\frac{1}{\sqrt{p}}\widetilde{H}_{ij}\right| \lesssim \frac{\mu}{np} + \frac{\widetilde{\sigma}_{\max}\sqrt{\log n}}{\sqrt{p}},$$
$$\mathbb{E}[H_{ij}^2] \lesssim \frac{\mu}{n^2\, p} + \widetilde{\sigma}_{\max}^2$$

with high probability, where $\widetilde{\sigma}_{\max} := \sigma_1 + (n-1)\delta_\sigma$. When the sampling rate exceeds $p \ge \frac{c_{\mathrm{p}}\mu\log^2 n}{n}$ and the noise size is below $\widetilde{\sigma}_{\max} \le \frac{1}{\sqrt{c_{\mathrm{p}} n\log n}}$ for some constant $c_{\mathrm{p}} \ge 1$, one has

$$|H_{ij}| \lesssim \frac{\mu}{np} + \frac{\widetilde{\sigma}_{\max}\sqrt{\log n}}{\sqrt{p}} \lesssim \frac{1}{c_{\mathrm{p}}\log n},$$
$$\sqrt{\mathbb{E}[H_{ij}^2]} \lesssim \frac{\sqrt{\mu}}{n\sqrt{p}} + \widetilde{\sigma}_{\max} \lesssim \frac{1}{\sqrt{c_{\mathrm{p}} n\log n}},$$

thus satisfying Assumption 2 for c_{p} sufficiently large. In the numerical experiments here, we shall choose $p = 0.1$ and

$$\sigma_1 = 1/\sqrt{10\ n\log n}, \quad \delta_\sigma = 9/((n-1)\sqrt{10\ n\log n}).$$

d) Estimation error vs. size of the ground-truth: To study the qualitative effect of the magnitude of the group-truth, various values of $|\boldsymbol{a}^\top\boldsymbol{u}_2^\star|$ are considered. For each configuration, the dimension n is set to be 500 and we run 100 trials for the box plot. The experiments are conducted for a variety of eigen-gaps and directions $\boldsymbol{a}$. A clear positive correspondence is seen between the size $|\boldsymbol{a}^\top\boldsymbol{u}_2^\star|$ and the estimation error; see Fig. 2.

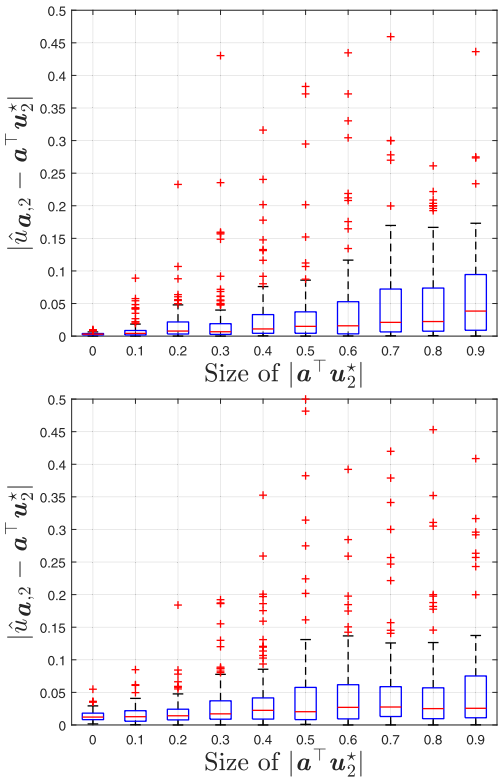


Fig. 2. Estimation error vs. size of the ground-truth $|\boldsymbol{a}^\top\boldsymbol{u}_2^\star|$. The left figure corresponds to the noisy observation case where $|\Delta_2^\star| = 0.01$ and the right figure corresponds to the missing observation case where $|\Delta_2^\star| = 0.05$. In both cases, $\boldsymbol{a}$ is chosen such that $\boldsymbol{a}^\top\boldsymbol{u}_1^\star = 0$.

e) Estimation error vs size of the interferer: To study the qualitative effect of the magnitude of the interferers, we consider a range of values for $|\boldsymbol{a}^\top\boldsymbol{u}_1^\star|$ while holding $|\boldsymbol{a}^\top\boldsymbol{u}_2^\star| = 0.5$ unchanged. Again, for each configuration, the dimension n is set to be 500 and we run 100 trials for the box plot. The experiments are run with various values of eigen-gaps. A negative dependency of the estimation error on the interferer is observed. In particular, Fig. 3 illustrates how the estimation errors grow as the interferer gets stronger.

f) Estimation error vs eigen-gap: In this part, we consider various magnitudes of the eigen-gap and study how the estimation error is influenced. Similar to what Theorem 1 and Theorem 3 predict, the estimation task becomes easier when the eigen-gap gets larger. Fig. 4 manifests this relationship in the case when $|\boldsymbol{a}^\top\boldsymbol{u}_1^\star| = 0.5$ and $|\boldsymbol{a}^\top\boldsymbol{u}_2^\star| = 0.2$.

C. Uncertainty Quantification

This subsection performs numerical experiments to validate our inferential procedures as well as the accompanying theory. Throughout this subsection, we consider the rank-2 case again as in the expression (11), where $\boldsymbol{M}^\star := \lambda_1^\star\boldsymbol{u}_1^\star\boldsymbol{u}_1^{\star\top} + \lambda_2^\star\boldsymbol{u}_2^\star\boldsymbol{u}_2^{\star\top}$. The ground truth $\boldsymbol{M}^\star$ is produced by setting $\lambda_1^\star = 1$ with the corresponding eigenvectors $\boldsymbol{u}_1^\star$ and $\boldsymbol{u}_2^\star$ randomly generated. We aim to construct 95% confidence intervals (namely, $\alpha = 0.05$) for

1. the linear form $\boldsymbol{a}^\top\boldsymbol{u}_2^\star$ for a given vector $\boldsymbol{a}$;
2. the eigenvalue $\lambda_2^\star$.

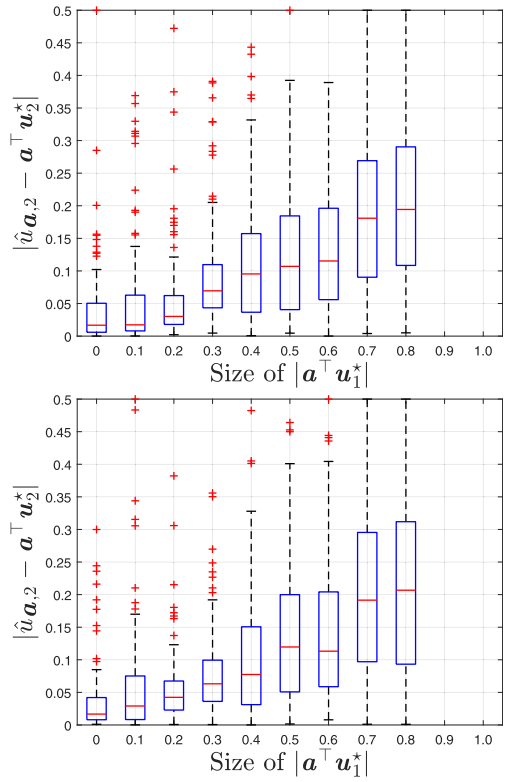


Fig. 3. Estimation error vs. size of the interferer $|\boldsymbol{a}^\top \boldsymbol{u}_1^\star|$. The left figure corresponds to the full observation case where $|\Delta_2^\star| = 0.01$, while the right figure corresponds to the missing observation case where $|\Delta_2^\star| = 0.05$.

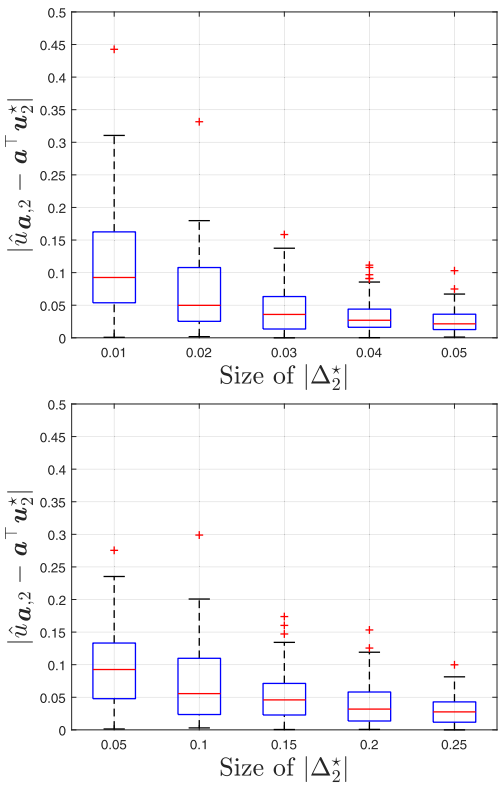


Fig. 4. Estimation error vs eigen-gap $|\Delta_2^\star|$. The left figure corresponds to the full observation case, while the right figure corresponds to the missing observation case.

The unit vector $\boldsymbol{a}$ is chosen such that $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0$ and $\boldsymbol{a}^\top \boldsymbol{u}_2^\star = 0.5$. Given that $|\boldsymbol{a}^\top \boldsymbol{u}_2^\star|$ is quite large, we are expected to have $|\boldsymbol{a}^\top \boldsymbol{u}_2| \geq c_{\mathrm{b}} \sqrt{\widehat{v}_{\boldsymbol{a},2}} \log^{1.5} n$ with high probability, and hence we can simply take $\widehat{u}_{\boldsymbol{a},2}^{\text{modified}} = \widehat{u}_{\boldsymbol{a},2}$ in these numerical experiments (according to (29) and (11)).

1) Heteroscedastic Gaussian Noise: The noise setting is the same as in expression (53). The numerical results are displayed in Fig. 5 and Tab. I. We also examine the necessity of our requirement on the "interferers" (i.e. $|\boldsymbol{a}^\top \boldsymbol{u}_k^\star| \lesssim |\lambda_l^\star - \lambda_k^\star|/(|\lambda_l^\star| \log n), k \neq l$) as stated in Theorem 4. Specifically, we make comparisons of the following two settings: (1) the "no interferer" case where $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0$ and $\boldsymbol{a}^\top \boldsymbol{u}_2^\star = 0.5$, as plotted in Fig. 5(a)-5(f); and (2) the case with a strong interferer where $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0.05$ and $\boldsymbol{a}^\top \boldsymbol{u}_2^\star = 0.5$, as plotted in Fig. 5(g)-5(i). Numerically, our distributional guarantees are inaccurate when there exists a strong interferer, which is consistent with what our theory predicts.

2) Heteroscedastic Bernoulli Noise: Consider a heteroscedastic Bernoulli noise matrix $\boldsymbol{H}$ with independent entries such that $H_{ij} = -\sigma_{ij}$ with probability $1/2$ and $H_{ij} = \sigma_{ij}$ otherwise. The variance matrix is also chosen to satisfy (53). The numerical results are plotted in Fig. 6 and Tab. I.

3) Missing Data Model: In the case when we only get to observe a fraction of the entries of $\boldsymbol{M}^\star$ as in model (54), in the same way, we aim to provide confidence intervals with precise coverages for both the linear functionals and the eigenvalues. Our numerical results are shown in Fig. 7 and Tab. I.

TABLE I

NUMERICAL COVERAGE RATES FOR OUR 95% CONFIDENCE INTERVALS OVER 10000 INDEPENDENT TRIALS

Settings	Target	Numerical coverage
heteroscedastic Gaussian	linear form $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$	0.9422
	eigenvalue $\lambda_2^\star$	0.9496
heteroscedastic Bernoulli	linear form $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$	0.9470
	eigenvalue $\lambda_2^\star$	0.9494
missing data	linear form $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$	0.9398
	eigenvalue $\lambda_2^\star$	0.9519

4) Conclusion: In all of the above numerical experiments, the confidence intervals and the Q-Q (quantile-quantile) plots we produce match the theoretical predictions in a reasonably well manner, thus corroborating the validity and practicability of our theoretical results. We have numerically verified the need of controlling the "interferers" in Fig. 5(g)-5(i).

VI. PRIOR ART

Recent years have witnessed a flurry of activity in noisy low-rank matrix estimation [2], [13], [21], [33]–[44]. Despite the fundamental importance of estimating linear functionals of eigenvectors (or singular vectors), how to accomplish this task remains largely unknown. Only until very recently, researchers started to understand estimation errors for a special type of linear functionals, namely, the entrywise estimation error for the leading eigenvector or the $\ell_{2,\infty}$ error for the rank-r eigenspace. Examples include [9], [17], [20], [33], [45]–[52], which have been motivated by various applications

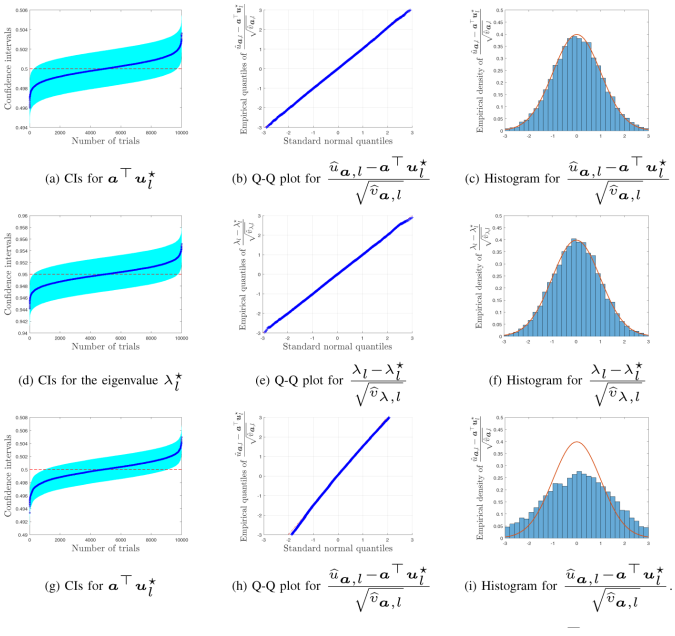


Fig. 5. Numerical results for inference for the linear form $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ and the eigenvalue $\lambda_l^\star = 0.95$ ($l = 2$) under heteroscedastic Gaussian noise. In (a)-(f), we take $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0$ (no interferer), while in (g), (h) and (i) $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0.05$ (with interferer). In both two settings, we take $n = 1000$, set $\sigma_1 = 0.1/\sqrt{n \log n}$ and $\delta_\sigma = 0.4/((n-1)\sqrt{n \log n})$ (cf. (53)), and run independent 10000 trials. In (a), (d) and (g), the confidence intervals are sorted respectively by the magnitudes of the estimators $\widehat{u}_{\boldsymbol{a},l}$ and λ_l in these trials. Here, $\boldsymbol{u}_2$ is chosen such that $\boldsymbol{u}_2^\top \boldsymbol{u}_2^\star \geq 0$. In (c), (f) and (i), the empirical densities are compared to the pdf of the standard normal (red curve).

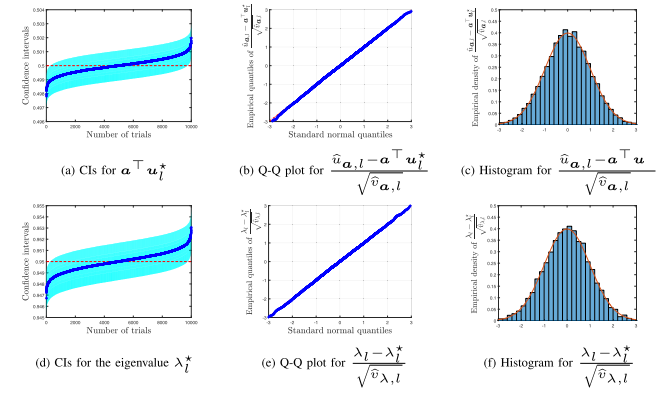


Fig. 6. Numerical results for inference for the linear form $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ and the eigenvalue $\lambda_l^\star = 0.95$ ($l = 2$) under heteroscedastic Bernoulli noise. We take $n = 1000$, set $\sigma_1 = 0.1/\sqrt{n \log n}$ and $\delta_\sigma = 0.4/((n-1)\sqrt{n \log n})$ (cf. (53)), and run independent 10000 trials. In (a) and (d), the confidence intervals are sorted respectively by the magnitudes of the estimators $\widehat{u}_{\boldsymbol{a},l}$ and λ_l in these trials. Here, $\boldsymbol{u}_2$ is chosen such that $\boldsymbol{u}_2^\top \boldsymbol{u}_2^\star \geq 0$. In (c) and (f), the empirical densities are compared to the pdf of the standard normal (red curve).

including matrix completion, community detection, top-K ranking, and so on. While tight ℓ_∞ eigenvector (resp. $\ell_{2,\infty}$ eigenspace) perturbation bounds in the presence of symmetric noise matrices have been derived in the prior works [9], [20], [52], all of these papers required the associated eigen-gap to exceed the spectral norm of the noise matrix, thereby falling short of addressing the scenarios with small eigen-gaps. The

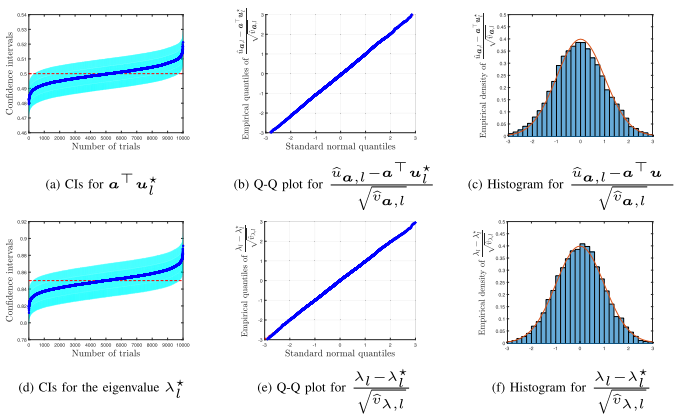


Fig. 7. Numerical results for inference for the linear form $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ and the eigenvalue $\lambda_l^\star = 0.85$ ($l = 2$) in the presence of missing data. We take $n = 1000$, set $p = 0.1$, $\sigma_1 = 1/\sqrt{10\, n \log n}$ and $\delta_\sigma = 0.4/((n-1)\sqrt{10\, n \log n})$ (cf. (54)), and run 10000 independent trials. In (a) and (g), the confidence intervals are sorted respectively by the magnitudes of the estimators $\widehat{u}_{\boldsymbol{a},l}$ and λ_l in these trials. Here, $\boldsymbol{u}_2$ is chosen such that $\boldsymbol{u}_2^\top \boldsymbol{u}_2^\star \geq 0$. In (c) and (f), the empirical densities are compared to the pdf of the standard normal (red curve).

work [33] also considered controlling the perturbation error of certain Fourier coefficients of the leading singular vector in a blind deconvolution problem, which, however, does not generalize to other linear functionals. [53] developed sharp concentration bounds for estimating linear forms of singular vectors under i.i.d. Gaussian noise, which, however, required the true singular values to be sufficiently separated (i.e. with a spacing much larger than the minimal eigen-gap studied herein). The recent work [28] is perhaps the only one that studied finite-grained eigenvector perturbation in the face of a small eigen-gap, which, however, is restricted to the case with i.i.d. Gaussian noise.

Given that estimation of linear functionals of eigenvectors is already largely under-explored, it is perhaps not surprising to see the lack of investigation about inference and uncertainty quantification for these quantities. This is in stark constrat to sparse estimation and learning problems, for which the construction of confidence regions has been extensively studied [54]–[64]. A few exceptions are worth mentioning: (1) [65]–[67] identified ℓ_2 confidence regions that are likely to cover the low-rank matrix of interest, which, however, might be loose in terms of the pre-constant; (2) focusing on low-rank matrix completion, the recent work [68] developed a de-biasing strategy that constructs both confidence regions for low-rank factors and entrywise confidence intervals for the unknown matrix, attaining statistical optimality in terms of both the pre-constant and the rate; an independent work by Xia et al. [69] analyzed a similar de-biasing strategy with the aid of double sample splitting, and shows asymptotic normality of linear forms of the matrix estimator; (3) [70], [71] developed a spectral projector to construct confidence regions for singular subspaces in the presence of i.i.d. additive noise; (4) [18] considered estimating linear forms of eigenvectors in a different covariance estimation model, whose analysis relies on the Gaussianity assumption; (5) [72] characterized the asymptotic normality of bilinear forms of eigenvectors, which

accommodates heterogeneous noise; and (6) [44] established the $\ell_{2,\infty}$ distributional guarantees for two spectral estimators (i.e., plain SVD and heteroskedastic PCA) tailored to PCA with heteroskedastic and missing data.

Additionally, the bulk distribution of the eigenvalues of i.i.d. random matrices has been studied in the physics literature (e.g. [73]–[78]), which falls short of the characterizing distributions of extreme eigenvalues. Several more recent papers started to consider a super-position of a low-rank matrix and an i.i.d. noise matrix, and studied the locations and distributions of the eigenvalues beyond the bulk [79]–[81]. These results did not cover a general class of heteroscadestic noise and did not allow the rank r to grow with n. What is more, none of these papers studied how to construct valid confidence intervals for extreme eigenvalues, let alone inference for individual eigenvectors.

VII. Discussion

The present paper contributes towards "fine-grained" statistical analysis, by developing guaranteed estimation and inference algorithms for linear functionals of the unknown eigenvectors and eigenvalues. The proposed procedures are model agnostic and are able to accommodate heteroscedastic noise, without the need of prior knowledge about the noise levels. The validity of our procedures is guaranteed even when the eigen-gap is extremely small, a condition that goes significantly beyond what we have learned from generic matrix perturbation theory. The key enabler of our findings lies in an appealing bias reduction feature of eigen-decomposition when coping with asymmetric noise matrices.

Our studies leave open several interesting questions worthy of future investigation. For instance, our current theory for confidence intervals falls short of accommodating the scenarios when quantity $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|$ far exceeds the associated eigen-gap — how to determine the fundamental inference limits for such scenarios and, perhaps more importantly, how to attain the limits efficiently? In addition, our theory is likely suboptimal in terms of the dependency on the rank r and the condition number κ. Can we further improve the theoretical support in these regards? Furthermore, our analysis framework shed some light on how to perform inference on functions of the eigenvectors. It would be interesting to develop a unified framework that leads to valid confidence intervals for a broader class of functions (e.g. quadratic functionals, or more general polynomials) of the eigenvectors. Moving beyond estimation and inference for individual eigenvectors, we remark that our current theory falls short of delivering useful eigenspace perturbation guarantees unless there exists a sufficient eigen-gap between any adjacent pair of eigenvalues. The challenges are at least two-fold when dealing with an asymmetric data matrix: (1) the eigenvectors are, in general, not orthogonal to each other, and (2) the eigenvectors might be complex-valued even when the observed data are real-valued, both of which are in stark contrast to what happens for a symmetric data matrix. How to extend our theory to accommodate more general eigenspace perturbation is left for future investigation.

Appendix A
Preliminaries

Denote by $\boldsymbol{u}_l$ (resp. λ_l) the lth leading right eigenvector (resp. eigenvalue) of $\boldsymbol{M} = \sum_{j=1}^r \lambda_j^\star \boldsymbol{u}_j^\star \boldsymbol{u}_j^{\star\top} + \boldsymbol{H}$. We make note of several facts about $\boldsymbol{u}_l$ and λ_l — previously established in [17] — that will prove useful throughout.

To begin with, a simple application of the Neumann series yields the following expansion [17, Theorem 2].

Lemma 1 (Neumann Expansion): Let $\boldsymbol{u}_l$ and λ_l be the lth leading right eigenvector and the lth leading eigenvalue of $\boldsymbol{M}$ (cf. (7)), respectively. If $\|\boldsymbol{H}\| < |\lambda_l^\star|$, then one has

$$\boldsymbol{u}_l = \sum_{j=1}^r \frac{\lambda_j^\star}{\lambda_l} \big(\boldsymbol{u}_j^{\star\top} \boldsymbol{u}_l\big) \left\{ \sum_{s=0}^\infty \frac{1}{\lambda_l^s} \boldsymbol{H}^s \boldsymbol{u}_j^\star \right\}. \tag{55}$$

As an immediate consequence, we have the following expansion for $\boldsymbol{a}^\top \boldsymbol{u}_l$, which forms the basis of our estimators:

$$\begin{aligned} &\frac{\lambda_l}{\lambda_l^\star \big(\boldsymbol{u}_l^{\star\top} \boldsymbol{u}_l\big)} \boldsymbol{a}^\top \boldsymbol{u}_l \\ &= \sum_{j=1}^r \frac{\lambda_j^\star}{\lambda_l^\star} \cdot \frac{\boldsymbol{u}_j^{\star\top} \boldsymbol{u}_l}{\boldsymbol{u}_l^{\star\top} \boldsymbol{u}_l} \left\{ \sum_{s=0}^\infty \frac{1}{\lambda_l^s} \boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_j^\star \right\} \\ &= \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{1}{\lambda_l} \boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_l^\star + \sum_{s=2}^\infty \frac{1}{\lambda_l^s} \boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_l^\star \\ &\quad + \sum_{j:j\neq l} \frac{\lambda_j^\star}{\lambda_l^\star} \frac{\boldsymbol{u}_j^{\star\top} \boldsymbol{u}_l}{\boldsymbol{u}_l^{\star\top} \boldsymbol{u}_l} \left\{ \sum_{s=0}^\infty \frac{1}{\lambda_l^s} \boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_j^\star \right\}. \end{aligned} \tag{56}$$

As we shall see shortly, our theory relies heavily on approximating (56) by the lower-order terms (w.r.t. $\boldsymbol{H}$). This requires controlling the influence of the high-order terms. Towards this end, we make note of several useful bounds about $\boldsymbol{H}$ that have been established in [17].

Lemma 2: Fix any vector $\boldsymbol{a} \in \mathbb{R}^n$, and suppose that $\|\boldsymbol{u}_l^\star\|_\infty \le \sqrt{\mu/n}$. Suppose the noise matrix $\boldsymbol{H}$ obeys Assumption 1, and assume the existence of some sufficiently small constant $c_1 > 0$ such that

$$\max\big\{\sigma_{\max}\sqrt{n\log n}, B\log n\big\} \le c_1 \lambda_{\min}^\star. \tag{57}$$

Then there exist some universal constants $c_2, c_3 > 0$ such that with probability at least $1 - O(n^{-10})$,

$$\big|\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_i^\star\big| \le \sqrt{\frac{\mu}{n}} \Big(c_2 \max\big\{\sigma_{\max}\sqrt{n\log n}, B\log n\big\}\Big)^s \|\boldsymbol{a}\|_2, \tag{58a}$$

$$\|\boldsymbol{H}\| \le c_3 \max\big\{\sigma_{\max}\sqrt{n\log n}, B\log n\big\}. \tag{58b}$$

Remark 8: The work [17] established the bound (58a) only for the case with $s \le 20\log n$. Fortunately, the case with $s > 20\log n$ follows immediately by combining the crude bound $\big|\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}^\star\big| \le \|\boldsymbol{H}\|^s \|\boldsymbol{a}\|_2$ and the inequality (58b) (by choosing $c_2 = 2c_3$ and using the fact that $(1/2)^s \ll \sqrt{1/n}$).

Combining Lemmas 1-2, the paper [17] establishes the following result.

Lemma 3: Consider a rank-r symmetric matrix $\boldsymbol{M}^\star = \sum_{i=1}^r \lambda_i^\star \boldsymbol{u}_i^\star \boldsymbol{u}_i^{\star\top} \in \mathbb{R}^{n\times n}$ with incoherence parameter μ (cf. Definition 1). Suppose that the noise matrix $\boldsymbol{H}$ obeys Assumption 1 and Condition (57). Then for any fixed vector

$\boldsymbol{a} \in \mathbb{R}^n$ and any $1 \le i \le r$, with probability at least $1 - O(n^{-10})$ one has

$$\left|\boldsymbol{a}^\top \left(\boldsymbol{u}_i - \sum_{k=1}^{r} \frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_i}{\lambda_i/\lambda_k^\star}\boldsymbol{u}_k^\star\right)\right| \le c_2 \frac{\max\left\{\sigma_{\max}\sqrt{n\log n}, B\log n\right\}}{\lambda_{\min}^\star}\sqrt{\frac{\mu\kappa^2 r}{n}}\|\boldsymbol{a}\|_2 \tag{59}$$

for some universal constant $c_2 > 0$. This result holds unchanged if the ith right eigenvector $\boldsymbol{u}_i$ is replaced by the ith left eigenvector $\boldsymbol{w}_i$.

Remark 9: Under the additional condition that $B\log n \le \sigma_{\max}\sqrt{n\log n}$, the preceding bounds simplify to

$$\left|\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_i^\star\right| \le \sqrt{\frac{\mu}{n}}\left(c_2\sigma_{\max}\sqrt{n\log n}\right)^s \|\boldsymbol{a}\|_2, \tag{60a}$$

$$\|\boldsymbol{H}\| \le c_3\sigma_{\max}\sqrt{n\log n}, \tag{60b}$$

$$\left|\boldsymbol{a}^\top \left(\boldsymbol{u}_i - \sum_{k=1}^{r} \frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_i}{\lambda_i/\lambda_k^\star}\boldsymbol{u}_k^\star\right)\right| \le c_2 \frac{\sigma_{\max}\sqrt{\mu\kappa^2 r\log n}}{\lambda_{\min}^\star}\|\boldsymbol{a}\|_2. \tag{60c}$$

Another immediate consequence under our assumptions is that, with probability at least $1 - O(n^{-10})$,

$$\|\boldsymbol{H}\| \le \lambda_{\min}^\star/10. \tag{61}$$

Appendix B
Proof for Eigenvector and Eigenvalue Estimation

In this section, we shall start by proving the eigenvalue perturbation bound stated in Theorem 2, which plays a pivotal role in establishing the eigenvector estimation guarantees in Theorem 1.

A. Proof of Theorem 2

This section aims to establish a slightly stronger version of Theorem 2, stated as follows.

Theorem 7: Assume that $\mu\kappa^2 r^4 \le c_3\, n$ for some sufficiently small constant $c_3 > 0$. Suppose that the noise parameters defined in Assumption 1 satisfy

$$\Delta_l^\star > 2c_4\kappa^2\, r^2 \max\left\{\sigma_{\max}\sqrt{n\log n}, B\log n\right\}\sqrt{\tfrac{\mu}{n}} \tag{62a}$$

$$\max\left\{\sigma_{\max}\sqrt{n\log n}, B\log n\right\} \le c_5\lambda_{\min}^\star \tag{62b}$$

for some sufficiently large (resp. small) constant $c_4 > 0$ (resp. $c_5 > 0$). Then given any integer $1 \le l \le r$, with probability $1 - O(n^{-8})$, the eigenvalue λ_l and the associated eigenvectors $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$ (see Notation 1) are all real-valued, and one has

$$|\lambda_l - \lambda_l^\star| \le c_4\kappa r^2 \max\left\{\sigma_{\max}\sqrt{n\log n}, B\log n\right\}\sqrt{\tfrac{\mu}{n}}. \tag{63}$$

A few immediate consequences of this theorem and the Bauer-Fike theorem are summarized as follows.

Corollary 2: Suppose that $\mu\kappa^2\, r^4 \lesssim n$, and that Assumptions 1-2 hold. With probability at least $1 - O(n^{-6})$,

$$|\lambda_l - \lambda_l^\star| \le \min\left\{\frac{\Delta_l^\star}{2}, c_3\sigma_{\max}\sqrt{n\log n}\right\} \quad \text{and} \quad \left|\frac{\lambda_l - \lambda_l^\star}{\lambda_l^\star}\right| \le \frac{1}{100} \tag{64}$$

for some sufficiently small constant $c_3 > 0$. In addition, if either $\sigma_{\max}\sqrt{n\log n} = o(\lambda_{\min}^\star)$ or $\Delta_l^\star = o(\lambda_{\min}^\star)$ holds, then with probability at least $1 - O(n^{-6})$,

$$\lambda_l = (1 + o(1))\lambda_l^\star. \tag{65}$$

1) Proof Outline: To establish this theorem, we borrow a powerful idea from [79] that converts eigenvalue analysis to zero counting of certain complex-valued functions. Specifically, consider the following functions

$$f(z) := \det\left(\boldsymbol{I} + \boldsymbol{U}^{\star\top}(\boldsymbol{H} - z\boldsymbol{I})^{-1}\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right), \tag{66a}$$

$$g(z) := \det\left(\boldsymbol{I} + \boldsymbol{U}^{\star\top}(-z)^{-1}\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right). \tag{66b}$$

The intimate connection between these two functions and the eigenvalues of $\boldsymbol{M}$ and $\boldsymbol{M}^\star$ is formalized in the following observation made by [79].

Claim 1: If $\lambda_{\min} > 2\|\boldsymbol{H}\|$, then the zeros of $f(\cdot)$ (resp. $g(\cdot)$) on the region $\mathcal{K} := \{z \in \mathbb{C} : |z| > \|\boldsymbol{H}\|\} \cup \{\infty\}$ are exactly the r leading eigenvalues of $\boldsymbol{M}$ (resp. $\boldsymbol{M}^\star$).

With this claim in mind, we turn attention to studying the zeros of $f(\cdot)$ and $g(\cdot)$. Given that $f(\cdot)$ can be viewed as a perturbed version of $g(\cdot)$ (since $f(\cdot)$ can be obtained by adding a perturbation $\boldsymbol{H}$ to $g(\cdot)$ in a certain way), we hope that the zeros of $f(\cdot)$ do not deviate by much from the zeros of $g(\cdot)$. Towards justifying this, we look at the following γ-neighborhood of $\{\lambda_1^\star, \lambda_2^\star, \cdots, \lambda_r^\star\}$:

$$\mathcal{D}(\gamma) := \bigcup_{k=1}^{r} \mathcal{B}(\lambda_k^\star, \gamma), \tag{67}$$

where $\mathcal{B}(\lambda, \gamma)$ is a ball of radius $\gamma > 0$ centered at λ, namely, $\mathcal{B}(\lambda, \gamma) := \{z \in \mathbb{C} : |z - \lambda| \le \gamma\}$. A crucial part of the proof is to demonstrate that all zeros of $f(\cdot)$ lie in $\mathcal{D}(\gamma)$ — or equivalently, in the γ-neighborhood of the zeros of $g(\cdot)$ — for sufficiently small γ.

Remark 10: Somewhat surprisingly, γ is allowed to be as small as $\frac{\mathrm{poly}\log(n)}{\sqrt{n}}\|\boldsymbol{H}\|$ when $r, \kappa, \mu \asymp 1$.

In what follows, we shall assume that

$$\gamma < \lambda_{\min}^\star/4 \quad \text{and} \quad \gamma < \Delta_l^\star/2. \tag{68}$$

In view of (61), one has $\|\boldsymbol{H}\| < \lambda_{\min}^\star/4$, thus indicating that

$$|z| \ge \lambda_{\min}^\star - \gamma > \lambda_{\min}^\star/2 > \|\boldsymbol{H}\| \quad \text{for all } z \in \mathcal{D}(\gamma). \tag{69}$$

Our proof is based on the following observations:

(i) Given that $\gamma < \Delta_l^\star/2$, one has $\mathcal{B}(\lambda_l^\star, \gamma) \cap \mathcal{B}(\lambda_k^\star, \gamma) = \emptyset$ for any $k \ne l$, and hence $g(\cdot)$ has exactly 1 zero in $\mathcal{B}(\lambda_l^\star, \gamma)$; in addition, it is clear that $g(\cdot)$ has $l - 1$ (resp. $r - l$) zeros in $\cup_{k:k<l}\mathcal{B}(\lambda_k^\star, \gamma)$ (resp. $\cup_{k:k>l}\mathcal{B}(\lambda_k^\star, \gamma)$).

(ii) Suppose that in each connected component of $\mathcal{D}(\gamma)$, the functions $f(\cdot)$ and $g(\cdot)$ always have the same number

of zeros. If this were true, then one would have (1) a unique zero of $f(\cdot)$ in $\mathcal{B}(\lambda_l^\star,\gamma)$; (2) $l-1$ zeros of $f(\cdot)$ in $\cup_{1\le k<l}\mathcal{B}(\lambda_k^\star,\gamma)$; (3) $r-l$ zeros of $f(\cdot)$ in $\cup_{k>l}\mathcal{B}(\lambda_k^\star,\gamma)$.

(iii) Recalling how we sort $\{\lambda_k^\star\}$ and $\{\lambda_k\}$, we see that the real part of any $z\in\cup_{1\le k<l}\mathcal{B}(\lambda_k^\star,\gamma)$ must exceed $\lambda_{l-1}^\star-\gamma>\lambda_l^\star+\gamma$ and, similarly, the real part of any $z\in\cup_{k>l}\mathcal{B}(\lambda_k^\star,\gamma)$ is at most $\lambda_{l+1}^\star+\gamma<\lambda_l^\star-\gamma$. Thus, the above arguments reveal that the unique zero of $f(\cdot)$ in $\mathcal{B}(\lambda_l^\star,\gamma)$ is λ_l, which obeys $|\lambda_l-\lambda_l^\star|\le\gamma$.

(iv) In addition, note that the complex conjugate $\overline{\lambda_l}$ of λ_l is also an eigenvalue of $\boldsymbol{M}$, given that $\boldsymbol{M}$ is a real-valued matrix. However, since there is only one zero of $f(\cdot)$ residing in $\mathcal{B}(\lambda_l^\star,\gamma)$, we necessarily have $\overline{\lambda_l}=\lambda_l$, meaning that λ_l is real-valued. A similar argument justifies that the associated right eigenvector $\boldsymbol{u}_l$ and left eigenvector $\boldsymbol{w}_l$ are real-valued as well.

Theorem 7 is thus established if we are allowed to pick

$$\gamma=c_4\kappa r^2\max\left\{B\log n,\sigma_{\max}\sqrt{n\log n}\right\}\sqrt{\frac{\mu}{n}}. \tag{70}$$

Clearly, this choice satisfies $\gamma<\lambda_{\min}^\star/4$ and $\gamma<\Delta_l^\star/2$ under the assumptions of this theorem.

The remaining proof then boils down to justifying that $f(\cdot)$ and $g(\cdot)$ have the same number of zeros in each connected component of $\mathcal{D}(\gamma)$. Towards this end, we resort to Rouché's theorem in complex analysis [82].

Theorem 8 (Rouché's Theorem): Let $f(\cdot)$ and $g(\cdot)$ be two complex-valued functions that are holomorphic inside a region $\mathcal{R}$ with closed contour $\partial\mathcal{R}$. If $|f(z)-g(z)|<|g(z)|$ for all $z\in\partial\mathcal{R}$, then $f(\cdot)$ and $g(\cdot)$ have the same number of zeros inside $\mathcal{R}$.

In order to invoke Rouché's theorem to justify the claim in (ii), it suffices to fulfill the requirement $|f(z)-g(z)|<|g(z)|$ on $\mathcal{D}(\gamma)$. Given that $|\boldsymbol{z}|>\|\boldsymbol{H}\|$ for all $z\in\mathcal{D}(\gamma)$ (see (69)), we can apply the Neumann series $(\boldsymbol{H}-z\boldsymbol{I})^{-1}=-\sum_{s=0}^{\infty}z^{-s-1}\boldsymbol{H}^s$ to reach

$$\begin{aligned}
&f(z)\\
&=\det\left(\boldsymbol{I}-\boldsymbol{U}^{\star\top}\left(\sum_{s=0}^{\infty}z^{-s-1}\boldsymbol{H}^s\right)\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right)\\
&=\det\left(\boldsymbol{I}+\boldsymbol{U}^{\star\top}(-z)^{-1}\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star-\sum_{s=1}^{\infty}z^{-s-1}\boldsymbol{U}^{\star\top}\boldsymbol{H}^s\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right)\\
&=\det\left(\left(\boldsymbol{I}+\boldsymbol{U}^{\star\top}(-z)^{-1}\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right)\left(\boldsymbol{I}\right.\right.\\
&\quad\left.\left.-\left(\boldsymbol{I}+\boldsymbol{U}^{\star\top}(-z)^{-1}\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right)^{-1}\sum_{s=1}^{\infty}z^{-s-1}\boldsymbol{U}^{\star\top}\boldsymbol{H}^s\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right)\right)\\
&=g(z)\det\left(\boldsymbol{I}-\left(\boldsymbol{I}+\boldsymbol{U}^{\star\top}(-z)^{-1}\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right)^{-1}\right.\\
&\quad\left.\cdot\sum_{s=1}^{\infty}z^{-s-1}\boldsymbol{U}^{\star\top}\boldsymbol{H}^s\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\right)\\
&=g(z)\det\Bigg(\underbrace{\boldsymbol{I}-(\boldsymbol{I}-\boldsymbol{\Sigma}^\star/z)^{-1}\sum_{s=1}^{\infty}z^{-s-1}\boldsymbol{U}^{\star\top}\boldsymbol{H}^s\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star}_{=:\,\boldsymbol{\Delta}}\Bigg),
\end{aligned}$$

where the last line holds since $\boldsymbol{U}^{\star\top}\boldsymbol{U}^\star=\boldsymbol{I}$. In addition, observe that the zeros of $g(\cdot)$ are in the interior of $\mathcal{D}$ and that $g(z)\neq0$ for all $z\in\partial\mathcal{D}(\gamma)$. Hence, on $\partial\mathcal{D}(\gamma)$ we have

$$\begin{aligned}
&|f(z)-g(z)|<|g(z)|\\
\iff\quad&|g(z)|\cdot|\det(\boldsymbol{I}-\boldsymbol{\Delta})-1|<|g(z)|\\
\iff\quad&|\det(\boldsymbol{I}-\boldsymbol{\Delta})-1|<1.
\end{aligned}\tag{71}$$

To justify (71) on $\partial\mathcal{D}$, we make the following observations:

Claim 2: The condition (71) can be guaranteed as long as $\|\boldsymbol{\Delta}\|<1/(2r)$.

Claim 3: There exists a universal constant $c_3>0$ such that with probability $1-O(n^{-8})$, one has

$$\|\boldsymbol{\Delta}\|\le c_3\frac{\max\left\{B\log n,\sigma_{\max}\sqrt{n\log n}\right\}}{\gamma}\sqrt{\frac{\mu\kappa^2r^2}{n}}. \tag{72}$$

In summary, in order to guarantee $|f(z)-g(z)|<|g(z)|$ on $\partial\mathcal{D}(\gamma)$, it suffices to ensure that $\|\boldsymbol{\Delta}\|\le1/(2r)$ (by (71) and Claim 2), which would hold as long as we take $\gamma=2c_3\max\left\{B\log n,\sigma_{\max}\sqrt{n\log n}\right\}\sqrt{\frac{\mu\kappa^2\ r^4}{n}}$ (by Claim 3). This in turn establishes Theorem 7 (and hence Theorem 2) as long as $c_4\ge2c_3$.

Finally, the proofs of the auxiliary claims are postponed to Appendix B-A2.

2) Proofs for Auxiliary Claims in Appendix B-A1:

a) Proof of Claim 1: Note that both $f(\cdot)$ and $g(\cdot)$ are holomorphic on $\mathcal{K}$. Making use of the identity $\det(\boldsymbol{I}+\boldsymbol{AB})=\det(\boldsymbol{I}+\boldsymbol{BA})$, we have

$$\begin{aligned}
f(z)&=\det\left(\boldsymbol{I}+\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{U}^{\star\top}(\boldsymbol{H}-\boldsymbol{z}\boldsymbol{I})^{-1}\right)\\
&=\det\left(\boldsymbol{I}+\boldsymbol{M}^\star(\boldsymbol{H}-\boldsymbol{z}\boldsymbol{I})^{-1}\right)\\
&=\det\left((\boldsymbol{H}-z\boldsymbol{I}+\boldsymbol{M}^\star)(\boldsymbol{H}-z\boldsymbol{I})^{-1}\right)\\
&=\frac{\det(\boldsymbol{M}-z\boldsymbol{I})}{\det(\boldsymbol{H}-z\boldsymbol{I})}
\end{aligned}\tag{73}$$

and

$$g(z)=\det\left(\boldsymbol{I}+\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{U}^{\star\top}/(-z)\right)=\prod_{i=1}^{r}\left(1-\frac{\lambda_i^\star}{z}\right). \tag{74}$$

These identities make clear that the zeros of $g(\cdot)$ (resp. $f(\cdot)$) on $\mathcal{K}$ are all eigenvalues of $\boldsymbol{M}^\star$ (resp. $\boldsymbol{M}$). In particular, the zeros of $g(\cdot)$ are precisely $\{\lambda_i^\star\}_{1\le i\le r}$.

It remains to show that there are exactly r zeros of $f(\cdot)$ lying in $\mathcal{K}$, and that they are exactly $\lambda_1,\cdots,\lambda_r$. This is equivalent to showing that the set of eigenvalues of $\boldsymbol{M}$ contained in the region $\mathcal{K}$ is $\{\lambda_i\}_{1\le i\le r}$. Towards this, define $\boldsymbol{M}(t)=\boldsymbol{M}^\star+t\boldsymbol{H}$. From the Bauer-Fike theorem, all eigenvalues of $\boldsymbol{M}(t)$ lie in the set

$$\mathcal{B}(0,t\|\boldsymbol{H}\|)\cup\underbrace{\left\{\cup_{k=1}^{r}\mathcal{B}(\lambda_k^\star,t\|\boldsymbol{H}\|)\right\}}_{=:\,\mathcal{D}(t\|\boldsymbol{H}\|)}. \tag{75}$$

Given that $\|\boldsymbol{H}\|<\lambda_{\min}/4$ (according to (61) under our assumptions), it is easily seen that $\mathcal{B}(0,t\|\boldsymbol{H}\|)$ does not intersect with $\mathcal{D}(t\|\boldsymbol{H}\|)$ for all $0\le t\le1$. Additionally, the set of the eigenvalues of $\boldsymbol{M}(t)$ depends continuously on t [83, Theorem 6], requiring $\boldsymbol{M}(t)$ to have the same number of

zeros in $\mathcal{B}(0, t\|\boldsymbol{H}\|)$ for all $0 \le t \le 1$. Hence, $\boldsymbol{M} = \boldsymbol{M}(1)$ has exactly r eigenvalues in $\mathcal{D}(\|\boldsymbol{H}\|)$ (since $\boldsymbol{M}^\star = \boldsymbol{M}(0)$ has r eigenvalues in this region). These eigenvalues necessarily have magnitudes larger than any point in $\mathcal{B}(0, \|\boldsymbol{H}\|)$ (since $\|\boldsymbol{H}\| < \lambda_{\min}/4$), thus indicating that the eigenvalues of $\boldsymbol{M}$ in $\mathcal{K}$ are precisely $\lambda_1, \cdots, \lambda_r$.

b) Proof of Claim 2: Denoting by $\mu_1, \cdots, \mu_r \in \mathbb{C}$ the r eigenvalues of $\boldsymbol{I} - \boldsymbol{\Delta} \in \mathbb{R}^{r\times r}$, one has

$$\big|\det(\boldsymbol{I}-\boldsymbol{\Delta}) - 1\big| = \Big|\prod_{i=1}^{r}\mu_i - 1\Big|.$$

By virtue of the elementary inequality[4] $|\prod_{i=1}^r(1+a_i)-1| \le \prod_{i=1}^r(1+|a_i|)-1$, we arrive at

$$\big|\det(\boldsymbol{I}-\boldsymbol{\Delta})-1\big| \le \prod_{i=1}^{r}(1+|\mu_i-1|)-1 \le (1+\|\boldsymbol{\Delta}\|)^r - 1,$$

where the last inequality follows from the Bauer-Fike theorem (which forces that $|\mu_i - 1| \le \|\boldsymbol{\Delta}\|$). This means that Condition (71) holds if $(1+\|\boldsymbol{\Delta}\|)^r < 2$; the latter condition is clearly guaranteed if $\|\boldsymbol{\Delta}\| \le 1/(2r)$.

c) Proof of Claim 3: Since $\gamma < \lambda_{\min}/4$, it always holds that $|z| > 3\lambda_{\min}/4$ for all $z \in \partial\mathcal{D}(\gamma)$. Hence

$$\begin{aligned}\|\boldsymbol{\Delta}\| &= \Big\|(z\boldsymbol{I}-\boldsymbol{\Sigma}^\star)^{-1}\sum_{s=1}^{\infty} z^{-s}\boldsymbol{U}^{\star\top}\boldsymbol{H}^s\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\Big\| \\ &\le \big\|(z\boldsymbol{I}-\boldsymbol{\Sigma}^\star)^{-1}\big\|\cdot\sum_{s=1}^{\infty}\frac{\big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}^s\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\big\|}{|z|^s} \\ &\le \|\boldsymbol{\Sigma}^\star\|\cdot\big\|(z\boldsymbol{I}-\boldsymbol{\Sigma}^\star)^{-1}\big\|\cdot\sum_{s=1}^{\infty}\frac{\big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}^s\boldsymbol{U}^\star\big\|}{|z|^s}. \quad (76)\end{aligned}$$

Recall that $\|\boldsymbol{\Sigma}^\star\| = \lambda_{\max}$. Also, on $\partial\mathcal{D}(\gamma)$ we have

$$\big\|(z\boldsymbol{I}-\boldsymbol{\Sigma}^\star)^{-1}\big\| \le \max_{1\le k\le r,\, z\in\partial\mathcal{D}(\gamma)}\left|\frac{1}{z-\lambda_k^\star}\right| \le \frac{1}{\gamma}. \quad (77)$$

Taken collectively, the above results yield

$$\begin{aligned}\|\boldsymbol{\Delta}\| &\le \frac{\lambda_{\max}}{\gamma}\sum_{s=1}^{\infty}\frac{\big\|\boldsymbol{U}^{\star\top}\boldsymbol{H}^s\boldsymbol{U}^\star\big\|}{\big(\frac{3}{4}\lambda_{\min}\big)^s} \\ &\overset{\text{(i)}}{\le} \frac{\lambda_{\max}r}{\gamma}\sum_{s=1}^{\infty}\frac{\max_{1\le i,j\le r}|\boldsymbol{u}_i^{\star\top}\boldsymbol{H}^s\boldsymbol{u}_j^\star|}{\big(\frac{3}{4}\lambda_{\min}\big)^s}, \quad (78)\end{aligned}$$

where (i) makes use of the elementary inequality $\|\boldsymbol{A}\| \le r\|\boldsymbol{A}\|_\infty$ for any $\boldsymbol{A} \in \mathbb{R}^{r\times r}$. Invoke Lemma 2 and a union bound to show that: with probability $1 - O(n^{-10}r^2)$,

$$\begin{aligned}&\|\boldsymbol{\Delta}\| \\ &\le \frac{\lambda_{\max}r}{\gamma}\sum_{s=1}^{\infty}\left(\frac{c_2\max\big\{\sigma_{\max}\sqrt{n\log n}, B\log n\big\}}{\frac{3}{4}\lambda_{\min}}\right)^s\sqrt{\frac{\mu}{n}} \\ &\overset{\text{(ii)}}{\le} \frac{\lambda_{\max}r}{\gamma}\cdot\frac{c_2\max\big\{\sigma_{\max}\sqrt{n\log n}, B\log n\big\}}{\frac{3}{4}\lambda_{\min}}\cdot\sum_{s=1}^{\infty}\left(\frac{1}{2}\right)^s\sqrt{\frac{\mu}{n}} \\ &\le \frac{4c_2\max\big\{\sigma_{\max}\sqrt{n\log n}, B\log n\big\}}{3\gamma}\sqrt{\frac{\mu\kappa^2r^2}{n}},\end{aligned}$$

where (ii) holds as long as $\frac{c_2\max\{B\log n, \sigma_{\max}\sqrt{n\log n}\}}{\frac{3}{4}\lambda_{\min}} \le \frac{1}{2}$.

B. Proof of Theorem 1

As mentioned previously, Theorem 2 enables us to prove the statistical guarantees stated in Theorem 1. In the sequel, we shall first establish perturbation bounds for the vanilla eigenvector estimator (i.e. $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$) as well as the estimator $\widehat{\boldsymbol{u}}_l = \frac{1}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}(\boldsymbol{u}_l+\boldsymbol{w}_l)$.

Theorem 9: Consider any $1 \le l \le r$. Suppose that $\mu\kappa^2\, r^4 \lesssim n$ and that Assumptions 1-2 hold. Then with probability at least $1 - O(n^{-6})$, we have:

1) (ℓ_2 perturbation of the lth eigenvector)

$$\begin{aligned}&\min\|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_2 \\ &\lesssim \frac{\kappa^2\sigma_{\max}\sqrt{\mu r^2\log n}}{\Delta_l^\star} + \frac{\kappa^2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\max}^\star} \quad (79a)\\ &|\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l| \\ &\lesssim \tfrac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\mu\kappa^4 r\log n}, \qquad k \ne l, \quad (79b)\\ &|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l| \\ &\ge 1 - O\left((\kappa^4\sigma_{\max}^2 n\log n)\left\{\tfrac{1}{(\Delta_l^\star)^2}\tfrac{\mu r^2}{n} + \tfrac{1}{(\lambda_{\max}^\star)^2}\right\}\right); \quad (79c)\end{aligned}$$

2) (perturbation of linear forms of the lth eigenvector) for any fixed vector $\boldsymbol{a}$ with $\|\boldsymbol{a}\|_2 = 1$,

$$\begin{aligned}&\min\big|\boldsymbol{a}^\top(\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star)\big| \\ &\lesssim \tfrac{\sigma_{\max}\sqrt{\mu\kappa^2 r^4\log n}}{\lambda_{\min}^\star} + \sigma_{\max}\sqrt{\mu\kappa^4 r^3\log n}\cdot\max_{k\ne l}\tfrac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star-\lambda_k^\star|} \\ &\quad + \tfrac{\sigma_{\max}^2\mu\kappa^4 r^2\log n}{(\Delta_l^\star)^2}|\boldsymbol{a}^\top\boldsymbol{u}_l^\star| + \tfrac{\sigma_{\max}^2\kappa^4 n\log n}{(\lambda_{\max}^\star)^2}|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|; \quad (79d)\end{aligned}$$

3) (ℓ_∞ perturbation of the lth eigenvector)

$$\min\|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_\infty \lesssim \tfrac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^4 r\log n} + \tfrac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\tfrac{\mu^2\kappa^4 r^3\log n}{n}}. \quad (79e)$$

In addition, the above results hold unchanged if $\boldsymbol{u}_l$ is replaced by either $\boldsymbol{w}_l$ or $\widehat{\boldsymbol{u}}_l = \frac{1}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}(\boldsymbol{u}_l+\boldsymbol{w}_l)$.

Proof: See Appendix B-B1. □

In words, Theorem 9 reveals appealing statistical performance for the estimators $\boldsymbol{u}_l$, $\boldsymbol{w}_l$ and $\widehat{\boldsymbol{u}}_l$. However, when estimating linear functionals of eigenvectors via, say, the plug-in estimator $\boldsymbol{a}^\top\boldsymbol{u}_l$, the bound (79d) suffers from an additional

[4]This holds since

$$\begin{aligned}\left|\prod_{i=1}^{r}(1+a_i)-1\right| &= \Big|\sum_{k=1}^{r}\sum_{1\le i_1,\cdots,i_k\le r} a_{i_1}\cdots a_{i_k}\Big| \\ &\le \sum_{k=1}^{r}\sum_{1\le i_1,\cdots,i_k\le r}|a_{i_1}|\cdots|a_{i_k}| = \prod_{i=1}^{r}(1+|a_i|)-1.\end{aligned}$$

term $\frac{\sigma_{\max}^2\kappa^4 n\log n}{(\lambda_{\max}^\star)^2}|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|$ compared to the desired bound (14c) in Theorem 1. As it turns out, this extra term arises due to a systematic bias of the plug-in estimates. To compensate for the bias, one needs to properly enlarge the plug-in estimator, thus leading us to the proposed estimators $\widehat{u}_{\boldsymbol{a},l}$. We now establish the claimed performance guarantees for these properly corrected estimators for $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$; the proof is deferred to Appendix B-B2, which builds heavily upon the analysis of Theorem 9.

Theorem 10: Instate the assumptions of Theorem 9. Fix any vector $\boldsymbol{a}$ with $\|\boldsymbol{a}\|_2=1$. With probability exceeding $1-O(n^{-6})$, the estimator $\widehat{u}_{\boldsymbol{a},l}:=\min\left\{\sqrt{\left|\frac{(\boldsymbol{a}^\top\boldsymbol{u}_l)(\boldsymbol{a}^\top\boldsymbol{w}_l)}{\boldsymbol{u}_l^\top\boldsymbol{w}_l}\right|},1\right\}$ satisfies

$$\begin{aligned}&\min\left|\widehat{u}_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\\&\lesssim\frac{\sigma_{\max}r^2\sqrt{\mu\kappa^2\log n}}{\lambda_{\min}^\star}+\frac{\sigma_{\max}^2\mu r^2\kappa^4\log n}{(\Delta_l^\star)^2}|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\\&\quad+\sigma_{\max}\sqrt{\mu\kappa^4r^3\log n}\max_{k\neq l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star-\lambda_k^\star|}.\end{aligned}\tag{80}$$

Theorems 9-10 taken together establish Theorem 1.

1) Proof of Theorem 9: We shall start by proving the results for $\boldsymbol{u}_l$; the proofs for the results w.r.t. $\boldsymbol{w}_l$ are clearly identical.

a) Proofs for ℓ_2 perturbation bounds: To derive the ℓ_2 bound for $\boldsymbol{u}_l$, we start by considering the distance between $\boldsymbol{u}_l$ and the subspace spanned by the true eigenvectors $\{\boldsymbol{u}_k^\star\}_{1\le k\le r}$. The Neumann series (cf. Lemma 1) tells us that: when $\|\boldsymbol{H}\|\le\lambda_{\min}^\star/4$,

$$\begin{aligned}&\left\|\boldsymbol{u}_l-\sum_{k=1}^r\frac{\lambda_k^\star\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l}{\lambda_l}\boldsymbol{u}_k^\star\right\|_2\\&=\left\|\sum_{k=1}^r\frac{\lambda_k^\star\left(\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\right)}{\lambda_l}\left\{\sum_{s=1}^\infty\frac{1}{\lambda_l^s}\boldsymbol{H}^s\boldsymbol{u}_k^\star\right\}\right\|_2\\&\le\sqrt{\sum_{k=1}^r\left|\frac{\lambda_k^\star\left(\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\right)}{\lambda_l}\right|^2}\left\|\left(\sum_{s=1}^\infty\frac{1}{\lambda_l^s}\boldsymbol{H}^s\right)[\boldsymbol{u}_1^\star,\cdots,\boldsymbol{u}_r^\star]\right\|\\&\le\sqrt{\frac{(\lambda_{\max}^\star)^2}{|\lambda_l|^2}\sum_{k=1}^r\left|\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\right|^2}\left\|\left(\sum_{s=1}^\infty\frac{1}{\lambda_l^s}\boldsymbol{H}^s\right)\right\|\left\|[\boldsymbol{u}_1^\star,\cdots,\boldsymbol{u}_r^\star]\right\|\\&\le\frac{\lambda_{\max}^\star}{|\lambda_l|}\sum_{s=1}^\infty\left(\frac{\|\boldsymbol{H}\|}{|\lambda_l|}\right)^s,\end{aligned}\tag{81}$$

where the last inequality holds since $\|[\boldsymbol{u}_1^\star,\cdots,\boldsymbol{u}_r^\star]\|=1$ and $\sum_{k=1}^r\left|\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\right|^2\le\|\boldsymbol{u}_l\|_2^2=1$ (given that the $\boldsymbol{u}_k^\star$'s are orthonormal). In view of the Bauer-Fike theorem and the bound $\|\boldsymbol{H}\|\le\lambda_{\min}^\star/4$, we have the crude lower bound

$$|\lambda_l|\ge\lambda_{\min}^\star-\|\boldsymbol{H}\|\ge3\lambda_{\min}^\star/4.\tag{82}$$

This taken collectively with the inequality (81) yields

$$\begin{aligned}&\left\|\boldsymbol{u}_l-\sum_{k=1}^r\frac{\lambda_k^\star\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l}{\lambda_l}\boldsymbol{u}_k^\star\right\|_2\\&\le\frac{\lambda_{\max}^\star}{|\lambda_l|}\cdot\frac{\|\boldsymbol{H}\|}{|\lambda_l|-\|\boldsymbol{H}\|}\le\frac{\lambda_{\max}^\star}{\frac34\lambda_{\min}^\star}\cdot\frac{\|\boldsymbol{H}\|}{\frac12\lambda_{\min}^\star}=\frac{8\kappa^2}{3}\cdot\frac{\|\boldsymbol{H}\|}{\lambda_{\max}^\star}.\end{aligned}\tag{83}$$

In addition, note that the Euclidean projection of $\boldsymbol{u}_l$ onto the subspace spanned by $\{\boldsymbol{u}_k^\star\}_{1\le k\le r}$ is given by $\sum_{k=1}^r(\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l)\boldsymbol{u}_k^\star$. This together with (83) implies that

$$\begin{aligned}&\frac{8\kappa^2}{3}\cdot\frac{\|\boldsymbol{H}\|}{\lambda_{\max}^\star}\\&\ge\left\|\boldsymbol{u}_l-\sum_{k=1}^r\frac{\lambda_k^\star\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l}{\lambda_l}\boldsymbol{u}_k^\star\right\|_2\ge\min_{\boldsymbol{z}\in\mathsf{span}\{\boldsymbol{u}_1^\star,\cdots,\boldsymbol{u}_r^\star\}}\|\boldsymbol{u}_l-\boldsymbol{z}\|_2\\&=\left\|\boldsymbol{u}_l-\sum_{k=1}^r(\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l)\boldsymbol{u}_k^\star\right\|_2=\sqrt{1-\sum_{k=1}^r\left|\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\right|^2},\end{aligned}\tag{84}$$

where the last identity arises from the Pythagorean theorem.

The above inequality (84) indicates that most of the energy of $\boldsymbol{u}_l$ lies in $\mathsf{span}\{\boldsymbol{u}_1^\star,\cdots,\boldsymbol{u}_r^\star\}$. In order to show that $|\boldsymbol{u}_l^\star{}^\top\boldsymbol{u}_l|\approx1$, it suffices to justify that $|\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l|$ is very small for any $k\neq l$. To this end, taking $\boldsymbol{a}=\boldsymbol{u}_k^\star$ in Lemma 3 and making use of Condition (13b), we arrive at

$$\left|\boldsymbol{u}_k^\star{}^\top\left(\boldsymbol{u}_l-\sum_{i=1}^r\frac{\boldsymbol{u}_i^\star{}^\top\boldsymbol{u}_l}{\lambda_l/\lambda_i^\star}\boldsymbol{u}_i^\star\right)\right|\le c_2\frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2r\log n}.\tag{85}$$

In addition, the orthonormality of $\{\boldsymbol{u}_i^\star\}_{1\le i\le r}$ indicates that

$$\begin{aligned}\boldsymbol{u}_k^\star{}^\top\left(\boldsymbol{u}_l-\sum_{i=1}^r\frac{\boldsymbol{u}_i^\star{}^\top\boldsymbol{u}_l}{\lambda_l/\lambda_i^\star}\boldsymbol{u}_i^\star\right)&=\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l-\frac{\lambda_k^\star}{\lambda_l}\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\\&=\left(1-\frac{\lambda_k^\star}{\lambda_l}\right)\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\end{aligned}\tag{86}$$

and, as a result,

$$\left|\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\right|\le\left|1-\frac{\lambda_k^\star}{\lambda_l}\right|^{-1}c_2\frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2r\log n}.\tag{87}$$

In order to control $\left|1-\frac{\lambda_k^\star}{\lambda_l}\right|^{-1}$, we note that: for all $k\neq l$, it follows from the triangle inequality and Theorem 7 that

$$\begin{aligned}&|\lambda_k^\star-\lambda_l|\ge|\lambda_k^\star-\lambda_l^\star|-|\lambda_l^\star-\lambda_l|\ge|\lambda_k^\star-\lambda_l^\star|-\Delta_l^\star/2\\&\ge|\lambda_k^\star-\lambda_l^\star|/2\ge\Delta_l^\star/2,\\&\implies\left|1-\frac{\lambda_k^\star}{\lambda_l}\right|^{-1}=\frac{|\lambda_l|}{|\lambda_k^\star-\lambda_l|}\le\frac{\lambda_{\max}^\star+\|\boldsymbol{H}\|}{\Delta_l^\star/2}\le\frac{4\lambda_{\max}^\star}{\Delta_l^\star},\end{aligned}\tag{88}$$

where we have also used the condition $\|\boldsymbol{H}\|\le\lambda_{\max}^\star$. Substitution into (87) yields

$$\begin{aligned}\left|\boldsymbol{u}_k^\star{}^\top\boldsymbol{u}_l\right|&\le\frac{4\lambda_{\max}^\star}{\Delta_l^\star}\cdot c_2\frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2r\log n}\\&=\frac{4c_2\sigma_{\max}}{\Delta_l^\star}\sqrt{\mu\kappa^4r\log n}\end{aligned}\tag{89}$$

for all $k \neq l$. Putting the above results together, we arrive at

$$\begin{aligned}
&\min\left\{\|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_2^2\right\} \\
&= 2 - 2|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l| \leq 2 - 2|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l|^2 \\
&\leq 2\left\{1 - \sum_{k=1}^{r} |\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l|^2\right\} + 2\left\{\sum_{k\neq l, k=1}^{r} |\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l|^2\right\} \\
&\lesssim \left(\kappa^2 \frac{\|\boldsymbol{H}\|}{\lambda_{\max}^\star}\right)^2 + \left(\frac{\sigma_{\max}}{\Delta_l^\star}\right)^2 \mu\kappa^4 r^2 \log n \\
&\lesssim \left(\kappa^2 \frac{\sigma_{\max}\sqrt{n\log n}}{\lambda_{\max}^\star}\right)^2 + \left(\frac{\sigma_{\max}}{\Delta_l^\star}\right)^2 \mu\kappa^4 r^2 \log n, \quad (90)
\end{aligned}$$

where the penultimate inequality relies on (84) and (89), and the last line follows from (60b).

Finally, the advertised bound on $|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l|$ can be immediately established by combining the identity $|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l| = 1 - \frac{1}{2}\min\{\|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_2^2\}$ and the above bound (90).

b) Proof for perturbation of linear forms of eigenvectors: Invoke the triangle inequality to obtain

$$\begin{aligned}
&\left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{u}_l^\star\right)\right| \\
&\leq \left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l - \sum_{k=1}^{r}\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}\boldsymbol{u}_k^\star\right)\right| + \sum_{k\neq l, k=1}^{r}\left|\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}\left(\boldsymbol{a}^\top\boldsymbol{u}_k^\star\right)\right|. \quad (91)
\end{aligned}$$

The first term on the right-hand side of (91) can be controlled via Lemma 3 and Condition (13b):

$$\left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l - \sum_{k=1}^{r}\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}\boldsymbol{u}_k^\star\right)\right| \lesssim \frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2 r\log n}\cdot\|\boldsymbol{a}\|_2.$$

Regarding the second term on the right-hand side of (91), we can invoke (87) to derive

$$\begin{aligned}
\left|\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}\right| &\lesssim \frac{|\lambda_k^\star|}{|\lambda_l - \lambda_k^\star|}\frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2 r\log n} \\
&\leq \frac{\lambda_{\max}^\star}{|\lambda_l^\star - \lambda_k^\star|/2}\frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2 r\log n} \\
&\lesssim \frac{\sigma_{\max}}{|\lambda_l^\star - \lambda_k^\star|}\sqrt{\mu\kappa^4 r\log n}, \quad (92)
\end{aligned}$$

where the penultimate inequality results from (88). As a consequence, one has

$$\begin{aligned}
&\sum_{k:k\neq l, 1\leq k\leq r}\left|\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}\left(\boldsymbol{a}^\top\boldsymbol{u}_k^\star\right)\right| \\
&\lesssim r\cdot\sigma_{\max}\sqrt{\mu\kappa^4 r\log n}\cdot\left\{\max_{k:k\neq l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star - \lambda_k^\star|}\right\}. \quad (93)
\end{aligned}$$

Substituting the above bounds into (91) and rearranging terms, we obtain

$$\begin{aligned}
&\left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{u}_l^\star\right)\right| \\
&\lesssim \sigma_{\max}\sqrt{\mu\kappa^2 r\log n}\left\{\frac{1}{\lambda_{\min}^\star}\|\boldsymbol{a}\|_2 + \kappa r\max_{k:k\neq l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star - \lambda_k^\star|}\right\}. \quad (94)
\end{aligned}$$

Next, recall from Theorem 7 that

$$\begin{aligned}
&|\lambda_l - \lambda_l^\star| \leq c_4\kappa r^2\sigma_{\max}\sqrt{\mu\log n} < \frac{1}{2}|\lambda_l^\star| \\
\Longrightarrow\quad &\frac{\lambda_l}{\lambda_l^\star} \in \left[1 - \frac{|\lambda_l - \lambda_l^\star|}{|\lambda_l^\star|},\, 1 + \frac{|\lambda_l - \lambda_l^\star|}{|\lambda_l^\star|}\right] \subset [0.5, 1.5],
\end{aligned}$$

provided that $\mu\kappa^2\, r^4 \lesssim n$ and that Condition (13b) holds. We then make use of the bounds (63) and (79a) to deduce that

$$\begin{aligned}
&\min\left|1 \pm \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\right| \\
&= \left|1 - \frac{\lambda_l^\star}{\lambda_l} + \frac{\lambda_l^\star}{\lambda_l} - \frac{|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l|}{\lambda_l/\lambda_l^\star}\right| \\
&\leq \left|1 - \frac{\lambda_l^\star}{\lambda_l}\right| + \left|\frac{\lambda_l^\star}{\lambda_l}\right|\cdot\left|1 - \left|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l\right|\right| \\
&\lesssim \frac{|\lambda_l - \lambda_l^\star|}{\lambda_{\min}^\star} + \left|1 - \left|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l\right|\right| \\
&\lesssim \frac{\kappa r^2\sigma_{\max}\sqrt{\mu\log n}}{\lambda_{\min}^\star} \\
&\quad + \left(\kappa^4\sigma_{\max}^2 n\log n\right)\left\{\frac{1}{(\Delta_l^\star)^2}\frac{\mu r^2}{n} + \frac{1}{(\lambda_{\max}^\star)^2}\right\}.
\end{aligned}$$

The above bounds taken collectively demonstrate that

$$\begin{aligned}
&\min\left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\right)\right| \\
&\leq \left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l - \mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{u}_l^\star\right)\right| \\
&\leq \left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{u}_l^\star\right)\right| \\
&\quad + \left|\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{a}^\top\boldsymbol{u}_l^\star - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| \\
&= \left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{u}_l^\star\right)\right| + \min\left|1 \pm \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\right|\cdot|\boldsymbol{a}^\top\boldsymbol{u}_l^\star| \\
&\lesssim \mathcal{E}_{\mathsf{au}}, \quad (95)
\end{aligned}$$

where

$$\begin{aligned}
&\mathcal{E}_{\mathsf{au}} \\
&:= \sigma_{\max}\sqrt{\mu\kappa^2 r\log n}\left\{\frac{1}{\lambda_{\min}^\star}\|\boldsymbol{a}\|_2 + \kappa r\max_{k:k\neq l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star - \lambda_k^\star|}\right\} \\
&\quad + \left\{\frac{\kappa r^2\sigma_{\max}\sqrt{\mu\log n}}{\lambda_{\min}^\star} + \left(\kappa^4\sigma_{\max}^2 n\log n\right)\left\{\frac{1}{(\Delta_l^\star)^2}\frac{\mu r^2}{n}\right.\right. \\
&\quad \left.\left. + \frac{1}{(\lambda_{\max}^\star)^2}\right\}\right\}|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|.
\end{aligned}$$

Finally, when it comes to the ℓ_∞ perturbation bound, we shall simply take $\boldsymbol{a} = \boldsymbol{e}_k$ $(1 \leq k \leq r)$ in the above inequality and use the incoherence condition $|\boldsymbol{e}_k^\top\boldsymbol{u}_l^\star| \leq \sqrt{\mu/n}$ $(1 \leq k \leq r)$ to obtain

$$\begin{aligned}
&\left|\boldsymbol{e}_k^\top\left(\boldsymbol{u}_l - \mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{u}_l^\star\right)\right| \\
&\lesssim \sigma_{\max}\sqrt{\mu\kappa^2 r\log n}\left\{\frac{1}{\lambda_{\min}^\star} + \frac{\kappa r}{\Delta_l^\star}\sqrt{\frac{\mu}{n}}\right\} \\
&\quad + \left\{\frac{\kappa r^2\sigma_{\max}\sqrt{\mu\log n}}{\lambda_{\min}^\star} + \left(\kappa^4\sigma_{\max}^2 n\log n\right)\left\{\frac{1}{(\Delta_l^\star)^2}\frac{\mu r^2}{n}\right.\right.
\end{aligned}$$

$$+\frac{1}{(\lambda_{\max}^\star)^2}\Big\}\Big\}\sqrt{\frac{\mu}{n}}$$
$$\lesssim \frac{\sigma_{\max}\sqrt{\mu\kappa^4 r\log n}}{\lambda_{\min}^\star}+\frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\frac{\mu^2\kappa^4 r^3\log n}{n}},$$

where the last line results from Condition (13a) and $\mu\kappa^2\ r^4\lesssim n$. This together with a union bound yields that: with high probability, one has

$$\begin{aligned}&\min\|\boldsymbol{u}_l\pm\boldsymbol{u}_l^\star\|_\infty\\&\le\Big\|\boldsymbol{u}_l-\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{u}_l^\star\Big\|_\infty\\&\lesssim\frac{\sigma_{\max}\sqrt{\mu\kappa^4 r\log n}}{\lambda_{\min}^\star}+\frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\frac{\mu^2\kappa^4 r^3\log n}{n}}\end{aligned}$$

as claimed.

c) Proof of the claims w.r.t. $\widehat{\boldsymbol{u}}_l:=\frac{\boldsymbol{u}_l+\boldsymbol{w}_l}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}$: With the above results in place, we can easily establish these claims w.r.t. $\widehat{\boldsymbol{u}}_l$ as well. In what follows, we demonstrate how to establish the bound (79d) regarding the linear form; the proofs for ℓ_2 and ℓ_∞ bounds follow from nearly identical arguments and are omitted for brevity.

To begin with, repeating the analysis (95) for $\boldsymbol{w}_l$ yields

$$\left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l-\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{u}_l^\star\right)\right|\lesssim\mathcal{E}_{\mathsf{au}},\tag{96}$$

One can therefore combine (95) and (96) to obtain

$$\begin{aligned}&\left|\boldsymbol{a}^\top\Big(\frac{\boldsymbol{u}_l+\boldsymbol{w}_l}{2}\Big)-\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\\&\le\frac{1}{2}\left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l-\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{u}_l^\star\right)\right|\\&\quad+\frac{1}{2}\left|\boldsymbol{a}^\top\left(\boldsymbol{w}_l-\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{u}_l^\star\right)\right|\\&\lesssim\mathcal{E}_{\mathsf{au}},\end{aligned}\tag{97}$$

where the first inequality arises from the triangle inequality as well as the fact $\mathsf{sign}(\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l)=\mathsf{sign}(\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l)$ (see (200) in Lemma 15). This in turn allows us to deduce that

$$\begin{aligned}&\left|\boldsymbol{a}^\top\Big(\frac{\boldsymbol{u}_l+\boldsymbol{w}_l}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}\Big)-\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\\&=\left|\frac{2}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}\boldsymbol{a}^\top\Big(\frac{\boldsymbol{u}_l+\boldsymbol{w}_l}{2}\Big)\right.\\&\quad-\frac{2}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{a}^\top\boldsymbol{u}_l^\star\\&\quad\left.+\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\Big)\left(\frac{2}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}-1\right)\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\\&\le\frac{2}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}\left|\boldsymbol{a}^\top\Big(\frac{\boldsymbol{u}_l+\boldsymbol{w}_l}{2}\Big)-\mathsf{sign}\Big(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\Big)\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\\&\quad+\left(\frac{2}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}-1\right)\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\\&\lesssim\mathcal{E}_{\mathsf{au}}+(2-\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2)\left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\lesssim\mathcal{E}_{\mathsf{au}},\end{aligned}\tag{98}$$

where the first inequality follows from the triangle inequality, the penultimate inequality relies on (97) and the fact $\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2\asymp 1$ (see Lemma 15), and the last inequality follows from the bound $\big|2-\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2\big|=O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2}+\frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right)$ (cf. Lemma 15) in addition to the expression of $\mathcal{E}_{\mathsf{au}}$. This establishes the bound (79d) when $\boldsymbol{u}_l$ is replaced by $\widehat{\boldsymbol{u}}_l$.

2) Proof of Theorem 10: We start by considering the coarse estimator $\widehat{u}_{\boldsymbol{a},l}$. Recall that $\boldsymbol{a}$ is a fixed unit vector obeying $\|\boldsymbol{a}\|_2=1$. The following trivial bound holds true:

$$\min\left|\widehat{u}_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right|\le 1.\tag{99}$$

Consequently, it suffices to focus on the scenario when the upper bound on the right-hand side of (80) is bounded above by some small constant; that is, the case where

$$\frac{\sigma_{\max}r^2\sqrt{\mu\kappa^4\log n}}{\lambda_{\min}^\star}\le c_7\tag{100a}$$
$$\frac{\sigma_{\max}^2\mu r^2\kappa^4\log n}{(\Delta_l^\star)^2}|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\le c_8\tag{100b}$$
$$\sigma_{\max}\sqrt{\mu\kappa^4 r^3\log n}\max_{k\neq l}\frac{|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|}{|\lambda_l^\star-\lambda_k^\star|}\le c_9\tag{100c}$$

for some sufficiently small constants $c_7,c_8,c_9>0$.

The key step is to invoke the Neumann series (cf. Lemma 1) for both $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$, which yields

$$\begin{aligned}&\big(\widehat{u}_{\boldsymbol{a},l}\big)^2\\&=\left|\frac{(\boldsymbol{a}^\top\boldsymbol{u}_l)(\boldsymbol{a}^\top\boldsymbol{w}_l)}{\Big(\sum_{k=1}^r\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_k^\star}\sum_{s=0}^\infty\frac{(\boldsymbol{H}^\top)^s\boldsymbol{u}_k^\star}{\lambda_l^s}\Big)^\top\Big(\sum_{k=1}^r\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}\sum_{s=0}^\infty\frac{\boldsymbol{H}^s\boldsymbol{u}_k^\star}{\lambda_l^s}\Big)}\right|\\&=\left|\frac{\Big(\boldsymbol{a}^\top\boldsymbol{u}_l^\star\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}+\delta_1\Big)\Big(\boldsymbol{a}^\top\boldsymbol{u}_l^\star\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}+\delta_2\Big)}{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}+\delta_3}\right|.\end{aligned}\tag{101}$$

Here, $\delta_1,\delta_2,\delta_3$ are defined as follows

$$\delta_1:=\boldsymbol{a}^\top\left(\boldsymbol{u}_l-\tfrac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{u}_l^\star\right),\tag{102a}$$
$$\delta_2:=\boldsymbol{a}^\top\left(\boldsymbol{w}_l-\tfrac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{u}_l^\star\right),\tag{102b}$$
$$\delta_3:=\mathcal{U}_1+\mathcal{U}_2,\tag{102c}$$

where the quantities $\mathcal{U}_1$ and $\mathcal{U}_2$ are given by

$$\mathcal{U}_1:=\sum_{k:k\neq l}\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_k^\star}\cdot\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star},\tag{103a}$$
$$\mathcal{U}_2:=\sum_{k_1=1}^r\sum_{k_2=1}^r\sum_{s_1,s_2:s_1+s_2\ge 1}\frac{\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_{k_1}^\star}\cdot\frac{\boldsymbol{u}_{k_2}^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_{k_2}^\star}\cdot\frac{\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{H}^{s_1+s_2}\boldsymbol{u}_{k_2}^\star}{\lambda_l^{s_1+s_2}}.\tag{103b}$$

This motivates us to control each term on the right-hand side of (101) separately.

- To begin with, it comes directly from the inequalities (13b) and (100c) that

$$\frac{\sigma_{\max}\sqrt{n\kappa^6\log n}}{\lambda_{\min}^\star}+\frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\mu\kappa^4 r^2\log n}\le c_5+\frac{c_9}{\sqrt{r}},$$

thus allowing us to invoke Theorem 9 to obtain, with probability at least $1-O(n^{-6})$, that

$$\min\left\{|\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l|,|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l|\right\}\ge 3/4.\tag{104}$$

- In view of (94), with probability $1-O(n^{-6})$ one has

$$\begin{aligned}&\max\{|\delta_1|,|\delta_2|\}\\&\lesssim\frac{\sigma_{\max}\sqrt{\mu\kappa^2 r\log n}}{\lambda^\star_{\min}}+\sigma_{\max}\sqrt{\mu\kappa^4 r^3\log n}\max_{k\neq l}\frac{\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_k\right|}{\left|\lambda^\star_l-\lambda^\star_k\right|}.\end{aligned}\tag{105}$$

- Next, substituting (92) into (103a) yields

$$\begin{aligned}|\mathcal{U}_1|&\le r\max_{k\neq l}\left|\frac{\boldsymbol{u}^{\star\top}_k\boldsymbol{w}_l}{\lambda_l/\lambda^\star_k}\cdot\frac{\boldsymbol{u}^{\star\top}_k\boldsymbol{u}_l}{\lambda_l/\lambda^\star_k}\right|\\&\lesssim r\left(\frac{\sigma_{\max}}{\Delta^\star_l}\sqrt{\mu\kappa^4 r\log n}\right)^2=\frac{\sigma^2_{\max}\mu r^2\kappa^4\log n}{\left(\Delta^\star_l\right)^2}.\end{aligned}\tag{106}$$

In addition, one has

$$\begin{aligned}&|\mathcal{U}_2|\\&\overset{\text{(i)}}{\lesssim}\sum_{k_1=1}^{r}\sum_{k_2=1}^{r}\\&\quad\sum_{s_1,s_2:s_1+s_2\ge1}\frac{1}{|\lambda^\star_l/\lambda^\star_{k_1}\cdot\lambda^\star_l/\lambda^\star_{k_2}|}\cdot\left|\frac{\boldsymbol{u}^{\star\top}_{k_1}\boldsymbol{H}^{s_1+s_2}\boldsymbol{u}^\star_{k_2}}{(\lambda^\star_l/2)^{s_1+s_2}}\right|\\&\overset{\text{(ii)}}{\le}\kappa^2r^2\sum_{s_1,s_2:s_1+s_2\ge1}\left(\frac{2}{\lambda^\star_{\min}}\right)^{s_1+s_2}\left|\boldsymbol{u}^{\star\top}_{k_1}\boldsymbol{H}^{s_1+s_2}\boldsymbol{u}^\star_{k_2}\right|\\&\overset{\text{(iii)}}{\le}\kappa^2r^2\sum_{s=1}^{\infty}\sum_{s_1,s_2:s_1+s_2=s}\left(\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}\right)^s\sqrt{\frac{\mu}{n}}\\&\le\sqrt{\frac{\mu}{n}}\kappa^2r^2\sum_{s=1}^{\infty}2^s\left(\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}\right)^s\\&=\sqrt{\frac{\mu}{n}}\kappa^2r^2\frac{\frac{4c_2\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}}{1-\frac{4c_2\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}}\\&\overset{\text{(iv)}}{\asymp}\frac{\sigma_{\max}\sqrt{\mu\kappa^4r^4\log n}}{\lambda^\star_{\min}},\end{aligned}$$

where (i) follows from Corollary 2 (so that $|\lambda_l|\ge|\lambda^\star_l|/2$), (ii) makes use of the fact $|\lambda^\star_k/\lambda^\star_l|\le\kappa$, (iii) arises from Lemma 2, and the last line relies on the assumption (13b) (so that $1-\frac{4c_2\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}\ge\frac12$). The above two bounds taken together yield

$$\begin{aligned}|\delta_3|&\le|\mathcal{U}_1|+|\mathcal{U}_2|\\&\lesssim\frac{\sigma^2_{\max}\mu r^2\kappa^4\log n}{\left(\Delta^\star_l\right)^2}+\frac{\sigma_{\max}r^2\sqrt{\mu\kappa^4\log n}}{\lambda^\star_{\min}}\le1/8,\end{aligned}\tag{107}$$

where the last inequality holds as long as the constants c_7 and c_8 in (100b) and (100c) are sufficiently small.

With the above bounds in mind, we are ready to control the estimation error of $\widehat{u}_{\boldsymbol{a},l}$. We divide the proof into two separate cases as follows.

a) Case 1: when $|\boldsymbol{a}^\top\boldsymbol{u}^\star_l|$ is "small": Consider the case where

$$\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|\le\max\left\{|\delta_1|,|\delta_2|\right\}.$$

Continue the bound in (101) to deduce that

$$\begin{aligned}\left(\widehat{u}_{\boldsymbol{a},l}\right)^2&=\left|\frac{\left(\boldsymbol{a}^\top\boldsymbol{u}^\star_l\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{u}_l}{\lambda_l/\lambda^\star_l}+\delta_1\right)\left(\boldsymbol{a}^\top\boldsymbol{u}^\star_l\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{w}_l}{\lambda_l/\lambda^\star_l}+\delta_2\right)}{\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{w}_l}{\lambda_l/\lambda^\star_l}\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{u}_l}{\lambda_l/\lambda^\star_l}+\delta_3}\right|\\&\overset{\text{(i)}}{\le}\frac{\left\{2|\boldsymbol{a}^\top\boldsymbol{u}^\star_l|+|\delta_1|\right\}\left\{2|\boldsymbol{a}^\top\boldsymbol{u}^\star_l|+|\delta_2|\right\}}{\left(\frac12\right)^2-\frac18}\\&\le72\left(\max\left\{|\delta_1|,|\delta_2|\right\}\right)^2.\end{aligned}\tag{108}$$

Here, the inequality (i) makes use of the bound (104), Corollary 2 (so that $1/2\le|\lambda_l/\lambda^\star_l|\le2$), and the inequality $|\delta_3|\le1/8$. Therefore, we can take the triangle inequality to conclude that

$$\min\left|\widehat{u}_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|\le|\widehat{u}_{\boldsymbol{a},l}|+\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|\lesssim\max\left\{|\delta_1|,|\delta_2|\right\},\tag{109}$$

which combined with (105) leads to the desired bound for this case.

b) Case 2: when $|\boldsymbol{a}^\top\boldsymbol{u}^\star_l|$ is "large": We now move on to the case where

$$\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|>\max\left\{|\delta_1|,|\delta_2|\right\}.\tag{110}$$

By the same argument above in (108), we have

$$\begin{aligned}&\left|\frac{(\boldsymbol{a}^\top\boldsymbol{u}_l)(\boldsymbol{a}^\top\boldsymbol{w}_l)}{\boldsymbol{u}_l{}^\top\boldsymbol{w}_l}-(\boldsymbol{a}^\top\boldsymbol{u}^\star_l)^2\right|\\&=\frac{1}{\left|\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{w}_l}{\lambda_l/\lambda^\star_l}\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{u}_l}{\lambda_l/\lambda^\star_l}+\delta_3\right|}\left|\left(\boldsymbol{a}^\top\boldsymbol{u}^\star_l\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{u}_l}{\lambda_l/\lambda^\star_l}+\delta_1\right)\right.\\&\quad\left.\cdot\left(\boldsymbol{a}^\top\boldsymbol{u}^\star_l\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{w}_l}{\lambda_l/\lambda^\star_l}+\delta_2\right)-(\boldsymbol{a}^\top\boldsymbol{u}^\star_l)^2\left\{\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{w}_l}{\lambda_l/\lambda^\star_l}\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{u}_l}{\lambda_l/\lambda^\star_l}+\delta_3\right\}\right|\\&=\left|\frac{\boldsymbol{a}^\top\boldsymbol{u}^\star_l\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{u}_l}{\lambda_l/\lambda^\star_l}\cdot\delta_2+\boldsymbol{a}^\top\boldsymbol{u}^\star_l\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{w}_l}{\lambda_l/\lambda^\star_l}\cdot\delta_1+\delta_1\delta_2-(\boldsymbol{a}^\top\boldsymbol{u}^\star_l)^2\cdot\delta_3}{\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{w}_l}{\lambda_l/\lambda^\star_l}\cdot\frac{\boldsymbol{u}^{\star\top}_l\boldsymbol{u}_l}{\lambda_l/\lambda^\star_l}+\delta_3}\right|\\&\lesssim\frac{\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|\max\{|\delta_1|,|\delta_2|\}+\left(\max\left\{|\delta_1|,|\delta_2|\right\}\right)^2+\left(\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right)^2|\delta_3|}{\left(\frac12\right)^2-\frac18}\\&\lesssim\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|\left(\frac{\sigma_{\max}\sqrt{\mu\kappa^2r\log n}}{\lambda^\star_{\min}}+\sigma_{\max}\sqrt{\mu\kappa^4r^3\log n}\max_{k\neq l}\frac{\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_k\right|}{\left|\lambda^\star_l-\lambda^\star_k\right|}\right)\\&\quad+\left(\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right)^2\cdot\left(\frac{\sigma^2_{\max}\mu r^2\kappa^4\log n}{\left(\Delta^\star_l\right)^2}+\frac{\sigma_{\max}r^2\sqrt{\mu\kappa^4\log n}}{\lambda^\star_{\min}}\right),\end{aligned}\tag{111}$$

where the last line relies on our upper bounds on $|\delta_1|,|\delta_2|$ and $|\delta_3|$ (see (105) and (107)) as well as the assumption (110).

Recognizing the trivial fact (here, define $\max|a\pm b|=\max\{|a+b|,|a-b|\}$)

$$\max\left|\widehat{u}_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|\ge\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|,$$

we have

$$\min\left|\widehat{u}_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|=\frac{\left|(\widehat{u}_{\boldsymbol{a},l})^2-(\boldsymbol{a}^\top\boldsymbol{u}^\star_l)^2\right|}{\max\left|\widehat{u}_{\boldsymbol{a},l}\pm\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|}\le\frac{\left|(\widehat{u}_{\boldsymbol{a},l})^2-(\boldsymbol{a}^\top\boldsymbol{u}^\star_l)^2\right|}{\left|\boldsymbol{a}^\top\boldsymbol{u}^\star_l\right|}.\tag{112}$$

If $\frac{(\boldsymbol{a}^\top \boldsymbol{u}_l)(\boldsymbol{a}^\top \boldsymbol{w}_l)}{\boldsymbol{u}_l^\top \boldsymbol{w}_l} \geq 0$, then

$$\left|(\widehat{u}_{\boldsymbol{a},l})^2 - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)^2\right| = \left|\frac{(\boldsymbol{a}^\top \boldsymbol{u}_l)(\boldsymbol{a}^\top \boldsymbol{w}_l)}{\boldsymbol{u}_l^\top \boldsymbol{w}_l} - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)^2\right|.$$

If instead $\frac{(\boldsymbol{a}^\top \boldsymbol{u}_l)(\boldsymbol{a}^\top \boldsymbol{w}_l)}{\boldsymbol{u}_l^\top \boldsymbol{w}_l} < 0$ holds, then

$$\begin{aligned}\left|(\widehat{u}_{\boldsymbol{a},l})^2 - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)^2\right| &\leq \left|(\widehat{u}_{\boldsymbol{a},l})^2 + (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)^2\right| \\ &= \left|\frac{(\boldsymbol{a}^\top \boldsymbol{u}_l)(\boldsymbol{a}^\top \boldsymbol{w}_l)}{\boldsymbol{u}_l^\top \boldsymbol{w}_l} - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)^2\right|.\end{aligned}$$

Therefore, putting these bounds together and invoking our bound (111), we deduce that

$$\begin{aligned}\min\left|\widehat{u}_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| &\leq \frac{\left|\frac{(\boldsymbol{a}^\top \boldsymbol{u}_l)(\boldsymbol{a}^\top \boldsymbol{w}_l)}{\boldsymbol{u}_l^\top \boldsymbol{w}_l} - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)^2\right|}{|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|} \\ &\lesssim \frac{\sigma_{\max}\sqrt{\mu\kappa^2 r \log n}}{\lambda_{\min}^\star} + \sigma_{\max}\sqrt{\mu\kappa^4 r^3 \log n}\max_{k\neq l}\frac{|\boldsymbol{a}^\top \boldsymbol{u}_k^\star|}{|\lambda_l^\star - \lambda_k^\star|} \\ &\quad + \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \cdot \left(\frac{\sigma_{\max}^2 \mu r^2 \kappa^4 \log n}{(\Delta_l^\star)^2} + \frac{\sigma_{\max} r^2\sqrt{\mu\kappa^4 \log n}}{\lambda_{\min}^\star}\right) \\ &\lesssim \frac{\sigma_{\max} r^2\sqrt{\mu\kappa^4 \log n}}{\lambda_{\min}^\star} + \sigma_{\max}\sqrt{\mu\kappa^4 r^3 \log n}\max_{k\neq l}\frac{|\boldsymbol{a}^\top \boldsymbol{u}_k^\star|}{|\lambda_l^\star - \lambda_k^\star|} \\ &\quad + \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| \cdot \frac{\sigma_{\max}^2 \mu r^2 \kappa^4 \log n}{(\Delta_l^\star)^2}. \end{aligned} \tag{113}$$

Taking collectively (109) and (113) establishes our estimation error bound for this case.

APPENDIX C
PROOF FOR DISTRIBUTIONAL GUARANTEES (THEOREM 5)

A. Distributional Theory for Linear Forms of Eigenvectors

By virtue of Theorem 7, we know that λ_l, $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$ are all real-valued. Without loss of generality, assume that $\|\boldsymbol{a}\|_2 = 1$ and that $\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l > 0$, which combined with Lemma 15 and Notation 1 yield

$$\boldsymbol{u}_l^\top \boldsymbol{u}_l^\star = 1 - o(1), \quad \boldsymbol{w}_l^\top \boldsymbol{u}_l^\star = 1 - o(1), \quad \boldsymbol{w}_l^\top \boldsymbol{u}_l = 1 - o(1). \tag{114}$$

The main step lies in establishing the following claim

$$\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}} = \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{(\boldsymbol{a}_l^\perp)^\top \left(\boldsymbol{H} + \boldsymbol{H}^\top\right)\boldsymbol{u}_l^\star}{2\lambda_l^\star} + o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right) \tag{115}$$

$$= \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{(\boldsymbol{a}_l^\perp)^\top \left(\boldsymbol{H} + \boldsymbol{H}^\top\right)\boldsymbol{u}_l^\star}{2\lambda_l^\star} + o\big(\sqrt{v_{\boldsymbol{a},l}^\star}\big) \tag{116}$$

with $\boldsymbol{a}_l^\perp := \boldsymbol{a} - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)\boldsymbol{u}_l^\star$, where the last line results from Lemma 12. Let us take Claim (115) as given for now and come back to its proof in the sequel. To establish Theorem 5, it suffices to pin down the distribution of $\frac{(\boldsymbol{a}_l^\perp)^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{2\lambda_l^\star}$ and show that it matches the distributional characterizations stated in Theorem 5. This follows immediately from the classical Berry-Esseen theorem, which we defer to Lemma 14.

The rest of the proof thus boils down to justifying the claim (115), for which we divide into two cases.

1) The Case When $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|$ Is Not "small": In this subsection, we focus on the scenario when $\sqrt{v_{\boldsymbol{a},l}^\star}\log n = o(|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|)$ which, according to Lemma 6, subsumes the case $\sqrt{\widehat{v}_{\boldsymbol{a},l}}\log n = o(|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|)$. Without loss of generality, we assume that $\boldsymbol{a}^\top \boldsymbol{u}_l^\star > 0$. By virtue of Lemma 12 and the assumption $\sigma_{\max} \asymp \sigma_{\min}$, we have

$$v_{\boldsymbol{a},l}^\star \asymp \frac{\sigma_{\max}^2(1 - |\boldsymbol{a}^\top \boldsymbol{u}_l^\star|^2)}{|\lambda_l^\star|^2} \asymp \frac{\sigma_{\max}^2}{|\lambda_l^\star|^2},$$

where the last line follows from our assumption that $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|$ is bounded away from 1. As a result, the regime considered in this subsection enjoys the following property:

$$\frac{\sigma_{\max}\log n}{|\lambda_l^\star|} = o(\boldsymbol{a}^\top \boldsymbol{u}_l^\star). \tag{117}$$

The key starting point is the following decomposition

$$\begin{aligned}&\left(\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}\right)^2 \\ &= \left|\frac{(\boldsymbol{a}^\top \boldsymbol{u}_l)(\boldsymbol{a}^\top \boldsymbol{w}_l)}{\boldsymbol{u}_l^\top \boldsymbol{w}_l}\right| \qquad (118) \\ &= \frac{1}{\left|\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(1 + \frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right) + \tau_3\right|}\left|\left\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\right.\right. \\ &\quad \cdot\left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right) + \tau_1\Big\}\Big\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star} \\ &\quad \left.\left.\cdot\left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}^\top \boldsymbol{u}_l^\star}{\lambda_l^\star}\right) + \tau_2\right\}\right|, \qquad (119)\end{aligned}$$

where

$$\begin{cases}\tau_1 &:= \boldsymbol{a}^\top \boldsymbol{u}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\cdot\left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right), \\ \tau_2 &:= \boldsymbol{a}^\top \boldsymbol{w}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right), \\ \tau_3 &:= \boldsymbol{u}_l^\top \boldsymbol{w}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(1 + \frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right).\end{cases} \tag{120}$$

Here, τ_1, τ_2, τ_3 encompass second- or higher-order terms in the Neumann series. As it turns out, these terms can be well-controlled, as stated in the following lemma.

Lemma 4: Instate the assumptions of Theorem 4. With probability at least $1 - O(n^{-6})$, we have

$$|\tau_1| = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right), \tag{121a}$$

$$|\tau_2| = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right), \tag{121b}$$

$$|\tau_3| \lesssim \left(\frac{\sigma_{\max}\kappa^2\, r\sqrt{\mu\log n}}{|\Delta_l^\star|}\right)^2 + o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right). \tag{121c}$$

Here, we recall that $\Delta_l^\star = \infty$ when $r = 1$, meaning that $|\tau_3| = o(\sigma_{\max}/|\lambda_l^\star|)$ when $r = 1$.

Recall from Theorem 7 and Lemma 15 that $|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l| = 1 - o(1), |\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l| = 1 - o(1)$ and $\lambda_l = (1 + o(1))\lambda_l^\star$. Therefore, Lemma 4 tells us that

$$|\tau_1| = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right) \leq o\left(|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|\right) = o\left(\left|\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\cdot\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\right);$$

$$|\tau_2| = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right) \leq o\left(|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|\right) = o\left(\left|\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\right).$$

In addition, Lemma 4 together with the assumption $\sigma_{\max}\kappa^2\ r\sqrt{\mu\log n} = o\left(|\Delta_l^\star|\right)$ implies that $|\tau_3| = o(1)$.

With these bounds in place, we see that τ_1 and $\frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}$ are indeed very small terms, meaning that the term in (119) involving τ_1 obeys $\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\cdot\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)+\tau_1$ and $\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ have the same signs. Similar conclusions hold as well for the terms involving τ_2 and τ_3. As a result,

$$\begin{aligned}&\mathsf{sign}\left\{\frac{(\boldsymbol{a}^\top\boldsymbol{u}_l)(\boldsymbol{a}^\top\boldsymbol{w}_l)}{\boldsymbol{u}_l^\top\boldsymbol{w}_l}\right\}\\&=\mathsf{sign}\left\{\frac{\left\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right\}\left\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right\}}{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}}\right\}\\&=\mathsf{sign}\left\{(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2\right\}=1.\end{aligned}$$

This indicates that we can safely remove the absolute value function in (119) to obtain

$$\begin{aligned}&\left(\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}\right)^2\\&=\frac{1}{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(1+\frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)+\tau_3}\\&\cdot\left\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)+\tau_1\right\}\\&\cdot\left\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)+\tau_2\right\}.\end{aligned}\tag{122}$$

Armed with this expression, the next lemma develops the distributional characterization of $\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}$.

Lemma 5: Instate the assumptions of Theorem 4. With probability at least $1-O(n^{-6})$, it follows that

$$\begin{aligned}&\left|\left(\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}\right)^2-\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}(\boldsymbol{a}_l^\perp)^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)^2\right|\\&\lesssim|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\cdot(|\tau_1|+|\tau_2|)+(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2\cdot|\tau_3|+O\left(\frac{\sigma_{\max}^2\log n}{\lambda_l^{\star2}}\right).\end{aligned}\tag{123}$$

The proof of this result is given in Section C-D.

Given our assumption $\boldsymbol{a}^\top\boldsymbol{u}_l^\star>0$, the expression (122) together with the bounds on τ_1, τ_2 and τ_3 immediately implies that

$$\begin{aligned}&\left|\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}+\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)\right|\\&\geq(1-o(1))\cdot\boldsymbol{a}^\top\boldsymbol{u}_l^\star.\end{aligned}$$

This combined with Lemma 5 leads to

$$\begin{aligned}&\left|\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}-\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)\right|\\&=\frac{\left|(\widehat{u}_{\boldsymbol{a},l})^2-\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)^2\right|}{\left|\widehat{u}_{\boldsymbol{a},l}+\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)\right|}\\&\lesssim|\tau_1|+|\tau_2|+|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\cdot|\tau_3|+O\left(\frac{\sigma_{\max}^2\log n}{\lambda_l^{\star2}\cdot|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|}\right).\end{aligned}$$

Taking this together with Lemma 4 as well as the condition (117), we arrive at

$$\begin{aligned}&\left|\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}-\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)\right|\\&\lesssim o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right)+|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\cdot\left(\frac{\sigma_{\max}\kappa^2\ r\sqrt{\mu\log n}}{|\Delta_l^\star|}\right)^2\\&=o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right),\end{aligned}\tag{124}$$

where the last line results from our assumption $\boldsymbol{a}^\top\boldsymbol{u}_l^\star=o\left(\frac{\Delta_l^{\star2}}{|\lambda_l^\star|\sigma_{\max}\kappa^4\ r^2\mu\log n}\right)$.

The proof of the claim (115) is thus complete for this case.

2) The Case When $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|$ Is "small": We then move on to the scenario where $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\lesssim\sqrt{v_{\boldsymbol{a},l}^\star}\log^{1.5}n$, which clearly subsumes the case with $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\lesssim\sqrt{\widehat{v}_{\boldsymbol{a},l}}\log^{1.5}n$ (according to Lemma 6). Once again, this restriction combined with Lemma 12 and the assumption $\sigma_{\max}/\sigma_{\min}=O(1)$ requires that

$$|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\lesssim\frac{\sigma_{\max}\log^{1.5}n}{|\lambda_l^\star|}.\tag{125}$$

The proof is built upon the following "first-order" approximations of $\boldsymbol{a}^\top\boldsymbol{u}_l$ and $\boldsymbol{a}^\top\boldsymbol{w}_l$.

Lemma 6: Instate the assumptions of Theorem 4. Fix any vector $\boldsymbol{a}\in\mathbb{R}^n$ with $\|\boldsymbol{a}\|_2=1$, and assume that

$$|\boldsymbol{a}^\top\boldsymbol{u}_k^\star|=o\left(\frac{|\lambda_l^\star-\lambda_k^\star|}{|\lambda_l^\star|\sqrt{\mu\kappa^4r^3\log n}}\right),\qquad\forall k\neq l.\tag{126}$$

Then under Assumption 3, with probability $1-O(n^{-6})$ one has

$$\left|\frac{\lambda_l}{\lambda_l^\star(\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l)}\boldsymbol{a}^\top\boldsymbol{u}_l-\boldsymbol{a}^\top\boldsymbol{u}_l^\star-\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right|=o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right)\tag{127}$$

If we further have $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|=o\left(1/\sqrt{\log n}\right)$, then with probability $1-O(n^{-6})$,

$$\begin{cases}\frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l=\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{(\boldsymbol{a}_l^\perp)^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}+o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right);\\\frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l=\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{(\boldsymbol{a}_l^\perp)^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_l^\star}+o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right).\end{cases}\tag{128}$$

Proof: See Appendix C-E. □

In addition, we claim that the following relations hold

$$\begin{cases}\frac{1}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l=\boldsymbol{a}^\top\boldsymbol{u}_l+o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right),\\\frac{1}{\boldsymbol{w}_l^\top\boldsymbol{u}_l^\star}\boldsymbol{a}^\top\boldsymbol{w}_l=\boldsymbol{a}^\top\boldsymbol{w}_l+o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right).\end{cases}\tag{129}$$

If these claims were valid, then one would have

$$\begin{aligned}\widehat{u}_{\boldsymbol{a},l}^{\mathsf{modified}}&=\left(\boldsymbol{a}^\top\boldsymbol{u}_l+\boldsymbol{a}^\top\boldsymbol{w}_l\right)/2\\&=\frac{1}{2\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\boldsymbol{w}_l^\top\boldsymbol{u}_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l^\star+o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right)\\&=\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{(\boldsymbol{a}_l^\perp)^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{2\lambda_l^\star}+o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right),\end{aligned}$$

where the last inequality follows from Lemma 6. This validates the relation (115) for this case, as long as the relations (129) hold true.

*Proof of the relations (129): As a direct consequence of Lemma 6, we can write

$$\begin{aligned}&\frac{1}{\boldsymbol{u}_l^\top \boldsymbol{u}_l^\star}\boldsymbol{a}^\top \boldsymbol{u}_l - \boldsymbol{a}^\top \boldsymbol{u}_l \\ &= \frac{1}{\boldsymbol{u}_l^\top \boldsymbol{u}_l^\star}\boldsymbol{a}^\top \boldsymbol{u}_l \cdot \left(1 - \boldsymbol{u}^{\star\top}\boldsymbol{u}_l\right) \\ &= \left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star} + o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big)\right)\left(1 - \boldsymbol{u}^{\star\top}\boldsymbol{u}_l\right). \quad (130)\end{aligned}$$

In view of Lemma 13, we have with probability at least $1 - O(n^{-10})$ that

$$\left|\frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right| = O\left(\frac{\sigma_{\max}\sqrt{\log n}}{|\lambda_l^\star|}\right).$$

In addition, Lemma 15 together with Assumption 3 guarantees that, with high probability,

$$\begin{aligned}|1 - \boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l| &= O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \\ &= o\Big(\frac{1}{\log^{1.5} n}\Big).\end{aligned}$$

Putting everything together and using the assumption $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \lesssim \frac{\sigma_{\max}\log^{1.5} n}{|\lambda_l^\star|}$, we conclude that

$$\frac{1}{\boldsymbol{u}_l^\top \boldsymbol{u}^\star}\boldsymbol{a}^\top \boldsymbol{u}_l - \boldsymbol{a}^\top \boldsymbol{u} = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right)$$

as claimed. The claim w.r.t. $\boldsymbol{w}_l$ follows from exactly the same argument.

B. Distributional Theory for Eigenvalues

Take $\boldsymbol{a} = \boldsymbol{u}_l^\star$. By definition, we have $\boldsymbol{a}^\top \boldsymbol{u}_k^\star = 0$ for any $k \neq l$. The expression (127) in Lemma 6 thus indicates that, with probability $1 - O(n^{-6})$, we have

$$\left|\frac{\lambda_l}{\lambda_l^\star(\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l)}\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l - \boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l^\star - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right| = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right).$$

Rearranging terms and using the fact $\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l^\star = 1$ further yield

$$|\lambda_l - \lambda_l^\star - \boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star| = o\left(\sigma_{\max}\right). \quad (131)$$

Consequently, it is sufficient to characterize the distribution of $\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star$. Setting $\boldsymbol{a} = \boldsymbol{u}_l^\star$ in Lemma 14, we see that $W_{\lambda,l} = \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star}{\sqrt{v_{\lambda,l}^\star}} = \frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{2\sqrt{v_{\lambda,l}^\star}}$ obeys

$$\sup_{z\in\mathbb{R}}\left|\,\mathbb{P}(W_{\lambda,l}\le z) - \Phi(z)\,\right| \le \frac{8}{\sqrt{\log n}} \quad (132)$$

as claimed.

C. Proof of Lemma 4

First of all, invoke the Neumann series (cf. Lemma 1 and (56)) and rearrange terms to obtain

$$\begin{aligned}\tau_1 &= \boldsymbol{a}^\top \boldsymbol{u}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right) \\ &= \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left[\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{1}{\lambda_l}\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star + \sum_{s=2}^{\infty}\frac{1}{\lambda_l^s}\boldsymbol{a}^\top \boldsymbol{H}^s\boldsymbol{u}_l^\star\right. \\ &\quad \left. + \sum_{k:k\neq l}\frac{\lambda_k^\star}{\lambda_l^\star}\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\left\{\sum_{s=0}^{\infty}\frac{1}{\lambda_l^s}\boldsymbol{a}^\top \boldsymbol{H}^s\boldsymbol{u}_k^\star\right\}\right] - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star} \\ &\quad \cdot\left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right) \\ &= \underbrace{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(\frac{1}{\lambda_l} - \frac{1}{\lambda_l^\star}\right)\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star}_{=:\tau_{1,1}} + \underbrace{\sum_{k:k\neq l}\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}\sum_{s=2}^{\infty}\frac{1}{\lambda_l^s}\boldsymbol{a}^\top \boldsymbol{H}^s\boldsymbol{u}_k^\star}_{=:\tau_{1,2}} \\ &\quad + \underbrace{\sum_{k=1}^{r}\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}\left\{\sum_{s=2}^{\infty}\frac{1}{\lambda_l^s}\boldsymbol{a}^\top \boldsymbol{H}^s\boldsymbol{u}_k^\star\right\}}_{=:\tau_{1,3}}\end{aligned}$$

and, similarly,

$$\begin{aligned}\tau_3 &= \boldsymbol{u}_l^\top \boldsymbol{w}_l - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(1 + \frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right) \\ &= \left(\sum_{j=1}^{r}\frac{\lambda_j^\star}{\lambda_l}(\boldsymbol{u}_j^{\star\top}\boldsymbol{u}_l)\left\{\sum_{s=0}^{\infty}\frac{1}{\lambda_l^s}\boldsymbol{H}^s\boldsymbol{u}_j^\star\right\}\right)\cdot \\ &\quad \cdot\left(\sum_{j=1}^{r}\frac{\lambda_j^\star}{\lambda_l}(\boldsymbol{u}_j^{\star\top}\boldsymbol{w}_l)\left\{\sum_{s=0}^{\infty}\frac{1}{\lambda_l^s}\boldsymbol{H}^s\boldsymbol{u}_j^\star\right\}\right) \\ &\quad - \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(1 + \frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right) \\ &= \underbrace{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\cdot\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\cdot\left(\frac{1}{\lambda_l} - \frac{1}{\lambda_l^\star}\right)}_{=:\tau_{3,1}} \\ &\quad + \underbrace{\sum_{k:k\neq l}\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_k^\star}\cdot\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_k^\star}}_{=:\tau_{3,2}} \\ &\quad + \underbrace{\sum_{(k_1,k_2)\neq(l,l)}\frac{\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_{k_1}^\star}\cdot\frac{\boldsymbol{u}_{k_2}^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_{k_2}^\star}\cdot\frac{2\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{H}\boldsymbol{u}_{k_2}^\star}{\lambda_l}}_{=:\tau_{3,3}} \\ &\quad + \underbrace{\sum_{k_1=1}^{r}\sum_{k_2=1}^{r}\sum_{s_1,s_2:s_1+s_2\ge 2}\frac{\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_{k_1}^\star}\cdot\frac{\boldsymbol{u}_{k_2}^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_{k_2}^\star}\cdot\frac{\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{H}^{s_1+s_2}\boldsymbol{u}_{k_2}^\star}{\lambda_l^{s_1+s_2}}}_{=:\tau_{3,4}}.\end{aligned}$$

In the special case where $r = 1$, one has $\tau_{1,2} = \tau_{3,2} = \tau_{3,3} = 0$. In what follows, we develop bounds for these terms separately.

a) Controlling $\tau_{1,1}$ and $\tau_{3,1}$: We have learned from Theorem 7, Lemma 13 and Lemma 15 that: under Assumption 3,

one has $\lambda_l = (1+o(1))\lambda_l^\star$, $\boldsymbol{u}_l^\top \boldsymbol{u}_l^\star = 1 - o(1)$,

$$\left|\frac{\lambda_l - \lambda_l^\star}{\lambda_l}\right| \lesssim \frac{\sigma_{\max}\sqrt{\mu\kappa^2 r^4 \log n}}{|\lambda_l^\star|} \quad \text{and} \tag{133}$$

$$\max\left\{|\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star|, |\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star|\right\} \lesssim \sigma_{\max}\sqrt{\log n}. \tag{134}$$

It then follows that

$$\begin{aligned}
|\tau_{1,1}| &\lesssim \left|\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_l^\star\right| \cdot \left|\frac{\lambda_l - \lambda_l^\star}{(\lambda_l^\star)^2}\right| \lesssim \frac{\sigma_{\max}^2 \kappa r^2 \sqrt{\mu}\log n}{|\lambda_l^\star|^2} \\
&\le \frac{\sigma_{\max}}{|\lambda_l^\star|} \cdot \frac{\sigma_{\max}\kappa r^2\sqrt{\mu}\log n}{|\lambda_{\min}^\star|} = o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big) \\
|\tau_{3,1}| &\lesssim \left|\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star\right| \cdot \left|\frac{\lambda_l - \lambda_l^\star}{(\lambda_l^\star)^2}\right| = o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big)
\end{aligned}$$

with the proviso that $\sigma_{\max}\kappa r^2\sqrt{\mu}\log n = o\left(\lambda_{\min}^\star\right)$.

b) Controlling $\tau_{1,2}$ and $\tau_{3,3}$: We make the observation that for any $1 \le k \le r$,

$$\begin{aligned}
\left|\frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_k^\star}{\lambda_l}\right| &\overset{\text{(i)}}{\lesssim} \frac{\sigma_{\max}\sqrt{\log n}}{|\lambda_l|} \asymp \frac{\sigma_{\max}\sqrt{\log n}}{|\lambda_l^\star|} \\
&\overset{\text{(ii)}}{=} o\left(\frac{\Delta_l^\star}{|\lambda_l^\star|\sqrt{\mu\kappa^6 r^3 \log^2 n}}\sqrt{\log n}\right) \\
&= o\left(\frac{|\lambda_l^\star - \lambda_k^\star|}{|\lambda_l^\star|\sqrt{\mu\kappa^4 r^3 \log n}}\right),
\end{aligned} \tag{135}$$

where (i) is a consequence of Lemma 13, and (ii) holds as long as $\sigma_{\max}\sqrt{\mu\kappa^6 r^3 \log^2 n} = o(\Delta_l^\star)$. Similarly, for any $1 \le k_1, k_2 \le r$,

$$\left|\frac{\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{H}\boldsymbol{u}_{k_2}^\star}{\lambda_l}\right| = o\left(\frac{\Delta_l^\star}{|\lambda_l^\star|\sqrt{\mu\kappa^6 r^3 \log n}}\right). \tag{136}$$

Additionally, one can invoke the inequality (87) to deduce that for any $k \ne l$,

$$\begin{aligned}
\left|\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l\right| &\lesssim \left|1 - \frac{\lambda_k^\star}{\lambda_l}\right|^{-1} \frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2 r\log n} \\
&\lesssim \frac{\lambda_l^\star}{|\lambda_l^\star - \lambda_k^\star|} \cdot \frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2 r\log n}.
\end{aligned} \tag{137}$$

Putting the above bounds together and using the conditions $|\boldsymbol{a}^\top \boldsymbol{u}_k^\star| = o\Big(\frac{|\lambda_l^\star - \lambda_k^\star|}{|\lambda_l^\star|\sqrt{\mu\kappa^4 r^3\log n}}\Big)$ and $|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l| = 1 - o(1)$ give

$$\begin{aligned}
|\tau_{1,2}| &\lesssim \sum_{k\ne l, 1\le k\le r} \left|\frac{\lambda_k^\star}{\lambda_l^\star}\right| \cdot \left|\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\right| \left(\left|\frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_k^\star}{\lambda_l}\right| + |\boldsymbol{a}^\top \boldsymbol{u}_k^\star|\right) \\
&= o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right) + o\Big(r \cdot \frac{\lambda_{\max}^\star}{|\lambda_l^\star - \lambda_k^\star|} \cdot \frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2 r\log n} \\
&\quad \cdot \frac{|\lambda_l^\star - \lambda_k^\star|}{|\lambda_l^\star|\sqrt{\mu\kappa^4 r^3\log n}}\Big) \\
&= o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right).
\end{aligned} \tag{138}$$

In addition, for every pair $(k_1, k_2) \ne (l, l)$, Theorem 9 implies that

$$\begin{aligned}
&\left(\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{u}_l\right)\left(\boldsymbol{u}_{k_2}^{\star\top}\boldsymbol{u}_l\right) \\
&= \begin{cases} O\left(\frac{\sigma_{\max}^2}{(\Delta_l^\star)^2}\mu\kappa^4 r\log n\right), & \text{if } k_1 \ne l, k_2 \ne l, \\ O\left(\frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\mu\kappa^4 r\log n}\right), & \text{else,} \end{cases}
\end{aligned} \tag{139}$$

which together with (136) and $\sigma_{\max}\sqrt{\mu\kappa^6 r^3\log^2 n} = o(\Delta_l^\star)$ indicates that

$$\begin{aligned}
I_1 &:= \left|\sum_{k_1\ne l}\sum_{k_2\ne l} \frac{\lambda_{k_1}^\star \lambda_{k_2}^\star}{\lambda_l^2}\left(\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{w}_l\right)\left(\boldsymbol{u}_{k_2}^{\star\top}\boldsymbol{u}_l\right) \cdot \frac{2\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{H}\boldsymbol{u}_{k_2}^\star}{\lambda_l}\right| \\
&\lesssim r^2\kappa^2 \max_{k\ne l, j\ne l} \left|\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l\right| \cdot \left|\boldsymbol{u}_j^{\star\top}\boldsymbol{w}_l\right| \cdot o\left(\frac{\Delta_l^\star}{|\lambda_l^\star|\sqrt{\mu\kappa^6 r^3\log n}}\right) \\
&= o\left(\left(\frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\mu\kappa^6 r^3\log n}\right)^2 \cdot \frac{\Delta_l^\star}{|\lambda_l^\star|\sqrt{\mu\kappa^6 r^3\log n}}\right) \\
&= o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|} \cdot \frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\mu\kappa^6 r^3\log n}\right) \\
&= o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right);
\end{aligned}$$

$$\begin{aligned}
I_2 &:= \left|\sum_{k_1\ne l} \frac{\lambda_{k_1}^\star \lambda_l^\star}{\lambda_l^2}\left(\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{w}_l\right)\left(\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l\right) \cdot \frac{2\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l}\right| \\
&\lesssim r\kappa \max_{k_1\ne l} \left|\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{w}_l\right| \cdot o\left(\frac{\Delta_l^\star}{|\lambda_l^\star|\sqrt{\mu\kappa^6 r^3\log n}}\right) \\
&= o\left(\frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\mu\kappa^6 r^3\log n} \cdot \frac{\Delta_l^\star}{|\lambda_l^\star|\sqrt{\mu\kappa^6 r^3\log n}}\right) \\
&= o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right);
\end{aligned}$$

$$I_3 := \left|\sum_{k_2\ne l} \frac{\lambda_l^\star \lambda_{k_2}^\star}{\lambda_l^2}\left(\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l\right)\left(\boldsymbol{u}_{k_2}^{\star\top}\boldsymbol{u}_l\right) \cdot \frac{2\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_{k_2}^\star}{\lambda_l}\right| = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right).$$

Putting the preceding bounds together yields

$$|\tau_{3,3}| \le I_1 + I_2 + I_3 = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right). \tag{140}$$

c) Controlling $\tau_{1,3}$ and $\tau_{3,4}$: It can be deduced from (133) and $|\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l| = 1 - o(1)$ that

$$\begin{aligned}
|\tau_{1,3}| &\lesssim \sum_{k=1}^r \left|\frac{\lambda_k^\star}{\lambda_l^\star}\right| \cdot \left|\frac{\boldsymbol{u}_k^{\star\top}\boldsymbol{u}_l}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\right| \sum_{s=2}^\infty \left|\frac{\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_k^\star}{\lambda_l^s}\right| \\
&\lesssim \kappa \sum_{k=1}^r \sum_{s=2}^\infty \left|\frac{\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_k^\star}{\lambda_l^s}\right|.
\end{aligned} \tag{141}$$

In addition,

$$\kappa \sum_{k=1}^r \sum_{s=2}^\infty \left|\frac{\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_k^\star}{\lambda_l^s}\right| \lesssim \frac{\kappa}{|\lambda_l^\star|} \sum_{k=1}^r \sum_{s=2}^\infty \left|\frac{2^s \boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}_k^\star}{(\lambda_{\min}^\star)^{s-1}}\right|, \tag{142}$$

where the last line follows since $|\lambda_l| \ge \lambda_{\min}^\star - \|\boldsymbol{H}\| \ge \lambda_{\min}^\star/2$ (by the Bauer-Fike theorem and (61)). Applying Lemma 2

(or (60a)) yields

$$\begin{aligned}
&\frac{\kappa}{|\lambda_l^\star|}\sum_{k=1}^{r}\sum_{s=2}^{\infty}\left|\frac{2^s\boldsymbol{a}^\top\boldsymbol{H}^s\boldsymbol{u}_k^\star}{(\lambda_{\min}^\star)^{s-1}}\right| \\
&\le \frac{\lambda_{\min}^\star\kappa}{|\lambda_l^\star|}\sqrt{\frac{\mu}{n}}\sum_{k=1}^{r}\sum_{s=2}^{\infty}\left(\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}\right)^s \\
&\le \frac{\lambda_{\min}^\star\kappa r}{|\lambda_l^\star|}\sqrt{\frac{\mu}{n}}\frac{\left(\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}\right)^2}{1-\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}} \\
&\asymp \frac{\lambda_{\min}^\star\kappa r}{|\lambda_l^\star|}\sqrt{\frac{\mu}{n}}\left(\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}\right)^2 = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right), \quad (143)
\end{aligned}$$

provided that $\sigma_{\max}\sqrt{\mu\kappa^2r^2n}\log n = o(\lambda_{\min}^\star)$. The above bounds thus imply that

$$|\tau_{1,3}| = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right). \quad (144)$$

Similarly, we can upper bound $\tau_{3,4}$ by

$$\begin{aligned}
|\tau_{3,4}| &\lesssim \kappa^2\sum_{k_1=1}^{r}\sum_{k_2=1}^{r}\sum_{s_1,s_2:s_1+s_2\ge 2}\left|\frac{\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{H}^{s_1+s_2}\boldsymbol{u}_{k_2}^\star}{\lambda_l^{s_1+s_2}}\right| \\
&\lesssim \frac{\kappa^2}{|\lambda_l^\star|}\sum_{k_1=1}^{r}\sum_{k_2=1}^{r}\sum_{s_1,s_2:s_1+s_2\ge 2}\left|\frac{2^{s_1+s_2}\boldsymbol{u}_{k_1}^{\star\top}\boldsymbol{H}^{s_1+s_2}\boldsymbol{u}_{k_2^\star}}{(\lambda_{\min}^\star)^{s_1+s_2-1}}\right| \\
&\le \frac{\lambda_{\min}^\star\kappa^2\ r^2}{|\lambda_l^\star|}\sqrt{\frac{\mu}{n}}\sum_{s_1,s_2:s_1+s_2\ge 2}\left(\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}\right)^{s_1+s_2} \\
&\overset{\text{(i)}}{=} \frac{\lambda_{\min}^\star\kappa^2\ r^2}{|\lambda_l^\star|}\sqrt{\frac{\mu}{n}}\left\{\left(\frac{1}{1-\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}}\right)^2-1\right. \\
&\quad \left.-2\cdot\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}\right\} \\
&\overset{\text{(ii)}}{=} \frac{\lambda_{\min}^\star\kappa^2r^2}{|\lambda_l^\star|}\sqrt{\frac{\mu}{n}}\cdot\left(\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}\right)^2 \\
&\quad\cdot\frac{3-\frac{4c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}}{\left(1-\frac{2c_2\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}\right)^2} \\
&\asymp \frac{\lambda_{\min}^\star\kappa^2r^2}{|\lambda_l^\star|}\sqrt{\frac{\mu}{n}}\left(\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^\star}\right)^2 = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right), \quad (145)
\end{aligned}$$

provided that $\sigma_{\max}\kappa^2\ r^2\sqrt{n\mu}\log n = o(\lambda_{\min}^\star)$. Here, (i) and (ii) make use of the elementary identity that $\sum_{s_1,s_2:s_1+s_2\ge 2}a^{s_1+s_2} = \frac{1}{(1-a)^2}-1-2a = \frac{3a^2(1-2a)}{(1-a)^2}$ for all $0<a<1$.

d) Controlling $\tau_{3,2}$: Clearly, one has $\tau_{3,2}=0$ if $r=1$. When $r\ge 2$, this term $\tau_{3,2}$ is the same term as $\mathcal{U}_1$ defined in (103a). Recalling our bound for inner product $|\boldsymbol{u}_k^{\star\top}\boldsymbol{w}_l|$ and $|\boldsymbol{u}_k^{\star\top}u_l|$ when $k\ne l$ from (137), we deduce that

$$\begin{aligned}
|\tau_{3,2}| &\lesssim \sum_{k:k\ne l}\left(\frac{\lambda_k^\star}{\lambda_l^\star}\right)^2\cdot\left(\frac{\lambda_l^\star}{|\lambda_l^\star-\lambda_k^\star|}\cdot\frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^2r\log n}\right)^2 \\
&\le r\left(\frac{\sigma_{\max}\kappa^2\sqrt{\mu r\log n}}{|\Delta_l^\star|}\right)^2 \le \left(\frac{\sigma_{\max}\kappa^2r\sqrt{\mu\log n}}{|\Delta_l^\star|}\right)^2.
\end{aligned}$$

Putting all this together, we arrive at

$$\begin{aligned}
|\tau_1| &\le |\tau_{1,1}|+|\tau_{1,2}|+|\tau_{1,3}| = o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right) \\
|\tau_3| &\le |\tau_{3,1}|+|\tau_{3,2}|+|\tau_{3,3}|+|\tau_{3,4}| \\
&= O\left(\frac{\sigma_{\max}\kappa^2r\sqrt{\mu\log n}}{|\Delta_l^\star|}\right)^2 + o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right)
\end{aligned}$$

as claimed. The bound on $|\tau_2|$ can be established using exactly the same way as for $|\tau_1|$.

D. Proof of Lemma 5

Direct calculation yields

$$\begin{aligned}
&\left(\widehat{u}_{\boldsymbol{a},l}^{\text{modified}}\right)^2-\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)^2 \\
&= \frac{\mathcal{R}}{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(1+\frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)+\tau_3}, \quad (146)
\end{aligned}$$

where $\mathcal{R}=\mathcal{R}_1-\mathcal{R}_2$ and

$$\begin{aligned}
\mathcal{R}_1 &= \left\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\cdot\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)+\tau_1\right\} \\
&\quad\cdot\left\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)+\tau_2\right\}, \\
\mathcal{R}_2 &= \left\{\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\left(1+\frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)+\tau_3\right\} \\
&\quad\cdot\left\{\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right\}^2.
\end{aligned}$$

Rearranging terms and using the above bounds, we can derive

$$\begin{aligned}
\mathcal{R}_1 &= \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\cdot\left\{\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right)^2+\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right)\cdot\frac{\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right. \\
&\quad\left.+\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\cdot\frac{\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_l^\star}\right\}+\boldsymbol{a}^\top\boldsymbol{u}_l^\star\cdot(\tau_1+\tau_2) \\
&\quad+o\left(|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\cdot(|\tau_1|+|\tau_2|)\right), \\
\mathcal{R}_2 &= \frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}{\lambda_l/\lambda_l^\star}\cdot\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}{\lambda_l/\lambda_l^\star}\cdot\left\{(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2+\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right)\right. \\
&\quad\left.\cdot\frac{\left(\boldsymbol{a}_l^\perp+(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)\boldsymbol{u}_l^\star\right)^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right. \\
&\quad+\left(1+\frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)\cdot\frac{\left(\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)^2}{4(\lambda_l^\star)^2} \\
&\quad\left.+\frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\cdot\frac{\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right\} \\
&\quad+(1+o(1))(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2\cdot\tau_3.
\end{aligned}$$

By definition, we have $\boldsymbol{a}_l^{\perp}+(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)\boldsymbol{u}_l^\star=\boldsymbol{a}$, which further gives

$$\begin{aligned}|\mathcal{R}| &= |\mathcal{R}_1-\mathcal{R}_2| \\ &= \bigg|(1+o(1))\bigg\{\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\cdot\frac{\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_l^\star} \\ &\quad -\left(1+\frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\right)\frac{\left(\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)^2}{4\lambda_l^{\star 2}}- \\ &\quad \frac{\boldsymbol{u}_l^{\star\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\cdot\frac{\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}{\lambda_l^\star}\bigg\} \\ &\quad +\boldsymbol{a}^\top\boldsymbol{u}_l^\star(\tau_1+\tau_2)+o\left(|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|(|\tau_1|+|\tau_2|)\right) \\ &\quad -(1+o(1))(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2\tau_3\bigg| \\ &\lesssim |\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\cdot(|\tau_1|+|\tau_2|)+(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2\cdot|\tau_3|+\frac{\sigma_{\max}^2\log n}{\lambda_l^{\star 2}}\end{aligned}$$

with probability exceeding $1-O(n^{-6})$. Here, the last line uses Lemma 13. Substitution into (146) yields

$$\begin{aligned}\left|\left(\widehat{u}_{\boldsymbol{a},l}^{\text{modified}}\right)^2-\left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star+\frac{1}{2\lambda_l^\star}\boldsymbol{a}_l^{\perp\top}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right)^2\right| &\le \frac{|\mathcal{R}|}{1-o(1)} \\ &\lesssim |\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\cdot(|\tau_1|+|\tau_2|)+(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2\cdot|\tau_3|+\frac{\sigma_{\max}^2\log n}{\lambda_l^{\star 2}}\end{aligned}$$

as claimed.

E. Proof of Lemma 6

The first claim

$$\left|\frac{\lambda_l}{\lambda_l^\star(\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l)}\boldsymbol{a}^\top\boldsymbol{u}_l-\boldsymbol{a}^\top\boldsymbol{u}_l^\star-\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right|=o\left(\frac{\sigma_{\max}}{|\lambda_l^\star|}\right) \tag{147}$$

follows immediately from the bound on $|\tau_1|$ in Lemma 4 and the facts $\lambda_l=(1+o(1))\lambda_l^\star$ and $\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star=1-o(1)$. As an immediate consequence, one has

$$\begin{aligned}\left|\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}\right| &\asymp \left|\frac{\lambda_l}{\lambda_l^\star}\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}\right| \\ &\le \left|\frac{\lambda_l}{\lambda_l^\star}\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}-\boldsymbol{a}^\top\boldsymbol{u}_l^\star-\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right|+|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|+\left|\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right| \\ &\lesssim o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big)+|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|+\frac{\sigma_{\max}\sqrt{\log n}}{|\lambda_l^\star|} \\ &\asymp |\boldsymbol{a}^\top\boldsymbol{u}_l^\star|+\frac{\sigma_{\max}\sqrt{\log n}}{|\lambda_l^\star|}.\end{aligned} \tag{148}$$

Next, the bound (131) tells us that

$$\frac{\lambda_l}{\lambda_l^\star}-1=\frac{\lambda_l-\lambda_l^\star}{\lambda_l^\star}=\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}+o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big).$$

This in turn allows us to bound

$$\begin{aligned}\left|\left(\frac{\lambda_l}{\lambda_l^\star}-1\right)\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}\right| &\lesssim \frac{\sigma_{\max}\sqrt{\log n}}{|\lambda_l^\star|}\left\{|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|+\frac{\sigma_{\max}\sqrt{\log n}}{|\lambda_l^\star|}\right\} \\ &= o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big),\end{aligned}$$

as long as $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|=o\left(1/\sqrt{\log n}\right)$. Putting the above bounds together, we arrive at the advertised bound

$$\begin{aligned}&\left|\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}-\boldsymbol{a}^\top\boldsymbol{u}_l^\star-\frac{(\boldsymbol{a}_l^\perp)^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right| \\ &=\left|\frac{\lambda_l}{\lambda_l^\star}\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}-\boldsymbol{a}^\top\boldsymbol{u}_l^\star-\frac{(\boldsymbol{a}_l^\perp)^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}-\left(\frac{\lambda_l}{\lambda_l^\star}-1\right)\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}\right| \\ &\le\left|\frac{\lambda_l}{\lambda_l^\star}\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}-\boldsymbol{a}^\top\boldsymbol{u}_l^\star-\frac{(\boldsymbol{a}_l^\perp)^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}-\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right| \\ &\quad +O\left(\frac{\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right)+o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big) \\ &\le\left|\frac{\lambda_l}{\lambda_l^\star}\frac{\boldsymbol{a}^\top\boldsymbol{u}_l}{\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star}-\boldsymbol{a}^\top\boldsymbol{u}_l^\star-\frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_l^\star}\right| \\ &\quad +O\left(\frac{\sigma_{\max}\sqrt{\log n}}{|\lambda_l^\star|}|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\right)+o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big) \\ &\le o\Big(\frac{\sigma_{\max}}{|\lambda_l^\star|}\Big),\end{aligned}$$

where the penultimate inequality holds with the proviso that $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|=o\left(1/\sqrt{\log n}\right)$.

The proof for the left eigenvector $\boldsymbol{w}_l$ is essentially the same and is thus omitted for brevity.

Appendix D
Proof for Variance Estimation (Theorem 6)

Without loss of generality, we assume $\|\boldsymbol{a}\|_2=1$ and $\lambda_l^\star=1$ in this section.

A. Proof Outline

e) The estimation accuracy of $\widehat{v}_{\boldsymbol{a},l}$: Before continuing, we note that under the assumption that $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|=o(\|\boldsymbol{a}\|_2)=o(1)$, we have learned from Lemma 12 (or (197)) that

$$\tfrac{1}{2}\sigma_{\min}^2\|\boldsymbol{a}_l^\perp\|_2^2\le v_{\boldsymbol{a},l}^\star\le\sigma_{\max}^2\|\boldsymbol{a}_l^\perp\|_2\le\sigma_{\max}^2\|\boldsymbol{a}\|_2^2. \tag{149}$$

In addition, the assumption $|\boldsymbol{a}^\top\boldsymbol{u}_l^\star|\le(1-\epsilon)\|\boldsymbol{a}\|_2$ for some constant $0<\epsilon<1$ implies that

$$1=\|\boldsymbol{a}\|_2^2\ge\|\boldsymbol{a}_l^\perp\|_2^2=\|\boldsymbol{a}\|_2^2-(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)^2\ge 2\epsilon-\epsilon^2>0. \tag{150}$$

Consequently, under the assumptions $\sigma_{\max}/\sigma_{\min}\asymp 1$ and $\epsilon\asymp 1$ one has

$$v_{\boldsymbol{a},l}^\star\asymp\sigma_{\max}^2. \tag{151}$$

Recall the construction of our variance estimator

$$\widehat{v}_{\boldsymbol{a},l}:=\frac{1}{4\lambda_l^2}\sum_{1\le i,j\le n}\left(\widehat{a}_{l,i}^\perp\widehat{u}_{l,j}+\widehat{a}_{l,j}^\perp\widehat{u}_{l,i}\right)^2\widehat{H}_{ij}^2, \tag{152}$$

where $\widehat{\boldsymbol{a}}_l^\perp:=\boldsymbol{a}-(\boldsymbol{a}^\top\widehat{\boldsymbol{u}}_l)\widehat{\boldsymbol{u}}_l$ and $\boldsymbol{a}_l^\perp:=\boldsymbol{a}-(\boldsymbol{a}^\top\boldsymbol{u}_l^\star)\boldsymbol{u}_l^\star$. To control the size of $\widehat{v}_{\boldsymbol{a},l}$, we find it convenient to introduce a surrogate quantity

$$\widetilde{v}_{\boldsymbol{a},l}:=\frac{1}{4}\sum_{1\le i,j\le n}\left(a_{l,i}^\perp u_{l,j}^\star+a_{l,j}^\perp u_{l,i}^\star\right)^2H_{ij}^2, \tag{153}$$

which turns out to be very close to our estimator $\widehat{v}_{\boldsymbol{a},l}$.

Lemma 7: With probability exceeding $1-O(n^{-5})$, one has $\big|\widehat{v}_{\boldsymbol{a},l}-\widetilde{v}_{\boldsymbol{a},l}\big|=o(v^{\star}_{\boldsymbol{a},l})$.

Proof: See Appendix D-B. □

This observation allows us to turn attention to bounding $\widetilde{v}_{\boldsymbol{a},l}$ instead. The result is this:

Lemma 8: With probability exceeding $1-O(n^{-10})$, one has $|\widetilde{v}_{\boldsymbol{a},l}-v^{\star}_{\boldsymbol{a},l}|=o(v^{\star}_{\boldsymbol{a},l})$.

Proof: See Appendix D-C. □

Putting Lemmas 7-8 together and invoking the triangle inequality, we establish the advertised bound

$$\widehat{v}_{\boldsymbol{a},l}=v^{\star}_{\boldsymbol{a},l}+O(|\widehat{v}_{\boldsymbol{a},l}-\widetilde{v}_{\boldsymbol{a},l}|+|\widetilde{v}_{\boldsymbol{a},l}-v^{\star}_{\boldsymbol{a},l}|)=(1+o(1))v^{\star}_{\boldsymbol{a},l}. \tag{154}$$

The rest of this section is thus mainly dedicated to establishing Lemma 7 and Lemma 8.

f) The estimation accuracy of $\widehat{v}_{\lambda,l}$: The proof for $\widehat{v}_{\lambda,l}=(1+o(1))v^{\star}_{\lambda,l}$ follows from a very similar argument, and is hence omitted.

B. Proof of Lemma 7

Without loss of generality, assume $\langle \boldsymbol{u}_l,\boldsymbol{u}^{\star}_l\rangle>0$, which combined with Lemma 15 and Notation III-A gives

$$\langle \widehat{\boldsymbol{u}}_l,\boldsymbol{u}^{\star}_l\rangle>0 \quad \text{and} \quad \langle \boldsymbol{w}_l,\boldsymbol{u}^{\star}_l\rangle>0. \tag{155}$$

To prove this lemma, let us first recall that, under the assumptions stated in Theorem 4 one has

$$\big\|\widehat{\boldsymbol{u}}_l-\boldsymbol{u}^{\star}_l\big\|_{\infty}\lesssim o\big(\tfrac{1}{\sqrt{\mu n\log n}}\big),\qquad \|\widehat{\boldsymbol{u}}_l\|_{\infty}\le 2\sqrt{\tfrac{\mu}{n}}, \tag{156a}$$

$$\big|\boldsymbol{a}^{\top}\widehat{\boldsymbol{u}}_l-\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l\big|\lesssim o\big(\tfrac{1}{\mu^2\log^2 n}\big), \tag{156b}$$

where (156a) comes from Lemma 15, and (156b) arises from (79d) and (98).

We also need the following properties that control the difference between each term of $\widehat{v}_{\boldsymbol{a},l}$ with that of $\widetilde{v}_{\boldsymbol{a},l}$. In particular, we introduce the following two lemmas.

Lemma 9: Suppose that Assumption 3 holds and that $32\kappa^2\sqrt{\mu r/n}\le 1$. Then one has

$$\xi:=\|\widehat{\boldsymbol{H}}-\boldsymbol{H}\|_{\infty}=\|\boldsymbol{M}_{\mathsf{svd}}-\boldsymbol{M}^{\star}\|_{\infty}\lesssim\sigma_{\max}\frac{\mu\kappa^4 r\sqrt{\log n}}{\sqrt{n}} \tag{157}$$

with probability exceeding $1-O(n^{-9})$.

Proof: See Appendix D-D. □

Lemma 10: With probability at least $1-O(n^{-8})$, for every $1\le i,j\le n$, one has

$$|a^{\perp}_{l,i}|\le\big|a_i\big|+\big|u^{\star}_{l,i}\big|\le|a_i|+\sqrt{\mu/n}; \tag{158}$$

$$\begin{aligned}|\zeta_{ij}|&:=\big|a_i\widehat{u}_{l,j}+a_j\widehat{u}_{l,i}-(a_iu^{\star}_{l,j}+a_ju^{\star}_{l,i})\big|\\&\le o\left(\frac{|a_i|+|a_j|+\big|u^{\star}_{l,i}\big|+\big|u^{\star}_{l,j}\big|}{\sqrt{\mu n\log n}}+\frac{1}{\mu n\log n}\right).\end{aligned} \tag{159}$$

Proof: Recalling that $\widehat{\boldsymbol{a}}^{\perp}_l=\boldsymbol{a}-(\boldsymbol{a}^{\top}\widehat{\boldsymbol{u}}_l)\widehat{\boldsymbol{u}}_l$ and $\boldsymbol{a}^{\perp}_l=\boldsymbol{a}-(\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l)\boldsymbol{u}^{\star}_l$, we obtain as a consequence of inequalities (156a) and (156b) that

$$\begin{aligned}\big|\widehat{a}^{\perp}_{l,i}-a^{\perp}_{l,i}\big|&=\big|\big[\boldsymbol{a}-(\boldsymbol{a}^{\top}\widehat{\boldsymbol{u}}_l)\widehat{\boldsymbol{u}}_l-\boldsymbol{a}+(\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l)\boldsymbol{u}^{\star}_l\big]_i\big|\\&=\big|(\boldsymbol{a}^{\top}\widehat{\boldsymbol{u}}_l)\widehat{u}_{l,i}-(\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l)u^{\star}_{l,i}\big|\\&\le\big|\boldsymbol{a}^{\top}\widehat{\boldsymbol{u}}_l\big|\cdot\big|\widehat{u}_{l,i}-u^{\star}_{l,i}\big|+\big|\boldsymbol{a}^{\top}\widehat{\boldsymbol{u}}_l-\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l\big|\cdot\big|u^{\star}_{l,i}\big|\\&\le o\Big(\tfrac{1}{\sqrt{\mu n\log n}}\Big)+o\Big(\tfrac{1}{\mu^2\log^2 n}\Big)\cdot\sqrt{\tfrac{\mu}{n}}\\&=o\Big(\tfrac{1}{\sqrt{\mu n\log n}}\Big)\end{aligned}$$

$$\begin{aligned}\text{and}\quad\big|a^{\perp}_{l,i}\big|&\le\big|a_i\big|+\big|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}_l\big|\cdot\big|u^{\star}_{l,i}\big|\le\big|a_i\big|+\big|u^{\star}_{l,i}\big|\\&\le\big|a_i\big|+\sqrt{\mu/n}.\end{aligned} \tag{160}$$

These in turn allow one to derive

$$\begin{aligned}&\big|\widehat{a}^{\perp}_{l,i}\widehat{u}_{l,j}+\widehat{a}^{\perp}_{l,j}\widehat{u}_{l,i}-(a^{\perp}_{l,i}u^{\star}_j+a^{\perp}_{l,j}u^{\star}_i)\big|\\&\le\big|\widehat{a}^{\perp}_{l,i}\widehat{u}_{l,j}+\widehat{a}^{\perp}_{l,j}\widehat{u}_{l,i}-(a^{\perp}_{l,i}\widehat{u}_{l,j}+a^{\perp}_{l,j}\widehat{u}_{l,i})\big|\\&\quad+\big|a^{\perp}_{l,i}\widehat{u}_{l,j}+a^{\perp}_{l,j}\widehat{u}_{l,i}-(a^{\perp}_{l,i}u^{\star}_j+a^{\perp}_{l,j}u^{\star}_i)\big|\\&\le\big(\big|\widehat{a}^{\perp}_{l,i}-a^{\perp}_{l,i}\big|+\big|\widehat{a}^{\perp}_{l,j}-a^{\perp}_{l,j}\big|\big)\big(\big|u^{\star}_{l,i}\big|+\big|u^{\star}_{l,j}\big|+\|\widehat{\boldsymbol{u}}_l-\boldsymbol{u}^{\star}_l\|_{\infty}\big)\\&\quad+\big(\big|a^{\perp}_{l,i}\big|+\big|a^{\perp}_{l,j}\big|\big)\|\widehat{\boldsymbol{u}}_l-\boldsymbol{u}^{\star}_l\|_{\infty}\\&\le o\left(\frac{1}{\sqrt{\mu n\log n}}\right)\left(\big|u^{\star}_{l,i}\big|+\big|u^{\star}_{l,j}\big|+o\left(\frac{1}{\sqrt{\mu n\log n}}\right)\right)\\&\quad+\big(|a_i|+|a_j|+\big|u^{\star}_{l,i}\big|+\big|u^{\star}_{l,j}\big|\big)\cdot o\left(\frac{1}{\sqrt{\mu n\log n}}\right)\\&=o\left(\frac{|a_i|+|a_j|+\big|u^{\star}_{l,i}\big|+\big|u^{\star}_{l,j}\big|}{\sqrt{\mu n\log n}}+\frac{1}{\mu n\log n}\right).\end{aligned}$$

□

We are ready to establish the claim. Invoking the above two lemmas yields

$$\begin{aligned}&\big|4\lambda_l^2\widehat{v}_{\boldsymbol{a},l}-4\widetilde{v}_{\boldsymbol{a},l}\big|\\&=\Big|\sum\nolimits_{i,j}\big(a^{\perp}_{l,i}u^{\star}_{l,j}+a^{\perp}_{l,j}u^{\star}_{l,i}+\zeta_{ij}\big)^2\widehat{H}^2_{ij}\\&\quad-\sum\nolimits_{i,j}\big(a^{\perp}_{l,i}u^{\star}_{l,j}+a^{\perp}_{l,j}u^{\star}_{l,i}\big)^2H^2_{ij}\Big|\\&\le\sum\nolimits_{i,j}2|\zeta_{ij}|\left(\big|a^{\perp}_{l,i}u^{\star}_{l,j}+a^{\perp}_{l,j}u^{\star}_{l,i}\big|+|\zeta_{ij}|\right)(|H_{ij}|+\xi)^2\\&\quad+\sum\nolimits_{i,j}2\left(\big|a^{\perp}_{l,i}u^{\star}_{l,j}+a^{\perp}_{l,j}u^{\star}_{l,i}\big|+|\zeta_{ij}|\right)^2\xi\,(|H_{ij}|+\xi)\\&\overset{\text{(i)}}{\lesssim}\frac{1}{n\sqrt{\log n}}\sum\nolimits_{i,j}\left(a_i^2+a_j^2+\big|u^{\star}_{l,i}\big|^2+\big|u^{\star}_{l,j}\big|^2+\frac{1}{\mu n\log n}\right)\\&\quad\cdot(|H_{ij}|+\xi)^2+\frac{\mu}{n}\sum\nolimits_{i,j}\xi\Big(a_i^2+a_j^2+\big|u^{\star}_{l,i}\big|^2+\big|u^{\star}_{l,j}\big|^2\\&\quad+\frac{1}{\mu n\log n}\Big)\cdot(|H_{ij}|+\xi)\\&\lesssim\frac{1}{n\sqrt{\log n}}\\&\quad\cdot\underbrace{\left\{\sum\nolimits_{i,j}\left(a_i^2+a_j^2+\big|u^{\star}_{l,i}\big|^2+\big|u^{\star}_{l,j}\big|^2+\frac{1}{\mu n\log n}\right)H^2_{ij}\right\}}_{=:\alpha_1}\\&\quad+\frac{1}{n\sqrt{\log n}}\\&\quad\cdot\underbrace{\left\{\sum\nolimits_{i,j}\left(a_i^2+a_j^2+\big|u^{\star}_{l,i}\big|^2+\big|u^{\star}_{l,j}\big|^2+\frac{1}{\mu n\log n}\right)\xi^2\right\}}_{=:\alpha_2}\end{aligned}$$

$$+\frac{\mu}{n}\cdot\underbrace{\sum\nolimits_{i,j}\left(a_i^2+a_j^2+\left|u_{l,i}^\star\right|^2+\left|u_{l,j}^\star\right|^2+\frac{1}{\mu n\log n}\right)\xi\left(|H_{ij}|+\xi\right)}_{=:\alpha_3},$$

where (i) follows since (according to (159) and the incoherence condition)

$$\begin{cases}\left|a_{l,i}^\perp u_j^\star+a_{l,j}^\perp u_i^\star\right|+\left|\zeta_{ij}\right| \\ \quad\lesssim\left(\left|a_i\right|+\left|a_j\right|+\left|u_{l,i}^\star\right|+\left|u_{l,j}^\star\right|+\sqrt{\frac{1}{\mu n\log n}}\right) \\ \quad\quad\cdot\left(\left|u_{l,i}^\star\right|+\left|u_{l,j}^\star\right|+\sqrt{\frac{1}{\mu n\log n}}\right) \\ \quad\lesssim\sqrt{a_i^2+a_j^2+\left|u_{l,i}^\star\right|^2+\left|u_{l,j}^\star\right|^2+\frac{1}{\mu n\log n}}\cdot\sqrt{\frac{\mu}{n}} \\ \quad\lesssim\sqrt{\frac{\mu}{n}\left(a_i^2+a_j^2+\left|u_{l,i}^\star\right|^2+\left|u_{l,j}^\star\right|^2+\frac{1}{\mu n\log n}\right)}; \\ \left|\zeta_{ij}\right|\left(\left|a_{l,i}^\perp u_j^\star+a_{l,j}^\perp u_i^\star\right|+\left|\zeta_{ij}\right|\right) \\ \quad\lesssim\left(\left|a_i\right|+\left|a_j\right|+\left|u_{l,i}^\star\right|+\left|u_{l,j}^\star\right|+\sqrt{\frac{1}{\mu n\log n}}\right)^2 \\ \quad\quad\cdot\sqrt{\frac{1}{\mu n\log n}}\cdot\sqrt{\frac{\mu}{n}} \\ \quad\lesssim\left(a_i^2+a_j^2+\left|u_{l,i}^\star\right|^2+\left|u_{l,j}^\star\right|^2+\frac{1}{\mu n\log n}\right)\frac{1}{n\sqrt{\log n}}.\end{cases}$$

We then control α_1, α_2 and α_3 separately.

- Regarding the term α_1, it is first seen (using the assumption $\|\boldsymbol{a}\|_2^2=1$) that

$$\mathbb{E}[\alpha_1]\lesssim\sum\nolimits_{i,j}\left(a_i^2+a_j^2+\left|u_{l,i}^\star\right|^2+\left|u_{l,j}^\star\right|^2+\frac{1}{\mu n\log n}\right)\sigma_{\max}^2 \asymp n\sigma_{\max}^2.$$

In addition, Bernstein's inequality tells us that with probability exceeding $1-O(n^{-10})$,

$$\begin{aligned}&|\alpha_1-\mathbb{E}[\alpha_1]| \\ &\lesssim\sqrt{\sum_{i,j}\mathsf{Var}\left[\left(a_i^2+a_j^2+\frac{\mu}{n}\right)H_{ij}^2\right]\log n} \\ &\quad+\max_{i,j}\left\{\left(a_i^2+a_j^2+\frac{\mu}{n}\right)H_{ij}^2\right\}\log n \\ &\lesssim\sqrt{\sum_{i,j}\left(a_i^2+a_j^2+\frac{\mu}{n}\right)^2B^2\sigma_{\max}^2\log n} \\ &\quad+\max_{i,j}\left(a_i^2+a_j^2+\frac{\mu}{n}\right)B^2\log n \\ &\lesssim\sqrt{\sum_{i,j}\left\{\left(a_i^2+a_j^2\right)^2+\left(\frac{\mu}{n}\right)^2\right\}B^2\sigma_{\max}^2\log n} \\ &\quad+B^2\log n \\ &\lesssim\sqrt{\sum_{i,j}\left\{a_i^2+a_j^2+\left(\frac{\mu}{n}\right)^2\right\}B^2\sigma_{\max}^2\log n} \\ &\quad+B^2\log n \\ &\lesssim B\sigma_{\max}\sqrt{n\log n}+B^2\log n\lesssim\sigma_{\max}^2n,\end{aligned}$$

where the fourth line follows since $a_i^2+a_j^2\le\|\boldsymbol{a}\|_2^2=1$, and the last inequality comes from Assumption 3. Consequently,

$$\alpha_1\le|\alpha_1-\mathbb{E}[\alpha_1]|+\mathbb{E}[\alpha_1]\lesssim\sigma_{\max}^2n. \tag{161}$$

- By virtue of Lemma 9 and the assumption $\|\boldsymbol{a}\|_2=1$, the second term α_2 can be upper bounded by

$$\begin{aligned}\alpha_2&=\sum\nolimits_{i,j}\left(a_i^2+a_j^2+\left|u_{l,i}^\star\right|^2+\left|u_{l,j}^\star\right|^2+\frac{1}{\mu n\log n}\right)\xi^2 \\ &=2n\left\{\|\boldsymbol{a}\|_2^2+\|\boldsymbol{u}_l^\star\|_2^2+\frac{1}{\mu\log n}\right\}\xi^2 \\ &\asymp n\xi^2\lesssim\sigma_{\max}^2\mu^2\kappa^8r^2\log n.\end{aligned}$$

- When it comes to α_3, we first make the observation that

$$\begin{aligned}&\xi\left(|H_{ij}|+\xi\right) \\ &\lesssim\begin{cases}\mu\xi^2\log n & \text{if } |H_{ij}|\lesssim\xi\mu\log n \\ \frac{1}{\mu\log n}\left(|H_{ij}|+\xi\right)^2 & \text{if } |H_{ij}|\gtrsim\xi\mu\log n\end{cases} \\ &\lesssim\mu\xi^2\log n+\frac{1}{\mu\log n}\left(|H_{ij}|+\xi\right)^2\mathbb{1}\{|H_{ij}|\gtrsim\xi\mu\log n\} \\ &\asymp\mu\xi^2\log n+\frac{1}{\mu\log n}H_{ij}^2.\end{aligned}$$

As a consequence, one can derive

$$\begin{aligned}\alpha_3&\lesssim\mu\log n\cdot\underbrace{\sum\nolimits_{i,j}\left(a_i^2+a_j^2+\left|u_{l,i}^\star\right|^2+\left|u_{l,j}^\star\right|^2+\frac{1}{\mu n\log n}\right)\xi^2}_{=\alpha_2} \\ &\quad+\frac{1}{\mu\log n}\cdot\underbrace{\sum\nolimits_{i,j}\left(a_i^2+a_j^2+\left|u_{l,i}^\star\right|^2+\left|u_{l,j}^\star\right|^2+\frac{1}{\mu n\log n}\right)H_{ij}^2}_{=\alpha_1} \\ &=\alpha_2\mu\log n+\frac{1}{\mu\log n}\alpha_1.\end{aligned}$$

Putting the above bounds together, we conclude that

$$\begin{aligned}&\left|4\lambda_l^2\widehat{v}_{\boldsymbol{a},l}-4\lambda_l^{\star2}\widetilde{v}_{\boldsymbol{a},l}\right| \\ &\lesssim\frac{1}{n\sqrt{\log n}}\alpha_1+\frac{1}{n\sqrt{\log n}}\alpha_2+\frac{\mu}{n}\left(\alpha_2\mu\log n+\frac{1}{\mu\log n}\alpha_1\right) \\ &\asymp\frac{1}{n\sqrt{\log n}}\alpha_1+\alpha_2\frac{\mu^2\log n}{n} \\ &\lesssim\left(\frac{1}{\sqrt{\log n}}+\frac{\mu^4\kappa^8r^2\log^2 n}{n}\right)\sigma_{\max}^2 \\ &=o(\sigma_{\max}^2),\end{aligned}$$

provided that $\mu^4\kappa^8\ r^2\log^2 n=o(n)$. This combined with the fact that $\lambda_l=(1+o(1))\lambda_l^\star$ (cf. Corollary 2) and the inequality (151) gives

$$\widehat{v}_{\boldsymbol{a},l}-\widetilde{v}_{\boldsymbol{a},l}=o(v_{\boldsymbol{a},l}^\star).$$

C. Proof of Lemma 8

The randomness in the quantity $\widetilde{v}_{\boldsymbol{a},l}$ purely comes from H_{ij}. Given that $\widetilde{v}_{\boldsymbol{a},l}$ is the sum of independent random variables, it is easily seen that $\mathbb{E}[\widetilde{v}_{\boldsymbol{a},l}]=v_{\boldsymbol{a},l}^\star$. Invoke Bernstein's

inequality [84] to guarantee that with probability at least $1-O(n^{-10})$,

$$\begin{aligned}
&\left|4(\widetilde{v}_{\boldsymbol{a},l}-v^\star_{\boldsymbol{a},l})\right| \\
&=\left|4(\widetilde{v}_{\boldsymbol{a},l}-\mathbb{E}[\widetilde{v}_{\boldsymbol{a},l}])\right| \\
&=\Big|\sum_{1\le i,j\le n}\left(a^\perp_{l,i}u^\star_{l,j}+a^\perp_{l,j}u^\star_{l,i}\right)^2\left(H^2_{ij}-\mathbb{E}[H^2_{ij}]\right)\Big| \\
&\lesssim \sqrt{\sum_{i,j}\mathsf{Var}\left[(a^\perp_{l,i}u^\star_{l,j}+a^\perp_{l,j}u^\star_{l,i})^2H^2_{ij}\right]\log n} \\
&\quad+\max_{i,j}\left\{\left|(a^\perp_{l,i}u^\star_{l,j}+a^\perp_{l,j}u^\star_{l,i})^2B^2\right|\right\}\log n \\
&\overset{\text{(i)}}{\lesssim}\sqrt{\sum_{i,j}(a^\perp_{l,i})^4(u^\star_{l,j})^4\sigma^2_{\max}B^2\log n}+\max_{i,j}(a^\perp_{l,i}u^\star_{l,j})^2B^2\log n \\
&\lesssim\sqrt{\Big\{\max_{i,j}(a^\perp_{l,i})^2(u^\star_{l,j})^2\Big\}\sum_{i,j}(a^\perp_{l,i}u^\star_{l,j})^2\sigma^2_{\max}B^2\log n} \\
&\quad+\max_{i,j}(a^\perp_{l,i}u^\star_{l,j})^2B^2\log n \\
&\overset{\text{(ii)}}{\lesssim} B\sigma_{\max}\sqrt{\frac{\mu\log n}{n}}+\frac{\mu B^2\log n}{n}\overset{\text{(iii)}}{=}o\big(\sigma^2_{\max}\big)=o(v^\star_{\boldsymbol{a},l}),
\end{aligned}$$

where the inequality (i) uses the fact that $\mathsf{Var}[H^2_{ij}]\le\mathbb{E}[H^4_{ij}]\le B^2\mathbb{E}[H^2_{ij}]\le B^2\sigma^2_{\max}$, (ii) follows since $\max_{i,j}(a^\perp_{l,i}u^\star_j)^2\le\|\boldsymbol{u}^\star\|^2_\infty\le\mu/n$ and $\|\boldsymbol{a}^\perp_l\|^2_2\le\|\boldsymbol{a}\|^2_2=\|\boldsymbol{u}^\star\|^2_2=1$, (iii) arises from Assumption 3, and the last identity follows from (151). Combining the above pieces, we obtain $|\widetilde{v}_{\boldsymbol{a},l}-v^\star_{\boldsymbol{a},l}|=o(v^\star_{\boldsymbol{a},l})$ as claimed.

D. Proof of Lemma 9

For notational convenience, we denote by $\boldsymbol{M}_{\mathsf{svd}}=\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{\Sigma}_{\mathsf{svd}}\boldsymbol{V}^\top_{\mathsf{svd}}$ the compact SVD of $\boldsymbol{M}_{\mathsf{svd}}$ (or equivalently, the rank-r SVD of $\boldsymbol{M}$). We shall also define the rotation matrix

$$\boldsymbol{Q}:=\mathop{\arg\min}_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\left\|\begin{bmatrix}\boldsymbol{U}_{\mathsf{svd}}\\ \boldsymbol{V}_{\mathsf{svd}}\end{bmatrix}\boldsymbol{R}-\begin{bmatrix}\boldsymbol{U}^\star\\ \boldsymbol{U}^\star\end{bmatrix}\right\|_{\mathrm{F}}, \tag{162}$$

where $\mathcal{O}^{r\times r}$ denotes the set of $r\times r$ orthonormal matrices.

We first record a perturbation bound regarding the singular subspace of $\boldsymbol{M}_{\mathsf{svd}}$.

Lemma 11: Instate the assumptions of Lemma 9. With probability exceeding $1-O(n^{-9})$ one has

$$\max\left\{\|\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{U}^\star\|_{2,\infty},\|\boldsymbol{V}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{U}^\star\|_{2,\infty}\right\}\lesssim\kappa^2\gamma\sqrt{\frac{\mu r}{n}}, \tag{163}$$

where $\gamma\asymp\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}$.

As immediate consequences of Lemma 11 and Assumption 3, we have

$$\begin{cases}\|\boldsymbol{U}_{\mathsf{svd}}\|_{2,\infty} &\le\|\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{U}^\star\|_{2,\infty}+\|\boldsymbol{U}^\star\|_{2,\infty}\lesssim\kappa\sqrt{\frac{\mu r}{n}},\\ \|\boldsymbol{V}_{\mathsf{svd}}\|_{2,\infty} &\le\|\boldsymbol{V}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{V}^\star\|_{2,\infty}+\|\boldsymbol{V}^\star\|_{2,\infty}\lesssim\kappa\sqrt{\frac{\mu r}{n}},\end{cases} \tag{164}$$

provided that $\kappa^2\sqrt{\mu r/n}\le 1$.

We are now ready to control the entrywise error of $\boldsymbol{M}_{\mathsf{svd}}$. Towards this, we make note of the following bound

$$\begin{aligned}
&\|\boldsymbol{M}_{\mathsf{svd}}-\boldsymbol{M}^\star\|_\infty \\
&\lesssim\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{\Lambda}^\star\|\cdot\|\boldsymbol{U}_{\mathsf{svd}}\|_{2,\infty}\|\boldsymbol{V}_{\mathsf{svd}}\|_{2,\infty} \\
&\quad+\|\boldsymbol{\Lambda}^\star\|\left(\|\boldsymbol{U}_{\mathsf{svd}}\|_{2,\infty}+\|\boldsymbol{V}^\star\|_{2,\infty}\right) \\
&\quad\cdot\left(\|\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{U}^\star\|_{2,\infty}+\|\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{U}^\star\|_{2,\infty}\right),
\end{aligned} \tag{165}$$

which can be obtained by combining the inequalities (C.17)-(C.18) in [9]. In addition, following the argument in [9, Appendix C.3.3], we obtain

$$\|\boldsymbol{Q}^\top\boldsymbol{\Sigma}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{\Lambda}^\star\|\lesssim\gamma\lambda^\star_{\min}, \tag{166}$$

where γ is defined in (168a). Substituting (163), (164) and (166) into (165) and combining terms, we reach

$$\begin{aligned}
&\|\boldsymbol{M}_{\mathsf{svd}}-\boldsymbol{M}^\star\|_\infty \\
&\lesssim\ \gamma\cdot\frac{\mu\kappa^2 r}{n}\lambda^\star_{\min}+\gamma\cdot\frac{\mu\kappa^4 r}{n}\lambda^\star_{\min}\asymp\gamma\cdot\frac{\mu\kappa^4 r}{n}\lambda^\star_{\min} \\
&\asymp\ \sigma_{\max}\sqrt{\frac{\mu^2\kappa^8 r^2\log n}{n}}.
\end{aligned} \tag{167}$$

Finally, it follows from our construction that $\widehat{\boldsymbol{H}}-\boldsymbol{H}=\boldsymbol{M}-\boldsymbol{M}_{\mathsf{svd}}-\boldsymbol{H}=\boldsymbol{M}^\star-\boldsymbol{M}_{\mathsf{svd}}$, which combined with (167) establishes this lemma.

Proof of Lemma 11: In order to invoke [9, Theorem 2.1] and the symmetric dilation trick in [9, Section 3.3], we need to first verify the assumptions required in [9, Section 2.1]. To this end, we introduce the following auxiliary quantity and function

$$\gamma:=c_\gamma\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}} \tag{168a}$$

$$\varphi(x):=\begin{cases}c_\varphi\left(\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}x+\frac{B\log n}{\lambda^\star_{\min}}\right), & \text{if } \frac{1}{\sqrt{n}}\le x\le 1\\ c_\varphi\left(\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}+\frac{B\sqrt{n}\log n}{\lambda^\star_{\min}}\right)x, & \text{if } 0\le x<\frac{1}{\sqrt{n}}\end{cases} \tag{168b}$$

for some sufficiently large constants $c_\gamma,c_\varphi>0$. We then make the following observations:

- To begin with, the two-to-infinity norm of the truth can be bounded by

$$\begin{aligned}
\|\boldsymbol{M}^\star\|_{2,\infty}&=\|\boldsymbol{U}^\star\boldsymbol{\Lambda}^\star\boldsymbol{U}^{\star\top}\|_{2,\infty}\le\|\boldsymbol{U}^\star\|_{2,\infty}\cdot\|\boldsymbol{\Lambda}^\star\|\cdot\|\boldsymbol{U}^\star\| \\
&\le\lambda^\star_{\max}\sqrt{\frac{\mu r}{n}}=\lambda^\star_{\min}\sqrt{\frac{\mu\kappa^2 r}{n}}.
\end{aligned} \tag{169}$$

- In view of the matrix Bernstein inequality [84], with probability exceeding $1-O(n^{-6})$ one has

$$\begin{aligned}
\|\boldsymbol{M}-\boldsymbol{M}^\star\|&=\|\boldsymbol{H}\|\lesssim\sigma_{\max}\sqrt{n\log n} \\
&=\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda^\star_{\min}}\lambda^\star_{\min}\le\gamma\lambda^\star_{\min}.
\end{aligned} \tag{170}$$

- Denote by $\boldsymbol{H}_{i,\cdot}$ the ith row of $\boldsymbol{H}$. For any fixed $\boldsymbol{W} \in \mathbb{R}^{n\times r}$, the matrix Bernstein inequality yields

$$\begin{aligned}\|\boldsymbol{H}_{i,\cdot}\boldsymbol{W}\|_2 &\lesssim \sqrt{\sum_{j=1}^{n}\sigma_{i,j}^2\|\boldsymbol{W}_{j,\cdot}\|_2^2\log n}+B\|\boldsymbol{W}\|_{2,\infty}\log n\\ &\lesssim \sigma_{\max}\|\boldsymbol{W}\|_{\mathrm{F}}\sqrt{\log n}+B\|\boldsymbol{W}\|_{2,\infty}\log n\\ &= \lambda_{\min}^{\star}\|\boldsymbol{W}\|_{2,\infty}\\ &\quad\cdot\left\{\frac{\|\boldsymbol{W}\|_{\mathrm{F}}}{\sqrt{n}\|\boldsymbol{W}\|_{2,\infty}}\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^{\star}}+\frac{B\log n}{\lambda_{\min}^{\star}}\right\}\end{aligned} \tag{171}$$

with probability exceeding $1-O(n^{-10})$. This combined with (168b) and $\frac{1}{\sqrt{n}} \le \frac{\|\boldsymbol{W}\|_{\mathrm{F}}}{\sqrt{n}\|\boldsymbol{W}\|_{2,\infty}} \le 1$ gives

$$\begin{aligned}&\mathbb{P}\left\{\|\boldsymbol{H}_{i,\cdot}\boldsymbol{W}\|_2 \ge \lambda_{\min}^{\star}\|\boldsymbol{W}\|_{2,\infty}\varphi\left(\frac{\|\boldsymbol{W}\|_{\mathrm{F}}}{\sqrt{n}\|\boldsymbol{W}\|_{2,\infty}}\right)\right\}\\ &\ge 1-O\big(n^{-10}\big).\end{aligned} \tag{172}$$

With the above observations in place, [9, Theorem 2.1] together with the dilation trick in [9, Section 3.3] implies that: with probability at least $1-O(n^{-5})$, one has

$$\begin{aligned}&\|\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{U}^{\star}-\boldsymbol{H}\boldsymbol{U}^{\star}(\boldsymbol{\Lambda}^{\star})^{-1}\|_{2,\infty}\\ &= \|\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{M}\boldsymbol{U}^{\star}(\boldsymbol{\Lambda}^{\star})^{-1}\|_{2,\infty}\\ &\lesssim \kappa^2\gamma\|\boldsymbol{U}^{\star}\|_{2,\infty}+\gamma\|\boldsymbol{M}^{\star}\|_{2,\infty}/\lambda_{\min}^{\star}\\ &\lesssim \gamma\kappa^2\sqrt{\frac{\mu r}{n}}+\gamma\sqrt{\frac{\mu\kappa^2 r}{n}} \asymp \gamma\sqrt{\frac{\mu\kappa^4 r}{n}},\end{aligned} \tag{173}$$

where the last line arises from the incoherence condition as well as the inequality (169). In addition,

$$\|\boldsymbol{U}^{\star}(\boldsymbol{\Lambda}^{\star})^{-1}\|_{2,\infty} \le \|\boldsymbol{U}^{\star}\|_{2,\infty}\|(\boldsymbol{\Lambda}^{\star})^{-1}\| \le \frac{1}{\lambda_{\min}^{\star}}\sqrt{\frac{\mu r}{n}},$$

$$\|\boldsymbol{U}^{\star}(\boldsymbol{\Lambda}^{\star})^{-1}\|_{\mathrm{F}} \le \|\boldsymbol{U}^{\star}\|_{\mathrm{F}}\|(\boldsymbol{\Lambda}^{\star})^{-1}\| \le \frac{\sqrt{r}}{\lambda_{\min}^{\star}},$$

which taken collectively with (171) (by setting $\boldsymbol{W} = \boldsymbol{U}^{\star}(\boldsymbol{\Lambda}^{\star})^{-1}$) demonstrate that

$$\begin{aligned}\|\boldsymbol{H}\boldsymbol{U}^{\star}(\boldsymbol{\Lambda}^{\star})^{-1}\|_{2,\infty} &\lesssim \frac{\sigma_{\max}\sqrt{r\log n}}{\lambda_{\min}^{\star}}+\frac{B}{\lambda_{\min}^{\star}}\sqrt{\frac{\mu r\log^2 n}{n}}\\ &\lesssim \frac{\sigma_{\max}\sqrt{\mu r\log n}}{\lambda_{\min}^{\star}}\end{aligned} \tag{174}$$

with probability $1-O(n^{-9})$. Taking it together with (173) and using the triangle inequality immediately yield

$$\begin{aligned}&\|\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{U}^{\star}\|_{2,\infty}\\ &\le \|\boldsymbol{U}_{\mathsf{svd}}\boldsymbol{Q}-\boldsymbol{U}^{\star}-\boldsymbol{H}\boldsymbol{U}^{\star}(\boldsymbol{\Lambda}^{\star})^{-1}\|_{2,\infty}+\|\boldsymbol{H}\boldsymbol{U}^{\star}(\boldsymbol{\Lambda}^{\star})^{-1}\|_{2,\infty}\\ &\lesssim \gamma\sqrt{\frac{\mu\kappa^4 r}{n}}+\frac{\sigma_{\max}\sqrt{\mu r\log n}}{\lambda_{\min}^{\star}} \asymp \gamma\sqrt{\frac{\mu\kappa^4 r}{n}}\end{aligned}$$

as claimed. The part concerning $\boldsymbol{V}_{\mathsf{svd}}$ follows from the same argument. □

Remark 11: The careful reader would note that the results in [9] require another relation between the incoherence condition and γ (namely, Assumption A1). This relation cannot be satisfied with the current choice of γ given in (168a), unless we increase it to

$$\gamma = c_{\gamma}\frac{\sigma_{\max}\sqrt{n\log n}}{\lambda_{\min}^{\star}}+\sqrt{\frac{\mu\kappa^2 r}{n}}. \tag{175}$$

Fortunately, the second term of (175) can be removed with slightly more refined analysis, provided that Assumption 3 holds. In short, the analysis of [9] is built upon a sequence of auxiliary leave-one-out estimates $\boldsymbol{M}^{(m)}$ $(1 \le m \le n)$ — obtained by zeroing out the mth row/column of $\boldsymbol{M}$ — which allows us to approximate $\boldsymbol{M}$ while being statistically independent of the data in the mth row/column. However, the approximation error $\boldsymbol{M}^{(m)}-\boldsymbol{M}$ does not decrease to zero even if $\boldsymbol{H}=\boldsymbol{0}$, leading to a bias term that is non-vanishing as $\sigma_{\max}\to 0$. To address this issue, it suffices to replace $\boldsymbol{M}^{(m)}$ by $\widetilde{\boldsymbol{M}}^{(m)}$, where $\widetilde{\boldsymbol{M}}^{(m)}$ is obtained by replacing all entries in the mth row/column of $\boldsymbol{M}$ by their expected values. This allows us to ensure that $\widetilde{\boldsymbol{M}}^{(m)}-\boldsymbol{M}\to\boldsymbol{0}$ as $\sigma_{\max}\to 0$, which in turn leads to the removal of the second term of (175). The refined proof is nearly identical to the original proof in [9], and is hence omitted here for the sake of brevity.

Appendix E
Proof for Minimax Lower Bounds (Theorem 3)

In the following, we prove each lower bound repectively.

A. Proof of Eqs. (21a) and (21b)

Without loss of generality, consider any $1 \le l \le r$ and any $k \ne l$. Consider the following hypotheses regarding the eigen-decomposition of $\boldsymbol{M}^{\star} \in \mathbb{R}^{n\times n}$:

$$\begin{aligned}\mathcal{H}_0:\ \ &\boldsymbol{M}=\boldsymbol{M}^{\star}+\boldsymbol{H}=\sum_{j=1}^{r}\lambda_j^{\star}\boldsymbol{u}_j^{\star}\boldsymbol{u}_j^{\star\top}+\boldsymbol{H};\\ \mathcal{H}_k:\ \ &\boldsymbol{M}=\boldsymbol{M}_k+\boldsymbol{H}=\lambda_l^{\star}\widetilde{\boldsymbol{u}}_l\widetilde{\boldsymbol{u}}_l^{\top}+\lambda_k^{\star}\widetilde{\boldsymbol{u}}_k\widetilde{\boldsymbol{u}}_k^{\top}\\ &\qquad\qquad+\sum_{j:j\ne k,j\ne l}\lambda_j^{\star}\boldsymbol{u}_j^{\star}\boldsymbol{u}_j^{\star\top}+\boldsymbol{H}.\end{aligned}$$

In words, $\mathcal{H}_k$ is obtained by perturbing the lth and the kth eigenvectors under $\mathcal{H}_0$, with the remaining eigenvectors unaltered. In particular, for any $k \ne l$, we shall pick $\widetilde{\boldsymbol{u}}_l$ and $\widetilde{\boldsymbol{u}}_k$ such that they are equivalent to $\boldsymbol{u}_l^{\star}$ and $\boldsymbol{u}_k^{\star}$ modulo global rotation, namely,

$$\boldsymbol{u}_l^{\star}\boldsymbol{u}_l^{\star\top}+\boldsymbol{u}_k^{\star}\boldsymbol{u}_k^{\star\top}=\widetilde{\boldsymbol{u}}_l\widetilde{\boldsymbol{u}}_l^{\top}+\widetilde{\boldsymbol{u}}_k\widetilde{\boldsymbol{u}}_k^{\top}; \tag{176}$$

as we shall see, this rotational invariance constraint (176) plays a pivotal role in understanding the effect of the eigen-gap upon estimation accuracy. In what follows, we let $\mathbb{P}_0$ (resp. $\mathbb{P}_k$) denote the distribution of $\boldsymbol{M}$ under $\mathcal{H}_0$ (resp. $\mathcal{H}_k$), and let $\mathbb{P}_{0,i,j}$ (resp. $\mathbb{P}_{k,i,j}$) stand for the distribution of M_{ij} under $\mathcal{H}_0$ (resp. $\mathcal{H}_k$).

g) Step 1: calculation of KL divergence: We first calculate the KL divergence of $\mathbb{P}_0$ from $\mathbb{P}_k$ as follows

$$\begin{aligned}\mathsf{KL}\left(\mathbb{P}_k \,\|\, \mathbb{P}_0\right) &\overset{\text{(i)}}{=} \sum_{1\le i,j\le n} \mathsf{KL}\left(\mathbb{P}_{k,i,j} \,\|\, \mathbb{P}_{0,i,j}\right) \\ &\overset{\text{(ii)}}{=} \sum_{1\le i,j\le n} \frac{\left(M^\star_{ij} - (\boldsymbol{M}_k)_{ij}\right)^2}{2\sigma_{ij}^2} \le \frac{\|\boldsymbol{M}^\star - \boldsymbol{M}_k\|_{\mathrm{F}}^2}{2\sigma_{\min}^2}.\end{aligned} \tag{177}$$

Here, (i) holds since KL divergence is additive for independent distributions [85, Chapter 2.4], and (ii) follows since

$$\mathsf{KL}\left(\mathcal{N}(\mu_1,\sigma^2) \,\|\, \mathcal{N}(\mu_2,\sigma^2)\right) = \frac{(\mu_1-\mu_2)^2}{2\sigma^2}.$$

In addition, a little algebra reveals that

$$\begin{aligned}&\boldsymbol{M}^\star - \boldsymbol{M}_k \\ &= \lambda_k^\star\left(\boldsymbol{u}_l^\star\boldsymbol{u}_l^{\star\top} + \boldsymbol{u}_k^\star\boldsymbol{u}_k^{\star\top}\right) + (\lambda_l^\star - \lambda_k^\star)\,\boldsymbol{u}_l^\star\boldsymbol{u}_l^{\star\top} \\ &\quad - \left\{\lambda_k^\star\left(\widetilde{\boldsymbol{u}}_l\widetilde{\boldsymbol{u}}_l^\top + \widetilde{\boldsymbol{u}}_k\widetilde{\boldsymbol{u}}_k^\top\right) + (\lambda_l^\star - \lambda_k^\star)\,\widetilde{\boldsymbol{u}}_l\widetilde{\boldsymbol{u}}_l^\top\right\} \\ &= (\lambda_l^\star - \lambda_k^\star)\left(\boldsymbol{u}_l^\star\boldsymbol{u}_l^{\star\top} - \widetilde{\boldsymbol{u}}_l\widetilde{\boldsymbol{u}}_l^\top\right),\end{aligned}$$

where the last relation makes use of the condition (176). Therefore, we continue the bound (177) to reach

$$\begin{aligned}\mathsf{KL}\left(\mathbb{P}_k \,\|\, \mathbb{P}_0\right) &\le \frac{\|\boldsymbol{M}^\star - \boldsymbol{M}_k\|_{\mathrm{F}}^2}{2\sigma_{\min}^2} \\ &= \frac{(\lambda_l^\star - \lambda_k^\star)^2}{2\sigma_{\min}^2}\left\|\boldsymbol{u}_l^\star\boldsymbol{u}_l^{\star\top} - \widetilde{\boldsymbol{u}}_l\widetilde{\boldsymbol{u}}_l^\top\right\|_{\mathrm{F}}^2 \\ &\le \frac{(\lambda_l^\star - \lambda_k^\star)^2}{\sigma_{\min}^2}\left\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\right\|_2^2,\end{aligned} \tag{178}$$

where the last line holds due to the following inequality that holds for any unit vectors $\boldsymbol{u}$ and $\boldsymbol{v}$:

$$\begin{aligned}&\|\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{v}\boldsymbol{v}^\top\|_{\mathrm{F}}^2 \\ &= \|\boldsymbol{u}\|_2^4 + \|\boldsymbol{v}\|_2^4 - 2\langle\boldsymbol{u}\boldsymbol{u}^\top, \boldsymbol{v}\boldsymbol{v}^\top\rangle = 2 - 2(\boldsymbol{u}^\top\boldsymbol{v})^2 \\ &= (2 - 2\boldsymbol{u}^\top\boldsymbol{v})(1 + \boldsymbol{u}^\top\boldsymbol{v}) = \tfrac{1}{2}\|\boldsymbol{u} - \boldsymbol{v}\|_2^2 \cdot \|\boldsymbol{u} + \boldsymbol{v}\|_2^2 \\ &\le 2\|\boldsymbol{u} - \boldsymbol{v}\|_2^2.\end{aligned} \tag{179}$$

h) Step 2: bounding minimax probability of error: Define the minimax probability of error as follows

$$\begin{aligned}&p_{\mathrm{e},k} \\ &:= \inf_{\psi} \max\Big\{\mathbb{P}\left\{\psi \text{ rejects } \mathcal{H}_0 \mid \mathcal{H}_0\right\}, \mathbb{P}\left\{\psi \text{ rejects } \mathcal{H}_k \mid \mathcal{H}_k\right\}\Big\},\end{aligned} \tag{180}$$

where the infimum is over all tests. Standard minimax lower bounds [85, Theorem 2] tell us that: if

$$\mathsf{KL}\left(\mathbb{P}_k \,\|\, \mathbb{P}_0\right) \le 1/16,$$

then one necessarily has $p_{\mathrm{e},k} \ge 1/5$. This taken collectively with the upper bound (178) implies that: if $\frac{(\lambda_l^\star - \lambda_k^\star)^2}{\sigma_{\min}^2}\left\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\right\|_2^2 \le 1/16$, or equivalently, if

$$\left\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\right\|_2 \le \frac{\sigma_{\min}}{4\left|\lambda_l^\star - \lambda_k^\star\right|},$$

then the minimax probability of error $p_{\mathrm{e},k}$ is lower bounded by $1/5$.

i) Step 3(a): establishing minimax ℓ_2 lower bounds: The above minimax probability of testing error has direct implications on ℓ_2 eigenvector estimation. Suppose that $\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2 = \frac{\sigma_{\min}}{4|\lambda_l^\star - \lambda_k^\star|}$, then the calculation in (178) indicates that

$$\|\boldsymbol{M}^\star - \boldsymbol{M}_k\|_{\mathrm{F}} \le \sqrt{2}\left|\lambda_l^\star - \lambda_k^\star\right| \cdot \|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2 < \tfrac{\sigma_{\min}}{2},$$

and hence $\boldsymbol{M}_k \in \mathcal{M}(\boldsymbol{M}^\star)$. Moreover, when $\sigma_{\min} \le \left|\lambda_l^\star - \lambda_k^\star\right|$, one has

$$\begin{aligned}&\|\boldsymbol{u}_l^\star + \widetilde{\boldsymbol{u}}_l\|_2 = \|2\boldsymbol{u}_l^\star - (\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l)\|_2 \ge \|2\boldsymbol{u}_l^\star\|_2 - \|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2 \\ &\ge 2 - \frac{\sigma_{\min}}{4\left|\lambda_l^\star - \lambda_k^\star\right|} \\ &> \frac{\sigma_{\min}}{4\left|\lambda_l^\star - \lambda_k^\star\right|} = \|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2,\end{aligned}$$

meaning that $\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2 = \min\|\boldsymbol{u}_l^\star \pm \widetilde{\boldsymbol{u}}_l\|_2$. Thus, the standard reduction scheme described in [85, Chapter 2.2] leads to

$$\begin{aligned}\inf_{\widehat{\boldsymbol{u}}_l}\sup_{\boldsymbol{A}\in\mathcal{M}_0(\boldsymbol{M}^\star)} \mathbb{E}\Big[\min\|\widehat{\boldsymbol{u}}_l \pm \boldsymbol{u}_l(\boldsymbol{A})\|_2\Big] &\gtrsim p_{\mathrm{e},k}\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2 \\ &\asymp \frac{\sigma_{\min}}{\left|\lambda_l^\star - \lambda_k^\star\right|},\end{aligned}$$

where the infimum is taken over all estimator for $\boldsymbol{u}_l(\boldsymbol{A})$. Since the preceding bound holds for all $k \ne l$, we conclude that

$$\begin{aligned}\inf_{\widehat{\boldsymbol{u}}_l}\sup_{\boldsymbol{A}\in\mathcal{M}_0(\boldsymbol{M}^\star)} \mathbb{E}\Big[\min\|\widehat{\boldsymbol{u}}_l \pm \boldsymbol{u}_l(\boldsymbol{A})\|_2\Big] &\gtrsim \max_{k:k\ne l}\frac{\sigma_{\min}}{\left|\lambda_l^\star - \lambda_k^\star\right|} \\ &= \frac{\sigma_{\min}}{\Delta_l^\star}.\end{aligned}$$

j) Step 3(b): establishing minimax lower bounds on estimating linear functionals of eigenvectors: The preceding minimax probability of error also has direct implications on estimating linear functionals of eigenvectors. In order to satisfy the rotational invariance constraint (176), we set

$$[\widetilde{\boldsymbol{u}}_l, \widetilde{\boldsymbol{u}}_k] = [\boldsymbol{u}_l^\star, \boldsymbol{u}_k^\star]\begin{bmatrix}\cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k\end{bmatrix}$$

for some $\theta_k \in [-\pi/2, \pi/2]$. Before continuing, we shall also make precise the connection between $\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2$ and θ_k. Specifically, the above relation allows one to derive

$$\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l = \boldsymbol{u}_l^\star(1 - \cos\theta_k) + \boldsymbol{u}_k^\star\sin\theta_k = 2\boldsymbol{u}_l^\star\sin^2\tfrac{\theta_k}{2} + \boldsymbol{u}_k^\star\sin\theta_k$$

and, as a consequence,

$$\begin{cases}\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2 \le (1 - \cos\theta_k) + |\sin\theta_k| = 2\sin^2\frac{\theta_k}{2} + |\sin\theta_k| \\ \qquad\qquad \le 3|\theta_k|, \\ \|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2 \ge |\sin\theta_k| \ge \frac{2}{\pi}|\theta_k|.\end{cases} \tag{181}$$

In what follows, we shall take $\mathsf{sign}(\theta_k) = \mathsf{sign}\big(\frac{\boldsymbol{a}^\top\boldsymbol{u}_k^\star}{\boldsymbol{a}^\top\boldsymbol{u}_l^\star}\big)$, and generate the magnitude $|\theta_k|$ as follows

$$|\theta_k| \sim \mathsf{Uniform}\Big(\Big[\tfrac{\sigma_{\min}}{120\left|\lambda_l^\star - \lambda_k^\star\right|}, \tfrac{\sigma_{\min}}{12\left|\lambda_l^\star - \lambda_k^\star\right|}\Big]\Big), \tag{182}$$

which combined with (181) guarantees that

$$\frac{\sigma_{\min}}{60\pi\left|\lambda_l^\star - \lambda_k^\star\right|} \le \|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2 \le \frac{\sigma_{\min}}{4\left|\lambda_l^\star - \lambda_k^\star\right|}. \tag{183}$$

We aim to translate the difficulty in distinguishing $\mathcal{H}_0$ and $\mathcal{H}_k$ into a fundamental lower bound on estimating the linear functional. Towards this, we are in need of computing the difference of the linear functional under these two hypotheses (namely, $\boldsymbol{a}^\top \boldsymbol{u}_l^\star - \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l$). The above expressions give

$$\begin{aligned}
&\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star - \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l\right| \\
&= \left|2\left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right)\sin^2\tfrac{\theta_k}{2} + \left(\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right)\sin\theta_k\right| \\
&\overset{\text{(i)}}{=} 2\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\sin^2\tfrac{\theta_k}{2} + \left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|\cdot|\sin\theta_k| \qquad (184)\\
&\asymp \theta_k^2\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| + |\theta_k|\cdot\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|,
\end{aligned}$$

where the identity (i) results from the condition $\mathsf{sign}(\theta_k) = \mathsf{sign}\big(\frac{\boldsymbol{a}^\top \boldsymbol{u}_k^\star}{\boldsymbol{a}^\top \boldsymbol{u}_l^\star}\big)$.

We shall also control $\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l\right|$, for which we divide into two cases:

- If $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \geq \frac{\sigma_{\min}}{|\lambda_l^\star - \lambda_k^\star|}|\boldsymbol{a}^\top \boldsymbol{u}_k^\star| \geq 12|\theta_k|\cdot|\boldsymbol{a}^\top \boldsymbol{u}_k^\star|$, then the identity $\boldsymbol{u}_l^\star + \widetilde{\boldsymbol{u}}_l = 2\boldsymbol{u}_l^\star - \left(\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\right)$ together with (184) yields

$$\begin{aligned}
&\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l\right| \\
&= \left|2|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| - 2|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|\sin^2\tfrac{\theta_k}{2} - |\boldsymbol{a}^\top \boldsymbol{u}_k^\star|\cdot|\sin\theta_k|\right| \\
&\geq 2\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\cos^2\tfrac{\theta_k}{2} - |\boldsymbol{a}^\top \boldsymbol{u}_k^\star|\cdot|\theta_k| \\
&\geq \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| - |\boldsymbol{a}^\top \boldsymbol{u}_k^\star|\cdot|\theta_k| \geq \tfrac{1}{2}|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| \\
&\gtrsim \theta_k^2\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| + |\theta_k|\cdot\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|,
\end{aligned}$$

  where we have used the fact that $\sigma_{\min} \leq |\lambda_l^\star - \lambda_k^\star|$ (so that $|\theta_k| \leq \frac{\sigma_{\min}}{12|\lambda_l^\star - \lambda_k^\star|} \leq \frac{1}{12}$ and hence $\cos^2\frac{\theta_k}{2} \geq \frac{1}{2}$).
- Suppose now that $|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| < \frac{\sigma_{\min}}{|\lambda_l^\star - \lambda_k^\star|}|\boldsymbol{a}^\top \boldsymbol{u}_k^\star|$. As one can easily verify (which we omit for brevity), the scheme (182) guarantees that with probability exceeding $1/2$, one has

$$\begin{aligned}
&\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l\right| \\
&= \left|2|\boldsymbol{a}^\top \boldsymbol{u}_l^\star| - 2|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|\sin^2\tfrac{\theta_k}{2} - |\boldsymbol{a}^\top \boldsymbol{u}_k^\star|\cdot|\sin\theta_k|\right| \\
&\gtrsim \max\left\{|\boldsymbol{a}^\top \boldsymbol{u}_l^\star|,\, |\boldsymbol{a}^\top \boldsymbol{u}_k^\star|\cdot|\theta_k|\right\} \\
&\gtrsim \theta_k^2\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| + |\theta_k|\cdot\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|.
\end{aligned}$$

Putting the above cases together reveals that

$$\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star + \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l\right| \gtrsim \theta_k^2\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| + |\theta_k|\cdot\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|$$

with probability exceeding $1/2$. Consequently, one can find $|\theta_k| \in \left[\frac{\sigma_{\min}}{120\left|\lambda_l^\star - \lambda_k^\star\right|}, \frac{\sigma_{\min}}{12\left|\lambda_l^\star - \lambda_k^\star\right|}\right]$ such that

$$\begin{aligned}
&\min\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star \pm \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l\right| \\
&\gtrsim \theta_k^2\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| + |\theta_k|\cdot\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right| \\
&\asymp \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\frac{\sigma_{\min}^2}{\left|\lambda_l^\star - \lambda_k^\star\right|^2} + \left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|\frac{\sigma_{\min}}{\left|\lambda_l^\star - \lambda_k^\star\right|},
\end{aligned}$$

where we recall that $\min|a \pm b| := \min\{|a-b|, |a+b|\}$.

Suppose for the moment that $\boldsymbol{M}_k \in \mathcal{M}_0(\boldsymbol{M}^\star)$. Applying the standard reduction scheme [85, Chapter 2.2] once again yields

$$\begin{aligned}
&\inf_{\widehat{u}_{\boldsymbol{a},l}} \sup_{\boldsymbol{A}\in\mathcal{M}_0(\boldsymbol{M}^\star)} \mathbb{E}\Big[\min\left|\widehat{u}_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{A})\right|\Big] \\
&\gtrsim p_{\mathrm{e},k}\min\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star \pm \boldsymbol{a}^\top \widetilde{\boldsymbol{u}}_l\right| \gtrsim \frac{\sigma_{\min}^2\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|}{\left|\lambda_l^\star - \lambda_k^\star\right|^2} + \frac{\sigma_{\min}\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star - \lambda_k^\star\right|},
\end{aligned}$$

where the infimum is taken over all estimators for $\boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{A})$. Recognizing that the preceding inequality holds for all $k \neq l$, we immediately arrive at the advertised claim

$$\begin{aligned}
&\inf_{\widehat{u}_{\boldsymbol{a},l}} \sup_{\boldsymbol{A}\in\mathcal{M}_0(\boldsymbol{M}^\star)} \mathbb{E}\Big[\min\left|\widehat{u}_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top \boldsymbol{u}_l(\boldsymbol{A})\right|\Big] \\
&\gtrsim \sigma_{\min}^2 \max_{k:k\neq l}\frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|}{\left|\lambda_l^\star - \lambda_k^\star\right|^2} + \sigma_{\min}\max_{k:k\neq l}\frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star - \lambda_k^\star\right|} \\
&= \left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right|\frac{\sigma_{\min}^2}{\Delta_l^{\star 2}} + \sigma_{\min}\max_{k:k\neq l}\frac{\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right|}{\left|\lambda_l^\star - \lambda_k^\star\right|}.
\end{aligned}$$

Finally, it remains to justify that $\boldsymbol{M}_k \in \mathcal{M}_0(\boldsymbol{M}^\star)$ for all $k \neq l$. When $|\theta_k| \leq \frac{\sigma_{\min}}{12|\lambda_l^\star - \lambda_k^\star|}$, invoking (178) and (181) yields

$$\begin{aligned}
\|\boldsymbol{M}^\star - \boldsymbol{M}_k\|_{\mathrm{F}}^2 &\leq 2\left(\lambda_l^\star - \lambda_k^\star\right)^2\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_l\|_2^2 \\
&\leq 18\left(\lambda_l^\star - \lambda_k^\star\right)^2|\theta_k|^2 M < \sigma_{\min}^2/4.
\end{aligned}$$

This means that $\boldsymbol{M}_k \in \mathcal{M}_0(\boldsymbol{M}^\star)$, thus concluding the proof.

B. Proof of Eq. (21c)

Consider the following two hypotheses

$$\begin{aligned}
\mathcal{H}_0:&\ \ \boldsymbol{M} = \boldsymbol{M}_0 + \boldsymbol{H} := \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} + \sum_{j:j\neq l}\lambda_j^\star \boldsymbol{v}_j \boldsymbol{v}_j^\top + \boldsymbol{H}, \\
\mathcal{H}_{\boldsymbol{a}}:&\ \ \boldsymbol{M} = \boldsymbol{M}_{\boldsymbol{a}} + \boldsymbol{H} := \lambda_l^\star \widetilde{\boldsymbol{u}}_{\boldsymbol{a}} \widetilde{\boldsymbol{u}}_{\boldsymbol{a}}^\top + \sum_{j:j\neq l}\lambda_j^\star \boldsymbol{v}_j \boldsymbol{v}_j^\top + \boldsymbol{H},
\end{aligned}$$

where the $\boldsymbol{v}_j$'s are orthonormal vectors obeying $\langle \boldsymbol{v}_j, \boldsymbol{a}\rangle = \langle \boldsymbol{v}_j, \boldsymbol{u}_l^\star\rangle = 0$ for any $j \neq l$, and $\widetilde{\boldsymbol{u}}_{\boldsymbol{a}}$ is defined as

$$\begin{aligned}
&\widetilde{\boldsymbol{u}}_{\boldsymbol{a}} := \frac{1}{\|\boldsymbol{u}_l^\star + \delta\boldsymbol{a}_\perp\|_2}(\boldsymbol{u}_l^\star + \delta\boldsymbol{a}_\perp), \quad \text{with} \\
&\boldsymbol{a}_\perp = \boldsymbol{a} - (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)\boldsymbol{u}_l^\star \ \text{ and } \ \delta = \frac{\sigma_{\min}}{12|\lambda_l^\star|\cdot\|\boldsymbol{a}_\perp\|_2}.
\end{aligned}$$

Recognizing the simple fact $\langle \boldsymbol{a}_\perp, \boldsymbol{u}_l^\star\rangle = 0$, we can derive

$$\begin{aligned}
&\|\boldsymbol{u}_l^\star + \delta\boldsymbol{a}_\perp\|_2 = \sqrt{1 + \delta^2\|\boldsymbol{a}_\perp\|_2^2} \ \text{ and} \\
&\widetilde{\boldsymbol{u}}_{\boldsymbol{a}} := \frac{1}{\sqrt{1 + \delta^2\|\boldsymbol{a}_\perp\|_2^2}}(\boldsymbol{u}_l^\star + \delta\boldsymbol{a}_\perp).
\end{aligned}$$

Our proof proceeds as follows. Without loss of generality, we shall assume that $\boldsymbol{a}^\top \boldsymbol{u}^\star \geq 0$.

- Let $\mathbb{P}_0$ (resp. $\mathbb{P}_{\boldsymbol{a}}$) denote the distribution of $\boldsymbol{M}$ under $\mathcal{H}_0$ (resp. $\mathcal{H}_{\boldsymbol{a}}$). Repeating the derivation in (177) gives

$$\begin{aligned}
\mathsf{KL}\left(\mathbb{P}_{\boldsymbol{a}} \,\|\, \mathbb{P}_0\right) &\leq \frac{\|\boldsymbol{M}_0 - \boldsymbol{M}_{\boldsymbol{a}}\|_{\mathrm{F}}^2}{2\sigma_{\min}^2} \\
&= \frac{(\lambda_l^\star)^2}{2\sigma_{\min}^2}\|\boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} - \widetilde{\boldsymbol{u}}_{\boldsymbol{a}}\widetilde{\boldsymbol{u}}_{\boldsymbol{a}}^\top\|_{\mathrm{F}}^2 \\
&\leq \frac{(\lambda_l^\star)^2}{\sigma_{\min}^2}\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_{\boldsymbol{a}}\|_2^2, \qquad (185)
\end{aligned}$$

where the last inequality comes from (179). In addition,

$$\begin{aligned}
&\|\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_{\boldsymbol{a}}\|_2 \\
&= \Big\|\boldsymbol{u}_l^\star - \frac{1}{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2}}(\boldsymbol{u}_l^\star + \delta\boldsymbol{a}_\perp)\Big\|_2 \\
&\le \big\|\boldsymbol{u}_l^\star - (\boldsymbol{u}_l^\star + \delta\boldsymbol{a}_\perp)\big\|_2 + \Big(1 - \frac{1}{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2}}\Big) \\
&\quad \cdot \big\|\boldsymbol{u}_l^\star + \delta\boldsymbol{a}_\perp\big\|_2 \\
&\le \delta\|\boldsymbol{a}_\perp\|_2 + \frac{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2} - 1}{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2}} \cdot (1+\delta\|\boldsymbol{a}_\perp\|_2) \\
&\le \delta\|\boldsymbol{a}_\perp\|_2 + (1+\delta\|\boldsymbol{a}_\perp\|_2)\,\delta\|\boldsymbol{a}_\perp\|_2 \le 3\delta\|\boldsymbol{a}_\perp\|_2,
\end{aligned} \tag{186}$$

where we have made use of the fact that $\delta\|\boldsymbol{a}_\perp\|_2 = \frac{\sigma_{\min}}{12|\lambda_l^\star|} \le \frac{1}{12}$. Combining this with (185) and our choice of δ, we arrive at

$$\mathsf{KL}\left(\mathbb{P}_{\boldsymbol{a}} \,\|\, \mathbb{P}_0\right) \le \frac{9\delta^2(\lambda_l^\star)^2\|\boldsymbol{a}_\perp\|_2^2}{\sigma_{\min}^2} = \frac{1}{16}. \tag{187}$$

- Define the minimax probability of error as follows

$$\begin{aligned}
&p_{\mathrm{e},\boldsymbol{a}} \\
&:= \inf_{\psi} \max\Big\{\mathbb{P}\{\psi \text{ rejects } \mathcal{H}_0 \,|\, \mathcal{H}_0\}, \mathbb{P}\{\psi \text{ rejects } \mathcal{H}_{\boldsymbol{a}} \,|\, \mathcal{H}_{\boldsymbol{a}}\}\Big\},
\end{aligned}$$

where the infimum is over all tests. It then follows from [85, Theorem 2] that: if

$$\mathsf{KL}\left(\mathbb{P}_k \,\|\, \mathbb{P}_0\right) \le 1/16,$$

one necessarily has $p_{\mathrm{e},\boldsymbol{a}} \ge 1/5$. This taken collectively with the upper bound (187) implies that: the minimax probability of error $p_{\mathrm{e},\boldsymbol{a}}$ in distinguishing $\mathcal{H}_0$ and $\mathcal{H}_{\boldsymbol{a}}$ is indeed lower bounded by $1/5$.

- Combining (186) with the fact $\delta\|\boldsymbol{a}_\perp\|_2 = \frac{\sigma_{\min}}{12|\lambda_l^\star|}$ gives

$$\|\widetilde{\boldsymbol{u}}_{\boldsymbol{a}} - \boldsymbol{u}_l^\star\|_2 \le \frac{\sigma_{\min}}{4|\lambda_l^\star|},$$

thus indicating that $\boldsymbol{M}_{\boldsymbol{a}} \in \mathcal{M}_1(\boldsymbol{M}^\star)$. Apply the standard reduction scheme [85, Chapter 2.2] to yield

$$\inf_{\widehat{u}_{\boldsymbol{a},l}} \sup_{\boldsymbol{A}\in\mathcal{M}_1(\boldsymbol{M}^\star)} \mathbb{E}\Big[\min\big|\widehat{u}_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top\boldsymbol{u}_l(\boldsymbol{A})\big|\Big] \gtrsim \min\big|\boldsymbol{a}^\top(\boldsymbol{u}_l^\star \pm \widetilde{\boldsymbol{u}}_{\boldsymbol{a}})\big|. \tag{188}$$

Everything then boils down to lower bounding $\min\big|\boldsymbol{a}^\top(\boldsymbol{u}_l^\star \pm \widetilde{\boldsymbol{u}}_{\boldsymbol{a}})\big|$. On the one hand, it is seen that

$$\begin{aligned}
&\big|\boldsymbol{a}^\top(\boldsymbol{u}_l^\star - \widetilde{\boldsymbol{u}}_{\boldsymbol{a}})\big| \\
&= \left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star - \frac{1}{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2}}(\boldsymbol{a}^\top\boldsymbol{u}_l^\star + \delta\boldsymbol{a}^\top\boldsymbol{a}_\perp)\right| \\
&\ge \frac{\delta\big|\boldsymbol{a}^\top\boldsymbol{a}_\perp\big|}{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2}} - \big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big|\left(1 - \frac{1}{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2}}\right) \\
&\ge \tfrac{1}{2}\delta\|\boldsymbol{a}_\perp\|_2^2 - \big|\boldsymbol{a}^\top\boldsymbol{u}_l^\star\big| \cdot \delta^2\|\boldsymbol{a}_\perp\|_2^2 \\
&\ge \tfrac{1}{4}\delta\|\boldsymbol{a}_\perp\|_2^2.
\end{aligned}$$

On the other hand, one can employ the assumption $\boldsymbol{a}^\top\boldsymbol{u}^\star \ge 0$ to derive

$$\begin{aligned}
&\big|\boldsymbol{a}^\top(\boldsymbol{u}_l^\star + \widetilde{\boldsymbol{u}}_{\boldsymbol{a}})\big| \\
&= \left|\boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{1}{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2}}(\boldsymbol{a}^\top\boldsymbol{u}_l^\star + \delta\boldsymbol{a}^\top\boldsymbol{a}_\perp)\right| \\
&\ge \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{1}{\sqrt{1+\delta^2\|\boldsymbol{a}_\perp\|_2^2}}\delta\|\boldsymbol{a}_\perp\|_2^2 \\
&\ge \tfrac{1}{2}\delta\|\boldsymbol{a}_\perp\|_2^2.
\end{aligned}$$

Putting all this together, we conclude that

$$\min\big|\boldsymbol{a}^\top(\boldsymbol{u}_l^\star \pm \widetilde{\boldsymbol{u}}_{\boldsymbol{a}})\big| \ge \tfrac{1}{4}\delta\|\boldsymbol{a}_\perp\|_2^2,$$

which taken collectively with (188) and our choice $\delta = \frac{\sigma_{\min}}{12|\lambda_l^\star|\cdot\|\boldsymbol{a}_\perp\|_2}$ yields

$$\begin{aligned}
&\inf_{\widehat{u}_{\boldsymbol{a},l}} \sup_{\boldsymbol{A}\in\mathcal{M}_1(\boldsymbol{M}^\star)} \mathbb{E}\Big[\min\big|\widehat{u}_{\boldsymbol{a},l} \pm \boldsymbol{a}^\top\boldsymbol{u}_l(\boldsymbol{A})\big|\Big] \\
&\gtrsim \min\big|\boldsymbol{a}^\top(\boldsymbol{u}_l^\star \pm \widetilde{\boldsymbol{u}}_{\boldsymbol{a}})\big| \gtrsim \frac{\sigma_{\min}}{|\lambda_l^\star|}\|\boldsymbol{a}_\perp\|_2.
\end{aligned}$$

C. Proof of Eq. (21d)

The proof of this part uses Fano's inequality. We start by constructing a proper packing set of the space: N unit vectors $\{\boldsymbol{v}_i\}_{1\le i\le N}$ within the subspace perpendicular to $\{\boldsymbol{u}_i^\star\}_{i\neq l}$ obeying

- $\langle \boldsymbol{v}_i, \boldsymbol{u}_j^\star\rangle = 0$ for all $1 \le i \le N$, $1 \le j \le r$ and $j \neq l$;
- there exists some sufficiently small constant $c_3 > 0$ such that

$$\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2 \ge c_3\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|}, \qquad 1 \le i \neq j \le N;$$

- there exists some sufficiently large constant $c_4 > 0$ such that

$$\|\boldsymbol{v}_i - \boldsymbol{u}_l^\star\|_2 \le c_4\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|}, \qquad 1 \le i \le N.$$

Standard packing arguments [1, Chapter 5.1] imply that N can be as large as

$$N = \exp\Big(n\log\frac{c_4}{2c_3}\Big). \tag{189}$$

In addition, when $4c_4\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|} < 1$, it follows that

$$\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2 \le \|\boldsymbol{v}_i - \boldsymbol{u}_l^\star\|_2 + \|\boldsymbol{v}_j - \boldsymbol{u}_l^\star\|_2 \le 2c_4\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|}, \tag{190}$$

$$\begin{aligned}
\|\boldsymbol{v}_i + \boldsymbol{v}_j\|_2 &\ge 2\|\boldsymbol{v}_i\|_2 - \|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2 \ge 2 - 2c_4\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|} \\
&> 2c_4\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|},
\end{aligned} \tag{191}$$

thus indicating that

$$\min\|\boldsymbol{v}_i \pm \boldsymbol{v}_j\|_2 \ge \min\{c_3, 2c_4\}\frac{\sigma_{\min}\sqrt{n}}{|\lambda_l^\star|}, \quad 1 \le i \neq j \le N. \tag{192}$$

The next step is to associate each vector $\boldsymbol{v}_i$ $(1 \le i \le N)$ with a hypothesis as follows

$$\mathcal{H}_i: \quad \boldsymbol{M} = \boldsymbol{M}_i + \boldsymbol{H} := \lambda_l^\star \boldsymbol{v}_i \boldsymbol{v}_i^\top + \sum_{j:j\neq l} \lambda_j^\star \boldsymbol{u}_j^\star \boldsymbol{u}_j^{\star\top} + \boldsymbol{H}.$$

If we denote by $\mathbb{P}_i$ the distribution of $\boldsymbol{M}$ under $\mathcal{H}_i$ $(1 \le i \le N)$, then we can invoke the argument in (177) and (179) to upper bound

$$\begin{aligned}\mathsf{KL}\left(\mathbb{P}_i \,\|\, \mathbb{P}_j\right) &\le \frac{\|\boldsymbol{M}_i - \boldsymbol{M}_j\|_{\mathrm{F}}^2}{2\sigma_{\min}^2} = \frac{(\lambda_l^\star)^2}{2\sigma_{\min}^2}\|\boldsymbol{v}_i\boldsymbol{v}_i^\top - \boldsymbol{v}_j\boldsymbol{v}_j^\top\|_{\mathrm{F}}^2 \\ &\le \frac{(\lambda_l^\star)^2}{\sigma_{\min}^2}\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2^2 \\ &\le 4c_4^2 n \end{aligned} \tag{193}$$

for any $i \neq j$, where the last inequality follows from (190). This upper bound on KL divergence plays an important role in invoking Fano's inequality. More specifically, recall that Fano's inequality [85, Corollary 2.6] asserts that if

$$\frac{1}{N}\sum_{i=2}^{N} \mathsf{KL}\left(\mathbb{P}_i \,\|\, \mathbb{P}_1\right) \le \frac{1}{4}\log N,$$

then the minimax probability of testing error must satisfy

$$p_{\mathrm{e},N} := \inf_{\psi} \max_{1\le i\le N} \mathbb{P}\left\{\psi \neq i \mid \mathcal{H}_i\right\} \ge 1/4,$$

where the infimum is over all tests. Combining this with the upper bound (193) and the packing number (189), we see that if

$$4c_4^2 n \le \frac{N}{4(N-1)}\log N = \frac{N}{4(N-1)} \cdot n \log\frac{c_4}{2c_3}, \tag{194}$$

then one would have $p_{\mathrm{e},N} \ge 1/4$. Clearly, the condition (194) would hold as long as c_4/c_3 is sufficiently large.

To finish up, it suffices to invoke the standard reduction scheme [85, Chapter 2.2] to obtain

$$\begin{aligned}&\inf_{\widehat{\boldsymbol{u}}_l} \sup_{\boldsymbol{A}\in\mathcal{M}_2} \mathbb{E}\Big[\min\|\widehat{\boldsymbol{u}}_l \pm \boldsymbol{u}_l(\boldsymbol{A})\|_2\Big] \\ &\gtrsim \min_{1\le i\neq j\le N}\left\{\min\|\boldsymbol{v}_i \pm \boldsymbol{v}_j\|_2\right\} \gtrsim \frac{\sigma_{\min}\sqrt{n}}{\left|\lambda_l^\star\right|},\end{aligned}$$

where the last inequality is a consequence of the condition (192). This concludes the proof.

APPENDIX F
A FEW MORE AUXILIARY LEMMAS

Lemma 12: Suppose that $\boldsymbol{H}$ satisfies Assumptions 1. Fix any $\boldsymbol{a} \in \mathbb{R}^n$. Then one has

$$\begin{aligned}\mathsf{Var}\left[\frac{1}{2\lambda_l^\star}\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right] &= \frac{1}{4(\lambda_l^\star)^2}\sum_{1\le i,j\le n}(a_i u_{l,j}^\star + a_j u_{l,i}^\star)^2\sigma_{ij}^2 \\ &=: v_{\boldsymbol{a},l}^\star,\end{aligned} \tag{195}$$

which satisfies

$$\frac{1}{2}\left(\|\boldsymbol{a}\|_2^2 + \left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right)^2\right)\frac{\sigma_{\min}^2}{(\lambda_l^\star)^2} \le v_{\boldsymbol{a},l}^\star \le \frac{1}{2}\left(\|\boldsymbol{a}\|_2^2 + \left(\boldsymbol{a}^\top\boldsymbol{u}_l^\star\right)^2\right)\frac{\sigma_{\max}^2}{(\lambda_l^\star)^2}. \tag{196}$$

Proof: See Appendix F-A. □

If we further have $\boldsymbol{a}^\top\boldsymbol{u}_l^\star = o(\|\boldsymbol{a}\|_2)$, then one has

$$\frac{1+o(1)}{2}\frac{\sigma_{\min}^2\|\boldsymbol{a}\|_2^2}{(\lambda_l^\star)^2} \le v_{\boldsymbol{a},l}^\star \le \frac{1+o(1)}{2}\frac{\sigma_{\max}^2\|\boldsymbol{a}\|_2^2}{(\lambda_l^\star)^2}. \tag{197}$$

Lemma 13: Suppose that $\boldsymbol{H}$ satisfies Assumptions 1 and obeys $B \le \sigma_{\max}\sqrt{\frac{n}{\mu\log n}}$. Fix any $\boldsymbol{a} \in \mathbb{R}^n$. Then with probability at least $1 - O(n^{-10})$, we have

$$\begin{aligned}\max\left\{\left|\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star\right|, \left|\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star\right|, \left|\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right|\right\} \\ \lesssim \sigma_{\max}\|\boldsymbol{a}\|_2\sqrt{\log n}\end{aligned} \tag{198}$$

for all $1 \le l \le r$.

Proof: See Appendix F-B. □

Lemma 14: Fix any unit vector $\boldsymbol{a} \in \mathbb{R}^n$, and consider another fixed unit vector $\boldsymbol{u}^\star$ obeying $\|\boldsymbol{u}^\star\|_\infty \le \sqrt{\mu/n}$. Suppose that $\boldsymbol{H}$ satisfies Assumption 1 and obeys $B \le \sigma_{\min}\sqrt{n/(\mu\log n)}$. Then the distribution of $W := \frac{\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}^\star}{2\sqrt{v^\star}}$ satisfies

$$\sup_{z\in\mathbb{R}}\left|\,\mathbb{P}(W\le z) - \Phi(z)\,\right| \le \frac{8}{\sqrt{\log n}}, \tag{199}$$

where $\Phi(\cdot)$ is the CDF of a standard Gaussian, and $v^\star := \mathsf{Var}\left(\frac{1}{2}\boldsymbol{a}^\top\left(\boldsymbol{H}+\boldsymbol{H}^\top\right)\boldsymbol{u}^\star\right)$.

Proof: See Appendix F-B. □

Lemma 15: Suppose that $\boldsymbol{H}$ satisfies Assumptions 1 and 2. Let $\boldsymbol{u}_l$ (resp. $\boldsymbol{w}_l$) be the lth right (resp. left) eigenvector of $\boldsymbol{M}$ obeying $\boldsymbol{u}_l^\top\boldsymbol{u}^\star > 0$ and $\boldsymbol{u}_l^\top\boldsymbol{w}_l > 0$, and set $\widehat{\boldsymbol{u}}_l := \frac{1}{\|\boldsymbol{u}_l+\boldsymbol{w}_l\|_2}(\boldsymbol{u}_l + \boldsymbol{w}_l)$. Then with probability exceeding $1 - O(n^{-10})$, one has

$$\begin{cases}\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l = 1 - O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \ge 49/50, \\ \boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l = 1 - O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \ge 49/50, \\ \widehat{\boldsymbol{u}}_l^\top\boldsymbol{u}_l^\star = 1 - O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \ge 49/50,\end{cases} \tag{200}$$

and

$$\begin{aligned}\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2 &= 2 + O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \\ &\ge 9/5.\end{aligned} \tag{201}$$

In addition, if Assumption 3 holds, then with probability at least $1 - O(n^{-10})$,

$$\begin{aligned}&\max\left\{\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_\infty, \|\boldsymbol{w}_l - \boldsymbol{u}_l^\star\|_\infty, \|\widehat{\boldsymbol{u}}_l - \boldsymbol{u}_l^\star\|_\infty\right\} \\ &\lesssim \frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^4 r\log n} + \frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\frac{\mu^2\kappa^4 r^3\log n}{n}} \\ &\lesssim o\left(\frac{1}{\sqrt{\mu n\log n}}\right).\end{aligned} \tag{202}$$

Proof: See Appendix F-C. □

A. Proof of Lemma 12

Without loss of generality, assume that $\lambda_l^\star = 1$ and that $\|\boldsymbol{a}\|_2 = 1$. The first claim (195) follows from direct calculations. Regarding the second claim (196), we first establish the lower bound

$$\begin{aligned} v_{\boldsymbol{a},l}^\star &\geq \frac{1}{4}\sum_{1\leq i,j\leq n}(a_i u_{l,j}^\star + a_j u_{l,i}^\star)^2\sigma_{\min}^2 \\ &= \frac{1}{4}\sum_{1\leq i,j\leq n}\Big(a_i^2(u_{l,j}^\star)^2 + a_j^2(u_{l,i}^\star)^2 + 2a_i u_{l,i}^\star a_j u_{l,j}^\star\Big)\sigma_{\min}^2 \\ &= \frac{1}{2}\left(1 + \left(\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right)^2\right)\sigma_{\min}^2, \end{aligned}$$

where we have used the fact that $\sum_{i,j} a_i^2(u_{l,j}^\star)^2 = \sum_i a_i^2 \sum_j (u_{l,j}^\star)^2 = 1$. A similar argument leads to the advertised upper bound $v_{\boldsymbol{a},l}^\star \leq \frac{1}{2}(1 + (\boldsymbol{a}^\top \boldsymbol{u}_l^\star)^2)\sigma_{\max}^2$.

B. Proofs of Lemma 13 and Lemma 14

Without loss of generality, we shall assume $\lambda_l^\star = 1$ throughout the proof. We shall also assume that $|H_{ij}| \leq B$ for all $1 \leq i,j \leq n$.[5] By direct calculations, the following quantity of interest can be expressed as

$$\frac{\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}^\star}{2} = \frac{1}{2}\sum_{i,j}(a_i u_j^\star + a_j u_{l,i}^\star)H_{ij} =: T_1. \quad (203)$$

The above quantity is the sum of n^2 independent zero-mean random variables, each obeying

$$\mathsf{Var}\left[(a_i u_{l,j}^\star + a_j u_{l,i}^\star)H_{ij}\right] = (a_i u_{l,j}^\star + a_j u_{l,i}^\star)^2\sigma_{ij}^2; \quad (204)$$

$$|(a_i u_{l,j}^\star + a_j u_{l,i}^\star)H_{ij}| \leq |a_i u_{l,j}^\star + a_j u_{l,i}^\star|\, B. \quad (205)$$

Proof of Lemma 13: We shall only establish the lemma for the quantity $\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star$; the bounds on $\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star$ and $\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star$ follow similarly. Invoking Bernstein's inequality for bounded random variables [84], we guarantee that with probability at least $1 - O(n^{-11})$,

$$\begin{aligned} &\Big|\sum_{i,j}(a_i u_{l,j}^\star + a_j u_{l,i}^\star)H_{ij}\Big| \\ &\lesssim \sqrt{\sum_{i,j}\mathsf{Var}\left[(a_i u_{l,j}^\star + a_i u_{l,j}^\star)H_{ij}\right]\log n} \\ &\quad + \max_{i,j}\left\{\left|(a_i u_{l,j}^\star + a_j u_{l,i}^\star)B\right|\right\}\log n \\ &\leq \sqrt{\sum_{i,j}(a_i u_{l,j}^\star + a_j u_{l,i}^\star)^2\sigma_{\max}^2\log n} \\ &\quad + \max_{i,j}\left\{\left|(a_i u_{l,j}^\star + a_j u_{l,i}^\star)B\right|\right\}\log n \\ &\lesssim \sqrt{\left(\sigma_{\max}^2\log n\right)\sum_{i,j}\left\{(a_i u_{l,j}^\star)^2 + (a_j u_{l,i}^\star)^2\right\}} \\ &\quad + B\|\boldsymbol{u}_l^\star\|_\infty\|\boldsymbol{a}\|_2\log n \\ &\asymp \sqrt{\left(\sigma_{\max}^2\log n\right)\|\boldsymbol{a}\|_2^2\|\boldsymbol{u}_l^\star\|_2^2} + B\|\boldsymbol{u}_l^\star\|_\infty\|\boldsymbol{a}\|_2\log n. \end{aligned}$$

[5]Otherwise, we can first look at the truncated version $\widetilde{H}_{ij} := H_{ij}\mathbb{1}_{\{|H_{ij}|\leq B\}}$ (which is also zero-mean with variance obeying $\mathsf{Var}[\widetilde{H}_{ij}] = (1+o(1))\sigma_{ij}^2$ under our assumptions), and then argue that $H_{ij} = \widetilde{H}_{ij}$ $(\forall i,j)$ with high probability. This argument is fairly standard and is hence omitted for brevity.

Now, as an immediate consequence of the incoherence condition and the identity $\|\boldsymbol{a}\|_2 = \|\boldsymbol{u}_l^\star\|_2 = 1$, we have

$$\Big|\sum_{i,j}(a_i u_{l,j}^\star + a_j u_{l,i}^\star)H_{ij}\Big| \lesssim \sigma_{\max}\sqrt{\log n} + B\log n\sqrt{\frac{\mu}{n}}.$$

Combining this with the expression (203) and the assumption $B \leq \sigma_{\max}\sqrt{\frac{n}{\mu\log n}}$, we complete the proof of the inequality (198) via the union bound.

Proof of Lemma 14: To establish this inequality, the key is to make use of the following Berry-Esseen type inequality [86, Theorem 3.7] .

Theorem 11: Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent zero-mean random variables satisfying $\sum_{i=1}^n \mathsf{Var}[\xi_i] = 1$. Then

$$\sup_{z\in\mathbb{R}}\Big|\mathbb{P}\Big(\sum_{i=1}^n \xi_i \leq z\Big) - \Phi(z)\Big| \leq 10\gamma, \text{ where } \gamma := \sum_{i=1}^n \mathbb{E}[|\xi_i|^3]. \quad (206)$$

In order to apply Theorem 11, let us define

$$\xi_{ij} := \frac{1}{\sqrt{\mathsf{Var}[T_1]}}\frac{1}{2}(a_i u_j^\star + a_i u_j^\star)H_{ij} \quad \text{and} \quad W = \sum_{i,j}\xi_{ij},$$

where it follows from the equality (204) that

$$v^\star := \mathsf{Var}[T_1] = \frac{1}{4}\sum_{i,j}(a_i u_j^\star + a_j u_i^\star)^2\sigma_{ij}^2. \quad (207)$$

By definition, it is easily seen that $\sum_{i,j}\mathsf{Var}[\xi_{ij}] = 1$. Therefore, the property (206) follows. In order to establish inequality (199), it suffices to upper bound γ in Theorem 11. To this end, we first make the observation that

$$\begin{aligned} \gamma &= \sum_{i,j}\mathbb{E}[|\xi_{ij}|^3] \\ &= \frac{1}{\mathsf{Var}[T_1]^{3/2}}\frac{1}{8}\sum_{i,j}\mathbb{E}\left[|(a_i u_j^\star + a_i u_j^\star)H_{ij}|^3\right] \\ &\leq \frac{1}{8(v^\star)^{3/2}} \\ &\quad \cdot \sum_{i,j}\mathbb{E}\left[\max_{i,j}|(a_i u_j^\star + a_i u_j^\star)H_{ij}| \cdot |(a_i u_j^\star + a_i u_j^\star)H_{ij}|^2\right] \\ &\leq \frac{1}{8(v^\star)^{3/2}}\sum_{i,j}\mathbb{E}\left[\|\boldsymbol{a}\|_2\|\boldsymbol{u}^\star\|_\infty B \cdot |(a_i u_j^\star + a_i u_j^\star)H_{ij}|^2\right] \\ &\leq \frac{1}{8(v^\star)^{3/2}}\|\boldsymbol{a}\|_2\|\boldsymbol{u}^\star\|_\infty B(4v^\star) = \frac{1}{2(v^\star)^{1/2}}\|\boldsymbol{a}\|_2\|\boldsymbol{u}^\star\|_\infty B. \end{aligned}$$

Re-organizing terms and using $\|\boldsymbol{a}\|_2 = \|\boldsymbol{u}^\star\|_2 = 1$ as well as the incoherence condition, we have

$$\gamma \leq B\sqrt{\frac{\mu}{4v^\star n}}. \quad (208)$$

It thus remains to lower bound $v^\star$. Again, use Lemma 12 and the fact $\|\boldsymbol{a}\|_2 = 1$ to arrive at 4 $v^\star \geq 2\sigma_{\min}^2$. Substitution into (208) allows us to further control γ as

$$\gamma \leq \frac{B}{\sigma_{\min}}\sqrt{\frac{\mu}{2n}} \leq \frac{1}{\sqrt{2\log n}},$$

where the last inequality arises since $B \leq \sigma_{\min}\sqrt{n/(\mu\log n)}$. Applying Theorem 11 concludes the proof.

C. Proof of Lemma 15

To begin with, Theorem 7 tells that that $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$ are both real-valued under Assumption 2. It has also been shown in Theorem 9 that

$$\begin{cases} \boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l = 1 - O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \geq 49/50 \\ \left|\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l\right| = 1 - O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \geq 49/50 \end{cases}$$

provided that $\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l > 0$ and that Assumption 2 holds. This immediately indicates that (i) $\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_2 = \sqrt{2-2\langle\boldsymbol{u}_l,\boldsymbol{u}_l^\star\rangle} \leq 1/5$ and $\|\boldsymbol{u}_l + \boldsymbol{u}_l^\star\|_2 = 2\|\boldsymbol{u}_l^\star\|_2 - \|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_2 \geq 9/5$, and (ii) one either has $\boldsymbol{w}_l^\top\boldsymbol{u}_l^\star \geq 49/50$ or $\boldsymbol{w}_l^\top\boldsymbol{u}_l^\star \leq -49/50$. If $\boldsymbol{w}_l^\top\boldsymbol{u}_l^\star \leq 49/50$, then one necessarily has

$$\begin{aligned} \boldsymbol{w}_l^\top\boldsymbol{u}_l &= \boldsymbol{w}_l^\top\boldsymbol{u}_l^\star + \boldsymbol{w}_l^\top(\boldsymbol{u}_l - \boldsymbol{u}_l^\star) \leq -49/50 + \|\boldsymbol{w}_l\|_2\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_2 \\ &\leq -49/50 + 1/5 < 0, \end{aligned}$$

thus contradicting the assumption that $\boldsymbol{w}_l^\top\boldsymbol{u}_l > 0$. As a result, one concludes that

$$\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l = 1 - O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \geq 49/50. \tag{209}$$

Further, it follows from Theorem 9 that

$$\begin{aligned} &\min\left\{\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_\infty, \|\boldsymbol{u}_l + \boldsymbol{u}_l^\star\|_\infty\right\} \\ &\lesssim \frac{\sigma_{\max}}{\lambda_{\min}^\star}\sqrt{\mu\kappa^4 r\log n} + \frac{\sigma_{\max}}{\Delta_l^\star}\sqrt{\frac{\mu^2\kappa^4 r^3\log n}{n}} \\ &\leq o\Big(\frac{1}{\sqrt{\mu n\log n}}\Big), \end{aligned}$$

where the last line holds under Assumption 3. Suppose that $\|\boldsymbol{u}_l + \boldsymbol{u}_l^\star\|_\infty \leq \|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_\infty$. Then the above inequality implies that

$$\begin{aligned} \|\boldsymbol{u}_l + \boldsymbol{u}_l^\star\|_2 &\leq \sqrt{n}\|\boldsymbol{u}_l + \boldsymbol{u}_l^\star\|_\infty \\ &= \sqrt{n}\min\left\{\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_\infty, \|\boldsymbol{u}_l + \boldsymbol{u}_l^\star\|_\infty\right\} \ll 1, \end{aligned}$$

which is contradictory to the relation $\|\boldsymbol{u}_l + \boldsymbol{u}_l^\star\|_2 \geq 9/5$ shown above. Therefore, we must have

$$\begin{aligned} \|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_\infty &= \min\left\{\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_\infty, \|\boldsymbol{u}_l + \boldsymbol{u}_l^\star\|_\infty\right\} \\ &\lesssim O\left(\frac{\sigma_{\max}\sqrt{\mu\log n}}{|\lambda^\star|}\right) = o\Big(\frac{1}{\sqrt{\mu n\log n}}\Big) \end{aligned} \tag{210}$$

as claimed. Similarly, we shall also have $\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_2 = \min\{\|\boldsymbol{u}_l \pm \boldsymbol{u}_l^\star\|_2\}$ under Assumption 2, which combined with Theorem 9 gives

$$\begin{aligned} &\max\left\{\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_2, \|\boldsymbol{w}_l - \boldsymbol{u}_l^\star\|_2\right\} \\ &\lesssim O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right). \end{aligned} \tag{211}$$

In addition, the above results further imply that

$$\begin{aligned} &\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2 - 2 \\ &= \frac{\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2^2 - 4}{\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2 + 2} \overset{\text{(i)}}{\asymp} \|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2^2 - 4 \\ &= \|\boldsymbol{u}_l\|_2^2 + \|\boldsymbol{w}_l\|_2^2 - 2 + 2\{\langle\boldsymbol{u}_l,\boldsymbol{w}_l\rangle - 1\} \\ &\overset{\text{(ii)}}{=} 2\{\langle\boldsymbol{u}_l,\boldsymbol{w}_l\rangle - \langle\boldsymbol{u}_l^\star,\boldsymbol{u}_l^\star\rangle\} \\ &= 2\{\langle\boldsymbol{u}_l - \boldsymbol{u}_l^\star,\boldsymbol{w}_l\rangle + \langle\boldsymbol{u}_l^\star,\boldsymbol{w}_l - \boldsymbol{u}_l^\star\rangle\} \\ &\overset{\text{(iii)}}{=} O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right), \end{aligned} \tag{212}$$

where (i) holds since $2 \leq \|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2 + 2 \leq 4$, (ii) follows since $\|\boldsymbol{u}_l\|_2 = \|\boldsymbol{w}_l\|_2 = 1$, and (iii) comes from Cauchy-Schwarz and the ℓ_2 bound (211).

To finish up, it remains to show that the claimed bounds hold when $\boldsymbol{u}_l$ is replaced by $\widehat{\boldsymbol{u}}_l$. Regarding the bound on $\widehat{\boldsymbol{u}}_l^\top\boldsymbol{u}_l^\star$, we make the observation that

$$\begin{aligned} &\widehat{\boldsymbol{u}}_l^\top\boldsymbol{u}_l^\star \\ &= \frac{1}{2}(\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star + \boldsymbol{w}_l^\top\boldsymbol{u}_l^\star) + \widehat{\boldsymbol{u}}_l^\top\boldsymbol{u}_l^\star - \frac{1}{2}(\boldsymbol{u}_l + \boldsymbol{w}_l)^\top\boldsymbol{u}_l^\star \\ &= \frac{1}{2}(\boldsymbol{u}_l^\top\boldsymbol{u}_l^\star + \boldsymbol{w}_l^\top\boldsymbol{u}_l^\star) + \left(\frac{1}{\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2} - \frac{1}{2}\right)(\boldsymbol{u}_l + \boldsymbol{w}_l)^\top\boldsymbol{u}_l^\star \\ &= 1 - O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right) \\ &\quad + \frac{2 - \|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2}{2\|\boldsymbol{u}_l + \boldsymbol{w}_l\|_2}\cdot(\boldsymbol{u}_l + \boldsymbol{w}_l)^\top\boldsymbol{u}_l^\star \\ &= 1 - O\left(\frac{\kappa^4\sigma_{\max}^2 n\log n}{(\lambda_{\max}^\star)^2} + \frac{\mu\kappa^4 r^2\sigma_{\max}^2\log n}{(\Delta_l^\star)^2}\right), \end{aligned}$$

where the last line comes from (212). The advertised ℓ_∞ norm bounds for $\widehat{\boldsymbol{u}}_l$ follow from the same arguments; we omit it for brevity.

APPENDIX G
EXAMPLE: A SYMMETRIC CASE WITH HOMOSCEDASTIC GAUSSIAN NOISE

In this section, we isolate a simple example to illustrate the potential applicability of our main results in the presence of certain symmetric noise matrices. Specifically, consider the symmetric and homoscedastic case with Gaussian noise, namely,

$$H_{ij} = H_{ji} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2), \qquad 1 \leq i \leq j \leq n; \tag{213a}$$

$$\boldsymbol{M}^\star = \sum_{l=1}^r \lambda_l^\star\boldsymbol{u}_l^\star\boldsymbol{u}_l^{\star\top}, \qquad \boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}. \tag{213b}$$

Clearly, this setting differs from the model in Assumption 1 in that both $\boldsymbol{H}$ and $\boldsymbol{M}$ are now symmetric matrices. To invoke our theorems, we need to first asymmetrize the data matrix. Towards this, we propose a procedure as follows, motivated by a simple asymmetrization trick pointed out by [17].

1) Generate $\widetilde{\boldsymbol{M}} = \boldsymbol{M} + \widetilde{\boldsymbol{H}}$, where $\widetilde{\boldsymbol{H}}$ is a skew-symmetric matrix obeying

$$\widetilde{H}_{ij} = -\widetilde{H}_{ji} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2), \qquad 1 \leq i < j \leq n. \tag{214}$$

2) Run Algorithm 1 with input data matrix $\widetilde{\boldsymbol{M}}$.

In short, the above scheme injects additional noise to the data matrix before implementing the inferential procedures. The theoretical support is this:

Corollary 3: Consider the model (213). Then Theorems 4, 5 and 6 continue to hold, if Algorithms 1 and 2 take $\widetilde{\boldsymbol{M}}$ as the input matrix.

Proof: The key observation is that: under our construction, the quantities $H_{ij}+\widetilde{H}_{ij}$ and $H_{ij}-\widetilde{H}_{ij}$ are uncorrelated and hence independent. Therefore, the effective noise matrix $\boldsymbol{H}_{\mathsf{eff}} := \boldsymbol{H}+\widetilde{\boldsymbol{H}}$ is in general asymmetric and contains i.i.d. entries drawn from $\mathcal{N}(0,2\sigma^2)$. Given that

$$\boldsymbol{H}_{\mathsf{eff}}+\boldsymbol{H}_{\mathsf{eff}}^\top = \boldsymbol{H}+\boldsymbol{H}^\top+\widetilde{\boldsymbol{H}}+\widetilde{\boldsymbol{H}}^\top = \boldsymbol{H}+\boldsymbol{H}^\top, \quad (215)$$

applying Theorems 4, 5 and 6 immediately concludes the proof. □

Interestingly, the above findings indicate that: a simple asymmetrization trick via proper noise injection allows one to construct confidence intervals under symmetric and homoscedastic Gaussian noise. We caution that the above procedure requires prior knowledge on the noise level σ, which can often be reliably estimated in the homoscedastic case.

While noise injection does not affect the first-order uncertainty term $\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star$, it does inflate the higher-order residual terms. For practical purposes, we recommend running the above procedure independently for multiple times and returning the "average" of them. This is summarized as follows, which might help improve practical performance.

1) Generate K matrices $\boldsymbol{M}^{(k)} = \boldsymbol{M}+\boldsymbol{H}^{(k)}$ $(1\le k\le K)$, where $\{\boldsymbol{H}^{(k)}\}$ are independent skew-symmetric matrices obeying

$$H_{ij}^{(k)} = -H_{ji}^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2), \qquad 1\le i<j\le n. \quad (216)$$

2) For each $1\le k\le K$, run Algorithm 1 with input data matrix $\boldsymbol{M}^{(k)}$. We denote by $\lambda_l^{(k)}$ the lth eigenvalue of $\boldsymbol{M}^{(k)}$, $\boldsymbol{u}_l^{(k)}$ the associated right eigenvector, $\widehat{u}_{\boldsymbol{a},l}^{(k)}$ the resulting estimator for $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$, respectively. Here, we calibrate the global signs of $\{\boldsymbol{u}_l^{(k)}\}$ by ensuring $\langle \boldsymbol{u}_l^{(k)}, \boldsymbol{u}_l^{(k+1)}\rangle \ge 0$ for all $1\le k<K$. We also let $\widehat{v}_{\boldsymbol{a},l}^{(k)}$ and $\widehat{v}_{\lambda,l}^{(k)}$ represent the resulting variance estimators.
3) Average the estimators

$$\widehat{u}_{\boldsymbol{a},l} = \frac{1}{K}\sum_{k=1}^K \widehat{u}_{\boldsymbol{a},l}^{(k)} \quad \text{and} \quad \lambda_l = \frac{1}{K}\sum_{k=1}^K \lambda_l^{(k)}. \quad (217)$$

4) For any prescribed coverage level $1-\alpha$, return the confidence intervals for $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ and $\lambda_l^\star$ as follows

$$\mathsf{CI}_{1-\alpha}^{\boldsymbol{a}} := \Big[\widehat{u}_{\boldsymbol{a},l} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{v}_{\boldsymbol{a},l}^{(1)}}\Big]; \quad (218a)$$

$$\mathsf{CI}_{1-\alpha}^{\lambda} := \Big[\lambda_l \pm \Phi^{-1}(1-\alpha/2)\sqrt{\widehat{v}_{\lambda,l}^{(1)}}\Big]. \quad (218b)$$

APPENDIX H
SYMMETRIZE OR NOT? SOME HIGH-LEVEL INTERPRETATION

The numerical experiments reported in Section V reveal the potential benefit of Spectral-asym compared to Spectral-sym. Here, we provide some informal interpretation.

If we hope the eigenvector $\boldsymbol{u}_2$ to behave as a reliable estimate of $\boldsymbol{u}_2^\star$ (meaning that $|\boldsymbol{u}_2^\top\boldsymbol{u}_2^\star| = 1-o(1)$), then we would necessarily require $|\boldsymbol{u}_2^\top\boldsymbol{u}_1^\star| = o(1)$. To develop an understanding about the size of $\boldsymbol{u}_2^\top\boldsymbol{u}_1^\star$, we recall from Neumann's series that (see [17] or Appendix A)

$$\boldsymbol{u}_1^{\star\top}\boldsymbol{u}_2 = \frac{\lambda_1^\star\left(\boldsymbol{u}_1^{\star\top}\boldsymbol{u}_2\right)}{\lambda_2} + \frac{\lambda_2^\star\left(\boldsymbol{u}_2^{\star\top}\boldsymbol{u}_2\right)}{\lambda_2}\sum_{s=1}^\infty \frac{\boldsymbol{u}_1^{\star\top}\boldsymbol{H}^s\boldsymbol{u}_2^\star}{\lambda_2^s} + \frac{\lambda_1^\star\left(\boldsymbol{u}_1^{\star\top}\boldsymbol{u}_2\right)}{\lambda_2}\sum_{s=1}^\infty \frac{\boldsymbol{u}_1^{\star\top}\boldsymbol{H}^s\boldsymbol{u}_1^\star}{\lambda_2^s}, \quad (219)$$

or equivalently,

$$\boldsymbol{u}_1^{\star\top}\boldsymbol{u}_2 = \frac{\lambda_2^\star\left(\boldsymbol{u}_2^{\star\top}\boldsymbol{u}_2\right)}{\lambda_2-\lambda_1^\star-\lambda_1^\star\sum_{s=1}^\infty \frac{\boldsymbol{u}_1^{\star\top}\boldsymbol{H}^s\boldsymbol{u}_1^\star}{\lambda_2^s}}\sum_{s=1}^\infty \frac{\boldsymbol{u}_1^{\star\top}\boldsymbol{H}^s\boldsymbol{u}_2^\star}{\lambda_2^s}. \quad (220)$$

Consequently, an important condition that allows one to control quantity $\boldsymbol{u}_1^{\star\top}\boldsymbol{u}_2$ is to ensure that each summand in expression (220) is as small as possible.

Towards this end, let us single out the second-order term $\boldsymbol{u}_1^{\star\top}\boldsymbol{H}^2\boldsymbol{u}_2^\star$, which suffices for us to develop some intuition (here, we assume that $\lambda_2\approx\lambda_2^\star$ so that the denominator λ_2^2 does not affect the order of this term).

- If $\boldsymbol{H}$ is asymmetric, then it has been shown in [17] (see also Appendix A) that

$$\left|\boldsymbol{u}_1^{\star\top}\boldsymbol{H}^2\boldsymbol{u}_2^\star\right| \lesssim \sigma_{\max}^2\sqrt{\mu n}\log n \quad (221)$$

  with high probability, which is exceedingly small given the assumption that $\sigma_{\max}\sqrt{n\log n}\ll 1$.
- If $\boldsymbol{H}$ is replaced by $\widetilde{\boldsymbol{H}} := \frac{1}{2}(\boldsymbol{H}+\boldsymbol{H}^\top)$, then in general there is no guarantee that $\boldsymbol{u}_1^{\star\top}\widetilde{\boldsymbol{H}}^2\boldsymbol{u}_2^\star$ can be equally well-controlled. Take the case (50) and (51) for example: straightforward calculations reveal

$$\mathbb{E}\Big[\boldsymbol{u}_1^{\star\top}\widetilde{\boldsymbol{H}}^2\boldsymbol{u}_2^\star\Big] = \boldsymbol{u}_1^{\star\top}\mathsf{Var}\big(\tfrac{1}{2}(\boldsymbol{H}+\boldsymbol{H}^\top)\big)\boldsymbol{u}_2^\star = \tfrac{1}{4}\sigma_1^2 n \asymp \sigma_{\max}^2 n, \quad (222)$$

  which far exceeds the order (221) with asymmetric data matrices. If this is the case, then one would have to require a better control of the multiplicative term $\frac{\lambda_2^\star}{\lambda_2-\lambda_1^\star-\lambda_1^\star\sum_{s=1}^\infty \frac{\boldsymbol{u}_1^{\star\top}\boldsymbol{H}^s\boldsymbol{u}_1^\star}{\lambda_2^s}}$ in (220), which is often equivalent to imposing a more stringent separation condition on the eigenvalue pair $(\lambda_1^\star,\lambda_2^\star)$.

The take-home message is this: Spectral-sym might sometimes be suboptimal when dealing with heteroscedastic noise, particularly when the variance structure of the noise matrix is somewhat "aligned" with the true eigen-structure. As a result, Spectral-sym might not be as robust as Spectral-asym when performing eigenvector estimation.

ACKNOWLEDGMENT

Chen Cheng would like to thank Lihua Lei for helpful discussions.

REFERENCES

[1] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, vol. 48. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[2] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[3] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[4] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6576–6601, Oct. 2014.

[5] A. Javanmard and A. Montanari, "Localization from incomplete noisy distance measurements," *Found. Comput. Math.*, vol. 13, no. 3, pp. 297–345, 2013.

[6] Y. Chen, L. Guibas, and Q. Huang, "Near-optimal joint object matching via convex relaxation," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 100–108.

[7] S. Péché, "The largest eigenvalue of small rank perturbations of Hermitian random matrices," *Probab. Theory Rel. Fields*, vol. 134, no. 1, pp. 127–173, 2006.

[8] T. T. Cai and A. Zhang, "Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics," *Ann. Statist.*, vol. 46, no. 1, pp. 60–89, 2018.

[9] E. Abbe, J. Fan, K. Wang, and Y. Zhong, "Entrywise eigenvector analysis of random matrices with low expected rank," *Ann. Statist.*, vol. 48, no. 3, p. 1452, Jun. 2020.

[10] S. O'Rourke, V. Vu, and K. Wang, "Random perturbation of low rank matrices: Improving classical bounds," *Linear Algebra Appl.*, vol. 540, pp. 26–59, Mar. 2018.

[11] V. Vu, "Singular vectors under random perturbation," *Random Struct. Algorithms*, vol. 39, no. 4, pp. 526–538, Dec. 2011.

[12] Y. Yu, T. Wang, and R. J. Samworth, "A useful variant of the Davis–Kahan theorem for statisticians," *Biometrika*, vol. 102, no. 2, pp. 315–323, 2015.

[13] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," *J. Mach. Learn. Res.*, vol. 11, pp. 2057–2078, Mar. 2010.

[14] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5239–5269, Oct. 2019.

[15] R. Wang, "Singular vector perturbation under Gaussian noise," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 1, pp. 158–177, Jan. 2015.

[16] Y. Zhong, "Eigenvector under random perturbation: A nonasymptotic Rayleigh-Schrödinger theory," 2017, *arXiv:1702.00139*. [Online]. Available: http://arxiv.org/abs/1702.00139

[17] Y. Chen, C. Cheng, and J. Fan, "Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices," *Ann. Statist.*, vol. 49, no. 1, pp. 435–458, Feb. 2021.

[18] V. Koltchinskii, M. Löffler, and R. Nickl, "Efficient estimation of linear functionals of principal components," *Ann. Statist.*, vol. 48, no. 1, pp. 464–490, 2020.

[19] J. Cho, D. Kim, and K. Rohe, "Asymptotic theory for estimating the singular vectors and values of a partially-observed low rank matrix with noise," *Statistica Sinica*, vol. 27, pp. 1921–1948, Oct. 2017.

[20] C. Cai, G. Li, Y. Chi, H. V. Poor, and Y. Chen, "Subspace estimation from unbalanced and incomplete data matrices: ℓ_2, ∞ statistical guarantees," *Ann. Statist.*, vol. 49, no. 2, pp. 944–967, Apr. 2021.

[21] A. Zhang, T. T. Cai, and Y. Wu, "Heteroskedastic PCA: Algorithm, optimality, and applications," *Ann. Statist.*, 2021.

[22] V. Koltchinskii and K. Lounici, "Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance," *Annales De l'Institut Henri Poincaré, Probabilités Et Statistiques*, vol. 52, no. 4, pp. 1976–2013, 2016.

[23] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Spectral methods for data science: A statistical perspective," 2020, *arXiv:2012.08496*. [Online]. Available: http://arxiv.org/abs/2012.08496

[24] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation. III," *SIAM J. Numer. Anal.*, vol. 7, no. 1, pp. 1–46, 1970.

[25] P.-Å. Wedin, "Perturbation bounds in connection with singular value decomposition," *BIT Numer. Math.*, vol. 12, no. 1, pp. 99–111, 1972.

[26] E. Abbe, "Community detection and stochastic block models: Recent developments," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6446–6531, Jan. 2017.

[27] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, Feb. 2014.

[28] G. Li, C. Cai, Y. Gu, H. V. Poor, and Y. Chen, "Minimax estimation of linear functions of eigenvectors in the face of small eigen-gaps," 2021, *arXiv:2104.03298*. [Online]. Available: http://arxiv.org/abs/2104.03298

[29] T. T. Cai and M. G. Low, "An adaptation theory for nonparametric confidence intervals," *Ann. Statist.*, vol. 32, no. 5, pp. 1805–1840, 2004.

[30] T. T. Cai and M. Low, "A framework for estimation of convex functions," *Statistica Sinica*, vol. 25, pp. 423–456, Apr. 2015.

[31] T. T. Cai, A. Guntuboyina, and Y. Wei, "Adaptive estimation of planar convex sets," *Ann. Statist.*, vol. 46, no. 3, pp. 1018–1049, Jun. 2018.

[32] C. Stein, "Efficient nonparametric testing and estimation," in *Proc. 3rd Berkeley Symp. Math. Statist. Probab., Contrib. Statist.*, vol. 1. Oakland, CA, USA: Regents of the Univ. of California, 1956, pp. 187–196.

[33] C. Ma, K. Wang, Y. Chi, and Y. Chen, "Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution," *Found. Comput. Math.*, vol. 20, no. 3, pp. 451–632, Jun. 2020.

[34] S. Negahban and M. J. Wainwright, "Restricted strong convexity and weighted matrix completion: Optimal bounds with noise," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 1665–1697, 2012.

[35] V. Koltchinskii, K. Lounici, and A. Tsybakov, "Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion," *Ann. Statist.*, vol. 39, no. 5, pp. 2302–2329, 2011, doi: 10.1214/11-AOS894.

[36] A. Montanari, F. Ruan, and J. Yan, "Adapting to unknown noise distribution in matrix denoising," 2018, *arXiv:1810.02954*. [Online]. Available: http://arxiv.org/abs/1810.02954

[37] D. Donoho and M. Gavish, "Minimax risk of matrix denoising by singular value thresholding," *Ann. Statist.*, vol. 42, no. 6, pp. 2413–2440, 2014.

[38] Y. Chen and M. J. Wainwright, "Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees," 2015, *arXiv:1509.03025*. [Online]. Available: http://arxiv.org/abs/1509.03025

[39] Y. Chen and E. J. Candès, "The projected power method: An efficient algorithm for joint alignment from pairwise differences," *Commun. Pure Appl. Math.*, vol. 71, no. 8, pp. 1648–1714, Aug. 2018.

[40] D. Donoho, M. Gavish, and I. Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model," *Ann. Statist.*, vol. 46, no. 4, p. 1742, Aug. 2018.

[41] L. T. Liu, E. Dobriban, and A. Singer, "ePCA: High dimensional exponential family PCA," *Ann. Appl. Statist.*, vol. 12, no. 4, pp. 2121–2150, 2018.

[42] I. M. Johnstone and D. Paul, "PCA in high dimensions: An orientation," *Proc. IEEE*, vol. 106, no. 8, pp. 1277–1292, Aug. 2018.

[43] Y. Chen, J. Fan, C. Ma, and Y. Yan, "Bridging convex and nonconvex optimization in robust PCA: Noise, outliers, and missing data," *Ann. Statist.*, 2020.

[44] Y. Yan, Y. Chen, and J. Fan, "Inference for heteroskedastic PCA with missing data," 2021, *arXiv:2107.12365*. [Online]. Available: http://arxiv.org/abs/2107.12365

[45] J. Fan, W. Wang, and Y. Zhong, "An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation," *J. Mach. Learn. Res.*, vol. 18, no. 207, pp. 1–42, 2018.

[46] Y. Chen, J. Fan, C. Ma, and K. Wang, "Spectral method and regularized MLE are both optimal for top-K ranking," *Ann. Statist.*, vol. 47, no. 4, pp. 2204–2235, Aug. 2019.

[47] J. Cape, M. Tang, and C. E. Priebe, "The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics," *Ann. Statist.*, vol. 47, no. 5, pp. 2405–2439, Oct. 2019.

[48] J. Eldridge, M. Belkin, and Y. Wang, "Unperturbed: Spectral analysis beyond Davis-Kahan," in *Proc. Algorithmic Learn. Theory*, 2018, pp. 321–358.

[49] Y. Chen, Y. Chi, J. Fan, C. Ma, and Y. Yan, "Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization," *SIAM J. Optim.*, vol. 30, no. 4, pp. 3098–3121, Jan. 2020.

[50] A. Pananjady and M. J. Wainwright, "Value function estimation in Markov reward processes: Instance-dependent ℓ_∞-bounds for policy evaluation," 2019, *arXiv:1909.08749*. [Online]. Available: https://arxiv.org/abs/1909.08749

[51] J. Chen, D. Liu, and X. Li, "Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5806–5841, Sep. 2020.

[52] L. Lei, "Unified $\ell_{2\to\infty}$ eigenspace perturbation theory for symmetric random matrices," 2019, *arXiv:1909.04798*. [Online]. Available: https://arxiv.org/abs/1909.04798
[53] V. Koltchinskii and D. Xia, "Perturbation of linear forms of singular vectors under Gaussian noise," in *High Dimensional Probability VII*. Birkhäuser, Cham: Springer, 2016, pp. 397–423.
[54] C.-H. Zhang and S. S. Zhang, "Confidence intervals for low dimensional parameters in high dimensional linear models," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 76, no. 1, pp. 217–242, 2014.
[55] S. van de Geer, P. Bühlmann, Y. Ritov, and R. Dezeure, "On asymptotically optimal confidence regions and tests for high-dimensional models," *Ann. Statist.*, vol. 42, no. 3, pp. 1166–1202, 2014.
[56] A. Javanmard and A. Montanari, "Confidence intervals and hypothesis testing for high-dimensional regression," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 2869–2909, Jan. 2014.
[57] T. T. Cai and Z. Guo, "Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity," *Ann. Statist.*, vol. 45, no. 2, pp. 615–646, 2017.
[58] A. Belloni, V. Chernozhukov, and C. Hansen, "Inference for high-dimensional sparse econometric models," 2011, *arXiv:1201.0220*. [Online]. Available: http://arxiv.org/abs/1201.0220
[59] Z. Ren, T. Sun, C.-H. Zhang, and H. H. Zhou, "Asymptotic normality and optimalities in estimation of large Gaussian graphical models," *Ann. Statist.*, vol. 43, no. 3, pp. 991–1026, Jun. 2015.
[60] Y. Ning and H. Liu, "A general theory of hypothesis tests and confidence regions for sparse high dimensional models," *Ann. Statist.*, vol. 45, no. 1, pp. 158–195, 2017.
[61] C. Ma, J. Lu, and H. Liu, "Inter-subject analysis: A partial Gaussian graphical model approach," *J. Amer. Stat. Assoc.*, vol. 116, pp. 1–57, Apr. 2020.
[62] J. Janková and S. van de Geer, "De-biased sparse PCA: Inference and testing for eigenstructure of large covariance matrices," 2018, *arXiv:1801.10567*. [Online]. Available: http://arxiv.org/abs/1801.10567
[63] L. Miolane and A. Montanari, "The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning," *Ann. Statist.*, 2020.
[64] M. Celentano, A. Montanari, and Y. Wei, "The lasso with general Gaussian designs with applications to hypothesis testing," 2020, *arXiv:2007.13716*. [Online]. Available: http://arxiv.org/abs/2007.13716
[65] A. Carpentier, J. Eisert, D. Gross, and R. Nickl, "Uncertainty quantification for matrix compressed sensing and quantum tomography problems," in *High Dimensional Probability VIII*. Birkhäuser, Cham: Springer, 2019, pp. 385–430.
[66] A. Carpentier and R. Nickl, "On signal detection and confidence sets for low rank inference problems," *Electron. J. Statist.*, vol. 9, no. 2, pp. 2675–2688, Jan. 2015.
[67] A. Carpentier, O. Klopp, and M. Löffler, "Constructing confidence sets for the matrix completion problem," in *Proc. Conf. Int. Soc. Non-Parametric Statist.* Cham, Switzerland: Springer, 2016, pp. 103–118.
[68] Y. Chen, J. Fan, C. Ma, and Y. Yan, "Inference and uncertainty quantification for noisy matrix completion," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 46, pp. 22931–22937, Nov. 2019.
[69] D. Xia and M. Yuan, "Statistical inferences of linear forms for noisy matrix completion," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 83, no. 1, pp. 58–77, Feb. 2021.
[70] D. Xia, "Confidence region of singular subspaces for low-rank matrix regression," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7437–7459, Nov. 2019.
[71] D. Xia, "Normal approximation and confidence region of singular subspaces," 2019, *arXiv:1901.00304*. [Online]. Available: http://arxiv.org/abs/1901.00304
[72] J. Fan, Y. Fan, X. Han, and J. Lv, "Asymptotic theory of eigenvectors for random matrices with diverging spikes," 2019, *arXiv:1902.06846*. [Online]. Available: http://arxiv.org/abs/1902.06846
[73] H. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, "Spectrum of large random asymmetric matrices," *Phys. Rev. Lett.*, vol. 60, no. 19, p. 1895, 1988.
[74] B. Khoruzhenko, "Large-N eigenvalue distribution of randomly perturbed asymmetric matrices," *J. Phys. A, Math. Gen.*, vol. 29, no. 7, p. L165, 1996.
[75] E. Brézin and A. Zee, "Non-Hermitean delocalization: Multiple scattering and bounds," *Nucl. Phys. B*, vol. 509, no. 3, pp. 599–614, Jan. 1998.
[76] J. T. Chalker and B. Mehlig, "Eigenvector statistics in non-Hermitian random matrix ensembles," *Phys. Rev. Lett.*, vol. 81, no. 16, p. 3367, 1998.
[77] J. Feinberg and A. Zee, "Non-Hermitian random matrix theory: Method of Hermitian reduction," *Nucl. Phys. B*, vol. 504, no. 3, pp. 579–608, Nov. 1997.
[78] A. Lytova and K. Tikhomirov, "On delocalization of eigenvectors of random non-Hermitian matrices," 2018, *arXiv:1810.01590*. [Online]. Available: http://arxiv.org/abs/1810.01590
[79] T. Tao, "Outliers in the spectrum of iid matrices with bounded rank perturbations," *Probab. Theory Rel. Fields*, vol. 155, nos. 1–2, pp. 231–263, 2013.
[80] A. B. Rajagopalan, "Outlier eigenvalue fluctuations of perturbed iid matrices," 2015, *arXiv:1507.01441*. [Online]. Available: http://arxiv.org/abs/1507.01441
[81] F. Benaych-Georges and J. Rochet, "Outliers in the single ring theorem," *Probab. Theory Rel. Fields*, vol. 165, pp. 313–363, Jun. 2016.
[82] J. B. Conway, *Functions of One Complex Variable II*, vol. 159. New York, NY, USA: Springer, 2012.
[83] M. Embree and L. N. Trefethen, "Generalizing eigenvalue theorems to pseudospectra theorems," *SIAM J. Sci. Comput.*, vol. 23, no. 2, pp. 583–590, Jan. 2001.
[84] J. A. Tropp, "An introduction to matrix concentration inequalities," *Found. Trends Mach. Learn.*, vol. 8, nos. 1–2, pp. 1–230, May 2015.
[85] A. B. Tsybakov, *Introduction to Nonparametric Estimation*. New York, NY, USA: Springer, 2008.
[86] L. H. Chen, L. Goldstein, and Q.-M. Shao, *Normal Approximation by Stein's Method*. Heidelberg, Germany: Springer, 2010.

Chen Cheng received the bachelor's degree from the School of Mathematical Sciences, Peking University. He is currently pursuing the Ph.D. degree with the Department of Statistics, Stanford University, jointly advised by Prof. John Duchi and Prof. Andrea Montanari. His research interests include statistical theory and algorithms for high-dimensional data, random matrix theory, machine learning theory, and reinforcement learning.

Yuting Wei received the Ph.D. degree in statistics from the University of California, Berkeley (UC Berkeley), advised by Martin Wainwright and Aditya Guntuboyina. She is currently an Assistant Professor with the Department of Statistics and Data Science, The Wharton School, University of Pennsylvania. Prior to that, she spent two years at Carnegie Mellon University as an Assistant Professor of statistics, and one year at Stanford University as a Stein Fellow. Her research interests include high-dimensional and non-parametric statistics, statistical machine learning, and reinforcement learning. She was a recipient of the 2018 Erich L. Lehmann Citation from the Department of Statistics, UC Berkeley, for her Ph.D. dissertation in theoretical statistics.

Yuxin Chen (Member, IEEE) received the B.E. degree (Hons.) from Tsinghua University in 2008, the M.S. degree in electrical and computer engineering from The University of Texas at Austin in 2010, and the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University in January 2015. He is currently an Assistant Professor with the Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. Prior to joining Princeton, he was a Post-Doctoral Scholar with the Department of Statistics, Stanford University. His research interests include high-dimensional data analysis, convex and non-convex optimization, reinforcement learning, statistical learning, statistical signal processing, and information theory. He was a recipient of the 2020 Princeton Graduate Mentoring Award, the 2019 AFOSR Young Investigator Award, the 2020 ARO Young Investigator Award, the ICCM Best Paper Award (Gold Medal), and the 2021 Princeton SEAS Junior Faculty Award. He was selected as a Finalist for the Best Paper Prize for Young Researchers in Continuous Optimization 2019.