ANALYSIS OF SINGULAR SUBSPACES UNDER RANDOM PERTURBATIONS

KE WANG

ABSTRACT. We present a comprehensive analysis of singular vector and singular subspace perturbations in the context of the signal plus random Gaussian noise matrix model. Assuming a low-rank signal matrix, we extend the Davis-Kahan-Wedin theorem in a fully generalized manner, applicable to any unitarily invariant matrix norm, extending previous results of O'Rourke, Vu and the author. We also obtain the fine-grained results, which encompass the ℓ_∞ analysis of singular vectors, the $\ell_{2,\infty}$ analysis of singular subspaces, as well as the exploration of linear and bilinear functions related to the singular vectors. Moreover, we explore the practical implications of these findings, in the context of the Gaussian mixture model and the submatrix localization problem.

1. INTRODUCTION

Matrix perturbation theory has emerged as a central and foundational subject within various disciplines, including probability, statistics, machine learning, and applied mathematics. Perturbation bounds, which quantify the influence of small noise on the spectral parameters of a matrix, are of paramount importance in numerous applications such as matrix completion [30, 31, 50], principal component analysis (PCA) [49], and community detection [71,73], to mention a few. This paper aims to present a comprehensive analysis establishing perturbation bounds for the singular vectors and singular subspaces of a low-rank signal matrix perturbed by additive random Gaussian noise.

Consider an unknown $N \times n$ data matrix A. Suppose we cannot observe A directly but instead have access to a corrupted version $\widetilde{A}$ given by

$$\widetilde{A} := A + E, \tag{1}$$

where E represents the noise matrix. In this paper, we focus on real matrices, and the extension to complex matrices is straightforward.

Assume that the $N \times n$ data matrix A has rank $r \geqslant 1$. The singular value decomposition (SVD) of A takes the form $A = UDV^T$, where $D = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$ is a diagonal matrix containing the non-zero singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ of A; the columns of the matrices $U = (u_1, \dots, u_r)$ and $V = (v_1, \dots, v_r)$ are the orthonormal left and right singular vectors of A, respectively. In other words, u_i and v_i are the left and right singular vectors corresponding to σ_i. It follows that $U^TU = V^TV = I_r$, where I_r is the $r \times r$ identity matrix. For convenience we will take $\sigma_{r+i} = 0$ for all $i \geqslant 1$. Denote the SVD of $\widetilde{A}$ given in (1) similarly by $\widetilde{A} = \widetilde{U}\widetilde{D}\widetilde{V}^{\mathrm{T}}$, where the diagonal entries of $\widetilde{D}$ are the singular values $\widetilde{\sigma}_1 \geqslant \widetilde{\sigma}_2 \geqslant$

K. Wang is partially supported by Hong Kong RGC grant GRF 16304222, GRF 16308219 and ECS 26304920.

$\cdots \geqslant \widetilde{\sigma}_{\min\{N,n\}} \geqslant 0$, and the columns of $\widetilde{U}$ and $\widetilde{V}$ are the orthonormal left and right singular vectors, denoted by $\widetilde{u}_i$ and $\widetilde{v}_i$, respectively.

The primary focus of this paper is the singular subspaces that are spanned by the leading singular vectors. For $1 \leqslant k \leqslant r$, let us denote

$$U_k := \mathrm{Span}\{u_1, \ldots, u_k\}, \quad V_k := \mathrm{Span}\{v_1, \ldots, v_k\},$$
$$\widetilde{U}_k := \mathrm{Span}\{\widetilde{u}_1, \ldots, \widetilde{u}_k\}, \quad \widetilde{V}_k := \mathrm{Span}\{\widetilde{v}_1, \ldots, \widetilde{v}_k\}.$$

With a slight abuse of notation, we also use $U_k = (u_1, \ldots, u_k)$ to represent the singular vector matrix. We employ the notation $V_k, \widetilde{U}_k, \widetilde{V}_k$ in a similar manner. Let $P_{U_k} = U_k U_k^{\mathrm{T}}$ (resp. $P_{V_k} = V_k V_k^{\mathrm{T}}$) be the orthogonal projection on the subspace U_k (resp. V_k). Denote the orthogonal complement of a subspace W as $W^{\perp}$.

The classical perturbation bounds related to the changes in singular values and singular vectors are detailed below. The matrix norm $\|\cdot\|$ on $\mathbb{R}^{N\times n}$ is said to be unitarily invariant if $\|A\| = \|UAV\|$ for all orthogonal matrices $U \in \mathbb{R}^{N\times N}$ and $V \in \mathbb{R}^{n\times n}$. In addition, we always consider the norm $\|\cdot\|$ to be *normalized*. This means that the norm always satisfies $\|A\| = 1$ if A has its $(1,1)$ entry equal to 1 and all other entries equal to zero. A more thorough exploration of the properties of unitarily invariant matrix norms can be found in Section 5.1.

Denote $\mathrm{diag}(\sigma_i - \widetilde{\sigma}_i) = \mathrm{diag}(\sigma_1 - \widetilde{\sigma}_2, \cdots, \sigma_{\min\{N,n\}} - \widetilde{\sigma}_{\min\{N,n\}})$. This represents the difference in singular values between A and $A+E$. The perturbations or changes in the singular values of A and $A+E$ are provided by Mirsky's theorem (see Theorem 4.11 in Chapter IV from [68]).

Theorem 1 (Mirsky)**.** *For any unitarily invariant norm* $\|\cdot\|$*,*

$$\|\mathrm{diag}(\sigma_i - \widetilde{\sigma}_i)\| \leqslant \|E\|.$$

When applied to the operator norm and eigenvalues of Hermitian matrices, the inequality stated can be recognized as the Weyl's inequality (see [21, Corollary III.2.6]).

The differences between subspaces U_k and $\widetilde{U}_k$ of A and $A+E$ can be quantified by calculating the separation between U_k and $\widetilde{U}_k$. This is achieved using k principal angles, defined as $0 \leqslant \theta_1 \leqslant \cdots \leqslant \theta_k \leqslant \pi/2$. These angles measure the distance between the two subspaces. For a detailed definition and further discussion of this concept, please refer to Section 5.2. Denote

$$\sin\angle(U_k, \widetilde{U}_k) := \mathrm{diag}(\sin\theta_1, \cdots, \sin\theta_k).$$

Define $\sin\angle(V_k, \widetilde{V}_k)$ analogously. The classical perturbation bound, which concerns the variations in the eigenspaces for symmetric matrices A and $A+E$, was initially investigated by Davis and Kahan [40]. Further generalizations to singular subspaces of rectangular matrices are encapsulated in Wedin's theorem (Eq. (3.11) from [76]).

Theorem 2 (Wedin [76])**.** *If* $\hat{\delta}_k := \sigma_k - \widetilde{\sigma}_{k+1} > 0$*, then for any unitarily invariant norm* $\|\cdot\|$*,*

$$\|\sin\angle(U_k, \widetilde{U}_k)\| \leqslant \frac{\max\{\|P_{U_k^{\perp}} E P_{\widetilde{V}_k}\|, \|P_{V_k^{\perp}} E^{\mathrm{T}} P_{\widetilde{U}_k}\|\}}{\hat{\delta}_k}. \tag{2}$$

The same result also holds for $\|\sin\angle(V_k, \widetilde{V}_k)\|$*.*

In the context of an unitarily invariant norm $\|\|\cdot\|\|$, there exist several well-established methods to quantify the separation between U_k and $\widetilde{U}_k$. These include using

$$\|\| \sin\angle(U_k, \widetilde{U}_k)\|\|,\ \|\|P_{U_k} - P_{\widetilde{U}_k}\|\| \text{ and } \min_{O\in\mathbb{O}^{k\times k}} \|\|U_k O - \widetilde{U}_k\|\|.$$

In Section 5.2, we provide a detailed discussion about the equivalence or relationships among these various methods.

The traditional bounds previously mentioned offer precise estimates, catering to worst-case scenarios. However, modern applications often operate under the premise that the data matrix A satisfies specific structural assumptions. A typical case is when A has a low rank r, where r remains constant or experiences slow growth relative to N and n. Moreover, the noise matrix E is generally assumed to be random.

In light of these additional assumptions about the data A and the noise E, we foresee significant enhancements over the traditional results. Our principal objective is to derive a stochastic variant of Wedin's theorem, under the assumption that A is low-rank and E is random. This paper builds upon the previous works [63, 64] by O'Rourke, Vu and the author, which initially stemmed from Vu's work [72]. Specifically, within the framework established in [64], we present a comprehensive extension of the Davis-Kahan-Wedin $\sin\Theta$ theorem. This extension applies to any unitarily invariant norm and operates under the assumption that E contains independent and identically distributed (i.i.d.) standard Gaussian entries. Additionally, we have enhanced the r-dependence in the bounds obtained from [64] and have eased some technical assumptions. A detailed discussion of this extension can be found in Section 2.1.

To extend the results beyond the scenario where E comprises i.i.d. Gaussian entries, we utilize a methodology similar to that in our previous work [63] and derive analogous perturbation of singular values and singular subspaces. Notably, these results hold true for random noise of any specific structure, provided the noise induces a negligible effect on the singular subspaces of the matrix A. Furthermore, these results alleviate the trio-concentration assumption for E imposed in our previous work [63]. Compared to the prior analysis where E is a Gaussian matrix, the bounds now include an additional term, which may not be optimal. Nonetheless, we posit that the generalized setting on E offers wider applicability in many practical scenarios. These findings are presented at the end of Section 2.1.

There is currently a surging interest in the field of ℓ_∞ analysis, also known as entrywise analysis of eigenvectors and singular vectors. This dynamic research area is dedicated to deriving rigorous bounds, such as those found in ℓ_∞ analysis [2, 20, 36, 43, 45, 84] for eigenvectors or singular vectors, or $\ell_{2,\infty}$ analysis for eigenspaces or singular subspaces [1, 4, 27, 32, 55], in relation to perturbed matrix models. The driving force behind these pursuits lies in the substantial impact and wide-ranging applications these analyses offer in statistics and machine learning.

Inspired by recent advancements, we have derived precise ℓ_∞ bounds for the perturbed singular vectors and the $\ell_{2,\infty}$ bounds for the perturbed singular subspaces of $A+E$. Beyond these specific bounds, we have also established results pertaining to the generalized components - also known as linear and bilinear forms - of the perturbed singular vectors and singular subspaces. We further investigate the $\ell_{2,\infty}$ bounds on the perturbed singular vectors, taking into account the weighting by their respective singular values. These new results are presented in Section 2.2.

In Section 4, we demonstrate the practical applications of our theoretical findings within two statistical models: the Gaussian Mixture Model and the submatrix localization problem. Our main goal is to use these results to examine how well spectral algorithms work and provide clear, straightforward proofs of their performance.

Organization: The paper is organized as follows. Section 2 presents our new matrix perturbation results. Notably, Section 2.1 extends Wedin's $\sin\Theta$ theorem to stochastic versions suitable for arbitrary unitarily invariant norms. Results focusing on the ℓ_∞ and $\ell_{2,\infty}$ norms of singular vectors and subspaces are consolidated in Section 2.2. Section 3 provides a concise survey of related literature. Applications of our perturbation results are demonstrated through the analysis of spectral algorithms for the Gaussian Mixture Model and the submatrix localization problem, discussed in Sections 4.1 and 4.2, respectively. Preliminaries and basic tools employed in our proofs are introduced in Section 5, which also includes an overview of the proofs for our main results. The subsequent sections, along with the appendices, are dedicated to the detailed proofs of our main results, as well as the proofs of basic tools related to them.

Notation: For a vector $v = (v_1, \cdots, v_n) \in \mathbb{R}^n$, the following norms are frequently used: $\|v\| = \sqrt{\sum_{i=1}^n v_i^2}$ and $\|v\|_\infty = \max_i |v_i|$. Also, $\|v\|_0$ is the number of non-zero elements in v. For a real matrix M, $\|M\|$ denotes its operator norm, while $\|M\|_F$ represents its Frobenius norm. The term $\|M\|_{\max}$ refers to the largest absolute value among its entries, and $\|M\|_{2,\infty}$ indicates the maximum length of its rows. For a set S, let $\mathbf{1}_S$ be the indicator function of this set. For two functions $f(n), g(n) > 0$, we use the asymptotic notations $f(n) \gg g(n)$ and $g(n) = o(f(n))$ if $f(n)/g(n) \to \infty$ as $n \to \infty$. The notation $f(n) = O(g(n))$ and $f(n) \lesssim g(n)$ are used when there exists some constant $C > 0$ such that $f(n) \leqslant Cg(n)$ for sufficiently large n. If $f(n) = O(g(n))$ and $g(n) = O(f(n))$, we denote $f(n) \asymp g(n)$. The set of $n \times n$ orthogonal matrices is denoted by $\mathbb{O}^{n\times n}$.

2. New results on the matrix perturbation bounds

2.1. **Stochastic Wedin's $\sin\Theta$ theorem.** We first generalize the previous results in [63,64] to an arbitrary unitarily invariant norm $|||\cdot|||$. Denote by $\sin\angle(U,V)$ the diagonal matrix whose diagonal entries are $\sin\theta_i$'s, where θ_i represents the principal angles between subspaces U and V. A detailed definition can be found in Section 5.2.

For any $1 \leqslant k \leqslant s \leqslant r$, denote

$$U_{k,s} := \text{Span}\{u_k, \ldots, u_s\}, \quad \widetilde{U}_{k,s} := \text{Span}\{\widetilde{u}_k, \ldots, \widetilde{u}_s\}, \quad P_{U_{k,s}} = U_{k,s}U_{k,s}^{\mathrm{T}}$$

and analogously for $V_{k,s}$, $\widetilde{V}_{k,s}$ and $P_{V_{k,s}}$. Denote

$$D_{k,s} = \text{diag}(\sigma_k, \cdots, \sigma_s)$$

and analogously for $\widetilde{D}_{k,s}$. If $k = 1$, we simply use $D_s, \widetilde{D}_s, U_s, \widetilde{U}_s, P_{U_s}$ and $V_s, \widetilde{V}_s, P_{V_s}$.

The spectral gap (or separation)

$$\delta_k := \sigma_k - \sigma_{k+1},$$

which refers to the difference between consecutive singular values of a matrix, will play a key role in the following results.

Theorem 3 (Unitarily invariant norms: simplified asymptotic version)**.** *Let A and E be $N \times n$ real matrices, where A is deterministic with rank $r \geqslant 1$ and the entries of E are i.i.d. standard Gaussian random variables. Let $\|\cdot\|$ be any normalized, unitarily invariant norm. Consider $1 \leqslant k \leqslant r$ such that $\delta_k \gtrsim r\sqrt{r + \log(N+n)}$. Denote $k_0 = \min\{k, r-k\}$. Then with probability $1-(N+n)^{-C}$ for some $C > 0$,*

$$\|\sin \angle(U_k, \widetilde{U}_k)\| \lesssim \sqrt{kk_0}\frac{\sqrt{r+\log(N+n)}}{\delta_k} + \frac{\|P_{U^\perp} E P_{\widetilde{V}_k}\| + \|P_{V^\perp} E^{\mathrm{T}} P_{\widetilde{U}_k}\|}{\sigma_k}. \quad (3)$$

Specially, for the operator norm, we have with probability $1-(N+n)^{-C}$,

$$\|\sin \angle(U_k, \widetilde{U}_k)\| \lesssim \sqrt{k}\frac{\sqrt{r+\log(N+n)}}{\delta_k}\mathbf{1}_{\{k\neq r\}} + \frac{\|E\|}{\sigma_k}. \quad (4)$$

The same conclusion also holds for $\sin\angle(V_k, \widetilde{V}_k)$.

This bound serves as a comprehensive generalization of the classical Wedin's bound in Theorem 2 when applied to the context of random noise. When $k = r$, the first term on the right-hand side of (3) vanishes, then (3) is essentially consistent with the Wedin's bound in Theorem 2. When $k < r$, it is worth noting that $P_{U_k^\perp} = P_{U_{k+1,r}} + P_{U^\perp}$ and $P_{V_k^\perp} = P_{V_{k+1,r}} + P_{V^\perp}$. Using Wedin's bound (2), one can deduce that

$$\begin{aligned} &\|\sin \angle(U_k, \widetilde{U}_k)\| \\ &\leqslant \frac{\|P_{U_{k+1,r}} E P_{\widetilde{V}_k}\| + \|P_{V_{k+1,r}} E^{\mathrm{T}} P_{\widetilde{U}_k}\|}{\hat{\delta}_k} + \frac{\|P_{U^\perp} E P_{\widetilde{V}_k}\| + \|P_{V^\perp} E^{\mathrm{T}} P_{\widetilde{V}_k}\|}{\hat{\delta}_k}. \end{aligned} \quad (5)$$

In the setting of a low-rank signal matrix A and random noise E, our result (3) improves the second term on the right-hand side of (5) by replacing the denominator $\hat{\delta}_k = \sigma_k - \widetilde{\sigma}_{k+1}$ with a usually much larger quantity σ_k. Additionally, we demonstrate that the first term on the right-hand side of (5) is essentially $C(r)/\delta_k$, where $C(r) \lesssim r^{3/2}$.

For the operator norm, (4) represents an improvement over the previous result in [63] by O'Rourke, Vu and the author in terms of the dependence on the rank r. In particular, when $k = 1$, we obtain that with probability $1-(N+n)^{-C}$,

$$\sin\angle(u_1, \widetilde{u}_1) \lesssim \frac{\sqrt{r+\log(N+n)}}{\delta_1} + \frac{\|E\|}{\sigma_1}. \quad (6)$$

We believe (6) is optimal up to the dependence on the constants.

In practice, computing the second term on the right-hand side of (3) precisely is challenging due to the dependence among $E, P_{\widetilde{U}_k}, P_{\widetilde{V}_k}$. Therefore, for practical applications, a simplified bound below offers convenience.

Corollary 4. *Under the assumptions of Theorem 3, the following holds with probability $1-(N+n)^{-C}$,*

$$\|\sin \angle(U_k, \widetilde{U}_k)\| \lesssim \sqrt{kk_0}\frac{\sqrt{r+\log(N+n)}}{\delta_k} + k\frac{\|E\|}{\sigma_k}. \quad (7)$$

Theorem 3 follows immediately from the next general and non-asymptotic result. For ease of notation, denote $\sigma_0 := \infty$ and $\delta_0 := \infty$. We define

$$\chi(b) := 1 + \frac{1}{4b(b-1)} \text{ for } b \geqslant 2.$$

Theorem 5 (Unitarily invariant norms: Gaussian noise)**.** *Let A and E be $N \times n$ real matrices, where A is deterministic and the entries of E are i.i.d. standard Gaussian random variables. Let $\|\|\cdot\|\|$ be any normalized, unitarily invariant norm. Assume A has rank $r \geqslant 1$. Let $K > 0$ and $b \geqslant 2$. Denote $\eta := \frac{11b^2}{(b-1)^2}\sqrt{2(\log 9)r + (K+7)\log(N+n)}$. Assume $(\sqrt{N}+\sqrt{n})^2 \geqslant 32(K+7)\log(N+n) + 64(\log 9)r$. Consider $1 \leqslant r_0 \leqslant r$ such that $\sigma_{r_0} \geqslant 2b(\sqrt{N}+\sqrt{n}) + 80b\eta r$ and $\delta_{r_0} \geqslant 75\chi(b)\eta r$. For any $1 \leqslant k \leqslant s \leqslant r_0$, if $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$, then*

$$\begin{aligned}\|\|\sin\angle(U_{k,s}, \widetilde{U}_{k,s})\|\| \leqslant & 6\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\sqrt{\min\{s-k+1, r-s+k-1\}}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1}, \delta_s\}} \\ &+ 2\frac{\|\|P_{U^\perp}EP_{\widetilde{V}_{k,s}} \oplus P_{V^\perp}E^{\mathrm{T}}P_{\widetilde{U}_{k,s}}\|\|}{\sigma_s}\end{aligned} \tag{8}$$

with probability at least $1-20(N+n)^{-K}$.

Specially, for the operator norm, we have with probability at least $1-20(N+n)^{-K}$

$$\|\sin\angle(U_{k,s}, \widetilde{U}_{k,s})\| \leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\mathbf{1}_{\{s-k+1\neq r\}}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1}, \delta_s\}} + 2\frac{\|E\|}{\sigma_s}.$$

The same conclusion also holds for $\sin\angle(V_{k,s}, \widetilde{V}_{k,s})$.

Remark 6. Throughout the proofs, we work on the event that $\|E\| \leqslant 2(\sqrt{N}+\sqrt{n})$. Lemma 32 below guarantees this event holds with very high probability. In Theorem 5, the parameter $b \geqslant 2$ represents the signal-to-noise ratio, and in the proof, we ensure that $\sigma_{r_0}/\|E\| \geqslant b$. The parameter b, which could depend on N and n, could account for a particularly strong signal. We have selected certain constants and expressions such as 80, $75\chi(b)$, and $\frac{(b+1)^2}{(b-1)^2}$ for the sake of convenience in our computations while our primary objective was not to optimize these constants within the proof. It is also feasible to conduct work on the event that $\|E\| \leqslant (1+\epsilon_1)(\sqrt{N}+\sqrt{n})$, and assume $b \geqslant 1+\epsilon_2$ for $\epsilon_1, \epsilon_2 > 0$. By following the same proof, one can arrive at refined constants and bounds.

Remark 7. Building upon the work of [64], which focused on bounds within operator norms, Theorem 5 achieves an improvement in terms of the dependency on r. Moreover, Theorem 5 eliminates a restrictive condition in [64], which requires the distinct singular values among $\sigma_k, \cdots, \sigma_s$ to be separated by a distance of order $r^2\sqrt{\log(N+n)}$. This condition, often challenging to verify for practical applications, is no longer necessary in our theorem.

To go beyond the i.i.d. Gaussian noise matrix, we record the following results on the perturbation of singular values and singular subspaces that is obtained using a similar approach as in the previous work by O'Rourke, Vu and the author [63]. In particular, these results remain valid for random noise of any specific structure, as long as the noise has a negligible effect on the singular subspaces of matrix A.

Theorem 8 (Singular value bounds: general noise)**.** *Assume A has rank r and E is random. Let $1 \leqslant k \leqslant r$. Consider any $\varepsilon \in (0,1)$.*

- *If there exists $t > 0$ such that $\|U_k^T E V_k\| \leqslant t$ with probability at least $1-\varepsilon$, then we have, with probability at least $1-\varepsilon$,*

$$\widetilde{\sigma}_k \geqslant \sigma_k - t. \tag{9}$$

- *If there exist* $L, B > 0$ *such that* $\|U^T EV\| \leqslant L$ *and* $\|E\| \leqslant B$ *with probability at least* $1-\varepsilon$*, then we have, with probability at least* $1-\varepsilon$*,*

$$\widetilde{\sigma}_k \leqslant \sigma_k + 2\sqrt{k}\frac{B^2}{\widetilde{\sigma}_k} + k\frac{B^3}{\widetilde{\sigma}_k^2} + L. \tag{10}$$

Theorem 9 (Singular subspace bounds: general noise)**.** *Assume* A *has rank* r *and* E *is random. Let* $1 \leqslant k \leqslant r$*. For* $\varepsilon > 0$*, assume there exist* $L, B > 0$ *such that* $\|U^{\mathrm{T}} EV\| \leqslant L$ *and* $\|E\| \leqslant B$ *with probability at least* $1-\varepsilon$*. Furthermore, assume* $\delta_k = \sigma_k - \sigma_{k+1} \geqslant 2L$*. Then for any normalized, unitarily invariant norm* $\|\!|\cdot|\!\|$*, the following holds with probability at least* $1-\varepsilon$*,*

$$\|\!|\sin\angle(U_k, \widetilde{U}_k)|\!\| \leqslant 2\sqrt{k\min\{k, r-k\}}\left(\frac{L}{\delta_k} + 2\frac{B^2}{\delta_k\sigma_k}\right) + 2k\frac{B}{\sigma_k}.$$

More specifically, for the operator norm,

$$\|\sin\angle(U_k, \widetilde{U}_k)\| \leqslant 2\sqrt{k}\left(\frac{L}{\delta_k} + 2\frac{B^2}{\delta_k\sigma_k}\right)\mathbf{1}_{\{k<r\}} + 2\frac{B}{\sigma_k}.$$

The same result also holds for $\sin\angle(V_k, \widetilde{V}_k)$*.*

To apply Theorems 8 and 9, it is necessary to obtain effective bounds on $\|E\|$ and $\|U^{\mathrm{T}} EV\|$. In general, matrix concentration inequalities (refer to [69] for example) can provide good upper bounds on $\|E\|$ for random noise E with heteroskedastic entries and even complex correlations among the entries. On the other hand, $U^{\mathrm{T}} EV$ is an $r \times r$ matrix, and $\|U^{\mathrm{T}} EV\|$ typically depends on r. Bounds on $\|U^{\mathrm{T}} EV\|$ can be obtained by applying concentration inequalities. In order to facilitate the application of our results, we provide a convenient version below that relaxes the trio-concentration assumption for E required in our previous work [63].

Consider a non-negative function $f(t)$ on $[0, \infty)$ which tends to zero with t tending to infinity. Given a matrix A, we say that a random matrix E is *f-bounded (with respect to A)* if for any non-trivial left singular vector u and non-trivial right singular vector v of A, we have

$$\mathbb{P}\left(|u^{\mathrm{T}} Ev| \geqslant t\right) \leqslant f(t).$$

Theorem 10 (f-bounded random noise)**.** *Assume* A *has rank* r *and* E *is* f*-bounded. For* $1 \leqslant k \leqslant r$*, if* $\delta_k > 0$*, then for any* $t > 0$*,*

$$\|\sin\angle(U_k, \widetilde{U}_k)\| \leqslant 2\sqrt{k}\left(\frac{t}{\delta_k} + 2\frac{\|E\|^2}{\delta_k\sigma_k}\right)\mathbf{1}_{\{k<r\}} + 2\frac{\|E\|}{\sigma_k}$$

holds with probability at least

$$1 - r^2 9^{2r} f\left(\frac{t}{2r}\right) - k^2 9^{2k} f\left(\frac{\delta_k}{4k}\right).$$

The same result also holds for $\sin\angle(V_k, \widetilde{V}_k)$*.*

The proofs of Theorems 8, 9 and 10 can be found in Section 7.

2.2. **ℓ_∞ and $\ell_{2,\infty}$ analysis.** Next, we present a result regarding the estimation of the singular vectors on an entrywise basis. In this context, a parameter known as the incoherence parameter of the singular vector matrices U and V, denoted as $\|U\|_{2,\infty}$ and $\|V\|_{2,\infty}$, is of central importance. Smaller values of $\|U\|_{2,\infty}$ and $\|V\|_{2,\infty}$ suggest that the information contained in the signal matrix A is less concentrated in just a few rows or columns.

In this section, we use $U_{k,s} = (u_k, \cdots, u_s)$ to denote the singular vector matrix for $1 \leqslant k \leqslant s \leqslant r$. We abbreviate $U_{k,s}$ to U_s when $k = 1$. Note that $P_{U_{k,s}} = U_{k,s}U_{k,s}^{\mathrm{T}}$. These notations also apply to $\widetilde{U}_{k,s}$ and $\widetilde{U}_s$. For simplicity, we only state the results for the left singular vectors $U_{k,s}$. The corresponding results for the right singular vectors can be derived by applying these results to the transposes of matrices A^{T} and $A^{\mathrm{T}} + E^{\mathrm{T}}$.

Let us make a temporary assumption that $\sigma_1 \leqslant n^2$. This assumption is reasonable because if $\sigma_k > n^2$, it indicates a highly significant signal, and the impact of noise becomes negligible in such cases. Denote $\sigma_0 := \infty$ and $\delta_0 := \infty$.

Theorem 11 (ℓ_∞ and $\ell_{2,\infty}$ bounds: simplified asymptotic version)**.** *Let A and E be $N \times n$ real matrices, where A is deterministic with rank $r \geqslant 1$ and the entries of E are i.i.d. standard Gaussian random variables. Let $1 \leqslant k \leqslant r$ and $\sigma_k \geqslant 2\|E\|$.*

- *If $\min\{\delta_{k-1}, \delta_k\} \gtrsim r\sqrt{r + \log(N+n)}$, then with probability $1 - (N+n)^{-C}$,*

$$\|\widetilde{u}_k - (\widetilde{u}_k^{\mathrm{T}} u_k)u_k\|_\infty \lesssim \frac{\sqrt{r + \log(N+n)}}{\min\{\delta_{k-1}, \delta_k\}}\|U\|_{2,\infty} + \frac{\sqrt{r\log(N+n)}}{\sigma_k}\left(1 + \|U\|_{2,\infty}\right).$$

- *If $\delta_k \gtrsim r\sqrt{r + \log(N+n)}$, then with probability $1 - (N+n)^{-C}$,*

$$\|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} \lesssim \sqrt{k}\frac{\sqrt{r + \log(N+n)}}{\delta_k}\|U\|_{2,\infty} + \sqrt{k}\frac{\sqrt{r\log(N+n)}}{\sigma_k}\left(1 + \|U\|_{2,\infty}\right).$$

In many applications, the primary interest lies in comparing $\widetilde{U}_k$ with $U_k O$, accounting for the non-uniqueness of singular vectors via an orthogonal matrix O. A suitable choice of O that aligns U_k with $\widetilde{U}_k$ effectively can be determined by examining the SVD of $U_k^{\mathrm{T}}\widetilde{U}_k$. By Proposition 24, the SVD of $U_k^{\mathrm{T}}\widetilde{U}_k$ is $U_k^{\mathrm{T}}\widetilde{U}_k = O_1 \cos\angle(U_k, \widetilde{U}_k)O_2^{\mathrm{T}}$ and we choose $O = O_1 O_2^{\mathrm{T}}$. Note that the discrepancy between $U_k^{\mathrm{T}}\widetilde{U}_k$ and O can be measured by the principal angles between the subspaces U_k and $\widetilde{U}_k$. In Proposition 25, we establish

$$\|\widetilde{U}_k - U_k O\|_{2,\infty} \leqslant \|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} + \|U_k\|_{2,\infty}\|\sin\angle(U_k, \widetilde{U}_k)\|^2. \tag{11}$$

Therefore, by combining Theorem 5 with Theorem 11, we obtain the next result.

Corollary 12. *Under the same assumption as Theorem 11, the following holds:*

- *If $\min\{\delta_{k-1}, \delta_k\} \gtrsim r\sqrt{r + \log(N+n)}$, then with probability $1 - (N+n)^{-C}$,*

$$\min_{\mathsf{s}\in\{\pm 1\}} \|u_k - \mathsf{s}\widetilde{u}_k\|_\infty \lesssim \frac{\sqrt{r + \log(N+n)}}{\min\{\delta_{k-1}, \delta_k\}}\|U\|_{2,\infty} + \frac{\sqrt{r\log(N+n)}}{\sigma_k}\left(1 + \|U\|_{2,\infty}\right) + \frac{\|E\|^2}{\sigma_k^2}\|u_k\|_\infty.$$

- *If* $\delta_k \gtrsim r\sqrt{r+\log(N+n)}$*, then with probability* $1-(N+n)^{-C}$*,*

$$\min_{O\in\mathbb{O}^{k\times k}}\|\widetilde{U}_k - U_kO\|_{2,\infty} \lesssim \sqrt{k}\frac{\sqrt{r+\log(N+n)}}{\delta_k}\|U\|_{2,\infty} + \sqrt{k}\frac{\sqrt{r\log(N+n)}}{\sigma_k}(1+\|U\|_{2,\infty}) + \frac{\|E\|^2}{\sigma_k^2}\|U_k\|_{2,\infty}.$$

Theorems 11 follows as a direct consequence of the next general and non-asymptotic result, which we will prove in Section 6.

Theorem 13. *Let A and E be $N\times n$ real matrices, where A is deterministic and the entries of E are i.i.d. standard Gaussian random variables. Assume A has rank $r \geqslant 1$. Let $K>0$ and $b\geqslant 2$. Denote $\eta := \frac{11b^2}{(b-1)^2}\sqrt{2(\log 9)r + (K+7)\log(N+n)}$. Assume $(\sqrt{N}+\sqrt{n})^2 \geqslant 32(K+7)\log(N+n) + 64(\log 9)r$. Consider $1\leqslant r_0\leqslant r$ such that $\sigma_{r_0} \geqslant 2b(\sqrt{N}+\sqrt{n}) + 80b\eta r$ and $\delta_{r_0}\geqslant 75\chi(b)\eta r$. For any $1\leqslant k\leqslant s\leqslant r_0$, if $\min\{\delta_{k-1},\delta_s\}\geqslant 75\chi(b)\eta r$, then with probability at least $1-40(N+n)^{-K}$,*

$$\|\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\|_{2,\infty} \leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|U\|_{2,\infty}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} + \frac{2\sqrt{2}b^2}{(b-1)^2}(1+\|U\|_{2,\infty})\sqrt{\sum_{i\in[\![k,s]\!],\sigma_i\leqslant n^2}\frac{\gamma^2}{\sigma_i^2} + \sum_{i\in[\![k,s]\!],\sigma_i>n^2}\frac{16n}{\sigma_i^2}}, \tag{12}$$

where $\gamma := \frac{9b^2}{(b-1)^2}\sqrt{r(K+7)\log(N+n)}$.

It should be noted that, as per the aforementioned result, the term

$$\sum_{k\leqslant i\leqslant s,\sigma_i>n^2}\frac{16n}{\sigma_i^2} < \frac{16}{n^2}$$

can always be considered negligible in comparison to the other terms. Indeed, when the signal is extremely strong, i.e. $\sigma_i > n^2 \gg \|E\| = \Theta(\sqrt{N}+\sqrt{n})$, the impact of noise becomes minimal.

More generally, we can establish the following result, which provides bounds for the singular subspaces in any arbitrary direction. For the simplicity of presentation, we assume all the singular values are no more than n^2. The proofs are provided in Section 6.3.

Theorem 14 (Bounds on linear and bilinear forms)**.** *Under the assumptions of Theorem 13 and further assuming that $\sigma_1\leqslant n^2$, for any unit vectors $x\in\mathbb{R}^N$ and $y=(y_k,\cdots,y_s)^{\mathrm{T}}\in\mathbb{R}^{s-k+1}$, the following holds with probability at least $1-40(N+n)^{-K}$:*

$$\|x^{\mathrm{T}}(\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s})\| \leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|x^{\mathrm{T}}U\|\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} + \frac{2\sqrt{2}b^2}{(b-1)^2}\gamma\left(1+\|x^{\mathrm{T}}U\|\right)\sqrt{\sum_{i=k}^{s}\frac{1}{\sigma_i^2}}$$

and

$$\left|x^{\mathrm{T}}(\widetilde{U}_{k,s}-P_{U_{k,s}}\widetilde{U}_{k,s})y\right| \leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|x^{\mathrm{T}}U\|\frac{\eta\sqrt{\|y\|_0}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} + \frac{2\sqrt{2}b^2}{(b-1)^2}\gamma\left(1+\|x^{\mathrm{T}}U\|\right)\sum_{i=k}^{s}\frac{|y_i|}{\sigma_i}, \tag{13}$$

where $\gamma = \frac{9b^2}{(b-1)^2}\sqrt{r(K+7)\log(N+n)}$.

Remark 15. When the focus is on comparing the linear (or bilinear) forms of $U_{k,s}$ and $\widetilde{U}_{k,s}$, in a manner analogous to Corollary 12, one can leverage the fact provided in Proposition 25:

$$\|x^{\mathrm{T}}(\widetilde{U}_{k,s}-U_{k,s}O)\| \leqslant \|x^{\mathrm{T}}(\widetilde{U}_{k,s}-P_{U_{k,s}}\widetilde{U}_{k,s})\| + \|x^{\mathrm{T}}U_{k,s}\|\|\sin\angle(U_{k,s},\widetilde{U}_{k,s})\|^2$$

and combine Theorems 5 and 14.

Specially, applying (13) with the canonical vectors, together with Proposition 25, we have with probability $1-(N+n)^{-C}$ that

$$\|\widetilde{U}_{k,s}-U_{k,s}O\|_{\max} \lesssim \frac{\sqrt{r+\log(N+n)}}{\min\{\delta_{k-1},\delta_s\}}\|U\|_{2,\infty} + \frac{\sqrt{r\log(N+n)}}{\sigma_s}(1+\|U\|_{2,\infty}) + \frac{\|E\|^2}{\sigma_s^2}\|U_{k,s}\|_{2,\infty} \tag{14}$$

for some orthogonal matrix O.

Building upon the proof of Theorem 12 and incorporating minor modifications, we obtain the subsequent bounds. These describe the extent to which the dominant singular vectors of the perturbed matrix, when weighted by their singular values, deviate from the original subspace. The proof can be found in Section 6.4.

Theorem 16 (Bounds on singular value-adjusted projection perturbation)**.** *Under the assumptions of Theorem 13, the following holds with probability at least* $1-40(N+n)^{-K}$*:*

$$\|\widetilde{U}_{k,s}\widetilde{D}_{k,s}-P_{U_{k,s}}\widetilde{U}_{k,s}\widetilde{D}_{k,s}\|_{2,\infty} \leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|U\|_{2,\infty}\frac{\eta\sigma_k\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} + \frac{2\sqrt{2}b^2}{(b-1)^2}(1+\|U\|_{2,\infty})\sqrt{\gamma^2(s-k+1)+16},$$

where $\gamma = \frac{9b^2}{(b-1)^2}\sqrt{r(K+7)\log(N+n)}$.

Directly comparing $\widetilde{U}_{k,s}\widetilde{D}_{k,s}$ with $U_{k,s}O\widetilde{D}_{k,s}$, with respect to the choice of an orthogonal matrix O, is not as straightforward as in the unweighted case given in Corollary 12. It requires a closer analysis of the interaction between $\widetilde{D}_{k,s}$ and the orthogonal matrix O.

To illustrate the main idea of such extension, we focus exclusively on the full singular vector matrix U and work with the following bound: for some orthogonal matrix O of size $r\times r$,

$$\|\widetilde{U}_r\widetilde{D}_r-UO\widetilde{D}_r\|_{2,\infty} \leqslant \|\widetilde{U}_r\widetilde{D}_r-P_U\widetilde{U}_r\widetilde{D}_r\|_{2,\infty} + \|U\|_{2,\infty}\|\sin\angle(U,\widetilde{U}_r)\|\|E\|. \tag{15}$$

The proof of (15) can be found in Appendix A.4. Using similar arguments, we can establish the comparison of $\widetilde{U}_{k,s}\widetilde{D}_{k,s}$ and $U_{k,s}O\widetilde{D}_{k,s}$, though this would result in a more complex version of the second term on the right-hand side of (15) (a generalization we will not pursue here). From (15), by combining Theorem 16 with Theorem 5, we arrive at the next result, which plays an important role in our applications.

Corollary 17. *Under the assumptions of Theorem 13, where we set $r_0 = s = r$ and $k = 1$, the following holds with probability at least $1 - 40(N+n)^{-K}$:*

$$\min_{O \in \mathbb{O}^{r\times r}} \|\widetilde{U}_r \widetilde{D}_r - UO\widetilde{D}_r\|_{2,\infty} \leqslant \frac{36b^4}{(b-1)^4} r\sqrt{(K+7)\log(N+n)}(1+\|U\|_{2,\infty}) + 2\|U\|_{2,\infty}\frac{\|E\|^2}{\sigma_r}.$$

The proof of Theorem 16 can be adapted to draw comparisons between $\widetilde{U}_{k,s}\widetilde{D}_{k,s}$ and $P_{U_{k,s}}\widetilde{U}_{k,s}\widetilde{D}_{k,s}$ in arbitrary directions, analogous to the approach taken in Theorem 14. However, we do not explore this generalization in the present work.

3. Related works

In the deterministic setting, several important studies have introduced variants and extensions of the classical Davis-Kahan-Wedin $\sin\Theta$ theorems [40, 76]. For instance, a study by Yu, Wang and Samworth [80] propose a variant of the Davis-Kahan-Wedin theorem for $\|\sin\Theta\|_F$ where the $\hat{\delta}_k := \sigma_k - \widetilde{\sigma}_{k+1}$ is replaced by $\delta_k := \sigma_k - \sigma_{k+1}$. Vu and Lei [74] present a variational form of the Davis-Kahan $\sin\Theta$ theorem for the perturbation of a positive semidefinite matrix A and obtained bounds on $\|\sin\Theta\|_F$ for the eigenvector space A and a subspace not necessarily corresponding to that of the perturbed matrix. Cai and Zhang [29] provide separate perturbation bounds on $\|\sin\Theta\|_F$ and $\|\sin\Theta\|$ for the left and right singular subspaces, specifically tailored to handle cases when N, n differ significantly. A robust perturbation analysis for symmetric low-rank plus perturbation matrices is proposed by [81] in a deterministic setting. Meanwhile, Luo, Han and Zhang [58] focus on low-rank matrix estimation and obtain bounds under the Schatten-q matrix norms, i.e. bounds on $\|\sin\angle(U,\widetilde{U})\|_q$ with $q \in [1,\infty]$ for the entire singular subspaces. Zhang and Zhou [82] have recently developed deterministic perturbation bounds for the singular subspaces in the Frobenius norm, specifically for pairs of matrices where one is derived from the other by omitting a single column. These bounds have been employed to analyze the performance of spectral clustering methods applied to mixture models.

Switching to the stochastic framework, the literature is more recent and equally rich. As outlined in the introduction, our main results in this paper improve upon the works [63, 64, 72]. In [75], Wang explores the non-asymptotic distribution of singular vectors when entries of X are i.i.d. standard Gaussian random variables. Meanwhile, Allez and Bouchaud [5] investigate the eigenvector dynamics of $A + E$ when both A and E are real symmetric matrices, and the entries of E are constructed from a family of independent real Brownian motions. A perturbative expansion of the coordinates of the eigenvectors is provided by Benaych-Georges, Enriquez and Michaïl [15]. In line with these studies, Zhong [83] develops a non-asymptotic Rayleigh-Schrdinger theory for symmetric low-rank plus random noise model. The results focus only on the perturbed leading eigenvector and investigate $\sin\angle(\widetilde{u}_1, u_k)$ with $k > 1$.

The field has recently witnessed significant growth in the literature focused on fine-grained perturbation analysis. This includes in-depth examinations of the perturbed singular vectors and singular subspaces, with particular attention to their ℓ_∞ and $\ell_{2,\infty}$ norms [1, 2, 4, 20, 27, 32, 36, 45, 55, 60, 78, 79, 84], as well as a broader exploration of their linear and bilinear forms [53, 54, 56, 77].

Among these works, in a deterministic setting, Fan, Wang, and Zhong [45] specifically focus on the scenario where the signal matrix is low-rank and exhibits incoherence. They establish bounds for $\|\widetilde{U}O - U\|_{\max}$ and $\max_{i\in[r]} \|\widetilde{u}_i - u_i\|_\infty$. A series of results in [1, 2, 20, 27, 36, 78, 79] has been obtained within the ingenious leave-one-out analysis framework, under varying assumptions about the noise matrix. In the context of symmetric matrices, Abbe, Fan, Wang, and Zhong [2] investigate the matrix $\widetilde{A} = A + E$, where A satisfies mild incoherence conditions and E is a random noise matrix. They establish $\ell_{2,\infty}$ norm bounds for the eigenspaces, considering a broad range of noise matrices E. Building on this work, Abbe, Fan, and Wang [1] further extend their previous work [2] by conducting a more comprehensive $\ell_{2,p}$ analysis of eigenspaces for a hollowed version of PCA. In another study, [27] considers a low-rank plus random noise matrix model, specifically in scenarios when the matrix dimensions are highly unbalanced. The authors examine the sample Gram matrix with diagonal deletion and achieve ℓ_2 and $\ell_{2,\infty}$ estimation accuracy. Lei [55] investigates the $\ell_{2,\infty}$ eigenspace perturbation bounds for symmetric random matrices, accounting for more complex dependency structures within the noise matrix. Meanwhile, Chen, Fan, Ma, and Wang [36] develop the ℓ_∞ eigenvector perturbation bounds for asymmetric probability transition matrices. Within the phase synchronization model, Zhong and Boumal [84] derive the ℓ_∞ perturbation bounds for the leading eigenvectors, as a by-product of their analysis of semidefinite programming relaxations and the generalized power method. A recent study by Bhardwaj and Vu [20] presents a stochastic variant of the Davis-Kahan-Wedin theorem, which quantifies the perturbation of eigenvectors and singular vectors in the ℓ_∞ norm. This study pertains to low-rank signal matrices in the presence of general random noise, analogous to the setting explored in [63] for ℓ_2 norm analysis. In a recent work by Yan and Wainwright [79], the authors explore the same low-rank matrix perturbation model as presented in this paper, where the noise E consists of independent sub-Gaussian entries. By employing a novel expansion of $\widetilde{U}_r O - U$ with respect to the noise E, the findings in [79] establish a foundation from which the $\ell_{2,\infty}$ bound for the full singular vector matrices can be derived, in addition to providing a distributional characterization of the error in the estimation. The results in [79] refine the estimates from an earlier work by Yan, Chen and Fan [78], which considers a more general noise matrix E.

In the context of symmetric matrices, Eldridge, Belkin, and Wang [43] utilize the Neumann series trick to bound $\|\widetilde{u} - u\|_\infty$. Cape, Tang, and Priebe [32] offer comprehensive perturbation bounds for the $\ell_{2,\infty}$ norm and discuss applications to matrices with specialized structures. Additionally, Agterberg, Lubbers, and Priebe [4] introduce an estimator for singular vectors of high-dimensional, low-rank matrices with additive heteroskedastic sub-Gaussian noise, proving finite-sample $\ell_{2,\infty}$ bounds and a Berry-Esseen theorem for the individual entries of the estimator.

Koltchinskii and Xia [54] have derived concentration bounds for linear and bilinear forms involving singular vectors and singular subspaces, under the same setting as the current paper. This work was later extended to tensors by Xia and

Zhou [77]. The concentration and asymptotic distributions of bilinear forms of the spectral projectors for principle components under Gaussian noise in a general Hilbert space have been explored in [53]. All these findings require the spectral gap to be of the same order as the size of noise matrix $\|E\|$. Improving upon the result in [54], Li, Cai, Poor and Chen in [56] consider the linear function $a^{\mathrm{T}}\widetilde{u}_k$ for the perturbed eigenvector $\widetilde{u}_k$ applicable to symmetric matrix denoising models and principal components under Gaussian noise. They derive bounds on the distance between $a^{\mathrm{T}}\widetilde{u}_k$ and a de-biased estimator, requiring only the eigen-gap to be larger than $\sqrt{r}\log n$. Other works by Cheng, Wei and Chen [38] and Agterberg [3] also highlight perturbation results in the presence of small eigen-gaps.

Comprehensive insights into the ℓ_2 and ℓ_∞ analyses of current perturbation results and their practical implications are available in the survey [35].

In the context of random matrix theory, significant efforts have been devoted to studying the spectral statistics of deformed random matrices, especially focusing on extreme eigenvalues and eigenvectors. These extreme eigenvalues show unique spectral behaviors that differ markedly across various transition regimes, a phenomenon known as the *BBP phase transition*, after Baik, Ben Arous, and Péché's foundational work [9]. The extreme eigenvectors also undergo a phase transition with initial results established by [17,18,66]. Extensive research has followed on extreme eigenvalues [8,10,16,23,42] and extreme eigenvectors [12,13,14,19,23,33,34,44] in these models. These studies largely investigate the limiting behavior of extreme eigenvalues and eigenvectors as the matrix size grows to infinity.

The selection of references cited herein represents a snapshot of a rapidly advancing field and is not intended to be exhaustive.

4. Applications

4.1. **Gaussian mixture model.**

The Gaussian Mixture Model (GMM) is a type of probabilistic model often used for clustering and density estimation. It assumes that the observed data are generated from a mixture of several Gaussian distributions, each characterized by a mean vector and a covariance matrix.

Consider observed data $X = (X_1, \cdots, X_n) \in \mathbb{R}^{p\times n}$, where each X_i is a p-dimensional vector. We assume there are k distinct clusters represented by the centers $\theta_1, \cdots, \theta_k \in \mathbb{R}^p$. Denote $[n] := \{1, \cdots, n\}$. Let $\mathbf{z} = (z_1, \cdots, z_n)^{\mathrm{T}} \in [k]^n$ be the latent variable that represents the true cluster labels for each observation X_i. The model assumes that each X_i is generated as a result of adding a Gaussian noise term ϵ_i to its corresponding center θ_{z_i}, with ϵ_i's being i.i.d. $\mathcal{N}(0, I_p)$. In particular, $X_i = \theta_{z_i} + \epsilon_i$ and we denote

$$X = \mathbb{E}(X) + E. \tag{16}$$

The goal of the GMM is to classify the observed data X into k clusters, and recover the latent variable $\mathbf{z}$. Let $\widetilde{\mathbf{z}}$ be the output of a clustering algorithm for the GMM and the accuracy of this algorithm can be evaluated using the misclassification rate, defined as:

$$\mathcal{M}(\mathbf{z}, \widetilde{\mathbf{z}}) := \frac{1}{n} \min_{\pi\in\mathcal{S}_k} |\{i \in [n] : z_i \neq \pi(\widetilde{z}_i)\}|\,,$$

where $\mathcal{S}_k$ is the set of all permutations of $[k]$.

To solve the clustering problem, typically, more satisfying outcomes can be obtained by beginning with an initial estimate and then refining it with other tools like iteration or semidefinite programming (SDP). However, our discussion will focus exclusively on the application of simple spectral methods to illustrate perturbation results. Such methods have recently received considerable attention in the literature, as seen in [1, 29, 57, 82], among others. Notably, the case of a two-cluster GMM with centers $\pm\mu$ for a fixed vector μ has been extensively studied in [1, 29].

In the context of a general k-cluster framework, it is important to recognize insights from [57] that establish spectral clustering as optimal for GMM. Our main goal is to show that the application of our perturbation results provides a succinct and effective proof for examining the theoretical performance of spectral algorithms.

Denote the minimum distance among centers as

$$\Delta := \min_{j,l\in[k]:j\neq l} \|\theta_j - \theta_l\|.$$

When the separation between cluster centers, denoted by Δ, is sufficiently large, distance-based clustering methods become particularly commendable.

The principle of spectral clustering is elegantly simple. Consider the SVD of $\mathbb{E}(X) = U\Sigma V^{\mathrm{T}}$, where Σ is a $k \times k$ diagonal matrix with potentially zero diagonal entries if the rank of $\mathbb{E}(X)$ is less than k. The matrices U and V respectively consist of k orthonormal vectors that contain the left and right singular vectors of $\mathbb{E}(X)$. Let us denote $(U^{\mathrm{T}}\mathbb{E}(X))_j$ the columns of $U^{\mathrm{T}}\mathbb{E}(X) \in \mathbb{R}^{k\times n}$. We can demonstrate, as elaborated in Section 8, that for any columns θ_i and θ_j of $\mathbb{E}(X) = (\theta_{z_1}, \cdots, \theta_{z_n})$,

$$\|\theta_i - \theta_j\| = \|(U^{\mathrm{T}}\mathbb{E}(X))_i - (U^{\mathrm{T}}\mathbb{E}(X))_j\|.$$

This indicates that the columns of $U^{\mathrm{T}}\mathbb{E}(X) = \Sigma V^{\mathrm{T}}$ preserve the geometric relationship among the centers.

Consider the SVD of $X = \widetilde{U}\widetilde{\Lambda}\widetilde{V}^{\mathrm{T}}$ and we use the previously defined notations $\widetilde{U}_s, \widetilde{\Lambda}_s, \widetilde{V}_s$. The crux of the analysis lies in proving that, with high probability, the following inequality holds:

$$\max_{1\leqslant j\leqslant n} \|(\widetilde{U}_k^T X)_j - (U^{\mathrm{T}}\mathbb{E}(X))_j\| < \frac{1}{5}\Delta. \tag{17}$$

If this is the case, then performing clustering based on the distances among the columns of $\widetilde{U}_r^T X$ will, with high probability, successfully recover the correct cluster labels. In light of the preceding analysis, we hereby present the following algorithm:

Algorithm 1 Spectral algorithm for GMM

Input: data matrix $X \in \mathbb{R}^{n\times p}$ and cluster number k.
Output: cluster labels $\tilde{\mathbf{z}} \in [k]^n$.
Step 1. Perform SVD on X and denote $\widetilde{U}_k \in \mathbb{R}^{p\times k}$ the singular vector matrix composed of the leading k left singular vectors of X.
Step 2. Perform k-means clustering on the columns of $\widetilde{U}_k^{\mathrm{T}} X$.

Algorithm 1 is identical to the algorithm proposed in [57] and [82]. This SVD-based algorithm has been widely adopted to address a variety of well-known problems in computer science and statistics, including the hidden clique, hidden bisection, hidden coloring, and matrix completion, among others (see for instance [57,73] and references therein for more discussion).

The use of k-means clustering in *Step 2* of Algorithm 1 is not a crucial component. The key requirement is to establish the inequality in (17); once this is achieved, alternative distance-based clustering algorithms may be employed in place of k-means.

For the output $\widetilde{\mathbf{z}}$ of Algorithm 1, we could show the following result:

Theorem 18. *Consider the GMM* (16) *with cluster number* k. *Let* $\sigma_{\min} > 0$ *be the smallest singular value of* $\mathbb{E}(X)$. *Denote the smallest cluster size by* $c_{\min}$. *Let* $L > 0$ *and assume* $(\sqrt{n} + \sqrt{p})^2 \geqslant 32(L + 7)\log(n + p) + 64(\log 9)k$. *If*

$$\Delta \geqslant \max\left\{\frac{40(\sqrt{n} + \sqrt{p})}{\sqrt{c_{\min}}}, 1800k\sqrt{(L + 7)\log(n + p)}\right\}, \tag{18}$$

$$\sigma_{\min} \geqslant 40(\sqrt{n} + \sqrt{p}) + 3.8 \times 10^4 k\sqrt{2(\log 9)k + (L + 7)\log(n + p)},$$

then $\mathbb{E}\mathcal{M}(\mathbf{z}, \widetilde{\mathbf{z}}) \leqslant 40(n + p)^{-L}$.

The proof of Theorem 18 is a direct application of Corollary 17 and is detailed in Section 8. By setting $L = (n+p)/\log(n+p)$, for instance, we achieve an exponential rate of misclassification.

Löffler, Zhang and Zhou [57] have demonstrated that for the output $\widetilde{\mathbf{z}}$ of Algorithm 1, provided that $\Delta \gg \frac{k^{10}(\sqrt{n}+\sqrt{p})}{\sqrt{c_{\min}}}$, the following bound holds:

$$\mathbb{E}\mathcal{M}(\mathbf{z}, \widetilde{\mathbf{z}}) \leqslant \exp\left(-(1 - o(1))\Delta^2/8\right) + \exp(-0.08n). \tag{19}$$

More recently, Zhang and Zhou [82] have developed another innovative approach to analyze the output $\widetilde{\mathbf{z}}$ and obtained the same asymptotic exponential error rate (19) for the GMM, assuming

$$c_{\min} \geqslant 100k^3 \quad \text{and} \quad \Delta \gg \frac{k^3(n + p)/\sqrt{n}}{\sqrt{c_{\min}}}.$$

Additionally, [82, Theorem 3.1] analyzes the estimator $\widetilde{\mathbf{z}}$ for the sub-Gaussian mixture model. For the output $\widetilde{\mathbf{z}}$ of Algorithm 1, where in *Step 2* the selection is made for $\widetilde{U}_r$ with $r = \operatorname{rank}(\mathbb{E}(X))$ (implying the use of all r singular vectors of $\mathbb{E}(X)$), an exponential error rate is attainable when

$$c_{\min} \geqslant 10k, \quad \Delta > C\frac{\sqrt{k}(\sqrt{n} + \sqrt{p})}{\sqrt{c_{\min}}} \quad \text{and} \quad \sigma_r > C(\sqrt{n} + \sqrt{p})$$

for some $C > 0$. Abbe, Fan, and Wang [1] also explored the sub-Gaussian mixture model, employing the eigenvectors of the hollowed Gram matrix $\mathcal{H}(X^\top X)$ for clustering. Their approach leverages the ℓ_p perturbation results formulated in their paper but necessitates stricter conditions on the number of clusters, their sizes, and the collinearity of the cluster centers.

It is noteworthy that in the context of the GMM, results in [57] and [82] do not require any assumptions regarding the smallest singular value $\sigma_{\min}$, due to the exploitation of the Gaussian nature of the noise matrix E. Our Theorem 18 aligns with the findings for the sub-Gaussian mixture model in [82]. Since our proof does not fully utilize the Gaussianality, we only employ the rotation invariance property to simplify the proof of isotropic local law, as given in Lemma 27. Our findings can be extended to scenarios where the entries of E are sub-Gaussian random variables. This extension is facilitated by a lemma analogous to Lemma 27, which can be proved using established random matrix theory methodologies. Due to the extensive technical details involved, we have reserved the discussion of the extension to sub-Gaussian cases for a separate paper.

4.2. **Submatrix localization.** The general formulation of the submatrix localization or recovery problem involves locating or recovering a $k \times s$ submatrix with entries sampled from a distribution $\mathcal{P}$ within a larger $m \times n$ matrix populated with samples from a different distribution $\mathcal{Q}$. Specially, when $\mathcal{P}$ and $\mathcal{Q}$ are both Bernoulli or Gaussian random matrices, the detection and recovery of the submatrix have been extensively studied. These investigations span various domains, including hidden clique, community detection, bi-clustering, and stochastic block models (see [6, 7, 11, 20, 24, 25, 26, 28, 37, 39, 41, 46, 47, 52, 59, 61, 62, 73] and references therein).

The task of recovering a single submatrix has been intensively explored (see for instance, [26, 28, 37, 47, 61, 73] and references therein), but research on locating a growing number of submatrices is comparatively limited [28, 37, 39]. In this section, we focus on the recovery of multiple (non-overlapping) submatrices within the model of size $m \times n$:

$$X = M + E, \tag{20}$$

where the entries of the noise matrix E are i.i.d. standard Gaussian random variable. The signal matrix is given by

$$M = \sum_{i=1}^{k} \lambda_i \mathbb{1}_{R_i} \mathbb{1}_{C_i}^{\mathrm{T}},$$

where $\{R_i\}_{i=1}^k$ are disjoint subsets in $[m]$ and $\{C_i\}_{i=1}^k$ are non-overlapping subsets in $[n]$. We denote $\mathbb{1}_{R_i}$ as a vector in $\mathbb{R}^m$ with entries equal to 1 for indices in the set R_i and 0 elsewhere, and $\mathbb{1}_{C_i}$ is defined analogously. Denote $|R_i| = r_i$ and $|C_i| = c_i$. Assume $\lambda_i \neq 0$ for all $1 \leqslant i \leqslant k$. The goal is to discover the pairs $\{(R_i, C_i)\}_{i=1}^k$ from the matrix X.

Observe that the SVD of M is given by $M = \sum_{i=1}^k \sigma_i u_i v_i^{\mathrm{T}} := UDV^{\mathrm{T}}$, where

$$\sigma_i := |\lambda_i|\sqrt{r_i c_i},\ u_i := \operatorname{sgn}(\lambda_i)\frac{\mathbb{1}_{R_i}}{\sqrt{r_i}},\ v_i := \frac{\mathbb{1}_{C_i}}{\sqrt{c_i}}.$$

The columns of U and V are composed of u_i's and v_i's respectively and $D = \operatorname{diag}(\sigma_1, \cdots, \sigma_k)$.

Note that $|X_{ij} - M_{ij}| = |E_{ij}|$ and with high probability, $\max_{i,j} |E_{ij}| \lesssim \sqrt{\log n}$. If $\min_{i,j} |M_{ij}| = \min_{1 \leqslant l \leqslant k} |\lambda_l| \gtrsim \sqrt{\log n}$ and is greater than $\max_{i,j} |E_{ij}|$, a simple element-wise thresholding proves effective for identifying the submatrices.

In general, as in Section 4.1, we apply the same spectral clustering method to locate the submatrices. Denote $C_0 := [n] \backslash \cup_{i=1}^k C_i$ the set of isolated column indices with size $|C_0| = c_0$; define R_0 and its size r_0 analagouly. Let $(U^{\mathrm{T}} M)_j$ represent the columns of $U^{\mathrm{T}} M$. From $U^{\mathrm{T}} M = DV^{\mathrm{T}}$ and the definitions of D and V, it follows that $(U^{\mathrm{T}} M)_j$ has only 1 non-zero entry $\lambda_l \sqrt{r_l}$ if $j \in C_l$ for some $l \in [k]$ and it is a zero vector if $j \in C_0$. In particular, if $i, j \in [n]$ belong to the same C_l for $0 \leqslant l \leqslant k$, it holds that $(U^{\mathrm{T}} M)_i = (U^{\mathrm{T}} M)_j$. For $i, j \in [n]$ from different submatrices, we have that

$$\min_{\substack{i \in C_l, j \in C_s, \\ 0 \leqslant l \neq s \leqslant k}} \|(U^{\mathrm{T}} M)_i - (U^{\mathrm{T}} M)_j\| = \min_{1 \leqslant i \leqslant k} |\lambda_i| \sqrt{r_i} := \Delta_R.$$

In particular, if Δ_R is sufficiently large, distance-based clustering can effectively be adapted to identify the column index sets of the submatrices.

Let $\widetilde{U}\widetilde{D}\widetilde{V}^{\mathrm{T}}$ be the SVD of $X = M + E$ and consider $\widetilde{U}_k^{\mathrm{T}} X$. The main objective is to show that, with high probability,

$$\max_{1 \leqslant j \leqslant n} \|(\widetilde{U}_k^{\mathrm{T}} X)_j - (U^{\mathrm{T}} M)_j\| < \frac{1}{5}\Delta_R.$$

Achieving this allows us to employ a standard clustering approach, such as k-means, based on distance to classify the columns of $\widetilde{U}_k^{\mathrm{T}} X$ and thus recover the column index subsets $\{C_i\}_{i=0}^k$. Similarly, to identify the row index subsets $\{R_i\}_{i=0}^k$, we utilize the parameter

$$\Delta_C := \min_{1 \leqslant i \leqslant k} |\lambda_i| \sqrt{c_i}$$

and apply k-means clustering to the rows of $X\widetilde{V}_r$. We propose the following spectral algorithm:

Algorithm 2 Spectral algorithm for submatrix localization

Input: data matrix $X \in \mathbb{R}^{m \times n}$ and submatrix number k.
Output: column index subsets $\{\widetilde{C}_i\}_{i=0}^k$ and row index subsets $\{\widetilde{R}_i\}_{i=0}^k$.
Step 1. Perform SVD on X and denote $\widetilde{U}_k \in \mathbb{R}^{m \times k}$ and $\widetilde{V}_k \in \mathbb{R}^{n \times k}$ the singular vector matrices composed of the leading k left and right singular vectors of X respectively.
Step 2. Perform $(k+1)$-means clustering on the columns of $\widetilde{U}_k^{\mathrm{T}} X$. Output the column index subsets $\{\widetilde{C}_i\}_{i=0}^k$.
Step 3. Perform $(k+1)$-means clustering on the rows of $X\widetilde{V}_r$. Output the row index subsets $\{\widetilde{R}_i\}_{i=0}^k$.

Define

$$\sigma_{\min} := \min_{1 \leqslant i \leqslant k} |\lambda_i| \sqrt{r_i c_i}, \quad r_{\min} := \min_{0 \leqslant i \leqslant k} r_i, \quad c_{\min} := \min_{0 \leqslant i \leqslant k} c_i.$$

Theorem 19. *Consider the submatrix localization model* (20) *with* k *submatrices. Let* $L > 0$ *and assume* $(\sqrt{n} + \sqrt{p})^2 \geqslant 32(L+7)\log(n+p) + 64(\log 9)k$. *Given that*

$$\Delta_R \geqslant \max\left\{ \frac{40(\sqrt{m} + \sqrt{n})}{\sqrt{r_{\min}}}, 1800k\sqrt{(L+7)\log(m+n)} \right\},$$
$$\Delta_C \geqslant \max\left\{ \frac{40(\sqrt{m} + \sqrt{n})}{\sqrt{c_{\min}}}, 1800k\sqrt{(L+7)\log(m+n)} \right\},$$
$$\sigma_{\min} \geqslant 40(\sqrt{m} + \sqrt{n}) + 3.8 \times 10^4 k \sqrt{2(\log 9)k + (L+7)\log(m+n)},$$

Algorithm 2 succeeds in finding $\widetilde{R}_i = R_i$ *and* $\widetilde{C}_i = C_{\pi(i)}$, $0 \leqslant i \leqslant k$ *for a bijection* $\pi : [k+1] \to [k+1]$ *with probability at least* $1 - 40(m+n)^{-L}$.

The proof of Theorem 19 parallels that of Theorem 18, and therefore we omit the details.

Previous research on the model (20) of multiple submatrix localization includes notable contributions such as those found in [28,37,39]. Chen and Xu [37] examine this problem across different regimes, each corresponding to unique statistical and computational complexities. They focus on scenarios where all k submatrices are identically sized at $K_R \times K_C$ and share a common positive value $\lambda_i = \lambda$. Their analysis of the Maximum Likelihood Estimator (MLE), a convexified version of

MLE, and a simple thresholding algorithm address the challenges specific to hard, easy, and simple regimes, respectively. In the work of Dadon, Huleihel and Bendory [39], the primary objective is to explore the computational and statistical limits associated with the detection and reconstruction of hidden submatrices. Under the same setting as [37] in the context of the multiple submatrix recovery problem, the authors introduce a MLE alongside an alternative peeling estimator and investigate the performance of these estimators.

Our Algorithm 2 is identical to Algorithm 3 presented in Cai, Liang and Rakhlin's paper [28]. The assumptions laid out in [28] include $r_i \asymp K_R, c_i \asymp K_C, \lambda_i \asymp \lambda$ for all $1 \leqslant i \leqslant k$ and $\min\{K_R, K_C\} \gtrsim \max\{\sqrt{m}, \sqrt{n}\}$. Given that

$$\lambda \gtrsim \frac{\sqrt{k}}{\min\{\sqrt{K_R}, \sqrt{K_C}\}} + \max\left\{\sqrt{\frac{\log m}{K_C}}, \sqrt{\frac{\log n}{K_R}}\right\} + \frac{\sqrt{m}+\sqrt{n}}{\sqrt{K_R K_C}}, \tag{21}$$

the authors of [28] demonstrate that Algorithm 2 successfully recovers the true submatrix row and column index sets with probability at least $1 - m^{-c} - n^{-c} - 2\exp(-c(m+n))$. The entries of the noise matrix E in [28] are assumed to be i.i.d zero-mean sub-Gaussian random variables.

While our method does not require that all row or column index sets have the same order of sizes, in the special case where $r_i \asymp K_R$, $c_i \asymp K_C$, and $\lambda_i \asymp \lambda$ for all $1 \leqslant i \leqslant k$, and furthermore $r_0 \gtrsim K_R$ and $c_0 \gtrsim K_C$, our analysis indicates that if

$$\lambda \gtrsim \frac{k\sqrt{\log(m+n)}}{\min\{\sqrt{K_R}, \sqrt{K_C}\}} + \frac{\sqrt{m}+\sqrt{n}}{\min\{K_R, K_C\}}, \tag{22}$$

then Algorithm 2 successfully recovers the submatrix index subsets with probability at least $1-(m+n)^{-c}$. It should be emphasized that the condition in (22) is more stringent than that in (21), a difference that becomes particularly pronounced in cases where K_R and K_C are highly unbalanced. An interesting direction for future research would be to improve our perturbation bounds to accommodate cases with unbalanced matrix dimensions.

5. Preliminary, basic tools and proof overview

5.1. **Matrix norms.** Consider an $N \times n$ matrix $A = (a_{ij})$ with singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_{\min\{N,n\}} \geqslant 0$. Let $|||A|||$ be a norm of A of certain interest.

The first type of matrix norms are the unitarily invariant norms. The norm $|||\cdot|||$ on $\mathbb{R}^{N\times n}$ is said to be *unitarily invariant* if $|||A||| = |||UAV|||$ for all orthogonal matrices $U \in \mathbb{R}^{N\times N}$ and $V \in \mathbb{R}^{n\times n}$. There is an intimate connection between the unitarily invariant norms and the singular values of matrices via the symmetric gauge functions (see [21, Section IV]).

Definition 20 (Symmetric gauge function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is a *symmetric gauge function* if

(i) f is a norm,
(ii) $f(Px) = f(x)$ for all $x \in \mathbb{R}^n$ and $P \in S_n$ (the set of permutation matrices),
(iii) $f(\epsilon_1 x_1, \cdots, \epsilon_n x_n) = f(x_1, \cdots, x_n)$ if $\epsilon_j = \pm 1$.

We say the symmetric gauge function f is *normalized* if $f(1, 0, \cdots, 0) = 1$.

Theorem 21 (Theorem IV.2.2 from [21])**.** *A norm $|||\cdot|||$ on $\mathbb{R}^{N\times n}$ is unitarily invariant if and only if $|||A||| = f(\sigma_1, \ldots, \sigma_{\min\{N,n\}})$ for all $A \in \mathbb{R}^{N\times n}$ for some symmetric gauge function f, where $\sigma_1, \ldots, \sigma_{\min\{N,n\}}$ are the singular values of A.*

If $\|\cdot\|$ is a unitarily invariant norm on $\mathbb{R}^{N\times n}$ and f is its associated symmetric gauge function, then for $k, s \leqslant \min\{N, n\}$, a unitarily invariant norm on $\mathbb{R}^{k\times s}$ can be defined by $|||A||| = f(\sigma_1, \cdots, \sigma_{\min\{k,s\}}, 0, \cdots, 0)$, where σ_i's are the singular values of $A \in \mathbb{R}^{k\times s}$. As a result, a family of matrix norms can be defined based on f that can be applied to matrices of varying dimensions. As such, we will not explicitly mention the dimensions of the unitarily invariant norm $|||\cdot|||$.

Moreover, a unitarily invariant norm $|||\cdot|||$ is said to be *normalized* if its associated symmetric gauge function f is normalized. Consequently, a normalized unitarily invariant norm always satisfies $|||\operatorname{diag}(1, 0, \cdots, 0)||| = 1$.

Another characterization of the unitarily invariant norm is given by the symmetric property.

Theorem 22 (Proposition IV.2.4 from [21]). *A norm $|||\cdot|||$ on $\mathbb{R}^{N\times n}$ is unitarily invariant if and only if the norm is symmetric, that is,*

$$|||ABC||| \leqslant \|A\| \cdot |||B||| \cdot \|C\|$$

for every $A \in \mathbb{R}^{N\times N}, B \in \mathbb{R}^{N\times n}$ and $C \in \mathbb{R}^{n\times n}$.

A wide range of matrix norms that are commonly used are part of the class of unitarily invariant norms. For instance, for $p \in [1, \infty]$, the Schatten p-norm of A is defined by

$$\|A\|_p = \left(\sum_{i=1}^{\min\{N,n\}} \sigma_i^p \right)^{1/p}.$$

In particular, the case $p = 2$ yields the Frobenius norm $\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$. The case $p = \infty$ yields the operator norm $\|A\| = \sigma_1$. The case $p = 1$ yields the nuclear (or trace) norm

$$\|A\|_* = \|A\|_1 = \sum_{i=1}^{\min\{N,n\}} \sigma_i = \operatorname{tr}\left(\sqrt{AA^{\mathrm{T}}}\right).$$

Note that

$$\|A\|_p^2 = \|A^{\mathrm{T}}A\|_{p/2} \quad \text{for } p \geqslant 2. \tag{23}$$

Another class of unitarily invariant norms is the Ky Fan k-norm

$$\|A\|_{(k)} = \sum_{i=1}^{k} \sigma_i, \quad 1 \leqslant k \leqslant \min\{N, n\}.$$

Hence, $\|A\|_{(1)} = \|A\|$ and $\|A\|_{(\min\{N,n\})} = \|A\|_*$. A highly significant result known as the Fan dominance theorem is connected to the Ky Fan norm:

Theorem 23 (Theorem IV.2.2 from [21]). *Let A, B be two $n \times n$ matrices. If*

$$\|A\|_{(k)} \leqslant \|B\|_{(k)} \quad \text{for } k = 1, 2, \cdots, n,$$

then $|||A||| \leqslant |||B|||$ for all unitarily invariant norms.

If $|||\cdot|||$ is also normalized, then a direct implication of Theorem 23 is that

$$\begin{aligned} \|A\| &\leqslant |||A||| \leqslant \|A\|_*, \\ \sigma_{\min}(A)|||B||| &\leqslant |||AB||| \leqslant \|A\||||B|||, \\ \sigma_{\min}(A)|||B||| &\leqslant |||BA||| \leqslant \|A\||||B|||. \end{aligned} \tag{24}$$

It also follows from Theorem 22 and (24) that $|||AB||| \leqslant |||A||||||B|||$.

We also consider the following norms, which do not belong to the class of unitarily invariant norms. Denote $A_{i,\cdot}$'s the rows of $A \in \mathbb{R}^{N\times n}$. The $\ell_{2,\infty}$ norm of A is

$$\|A\|_{2,\infty} = \max_i \|A_{i,\cdot}\|_2 = \max_{1\leqslant i\leqslant N} \|e_i^{\mathrm{T}} A\|.$$

Finally, denote $\|A\|_{\max} = \max_{i,j} |a_{ij}|$. Note that $\|\cdot\|_{\max}$ is not sub-multiplicative.

5.2. **Distance between subspaces.** Using the angles between subspaces and using the orthogonal projections to describe their separation are two popular approaches for measuring the distance between subspaces. These two methods are essentially equivalent when it comes to any unitarily invariant norm $|||\cdot|||$. We start with some basic notions.

If U and V are two subspaces of the same dimension r, then one could define the principal angles $0 \leqslant \theta_1 \leqslant \cdots \leqslant \theta_r \leqslant \pi/2$ between them recursively:

$$\cos(\theta_i) = \max_{u\in U, v\in V} u^T v = u_i^T v_i, \qquad \|u\| = \|v\| = 1$$

subject to the constraint

$$u_i^T u_l = 0, \quad v_i^T v_l = 0 \quad \text{for } l = 1, \ldots, i-1.$$

Denote $\angle(U,V) := \operatorname{diag}(\theta_1, \cdots, \theta_r)$. Further, let

$$\sin\angle(U,V) := \operatorname{diag}(\sin\theta_1, \cdots, \sin\theta_r),$$
$$\cos\angle(U,V) := \operatorname{diag}(\cos\theta_1, \cdots, \cos\theta_r).$$

With abuse of notation, we also let $U = (u_1, \cdots, u_r)$ and $V = (v_1, \cdots, v_r)$ be matrices of size $n \times r$ whose columns are orthonormal bases of subspaces U and V respectively. Then $P_U = UU^T$ (resp. $P_V = VV^T$) is the orthogonal projection matrix onto the subspace U (resp. V). For a subspace W, denote its complement by $W^\perp$.

The following facts are collected from [21, Exercises VII. 1. 9 – 1.11].

Proposition 24. *Let $U, V, P_U, P_V, \sin\angle(U,V), \cos\angle(U,V)$ be as above.*

(i) The nonzero singular values of $P_U P_V$ are the same as the nonzero singular values of $U^T V$.
(ii) The singular values of $P_U P_V$ are $\cos\theta_1, \cdots, \cos\theta_r$. The nonzero singular values of $P_{U^\perp} P_V$ are the nonzero values of $\sin\theta_1, \cdots, \sin\theta_r$.
(iii) The nonzero singular values of $P_U - P_V$ are the nonzero singular values $P_{U^\perp} P_V$, each counted twice; i.e., these are the nonzero numbers in

$$\sin\theta_1, \sin\theta_1, \sin\theta_2, \sin\theta_2, \cdots, \sin\theta_r, \sin\theta_r.$$

For any unitarily invariant norm $|||\cdot|||$, by Proposition 24, we observe

$$|||\sin\angle(U,V)||| = |||P_{U^\perp} P_V||| = |||P_{V^\perp} P_U||| \tag{25}$$

and

$$|||P_U - P_V||| = |||P_{U^\perp} P_V \oplus P_U P_{V^\perp}|||.$$

This suggests the (near) equivalence of $|||\sin\angle(U,V)|||$ and $|||P_U - P_V|||$. For instance, for the Schatten p-norm, we have

$$\|P_U - P_V\|_p = 2^{\frac{1}{p}} \|P_{U^\perp} P_V\|_p = 2^{\frac{1}{p}} \|\sin\angle(U,V)\|_p.$$

For the Ky Fan k-norm, denote $\|A\|_{(0)} = 0$. Then

$$\|P_U - P_V\|_{(k)} = \begin{cases} \|P_{U^\perp} P_V\|_{(\frac{k-1}{2})} + \|P_{U^\perp} P_V\|_{(\frac{k+1}{2})}, & \text{if } k \text{ is odd;} \\ 2\|P_{U^\perp} P_V\|_{(\frac{k}{2})}, & \text{if } k \text{ is even.} \end{cases}$$

Another method that is commonly used to quantify the distance between U and V is to use

$$\min_{O \in \mathbb{O}^{r\times r}} |||UO - V|||.$$

It is shown in [35, Lemma 2.6] that the above distance is (near) equivalent to the $|||P_U - P_V|||$ for the Frobenius norm and the operator norm. In fact, we can demonstrate that these distances are (near) equivalent when considering the Schatten-p norm for any $p \in [2, \infty]$:

$$\|\sin\angle(U,V)\|_p \leqslant \min_{O \in \mathbb{O}^{r\times r}} \|UO - V\|_p \leqslant \sqrt{2}\|\sin\angle(U,V)\|_p. \tag{26}$$

The proof of (26) is given in Appendix A.1.

More generally, for any unitarily invariant norm $|||\cdot|||$ on $\mathbb{R}^{N\times n}$, we have

$$\min_{O \in \mathbb{O}^{r\times r}} |||UO - V||| \leqslant \sqrt{2}|||\sin\angle(U,V)|||. \tag{27}$$

The proof of (27) is given in Appendix A.2.

In certain applications, the primary focus is to compare the matrices $U = (u_1, \cdots, u_r)$ and $V = (v_1, \cdots, v_r)$ of size $n \times r$ with respect to specific directions. According to Proposition 24, the SVD of $U^{\mathrm{T}}V$ is given by

$$U^{\mathrm{T}}V = O_1 \cos\angle(U,V) O_2^{\mathrm{T}}, \tag{28}$$

where $O_1, O_2 \in \mathbb{O}^{r\times r}$. Denote $O := O_1 O_2^{\mathrm{T}} \in \mathbb{O}^{r\times r}$. We highlight the following deterministic result, the proof of which can be found in Appendix A.3.

Proposition 25. *Let x be any unit vector in $\mathbb{R}^n$ and y be any unit vector in $\mathbb{R}^r$. We have*

$$\|x^{\mathrm{T}}(V - UO)\| \leqslant \|x^{\mathrm{T}}(V - P_U V)\| + \|x^{\mathrm{T}}U\|\|\sin\angle(U,V)\|^2$$

and

$$\left|x^{\mathrm{T}}(V - UO)y\right| \leqslant \left|x^{\mathrm{T}}(V - P_U V)y\right| + \|x^{\mathrm{T}}U\|\|\sin\angle(U,V)\|^2.$$

In particular,

$$\|V - UO\|_{2,\infty} \leqslant \|V - P_U V\|_{2,\infty} + \|U\|_{2,\infty}\|\sin\angle(U,V)\|^2.$$

Finally, it can be verified from the definition that for any orthogonal matrix O,

$$\|V - UO\|_{2,\infty} = \|VO^{\mathrm{T}} - U\|_{2,\infty}.$$

5.3. **Basic tools.** This section presents the basic tools necessary for the proofs of our main results, many of which build upon the previous work by O'Rourke, Vu and the author [64].

We start with the standard linearization of the perturbation model (1). Consider the $(N+n)\times(N+n)$ matrices

$$\mathcal{A} := \begin{pmatrix} 0 & A \\ A^{\mathrm{T}} & 0 \end{pmatrix} \quad \text{and} \quad \mathcal{E} := \begin{pmatrix} 0 & E \\ E^{\mathrm{T}} & 0 \end{pmatrix}$$

in block form. Define

$$\widetilde{\mathcal{A}} := \mathcal{A} + \mathcal{E}.$$

The non-zero eigenvalues of $\mathcal{A}$ are given by $\lambda_j = \sigma_j$ and $\lambda_{j+r} = -\sigma_j$ for $1 \leqslant j \leqslant r$. Then $\mathbf{u}_j := \frac{1}{\sqrt{2}}(u_j^{\mathrm{T}}, v_j^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{u}_{j+r} := \frac{1}{\sqrt{2}}(u_j^{\mathrm{T}}, -v_j^{\mathrm{T}})^{\mathrm{T}}$ for $1 \leqslant j \leqslant r$ are their corresponding orthonormal eigenvectors. The spectral decomposition of $\mathcal{A}$ is

$$\mathcal{A} = \mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}, \tag{29}$$

where $\mathcal{U} := (\mathbf{u}_1, \ldots, \mathbf{u}_{2r})$ and $\mathcal{D} := \mathrm{diag}(\lambda_1, \cdots, \lambda_{2r})$. It follows that $\mathcal{U}^{\mathrm{T}}\mathcal{U} = I_{2r}$. Similarly, the non-zero eigenvalues of $\widetilde{\mathcal{A}}$ are denoted by $\widetilde{\lambda}_j = \widetilde{\sigma}_j$ and $\widetilde{\lambda}_{j+\min\{N,n\}} = -\widetilde{\sigma}_j$ for $1 \leqslant j \leqslant \min\{N, n\}$. The eigenvector corresponding to $\widetilde{\lambda}_j$ is dentoed by $\tilde{\mathbf{u}}_j$ and is formed by the right and left singular vectors of $\widetilde{A}$.

For $z \in \mathbb{C}$ with $|z| > \|\mathcal{E}\|$, we define the resolvent of $\mathcal{E}$ as

$$G(z) := (zI - \mathcal{E})^{-1}.$$

Often we will drop the identity matrix and simply write $(z - \mathcal{E})^{-1}$ for this matrix. We use $G_{ij}(z)$ to denote the (i, j)-entry of $G(z)$.

The key observation is that $G(z)$ can be approximated by a diagonal matrix. Consider a random diagonal matrix

$$\Phi(z) := \begin{pmatrix} \frac{1}{\phi_1(z)} I_N & 0 \\ 0 & \frac{1}{\phi_2(z)} I_n \end{pmatrix}, \tag{30}$$

where

$$\phi_1(z) := z - \sum_{t \in [\![N+1, N+n]\!]} G_{tt}(z), \qquad \phi_2(z) := z - \sum_{s \in [\![1, N]\!]} G_{ss}(z). \tag{31}$$

By setting

$$\mathcal{I}^{\mathrm{u}} := \begin{pmatrix} I_N & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathcal{I}^{\mathrm{d}} := \begin{pmatrix} 0 & 0 \\ 0 & I_n \end{pmatrix},$$

one can rewrite (31) as

$$\phi_1(z) = z - \operatorname{tr} \mathcal{I}^{\mathrm{d}} G(z), \qquad \phi_2(z) = z - \operatorname{tr} \mathcal{I}^{\mathrm{u}} G(z). \tag{32}$$

By elementary linear algebra, it can be verified that

$$\phi_1(z) = \phi_2(z) - \frac{1}{z}(n - N). \tag{33}$$

From the definition of $\mathcal{U}$ in (29), it is easy to verify that

$$\mathcal{U}^{\mathrm{T}} \Phi(z) \mathcal{U} = \begin{pmatrix} \alpha(z) I_r & \beta(z) I_r \\ \beta(z) I_r & \alpha(z) I_r \end{pmatrix}, \tag{34}$$

where we denote

$$\alpha(z) := \frac{1}{2}\left(\frac{1}{\phi_1(z)} + \frac{1}{\phi_2(z)}\right) \quad \text{and} \quad \beta(z) := \frac{1}{2}\left(\frac{1}{\phi_1(z)} - \frac{1}{\phi_2(z)}\right)$$

for notational brevity. It follows that

$$\|\mathcal{U}^{\mathrm{T}} \Phi(z) \mathcal{U}\| = \max\left\{\frac{1}{|\phi_1(z)|}, \frac{1}{|\phi_2(z)|}\right\}. \tag{35}$$

The next lemma offers bounds for the resolvent and the functions $\phi_i(z)$'s. The proof follows similarly to that of Lemma 16 in [64] and is omitted for brevity.

Lemma 26. *On the event where* $\|E\| \leqslant 2(\sqrt{N}+\sqrt{n})$,

$$\|G(z)\| \leqslant \frac{b}{b-1}\frac{1}{|z|}$$

and

$$\left(1-\frac{1}{4b(b-1)}\right)|z| \leqslant |\phi_i(z)| \leqslant \left(1+\frac{1}{4b(b-1)}\right)|z| \quad \text{for } i=1,2 \tag{36}$$

for any $z \in \mathbb{C}$ *with* $|z| \geqslant 2b(\sqrt{N}+\sqrt{n})$ *and for any* $k \in [\![1, N+n]\!]$.

Consequently, by Lemma 26, we obtain

$$\max\{|\operatorname{tr} G(z)|, |\operatorname{tr} \mathcal{I}^u G(z)|, |\operatorname{tr} \mathcal{I}^d G(z)|\} \leqslant (N+n)\|G(z)\| \leqslant \frac{b}{b-1}\frac{N+n}{|z|}. \tag{37}$$

The subsequent isotropic local law is derived using a proof similar to that of [64, Lemma 27]. For completeness, we briefly describe the proof in Appendix B.1.

Lemma 27. *Let* $K > 0$ *and assume* $(\sqrt{N}+\sqrt{n})^2 \geqslant 32(K+1)\log(N+n)$. *For any unit vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N+n}$ *and for any* $z \in \mathbb{C}$ *with* $|z| \geqslant 2b(\sqrt{N}+\sqrt{n})$,

$$\left|\mathbf{x}^{\mathrm{T}}\left(G(z)-\Phi(z)\right)\mathbf{y}\right| \leqslant \frac{5b^2}{(b-1)^2}\frac{\sqrt{(K+1)\log(N+n)}}{|z|^2} \tag{38}$$

with probability at least $1-9(N+n)^{-(K+1)}$.

Recall

$$\eta = \frac{11b^2}{(b-1)^2}\sqrt{(K+7)\log(N+n)+(\log 9)r}.$$

Denote

$$\mathsf{D} := \{z \in \mathbb{C} : 2b(\sqrt{N}+\sqrt{n}) \leqslant |z| \leqslant 2n^3\}.$$

Using the previous lemma and a standard ε-net argument, we obtain the following result that is analogous to [64, Lemma 9]:

Lemma 28. *Under the assumptions of Theorem 5, one has*

$$\max_{z\in\mathsf{D}} |z|^2 \left\|\mathcal{U}^{\mathrm{T}}\left(G(z)-\Phi(z)\right)\mathcal{U}\right\| \leqslant \eta$$

with probability at least $1-9(N+n)^{-K}$.

Lemma 28 improves the rank r-dependence in the bound of [64, Lemma 9]. The proof of Lemma 28 is included in Appendix B.2. For the case $2b(\sqrt{N}+\sqrt{n}) > 2n^3$ where D is empty, $G(z)$ can be approximately be even simpler matrices (see Lemma 30 below).

The following result on the location of perturbed singular values is obtained using Lemma 28. Consider the random function

$$\varphi(z) := \phi_1(z)\phi_2(z), \tag{39}$$

where $\phi_1(z)$ and $\phi_2(z)$ are defined in (31). Define the auxiliary functions for $b \geqslant 2$:

$$\xi(b) := 1+\frac{1}{2(b-1)^2} \quad \text{and} \quad \chi(b) := 1+\frac{1}{4b(b-1)}.$$

Define a set in the complex plane in the neighborhood of any $\sigma \in \mathbb{R}$ by

$$S_\sigma := \{w \in \mathbb{C} : |\operatorname{Im}(w)| \leqslant 20\chi(b)\eta r, \sigma - 20\chi(b)\eta r \leqslant \operatorname{Re}(w) \leqslant \chi(b)\sigma + 20\chi(b)\eta r\}. \tag{40}$$

Theorem 29 (Singular value locations)**.** *Let A and E be $N \times n$ real matrices, where A is deterministic and the entries of E are i.i.d. standard Gaussian random variables. Assume A has rank $r \geqslant 1$. Let $K > 0$ and $b \geqslant 2$. Denote $\eta := \frac{11b^2}{(b-1)^2}\sqrt{2(\log 9)r + (K+7)\log(N+n)}$. Assume $(\sqrt{N}+\sqrt{n})^2 \geqslant 32(K+7)\log(N+n) + 64(\log 9)r$. Let $1 \leqslant r_0 \leqslant r$ such that $\sigma_{r_0} \geqslant 2b(\sqrt{N}+\sqrt{n}) + 80b\eta r$ and $\delta_{r_0} \geqslant 75\chi(b)\eta r$. Consider any $1 \leqslant k \leqslant s \leqslant r_0$ satisfying $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$. For any $j \in [\![k, s]\!]$, there exists $j_0 \in [\![k, s]\!]$ such that $\widetilde{\sigma}_j \in S_{\sigma_{j_0}}$, and*

$$|\varphi(\widetilde{\sigma}_j) - \sigma_{j_0}^2| \leqslant 20\xi(b)\chi(b)\eta r\,(\widetilde{\sigma}_j + \chi(b)\sigma_{j_0}) \tag{41}$$

with probability at least $1 - 10(N+n)^{-K}$.

Theorem 29 shares similarities with Theorem 12 in [64], albeit with a relaxed requirement regarding the separation of distinct singular values of A. The derivation of Theorem 29 is largely based on the proof of Theorem 12 in [64]. To ensure our paper is self-contained, we have provided the proof in the Appendix D.

The next result suggests that when $|z|$ is large, the resolvent $G(z)$ can be approximated by simpler matrices. The proof is analogous to that of Lemma 17 in [64], and is omitted.

Lemma 30. *On the event where $\|E\| \leqslant 2(\sqrt{N}+\sqrt{n})$,*

$$\left\| G(z) - \frac{1}{z} I_{N+n} \right\| \leqslant \frac{b}{b-1} \frac{\|E\|}{|z|^2}$$

and

$$\left\| G(z) - \frac{1}{z} I_{N+n} - \frac{\mathcal{E}}{z^2} \right\| \leqslant \frac{b}{b-1} \frac{\|E\|^2}{|z|^3}$$

for any $z \in \mathbb{C}$ with $|z| \geqslant 2b(\sqrt{N}+\sqrt{n})$.

Lemma 31 (Lemma 13 from [64])**.** *Under the assumptions of Theorem 5,*

$$\max_{l \in [\![1, r_0]\!] : \sigma_l > \frac{1}{2} n^2} |\widetilde{\sigma}_l - \sigma_l| \leqslant \eta r$$

with probability at least $1 - (N+n)^{-1.5r^2(K+4)}$.

The next result provides a non-asymptotic bound on the operator bound of $\|E\|$.

Lemma 32 (Spectral norm bound; see (2.3) from [67])**.** *Let E be an $N \times n$ matrix whose entries are independent standard Gaussian random variables. Then*

$$\|E\| \leqslant 2(\sqrt{N}+\sqrt{n})$$

with probability at least $1 - 2e^{-(\sqrt{N}+\sqrt{n})^2/2}$.

5.4. **Proof overview.** In this section, we outline our proof strategy, which leverages techniques from random matrix theory, particularly the resolvent method, to analyze the eigenvalues and eigenvectors of the symmetric matrices $\mathcal{A}$ and $\widetilde{\mathcal{A}} = \mathcal{A} + \mathcal{E}$ detailed in Section 5.3.

At the heart of our analysis is the isotropic local law (Lemma 27), which asserts that the resolvent $G(z) = (zI - \mathcal{E})^{-1}$ can be approximated by a simpler matrix $\Phi(z)$. This approximation streamlines complex calculations involving $G(z)$ and is a technique commonly used to study extreme eigenvalues and eigenvectors in random matrix theory, as seen in, for instance, [12, 13, 22, 51]. Our work diverges from these prior approaches by selecting $\Phi(z)$ as a random matrix derived from $G(z)$

itself, which better suits the finite sample context, compared to the deterministic approximations used in previous studies that rely on Stieltjes transforms in the asymptotic regime.

Building upon the isotropic local law, we determine the singular value locations of $\widetilde{A}$ in Theorem 29 and achieve the control of $\left\|\mathcal{U}^{\mathrm{T}}\left(G(z)-\Phi(z)\right)\mathcal{U}\right\|$ as given in Lemma 28. These instruments have been previously explored in the previous work by O'Rourke, Vu and the author [64]. In this paper, we refine these estimations and ease the conditions in our earlier work [64]. Furthermore, we deploy these refined tools to derive a variety of new perturbation bounds.

To explain the idea of deriving perturbation bounds, we simply focus on the largest eigenvector $\widetilde{\mathbf{u}}_1$ of $\widetilde{\mathcal{A}}$. In practice, it is necessary to consider both $\widetilde{\mathbf{u}}_1$ and $\widetilde{\mathbf{u}}_{r+1}$, as they jointly involve the largest singular vectors u_1 and v_1. Nonetheless, for the sake of clarity, we will momentarily disregard the terms related to $\mathbf{u}_{r+1}$ and $\widetilde{\mathbf{u}}_{r+1}$ to illustrate the main ideas.

We start with the decomposition

$$\widetilde{\mathbf{u}}_1 = (\mathbf{u}_1\mathbf{u}_1^{\mathrm{T}})\widetilde{\mathbf{u}}_1 + \mathcal{P}_1\widetilde{\mathbf{u}}_1 + \mathcal{Q}\widetilde{\mathbf{u}}_1, \tag{42}$$

where $\mathcal{P}_1 = \mathcal{U}_1\mathcal{U}_1^{\mathrm{T}}$ and $\mathcal{U}_1$ is the matrix of eigenvectors of $\mathcal{A}$ excluding $\mathbf{u}_1$. Meanwhile, $\mathcal{Q}$ is the orthogonal projection matrix onto the null space of $\mathcal{A}$. The challenge in establishing perturbation bounds for $\widetilde{\mathbf{u}}_1$ lies in quantifying the latter two terms on the right-hand side of (42).

First, for the ℓ_2 analysis, we aim to bound $\sin\angle(\mathbf{u}_1, \widetilde{\mathbf{u}}_1)$. By taking the Frobenius norm on both sides of (42), we obtain

$$1 = \cos^2\angle(\mathbf{u}_1, \widetilde{\mathbf{u}}_1) + \|\mathcal{P}_1\widetilde{\mathbf{u}}_1\|^2 + \|\mathcal{Q}\widetilde{\mathbf{u}}_1\|^2.$$

Rearranging the terms yields

$$\sin^2\angle(\mathbf{u}_1, \widetilde{\mathbf{u}}_1) = \|\mathcal{P}_1\widetilde{\mathbf{u}}_1\|^2 + \|\mathcal{Q}\widetilde{\mathbf{u}}_1\|^2.$$

A straightforward linear algebra argument allows us to bound $\|\mathcal{Q}\widetilde{\mathbf{u}}_1\|$ by the noise-to-signal ratio $\|E\|/\sigma_1$. The main task is then to establish a bound for $\|\mathcal{P}_1\widetilde{\mathbf{u}}_1\| \leqslant \|\mathcal{U}_1^{\mathrm{T}}\widetilde{\mathbf{u}}_1\|$, which effectively comes down to bounding $|\mathbf{u}_j^{\mathrm{T}}\widetilde{\mathbf{u}}_1|$ for $j \neq 1$. We explain how to achieve this bound below.

From the equation $\widetilde{\mathcal{A}}\widetilde{\mathbf{u}}_1 = (\mathcal{A}+\mathcal{E})\widetilde{\mathbf{u}}_1 = \widetilde{\lambda}_1\widetilde{\mathbf{u}}_1$, we can express $\widetilde{\mathbf{u}}_1$ as

$$\widetilde{\mathbf{u}}_1 = (\widetilde{\lambda}_1 I - \mathcal{E})^{-1}\mathcal{A}\widetilde{\mathbf{u}}_1 = G(\widetilde{\lambda}_1)\mathcal{A}\widetilde{\mathbf{u}}_1$$

and further rewrite it as

$$\widetilde{\mathbf{u}}_1 = \Phi(\widetilde{\lambda}_1)\mathcal{A}\widetilde{\mathbf{u}}_1 + \left(G(\widetilde{\lambda}_1) - \Phi(\widetilde{\lambda}_1)\right)\mathcal{A}\widetilde{\mathbf{u}}_1.$$

Hence, for $j \neq 1$, we have

$$\mathbf{u}_j^{\mathrm{T}}\widetilde{\mathbf{u}}_1 = \mathbf{u}_j^{\mathrm{T}}\Phi(\widetilde{\lambda}_1)\mathcal{A}\widetilde{\mathbf{u}}_1 + \mathbf{u}_j^{\mathrm{T}}\left(G(\widetilde{\lambda}_1) - \Phi(\widetilde{\lambda}_1)\right)\mathcal{A}\widetilde{\mathbf{u}}_1 \tag{43}$$

Calculations similar to those in (34) indicate that the first term on the right-hand side of (43), $\mathbf{u}_j^{\mathrm{T}}\Phi(\widetilde{\lambda}_1)\mathcal{A}\widetilde{\mathbf{u}}_1$, is exactly $\lambda_j\alpha(\widetilde{\lambda}_1)\mathbf{u}_j^{\mathrm{T}}\widetilde{\mathbf{u}}_1$ (omitting the term containing $\mathbf{u}_{r+1}$). We continue from (43) to get

$$\left(1-\lambda_j\alpha(\widetilde{\lambda}_1)\right)\mathbf{u}_j^{\mathrm{T}}\widetilde{\mathbf{u}}_1 \approx \mathbf{u}_j^{\mathrm{T}}\left(G(\widetilde{\lambda}_1) - \Phi(\widetilde{\lambda}_1)\right)\mathcal{A}\widetilde{\mathbf{u}}_1.$$

To control $|\mathbf{u}_j^{\mathrm{T}}\widetilde{\mathbf{u}}_1|$, we apply Theorem 29 to analyze the coefficient $1-\lambda_j\alpha(\widetilde{\lambda}_1)$ that precedes it. Lemma 28 is applied to manage the term on the right-hand side.

Next, for the ℓ_∞ analysis, from (42), we obtain

$$\|\tilde{\mathbf{u}}_1 - (\mathbf{u}_1\mathbf{u}_1^{\mathrm{T}})\tilde{\mathbf{u}}_1\|_\infty \leqslant \|\mathcal{P}_1\tilde{\mathbf{u}}_1\|_\infty + \|\mathcal{Q}\tilde{\mathbf{u}}_1\|_\infty \leqslant \|\mathcal{U}\|_{2,\infty}\|\mathcal{U}_1^{\mathrm{T}}\tilde{\mathbf{u}}_1\| + \|\mathcal{Q}\tilde{\mathbf{u}}_1\|_\infty.$$

The bound for $\|\mathcal{U}_1^{\mathrm{T}}\tilde{\mathbf{u}}_1\|$ has already been established in the preceding ℓ_2 analysis. The second term, $\|\mathcal{Q}\tilde{\mathbf{u}}_1\|_\infty$, can be bounded by considering the fact

$$\mathcal{Q}\tilde{\mathbf{u}}_1 = \mathcal{Q}\left(G(\tilde{\lambda}_1) - \Phi(\tilde{\lambda}_1)\right)\mathcal{A}\tilde{\mathbf{u}}_1$$

and then applying Lemma 28.

These are the main ideas that we have incorporated in our proofs. Before concluding this section, we would like to highlight that the results presented in this paper can be extended to scenarios where the noise matrix E contains independent sub-Gaussian entries. This extension would rely on a lemma similar to Lemma 27, which can be demonstrated using the tools provided by random matrix theory. However, due to the technical complexities involved, we have chosen to reserve the discussion of this extension to sub-Gaussian cases for a forthcoming paper. It remains a highly interesting direction to further establish these perturbation bounds when the noise matrix E comprises heteroskedastic random variables. We believe that new tools and insights, extending beyond the scope of the methods presented in this paper, will be required to rigorously establish such extensions.

6. Proofs of Theorems 5, 13, 14 and 16

In the proofs below, we always work on the event where $\|E\| \leqslant 2(\sqrt{N}+\sqrt{n})$; Lemma 32 shows this event holds with probability at least $1 - 2e^{-(\sqrt{N}+\sqrt{n})^2/2} \geqslant 1 - 2(N+n)^{-16(K+7)}$ since $(\sqrt{N}+\sqrt{n})^2 \geqslant 32(K+7)\log(N+n)$ by assumption.

Denote

$$I := [\![k, s]\!] \cup [\![r+k, r+s]\!]$$

and

$$J := [\![1, 2r]\!]\backslash I = [\![1, k-1]\!] \cup [\![s+1, r]\!] \cup [\![r+1, r+k-1]\!] \cup [\![r+s+1, 2r]\!].$$

We first obtain an identity for the eigenvector $\tilde{\mathbf{u}}_i$. For each $i \in I$, by Weyl's inequality, $|\tilde{\lambda}_i| \geqslant \tilde{\sigma}_{r_0} \geqslant \sigma_{r_0} - \|E\| > \|E\| = \|\mathcal{E}\|$ by supposition on σ_{r_0} and thus $G(\tilde{\lambda}_i)$ and $\Phi(\tilde{\lambda}_i)$ are well-defined. As $(\mathcal{A}+\mathcal{E})\tilde{\mathbf{u}}_i = \tilde{\lambda}_i\tilde{\mathbf{u}}_i$, we solve for $\tilde{\mathbf{u}}_i$ to obtain

$$\tilde{\mathbf{u}}_i = (\tilde{\lambda}_i I - \mathcal{E})^{-1}\mathcal{A}\tilde{\mathbf{u}}_i = G(\tilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i.$$

In fact, we can approximate $G(\tilde{\lambda}_i)$ with a simpler random matrix, denoted as $\Pi(z)$ below. Depending on the magnitude of $\tilde{\lambda}_i$ under consideration (the specifics of which will become clear in subsequent context), we choose $\Pi(\tilde{\lambda}_i)$ to be either $\Phi(\tilde{\lambda}_i)$ if $|\tilde{\lambda}_i|$ is relatively small and $\frac{1}{\tilde{\lambda}_i}I_{N+n} + \frac{1}{\tilde{\lambda}_i^2}\mathcal{E}$ or even simply $\frac{1}{\tilde{\lambda}_i}I_{N+n}$ if $|\tilde{\lambda}_i|$ is sufficiently large. Furthermore, denote

$$\Xi(\tilde{\lambda}_i) := G(\tilde{\lambda}_i) - \Pi(\tilde{\lambda}_i). \tag{44}$$

Hence, we rewrite

$$\tilde{\mathbf{u}}_i = \Pi(\tilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i + \Xi(\tilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i \tag{45}$$

This decomposition of $\tilde{\mathbf{u}}_i$ is critical in facilitating the extraction of its desired property information. Note that Lemma 28 and Lemma 30 provides precise control on the size of $\Xi(\tilde{\lambda}_i)$.

For $J \subset [\![1, 2r]\!]$, we introduce the notation $\mathcal{U}_J$ to denote the $(N+n) \times |J|$ matrix formed from $\mathcal{U}$ by removing the columns containing $\mathbf{u}_i$ for $i \notin J$. Similarly, $\mathcal{D}_J$ will denote the $|J| \times |J|$ matrix formed from $\mathcal{D}$ by removing the rows and columns containing λ_i for $i \notin J$. Let $I := [\![1, 2r]\!] \backslash J$. In this way, we can decompose A as

$$\mathcal{A} = \mathcal{U}\mathcal{D}\mathcal{U}^T = \mathcal{U}_J \mathcal{D}_J \mathcal{U}_J^{\mathrm{T}} + \mathcal{U}_I \mathcal{D}_I \mathcal{U}_I^{\mathrm{T}}. \tag{46}$$

Let $\mathcal{P}_J$ be the orthogonal projection onto the subspace $\operatorname{Span}\{\mathbf{u}_k : k \in J\}$. Clearly, $\mathcal{P}_J = \mathcal{U}_J \mathcal{U}_J^{\mathrm{T}}$. If $J = [\![1, k]\!]$, we sometimes simply write $\mathcal{U}_k$ for $\mathcal{U}_J$ and $\mathcal{P}_k$ for $\mathcal{P}_J$. Analogous notations $\widetilde{\mathcal{U}}_J, \widetilde{P}_J, \widetilde{\mathcal{D}}_J$ are also defined for $\widetilde{\mathcal{A}}$. We also use $\mathcal{P}_{[\![2r+1, N+n]\!]} = \mathcal{U}_{[\![2r+1, N+n]\!]} \mathcal{U}_{[\![2r+1, N+n]\!]}^{\mathrm{T}}$ to denote the orthogonal projection onto the null space of $\mathcal{A}$.

Now we proceed to the proofs of the main results.

6.1. **Proof of Theorem 5.** From (25), we start by observing

$$\||\sin \angle(\mathcal{U}_I, \widetilde{\mathcal{U}}_I)\|| = \||\mathcal{P}_{I^c} \widetilde{\mathcal{P}}_I\|| \leqslant \||\mathcal{P}_{[\![2r+1, N+n]\!]} \widetilde{\mathcal{P}}_I\|| + \||\mathcal{P}_J \widetilde{\mathcal{P}}_I\||.$$

We bound the two terms $\||\mathcal{P}_{[\![2r+1, N+n]\!]} \widetilde{\mathcal{P}}_I\||$ and $\||\mathcal{P}_J \widetilde{\mathcal{P}}_I\||$ respectively.

Lemma 33. *With probability 1,*

$$\||\mathcal{P}_{[\![2r+1, N+n]\!]} \widetilde{\mathcal{P}}_I\|| \leqslant 2 \frac{\||\mathcal{P}_{[\![2r+1, N+n]\!]} \mathcal{E} \widetilde{\mathcal{P}}_I\||}{\sigma_s}. \tag{47}$$

Proof. By Proposition 24 (i) and Theorem 21,

$$\||\mathcal{P}_{[\![2r+1, N+n]\!]} \widetilde{\mathcal{P}}_I\|| = \||\mathcal{U}_{[\![2r+1, N+n]\!]}^{\mathrm{T}} \widetilde{\mathcal{U}}_I\||.$$

From the spectral decomposition of $\widetilde{\mathcal{A}}$, we have

$$(\mathcal{A} + \mathcal{E}) \widetilde{\mathcal{U}}_I = \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I.$$

Multiplying by $\mathcal{U}_{[\![2r+1, N+n]\!]}^{\mathrm{T}}$ on the left of the equation above, we further have

$$\mathcal{U}_{[\![2r+1, N+n]\!]}^{\mathrm{T}} \mathcal{E} \widetilde{\mathcal{U}}_I = \mathcal{U}_{[\![2r+1, N+n]\!]}^{T} \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I.$$

As $\sigma_{r_0} \geqslant b\|E\| \geqslant 2\|E\|$ by supposition, Weyl's inequality implies that

$$\widetilde{\sigma}_i \geqslant \sigma_i - \|E\| \geqslant \frac{1}{2} \sigma_i \tag{48}$$

for $k \leqslant i \leqslant s$. Hence, $\widetilde{\mathcal{D}}_I$ is invertible since $|\widetilde{\lambda}_i| = \widetilde{\sigma}_i > 0$ for $i \in [\![k, s]\!]$ and $|\widetilde{\lambda}_i| = \widetilde{\sigma}_{i-r} > 0$ for $i \in [\![r+k, r+s]\!]$. It follows from Theorem 22 that

$$\begin{aligned}
\||\mathcal{U}_{[\![2r+1, N+n]\!]}^{\mathrm{T}} \widetilde{\mathcal{U}}_I\|| &= \||\mathcal{U}_{[\![2r+1, N+n]\!]}^{\mathrm{T}} \mathcal{E} \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I^{-1}\|| \\
&\leqslant \||\mathcal{U}_{[\![2r+1, N+n]\!]}^{\mathrm{T}} \mathcal{E} \widetilde{\mathcal{U}}_I\|| \|\widetilde{\mathcal{D}}_I^{-1}\| \\
&= \frac{\||\mathcal{P}_{[\![2r+1, N+n]\!]} \mathcal{E} \widetilde{\mathcal{P}}_I\||}{\widetilde{\sigma}_s}.
\end{aligned}$$

The last equation above follows from the fact that for U with orthonormal columns, $U^{\mathrm{T}} B$ and $U U^{\mathrm{T}} B$ share the same singular values.

Thus by another application of (48), we get

$$\||\mathcal{P}_{[\![2r+1, N+n]\!]} \widetilde{\mathcal{P}}_I\|| \leqslant \frac{\||\mathcal{P}_{[\![2r+1, N+n]\!]} \mathcal{E} \widetilde{\mathcal{P}}_I\||}{\widetilde{\sigma}_s} \leqslant 2 \frac{\||\mathcal{P}_{[\![2r+1, N+n]\!]} \mathcal{E} \widetilde{\mathcal{P}}_I\||}{\sigma_s}. \tag{49}$$

as desired. □

It remains to bound $\|\mathcal{P}_J\widetilde{\mathcal{P}}_I\| = \|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\|$. We apply (24) to obtain

$$\begin{aligned}\|\mathcal{P}_J\widetilde{\mathcal{P}}_I\| = \|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\| &\leqslant \|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\|_* \\ &\leqslant \sqrt{\operatorname{rank}(\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I)}\cdot\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\|_F \\ &\leqslant 2\sqrt{\min\{s-k+1, r-s+k-1\}}\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\|_F \\ &= 2\sqrt{\min\{s-k+1, r-s+k-1\}}\sqrt{\sum_{i\in I}\|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\|^2}. \end{aligned} \tag{50}$$

In particular, for the operator norm, when $|J| \neq 0$, we simply have

$$\|\mathcal{P}_J\widetilde{\mathcal{P}}_I\| = \|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\| \leqslant \|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\|_F = \sqrt{\sum_{i\in I}\|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\|^2}. \tag{51}$$

It remains to bound $\|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\|$ for each $i \in I$. We have the following estimates

Lemma 34. *For every $i \in I$,*

$$\|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\| \leqslant 3\frac{(b+2)^2}{(b-1)^2}\frac{\eta}{\min\{\delta_{k-1},\delta_s\}} \tag{52}$$

with probability at least $1-20(N+n)^{-K}$.

The proof of Lemma 34 closely mirrors the strategy employed in Lemma 20 from [64]. For the sake of completeness, we have included the proof of Lemma 34 in Appendix C.

It follows from (50) and Lemma 34 that

$$\|\mathcal{P}_J\widetilde{\mathcal{P}}_I\| \leqslant 6\sqrt{2}\frac{(b+2)^2}{(b-1)^2}\sqrt{\min\{s-k+1, r-s+k-1\}}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}$$

with probability at least $1-20(N+n)^{-K}$.

Consequently, we arrive at

$$\begin{aligned}\|\sin\angle(\mathcal{U}_I,\widetilde{\mathcal{U}}_I)\| &= \|\mathcal{P}_{I^c}\widetilde{\mathcal{P}}_I\| \\ &\leqslant 6\sqrt{2}\frac{(b+2)^2}{(b-1)^2}\sqrt{\min\{s-k+1, r-s+k-1\}}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}} \\ &\quad + 2\frac{\|\mathcal{P}_{[\![2r+1,N+n]\!]}\mathcal{E}\widetilde{\mathcal{P}}_I\|}{\sigma_s}. \end{aligned} \tag{53}$$

Specially, the following bound holds for the operator norm:

$$\|\sin\angle(\mathcal{U}_I,\widetilde{\mathcal{U}}_I)\| \leqslant 3\sqrt{2}\frac{(b+2)^2}{(b-1)^2}\mathbf{1}_{\{|J|\neq 0\}}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}} + 2\frac{\|E\|}{\sigma_s}. \tag{54}$$

Let $\gamma_1,\cdots,\gamma_{2(s-k+1)}$ be the principle angles of the subspaces $\mathcal{U}_I,\widetilde{\mathcal{U}}_I$. Denote $\alpha_1,\cdots,\alpha_{s-k+1}$ (resp. $\beta_1,\cdots,\beta_{s-k+1}$) the principle angles of $U_{k,s},\widetilde{U}_{k,s}$ (resp. $V_{k,s},\widetilde{V}_{k,s}$). From the proof of [64, Proposition 8], we see that the singular values of $\mathcal{U}_I^{\mathrm{T}}\widetilde{\mathcal{U}}_I$, given by $\cos\gamma_1,\cdots,\cos\gamma_{2(s-k+1)}$, are exactly

$$\cos\alpha_1,\cdots,\cos\alpha_{s-k+1},\cos\beta_1,\cdots,\cos\beta_{s-k+1}.$$

Hence,

$$\|\sin\angle(\mathcal{U}_I,\widetilde{\mathcal{U}}_I)\| = \|\sin\angle(U_{k,s},\widetilde{U}_{k,s})\oplus\sin\angle(V_{k,s},\widetilde{V}_{k,s})\|.$$

Note that by the definitions of $\mathcal{E}, \widetilde{\mathcal{P}}_I$ and $\mathcal{P}_{[\![2r+1,N+n]\!]}$,

$$\mathcal{P}_{[\![2r+1,N+n]\!]}\mathcal{E}\widetilde{\mathcal{P}}_I = \begin{pmatrix} 0 & P_{U^\perp}EP_{\widetilde{V}_{k,s}} \\ P_{V^\perp}E^{\mathrm{T}}P_{\widetilde{U}_{k,s}} & 0 \end{pmatrix}.$$

Using the unitary equivalence, we find

$$|\!|\!|\mathcal{P}_{[\![2r+1,N+n]\!]}\mathcal{E}\widetilde{\mathcal{P}}_I|\!|\!| = |\!|\!|P_{U^\perp}EP_{\widetilde{V}_{k,s}} \oplus P_{V^\perp}E^{\mathrm{T}}P_{\widetilde{U}_{k,s}}|\!|\!|.$$

Hence, from (53), we conclude that

$$\begin{aligned} &|\!|\!|\sin\angle(U_{k,s},\widetilde{U}_{k,s}) \oplus \sin\angle(V_{k,s},\widetilde{V}_{k,s})|\!|\!| \\ &\leqslant 6\sqrt{2}\frac{(b+2)^2}{(b-1)^2}\sqrt{\min\{s-k+1, r-s+k-1\}}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}} \\ &\quad + 2\frac{|\!|\!|P_{U^\perp}EP_{\widetilde{V}_{k,s}} \oplus P_{V^\perp}E^{\mathrm{T}}P_{\widetilde{U}_{k,s}}|\!|\!|}{\sigma_s}. \end{aligned} \tag{55}$$

The conclusion of Theorem 5 follows immediately from the fact that

$$\max\{|\!|\!|\sin\angle(U_{k,s},\widetilde{U}_{k,s})|\!|\!|, |\!|\!|\sin\angle(V_{k,s},\widetilde{V}_{k,s})|\!|\!|\} \leqslant |\!|\!|\sin\angle(U_{k,s},\widetilde{U}_{k,s})\oplus\sin\angle(V_{k,s},\widetilde{V}_{k,s})|\!|\!|$$

by Theorem 23.

Specially, for the operator norm, from (54), we see that

$$\begin{aligned} &\max\{\|\sin\angle(U_{k,s},\widetilde{U}_{k,s})\|, \|\sin\angle(V_{k,s},\widetilde{V}_{k,s})\|\} \\ &\leqslant \|\sin\angle(U_{k,s},\widetilde{U}_{k,s}) \oplus \sin\angle(V_{k,s},\widetilde{V}_{k,s})\| = \|\sin\angle(\mathcal{U}_I,\widetilde{\mathcal{U}}_I)\| \\ &\leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\mathbf{1}_{\{s-k+1\neq r\}}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}} + 2\frac{\|E\|}{\sigma_s}. \end{aligned}$$

This completes the proof.

6.2. **Proof of Theorem 13.** We start with the decomposition (45):

$$\widetilde{\mathbf{u}}_i = \Pi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i + \Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i.$$

In the proof, we set

$$\Pi(\widetilde{\lambda}_i) := \begin{cases} \Phi(\widetilde{\lambda}_i), & \text{if } |\lambda_i| \leqslant n^2; \\ \frac{1}{\widetilde{\lambda}_i}I_{N+n}, & \text{if } |\lambda_i| > n^2, \end{cases} \tag{56}$$

and recall that

$$\Xi(\widetilde{\lambda}_i) = G(\widetilde{\lambda}_i) - \Pi(\widetilde{\lambda}_i).$$

Let $\mathcal{Q} = I - \mathcal{P}_r$ be the orthogonal projection matrix onto the null space of $\mathcal{A}$. It is elementary to verify that $\mathcal{P}_r\Pi(\widetilde{\lambda}_i)\mathcal{A} = \Pi(\widetilde{\lambda}_i)\mathcal{A}$ using the definitions of $\mathcal{U}$ and $\mathcal{P}_r = \mathcal{U}\mathcal{U}^{\mathrm{T}}$. Hence, continuing from (45), we can derive the following expression

$$\mathcal{Q}\widetilde{\mathbf{u}}_i = \mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i.$$

Furthermore, we obtain the decomposition

$$\begin{aligned} \widetilde{\mathbf{u}}_i &= \mathcal{P}_I\widetilde{\mathbf{u}}_i + \mathcal{P}_J\widetilde{\mathbf{u}}_i + \mathcal{Q}\widetilde{\mathbf{u}}_i \\ &= \mathcal{P}_I\widetilde{\mathbf{u}}_i + \mathcal{P}_J\widetilde{\mathbf{u}}_i + \mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i. \end{aligned}$$

It follows that

$$\widetilde{\mathcal{U}}_I - \mathcal{P}_I\widetilde{\mathcal{U}}_I = \mathcal{P}_J\widetilde{\mathcal{U}}_I + (\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i)_{i\in I}. \tag{57}$$

We aim to bound

$$\|\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\|_{2,\infty} = \max_{1\leqslant l\leqslant N} \|\mathsf{e}_l^{\mathrm{T}}(U_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s})\|.$$

where e_l's are the canonical vectors in $\mathbb{R}^N$. From the definition of

$$\mathcal{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} U & U \\ V & -V \end{pmatrix}$$

from Section 5.3, it is elementary to check that

$$\widetilde{\mathcal{U}}_I - P_I\widetilde{\mathcal{U}}_I = \frac{1}{\sqrt{2}}\begin{pmatrix} \widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s} & \widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s} \\ \widetilde{V}_{k,s} - P_{V_{k,s}}\widetilde{V}_{k,s} & -(\widetilde{V}_{k,s} - P_{V_{k,s}}\widetilde{V}_{k,s}) \end{pmatrix}.$$

Hence,

$$\|\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\|_{2,\infty} = \max_{1\leqslant l\leqslant N} \left\| e_l^{\mathrm{T}}\left(\widetilde{\mathcal{U}}_I - \mathcal{P}_I\widetilde{\mathcal{U}}_I\right)\right\|,$$

where e_l's are the canonical vectors in $\mathbb{R}^{N+n}$. Continuing from (57), we see

$$\begin{aligned}\|\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\|_{2,\infty} &= \max_{1\leqslant l\leqslant N} \|e_l^{\mathrm{T}}\left(\widetilde{\mathcal{U}}_I - \mathcal{P}_I\widetilde{\mathcal{U}}_I\right)\| \\ &\leqslant \max_{1\leqslant l\leqslant N}\|e_l^{\mathrm{T}}\mathcal{P}_J\widetilde{\mathcal{U}}_I\|\cdot \mathbf{1}_{\{|J|\neq 0\}} + \max_{1\leqslant l\leqslant N}\|e_l^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i)_{i\in I}\|.\end{aligned} \tag{58}$$

Provided that $|J| \neq 0$ or equivalently, $s-k+1 \neq r$, the first term on the right-hand side of (58) can be bounded by

$$\begin{aligned}\max_{1\leqslant l\leqslant N}\|e_l^{\mathrm{T}}\mathcal{P}_J\widetilde{\mathcal{U}}_I\| &= \max_{1\leqslant l\leqslant N}\|e_l^{\mathrm{T}}\mathcal{U}_J\cdot\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\| \\ &\leqslant \max_{1\leqslant l\leqslant N}\|e_l^{\mathrm{T}}\mathcal{U}_J\|\cdot\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\| \\ &\leqslant \|U\|_{2,\infty}\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\|_F = \|U\|_{2,\infty}\sqrt{\sum_{i\in I}\|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\|^2}.\end{aligned}$$

By Lemma 34, we further obtain

$$\max_{1\leqslant l\leqslant N}\|e_l^{\mathrm{T}}\mathcal{P}_J\widetilde{\mathcal{U}}_I\| \leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|U\|_{2,\infty}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}} \tag{59}$$

with probability at least $1-20(N+n)^{-K}$.

Next, we bound the second term on the right-hand side of (58):

$$\max_{1\leqslant l\leqslant N}\|e_l^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i)_{i\in I}\| = \max_{1\leqslant l\leqslant N}\sqrt{\sum_{i\in I}(e_l^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i)^2}. \tag{60}$$

For each $i\in I$,

$$\begin{aligned}\left|e_l^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i\right| &= \left|e_l^{\mathrm{T}}(I-\mathcal{U}\mathcal{U}^{\mathrm{T}})\Xi(\widetilde{\lambda}_i)\mathcal{U}\cdot\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\right| \\ &\leqslant \left\|e_l^{\mathrm{T}}(I-\mathcal{U}\mathcal{U}^{\mathrm{T}})\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\|\cdot\left\|\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\right\|.\end{aligned} \tag{61}$$

Observe from $(\mathcal{A}+\mathcal{E})\tilde{\mathbf{u}}_i = \widetilde{\lambda}_i\tilde{\mathbf{u}}_i$ that $\mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i = (\widetilde{\lambda}_iI-\mathcal{E})\tilde{\mathbf{u}}_i$. Multiplying $\mathcal{U}^{\mathrm{T}}$ on both sides, we get the bound

$$\left\|\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\right\| \leqslant \|\mathcal{E}\| + |\widetilde{\lambda}_i| \leqslant \left(1+\frac{1}{b-1}\right)|\widetilde{\lambda}_i| = \frac{b}{b-1}|\widetilde{\lambda}_i| \tag{62}$$

using the assumption $\|\mathcal{E}\| = \|E\| \leqslant \frac{1}{b}|\lambda_i|$ and the Weyl's inequality $|\widetilde{\lambda}_i| \geqslant |\lambda_i| - \|E\| \geqslant (b-1)\|E\|$.

To estimate

$$\begin{aligned}\left\|e_l^{\mathrm{T}}(I-\mathcal{U}\mathcal{U}^{\mathrm{T}})\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| &\leqslant \left\|e_l^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| + \left\|e_l^{\mathrm{T}}\mathcal{U}\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| \\ &\leqslant \left\|e_l^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| + \left\|U\|_{2,\infty}\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\|, \end{aligned} \tag{63}$$

we split the index set I into two disjoint sets:

$$\mathcal{I}_s := \left\{i \in I : |\lambda_i| \leqslant n^2\right\} \quad \text{and} \quad \mathcal{I}_b := \left\{i \in I : |\lambda_i| > n^2\right\}.$$

Note that $\mathcal{I}_s$ or $\mathcal{I}_b$ could be the empty set.

Case (1): $i \in I \cap \mathcal{I}_s$. In this case,

$$\Xi(\widetilde{\lambda}_i) = G(\widetilde{\lambda}_i) - \Phi(\widetilde{\lambda}_i).$$

Note that if $z \in S_{\sigma_i}$ specified in (40) for any $1 \leqslant i \leqslant r_0$, then $|z| \geqslant 2b(\sqrt{N}+\sqrt{n})$ by the supposition of σ_i. Recall

$$\eta = \frac{11b^2}{(b-1)^2}\sqrt{(K+7)\log(N+n) + 2(\log 9)r}.$$

We work on the event $\mathsf{E} := \cap_{i \in [\![k,s]\!] \cap \mathcal{I}_s} \mathsf{E}_i$ where

$$\begin{aligned}\mathsf{E}_i := &\left\{\widetilde{\sigma}_i \in S_{\sigma_{l_i}} \text{ for some } l_i \in [\![1, r_0]\!]\right\} \cap \left\{\left\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\sigma}_i)\mathcal{U}\right\| \leqslant \frac{\eta}{\widetilde{\sigma}_i^2}\right\} \\ &\cap \left\{\left|e_l^{\mathrm{T}}\Xi(\widetilde{\sigma}_i)\mathbf{u}_s\right| \leqslant \frac{5b^2}{(b-1)^2}\frac{\sqrt{(K+7)\log(N+n)}}{\widetilde{\sigma}_i^2} \text{ for } 1 \leqslant l \leqslant N+n, 1 \leqslant s \leqslant r\right\}.\end{aligned} \tag{64}$$

By Theorem 29, Lemma 28 and Lemma 27, the event E holds with probability at least $1 - 20(N+n)^{-K}$. For $i \in [\![k,s]\!] \cap \mathcal{I}_s$, $\widetilde{\lambda}_i = \widetilde{\sigma}_i$. It follows immediately that

$$\begin{aligned}&\left\|e_l^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| + \|U\|_{2,\infty}\left\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| \\ &= \sqrt{\sum_{s=1}^{2r}\left(e_l^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathbf{u}_s\right)^2} + \|U\|_{2,\infty}\left\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| \\ &< \frac{18b^2}{(b-1)^2}\frac{\sqrt{r(K+7)\log(N+n)}}{\widetilde{\lambda}_i^2}(1+\|U\|_{2,\infty}).\end{aligned} \tag{65}$$

Continuing from (61), (62) and (63), we further have for any $i \in I \cap \mathcal{I}_s$,

$$\left|e_l^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i\right| \leqslant \frac{2b}{b-1}\frac{\gamma}{|\widetilde{\lambda}_i|}(1+\|U\|_{2,\infty}) \tag{66}$$

where we define

$$\gamma := \frac{9b^2}{(b-1)^2}\sqrt{r(K+7)\log(N+n)}$$

for the sake of brevity. For $i \in [\![r+k, r+s]\!] \cap \mathcal{I}_s$, $\widetilde{\lambda}_i = -\widetilde{\sigma}_{i-r}$. Note that $\Xi(\widetilde{\lambda}_i) \sim -\Xi(\widetilde{\sigma}_{i-r})$ since the distribution of E is symmetric. The bound (66) still holds.

Case (2): $i \in I \cap \mathcal{I}_b$. In this case,

$$\Xi(\widetilde{\lambda}_i) = G(\widetilde{\lambda}_i) - \frac{1}{\widetilde{\lambda}_i}I_{N+n}.$$

By Weyl's inequality, $|\widetilde{\lambda}_i| \geqslant n^2 - \|\mathcal{E}\| \geqslant 4(\sqrt{N} + \sqrt{n})$ for every $i \in \mathcal{I}_b$, we apply Lemma 30 to get

$$\|\Xi(\widetilde{\lambda}_i)\| \leqslant \frac{2\|\mathcal{E}\|}{\widetilde{\lambda}_i^2}.$$

As a result,

$$\left\|e_l^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| + \|U\|_{2,\infty}\left\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| \leqslant (1 + \|U\|_{2,\infty})\|\Xi(\widetilde{\lambda}_i)\| \leqslant \frac{2\|\mathcal{E}\|}{\widetilde{\lambda}_i^2}(1 + \|U\|_{2,\infty}).$$

Continuing from (61) and (63), we further have

$$\left|e_l^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i\right| \leqslant \frac{2b}{b-1}\frac{\|\mathcal{E}\|}{|\widetilde{\lambda}_i|}(1 + \|U\|_{2,\infty}). \tag{67}$$

Note by Weyl's inequality, for $i \in [\![k, s]\!]$, $|\widetilde{\lambda}_i| \geqslant \frac{b-1}{b}\sigma_i$ and for $i \in [\![r+k, r+s]\!]$, $|\widetilde{\lambda}_i| \geqslant \frac{b-1}{b}\sigma_{i-r}$. Continuing from (60) with (66) and (67), we conclude that

$$\begin{aligned}
&\max_{1\leqslant l\leqslant N}\left\|e_l^{\mathrm{T}}(Q\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i)_{i\in I}\right\| \\
&\leqslant 2\sqrt{2}\frac{b^2}{(b-1)^2}(1 + \|U\|_{2,\infty})\sqrt{\sum_{i\in[\![k,s]\!],\sigma_i\leqslant n^2}\frac{\gamma^2}{\sigma_i^2} + \sum_{i\in[\![k,s]\!],\sigma_i>n^2}\frac{\|E\|^2}{\sigma_i^2}}.
\end{aligned}$$

Note that $\|E\| \leqslant 2(\sqrt{N} + \sqrt{n}) \leqslant 4\sqrt{n}$. Inserting the above estimate and (59) into (58) yields that

$$\begin{aligned}
\left\|\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\right\|_{2,\infty} &\leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|U\|_{2,\infty}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} \\
&\quad + \frac{2\sqrt{2}b^2}{(b-1)^2}(1 + \|U\|_{2,\infty})\sqrt{\sum_{i\in[\![k,s]\!],\sigma_i\leqslant n^2}\frac{\gamma^2}{\sigma_i^2} + \sum_{i\in[\![k,s]\!],\sigma_i>n^2}\frac{16n}{\sigma_i^2}}.
\end{aligned}$$

This concludes the proof.

6.3. **Proof of Theorem 14.** The proof strategy for Theorem 14 mirrors that of Theorem 13. We provide a brief outline below.

First, we estimate $\left\|x^{\mathrm{T}}(\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s})\right\|$ for a unit vector $x \in \mathbb{R}^N$. Let $\mathbf{a} = (x^{\mathrm{T}}, 0)^{\mathrm{T}}$ be a unit vector in $\mathbb{R}^{N+n}$. Following the same line of the above proof, we first observe that

$$\left\|x^{\mathrm{T}}(\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s})\right\| = \left\|\mathbf{a}^{\mathrm{T}}(\widetilde{\mathcal{U}}_I - P_I\widetilde{\mathcal{U}}_I)\right\|.$$

Using the same proof as that of (58), one gets

$$\left\|\mathbf{a}^{\mathrm{T}}(\widetilde{\mathcal{U}}_I - P_I\widetilde{\mathcal{U}}_I)\right\| \leqslant \|\mathbf{a}^{\mathrm{T}}\mathcal{P}_J\widetilde{\mathcal{U}}_I\|\mathbf{1}_{\{|J|\neq 0\}} + \|\mathbf{a}^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i)_{i\in I}\|. \tag{68}$$

For the first term on the right-hand side of (68), following the same line as (59), we have with probability at least $1 - 20(N+n)^{-K}$ that

$$\|\mathbf{a}^{\mathrm{T}}\mathcal{P}_J\widetilde{\mathcal{U}}_I\| \leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|\mathbf{a}^{\mathrm{T}}U\|\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{|J|\neq 0\}}.$$

For the second term on the right-hand side of (68), using the similar arguments as (60), we also get with probability at least $1-20(N+n)^{-K}$ that

$$\|\mathbf{a}^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i)_{i\in I}\| \leqslant \frac{2\sqrt{2}b^2}{(b-1)^2}(1+\|\mathbf{a}^{\mathrm{T}}\mathcal{U}\|)\sqrt{\sum_{i=k}^{s}\frac{\gamma^2}{\sigma_i^2}}.$$

Combining the above estimates and noting that $\|\mathbf{a}^{\mathrm{T}}\mathcal{U}\| = \|x^{\mathrm{T}}U\|$, we conclude

$$\begin{aligned}\left\|x^{\mathrm{T}}(\widetilde{U}_{k,s}-P_{U_{k,s}}\widetilde{U}_{k,s})\right\| \leqslant & 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|x^{\mathrm{T}}U\|\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} \\ & + \frac{2\sqrt{2}b^2}{(b-1)^2}(1+\|x^{\mathrm{T}}U\|)\sqrt{\sum_{i=k}^{s}\frac{\gamma^2}{\sigma_i^2}}\end{aligned}$$

with probability at least $1-40(N+n)^{-K}$.

Next, we turn to the estimation of $\left|x^{\mathrm{T}}(\widetilde{U}_{k,s}-P_{U_{k,s}}\widetilde{U}_{k,s})y\right|$. Set $\mathbf{b}=(y^{\mathrm{T}},0)^{\mathrm{T}}\in\mathbb{R}^{2(k-s+1)}$. It is elementary to check that

$$\left|x^{\mathrm{T}}(\widetilde{U}_{k,s}-P_{U_{k,s}}\widetilde{U}_{k,s})y\right| = \sqrt{2}\left|\mathbf{a}^{\mathrm{T}}(\widetilde{\mathcal{U}}_I-P_I\widetilde{\mathcal{U}}_I)\mathbf{b}\right|.$$

Using the decomposition in (57), we get

$$\left|\mathbf{a}^{\mathrm{T}}(\widetilde{\mathcal{U}}_I-P_I\widetilde{\mathcal{U}}_I)\mathbf{b}\right| \leqslant \left|\mathbf{a}^{\mathrm{T}}\mathcal{P}_J\widetilde{\mathcal{U}}_I\mathbf{b}\right|\mathbf{1}_{\{|J|\neq 0\}} + \left|\mathbf{a}^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i)_{i\in I}\mathbf{b}\right|. \tag{69}$$

When $|J|\neq 0$, taking the definitions of $\mathbf{a},\mathbf{b}$ into consideration, we can derive an upper bound for the first term on the right-hand side of equation (69):

$$\left|\mathbf{a}^{\mathrm{T}}\mathcal{P}_J\widetilde{\mathcal{U}}_I\mathbf{b}\right| = \left|\mathbf{a}^{\mathrm{T}}\mathcal{U}_J\cdot\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\mathbf{b}\right| \leqslant \|\mathbf{a}^{\mathrm{T}}\mathcal{U}_J\|\cdot\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathcal{U}}_I\mathbf{b}\| = \|x^{\mathrm{T}}U_{J_0}\|\left\|\sum_{i\in I_0}y_i\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathbf{u}}_i\right\|$$

where $J_0 := [\![1,k-1]\!]\cup[\![s+1,r]\!]$ and $I_0 := [\![k,s]\!]$. By Lemma 34, we further obtain for each $i\in I_0$,

$$\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathbf{u}}_i\| \leqslant 3\frac{(b+1)^2}{(b-1)^2}\frac{\eta}{\min\{\delta_{k-1},\delta_s\}}$$

with probability at least $1-20(N+n)^{-K}$. By Cauchy-Schwartz inequality, we further obtain

$$\left|\mathbf{a}^{\mathrm{T}}\mathcal{P}_J\widetilde{\mathcal{U}}_I\mathbf{b}\right| \leqslant 3\frac{(b+1)^2}{(b-1)^2}\|x^{\mathrm{T}}U\|\frac{\eta\sqrt{\|y\|_0}}{\min\{\delta_{k-1},\delta_s\}} \tag{70}$$

with probability at least $1-20(N+n)^{-K}$.

For the second term on the right-hand side of equation (69), we start with

$$\left|\mathbf{a}^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i)_{i\in I}\mathbf{b}\right| = \left|\sum_{i\in I_0}y_i\mathbf{a}^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i\right|. \tag{71}$$

For each $i\in I_0$, similar to (61), we have

$$\begin{aligned}\left|\mathbf{a}^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i\right| &= \left|\mathbf{a}^{\mathrm{T}}(I-\mathcal{U}\mathcal{U}^{\mathrm{T}})\Xi(\widetilde{\lambda}_i)\mathcal{U}\cdot\mathcal{D}\mathcal{U}^{\mathrm{T}}\widetilde{\mathbf{u}}_i\right| \\ &\leqslant \left\|\mathbf{a}^{\mathrm{T}}(I-\mathcal{U}\mathcal{U}^{\mathrm{T}})\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\|\cdot\left\|\mathcal{D}\mathcal{U}^{\mathrm{T}}\widetilde{\mathbf{u}}_i\right\|.\end{aligned}$$

Note that, by (62),

$$\left\|\mathcal{D}\mathcal{U}^{\mathrm{T}}\widetilde{\mathbf{u}}_i\right\| \leqslant \frac{b}{b-1}|\widetilde{\lambda}_i|.$$

The estimation of

$$\begin{aligned}\left\|\mathbf{a}^{\mathrm{T}}(I-\mathcal{U}\mathcal{U}^{\mathrm{T}})\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| &\leqslant \left\|\mathbf{a}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| + \left\|\mathbf{a}^{\mathrm{T}}\mathcal{U}\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| \\ &\leqslant \left\|\mathbf{a}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| + \|x^{\mathrm{T}}U\|\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\| \\ &= \sqrt{\sum_{s=1}^{2r}\left(\mathbf{a}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathbf{u}_s\right)^2} + \|x^{\mathrm{T}}U\|\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\| \\ &< \frac{18b^2}{(b-1)^2}\frac{\sqrt{r(K+7)\log(N+n)}}{\widetilde{\lambda}_i^2}(1+\|x^{\mathrm{T}}U\|) = \frac{2\gamma}{\widetilde{\lambda}_i^2}(1+\|x^{\mathrm{T}}U\|)\end{aligned}$$

follows the same line as those of (63) and (65) with e_l replaced by the unit vector $\mathbf{a}$. In particular, combined with $|\widetilde{\lambda}_i| \geqslant \frac{b-1}{b}\sigma_i$ for each $i \in I_0$ from the Weyl's inequality, we obtain

$$\begin{aligned}\left|\mathbf{a}^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i\right| &\leqslant \left\|\mathbf{a}^{\mathrm{T}}(I-\mathcal{U}\mathcal{U}^{\mathrm{T}})\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| \cdot \left\|\mathcal{D}\mathcal{U}^{\mathrm{T}}\widetilde{\mathbf{u}}_i\right\| \\ &\leqslant \frac{2b}{b-1}\frac{\gamma}{|\widetilde{\lambda}_i|}(1+\|x^{\mathrm{T}}U\|) \leqslant \frac{2b^2}{(b-1)^2}\frac{\gamma}{\sigma_i}(1+\|x^{\mathrm{T}}U\|)\end{aligned}$$

holds with probability at least $1-20(N+n)^{-K}$. Continuing from (71), we have

$$\left|\mathbf{a}^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i)_{i\in I}\mathbf{b}\right| \leqslant \frac{2b^2}{(b-1)^2}\gamma(1+\|x^{\mathrm{T}}U\|)\sum_{i=k}^{s}\frac{|y_i|}{\sigma_i}. \tag{72}$$

Finally, inserting (70) and (72) back into (69), we obtain that

$$\begin{aligned}\left|x^{\mathrm{T}}(\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s})y\right| &= \sqrt{2}\left|\mathbf{a}^{\mathrm{T}}(\widetilde{\mathcal{U}}_I - P_I\widetilde{\mathcal{U}}_I)\mathbf{b}\right| \\ &\leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|x^{\mathrm{T}}U\|\frac{\eta\sqrt{\|y\|_0}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} \\ &\quad + \frac{2\sqrt{2}b^2}{(b-1)^2}\gamma(1+\|x^{\mathrm{T}}U\|)\sum_{i=k}^{s}\frac{|y_i|}{\sigma_i}\end{aligned}$$

holds with probability at least $1-40(N+n)^{-K}$.

6.4. **Proof of Theorem 16.** The proof of Theorem 16 follows largely the proof of Theorem 13. We sketch the proof and focus on the difference. We start with the decomposition (45):

$$\widetilde{\mathbf{u}}_i = \Pi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i + \Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i$$

and set

$$\Pi(\widetilde{\lambda}_i) := \begin{cases}\Phi(\widetilde{\lambda}_i), & \text{if } |\lambda_i| \leqslant n^2; \\ \frac{1}{\widetilde{\lambda}_i}I_{N+n} + \frac{\mathcal{E}}{\widetilde{\lambda}_i^2}, & \text{if } |\lambda_i| > n^2.\end{cases} \tag{73}$$

The definition of $\Pi(\widetilde{\lambda}_i)$ differs from the one given in (56) when $|\lambda_i| > n^2$. We adopt this definition to achieve more precise control over the error term by considering the weights $|\widetilde{\lambda}_i|$.

Let $\mathcal{Q} = I - \mathcal{P}_r$ be the orthogonal projection matrix onto the null space of $\mathcal{A}$. Using the same derivation as in the beginning of the proof of Theorem 13, we obtain the decomposition

$$\widetilde{\lambda}_i \widetilde{\mathbf{u}}_i = \mathcal{P}_I \widetilde{\lambda}_i \widetilde{\mathbf{u}}_i + \mathcal{P}_J \widetilde{\lambda}_i \widetilde{\mathbf{u}}_i + \mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\lambda}_i \widetilde{\mathbf{u}}_i$$

and hence

$$\widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I - \mathcal{P}_I \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I = \mathcal{P}_J \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I + (\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\lambda}_i \widetilde{\mathbf{u}}_i)_{i\in I}. \tag{74}$$

It can be verified using the definitions of $\widetilde{\mathcal{U}}_I, \widetilde{\mathcal{D}}_I$ and $\mathcal{P}_I = \widetilde{\mathcal{U}}_I \widetilde{\mathcal{U}}_I^{\mathrm{T}}$ that

$$\widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I - \mathcal{P}_I \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I = \frac{1}{\sqrt{2}} \begin{pmatrix} \widetilde{U}_{k,s}\widetilde{D}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\widetilde{D}_{k,s} & -(\widetilde{U}_{k,s}\widetilde{D}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\widetilde{D}_{k,s}) \\ \widetilde{V}_{k,s}\widetilde{D}_{k,s} - P_{V_{k,s}}\widetilde{V}_{k,s}\widetilde{D}_{k,s} & \widetilde{V}_{k,s}\widetilde{D}_{k,s} - P_{V_{k,s}}\widetilde{V}_{k,s}\widetilde{D}_{k,s}. \end{pmatrix}$$

Therefore, combining (74), we observe that

$$\begin{aligned} \|\widetilde{U}_{k,s}\widetilde{D}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\widetilde{D}_{k,s}\|_{2,\infty} &= \max_{1\leqslant i\leqslant N} \|e_l^{\mathrm{T}}(\widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I - \mathcal{P}_I \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I)\| \\ &= \max_{1\leqslant i\leqslant N} \|e_l^{\mathrm{T}}(\mathcal{P}_J \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I + (\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\lambda}_i \widetilde{\mathbf{u}}_i)_{i\in I})\| \\ &\leqslant \max_{1\leqslant i\leqslant N} \|e_l^{\mathrm{T}}\mathcal{P}_J \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I\| \mathbf{1}_{|J|\neq 0} + \max_{1\leqslant i\leqslant N} \|e_l^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\lambda}_i \widetilde{\mathbf{u}}_i)_{i\in I})\|. \end{aligned} \tag{75}$$

The first term on the right-hand side of (75) can be bounded similarly as that of (59) using Lemma 34: if $|J| \neq 0$, then

$$\begin{aligned} \max_{1\leqslant l\leqslant N} \|e_l^{\mathrm{T}}\mathcal{P}_J \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I\| &= \max_{1\leqslant l\leqslant N} \|e_l^{\mathrm{T}}\mathcal{U}_J \cdot \mathcal{U}_J^{\mathrm{T}} \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I\| \\ &\leqslant \max_{1\leqslant l\leqslant N} \|e_l^{\mathrm{T}}\mathcal{U}_J\| \cdot \|\mathcal{U}_J^{\mathrm{T}} \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I\| \\ &\leqslant \|U\|_{2,\infty} \|\mathcal{U}_J^{\mathrm{T}} \widetilde{\mathcal{U}}_I \widetilde{\mathcal{D}}_I\|_F = \|U\|_{2,\infty} \sqrt{\sum_{i\in I} \widetilde{\lambda}_i^2 \|\mathcal{U}_J^{\mathrm{T}} \widetilde{\mathbf{u}}_i\|^2} \\ &\leqslant 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2} \|U\|_{2,\infty} \frac{\eta\sigma_k\sqrt{s-k+1}}{\min\{\delta_{k-1}, \delta_s\}} \end{aligned}$$

with probability at least $1 - 20(N+n)^{-K}$.

The bound of the second term on the right-hand side of (75) proceeds in a similar manner to that of (60):

$$\max_{1\leqslant i\leqslant N} \|e_l^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\lambda}_i \widetilde{\mathbf{u}}_i)_{i\in I})\| = \max_{1\leqslant l\leqslant N} \sqrt{\sum_{i\in I} \widetilde{\lambda}_i^2 (e_l^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i)^2}. \tag{76}$$

For each $i \in I$, we first establish the following bound using (61), (62) and (63):

$$|\widetilde{\lambda}_i| \cdot \left| e_l^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i \right| \leqslant \frac{b}{b-1}\widetilde{\lambda}_i^2 \left(\|e_l^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\| + \|U\|_{2,\infty}\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\| \right). \tag{77}$$

We then differentiate the cases by splitting the discussion according to whether $i \in I \cap \mathcal{I}_s$ or $i \in I \cap \mathcal{I}_b$, where

$$\mathcal{I}_s := \left\{ i \in I : |\lambda_i| \leqslant n^2 \right\} \quad \text{and} \quad \mathcal{I}_b := \left\{ i \in I : |\lambda_i| > n^2 \right\}.$$

If $i \in I \cap \mathcal{I}_s$, then by (66), we immediately obtain

$$|\widetilde{\lambda}_i| \cdot \left| e_l^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i \right| \leqslant \frac{2b}{b-1}\gamma(1 + \|U\|_{2,\infty}) \tag{78}$$

with probability at least $1-20(N+n)^{-K}$.

If $i \in I \cap \mathcal{I}_b$, then the only difference from the proof of (67) is that

$$\Xi(\widetilde{\lambda}_i) = G(\widetilde{\lambda}_i) - \frac{1}{\widetilde{\lambda}_i} I_{N+n} - \frac{\mathcal{E}}{\widetilde{\lambda}_i^2}$$

and by Lemma 30, we have

$$\|\Xi(\widetilde{\lambda}_i)\| \leqslant \frac{2\|\mathcal{E}\|^2}{|\widetilde{\lambda}_i|^3}.$$

It follows that

$$\left\|e_l^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| + \|U\|_{2,\infty}\left\|\mathcal{U}^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\right\| \leqslant (1+\|U\|_{2,\infty})\|\Xi(\widetilde{\lambda}_i)\| \leqslant \frac{2\|\mathcal{E}\|^2}{|\widetilde{\lambda}_i|^3}(1+\|U\|_{2,\infty}).$$

Continuing from (77), we further get

$$|\widetilde{\lambda}_i| \cdot \left|e_l^{\mathrm{T}}\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i\right| \leqslant \frac{2b}{b-1}\frac{\|E\|^2}{|\widetilde{\lambda}_i|}(1+\|U\|_{2,\infty}). \tag{79}$$

Note by Weyl's inequality, for $i \in [\![k,s]\!]$, $|\widetilde{\lambda}_i| \geqslant \frac{b-1}{b}\sigma_i$ and for $i \in [\![r+k, r+s]\!]$, $|\widetilde{\lambda}_i| \geqslant \frac{b-1}{b}\sigma_{i-r}$. Inserting (78) and (79) back into (76), we obtain that with probability at least $1-20(N+n)^{-K}$,

$$\begin{aligned}
&\max_{1\leqslant i\leqslant N} \|e_l^{\mathrm{T}}(\mathcal{Q}\Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\lambda}_i\tilde{\mathbf{u}}_i)_{i\in I})\| \\
&\leqslant \frac{2\sqrt{2}b^2}{(b-1)^2}(1+\|U\|_{2,\infty})\sqrt{\gamma^2(s-k+1) + \sum_{i\in[\![k,s]\!],\sigma_i>n^2}\frac{\|E\|^2}{\sigma_i}} \\
&\leqslant \frac{2\sqrt{2}b^2}{(b-1)^2}(1+\|U\|_{2,\infty})\sqrt{\gamma^2(s-k+1)+16},
\end{aligned}$$

where we used the crude estimate

$$\sum_{i\in[\![k,s]\!],\sigma_i>n^2}\frac{\|E\|^2}{\sigma_i} \leqslant n\frac{16n}{n^2} = 16.$$

This concludes the proof.

7. Proofs of Theorems 8, 9 and 10

7.1. **Proof of Theorem 8.** By the min-max theorem, for an $N \times n$ matrix S, the jth largest singular value of S is

$$\sigma_j(S) = \max_{\substack{W\in\mathbb{R}^N,\dim(W)=j\\ K\in\mathbb{R}^n,\dim(K)=j}} \min_{\substack{(x,y)\in W\times K\\ \|x\|=\|y\|=1}} x^T S y. \tag{80}$$

For the lower bound (9) of $\widetilde{\sigma}_k$, by (80),

$$\widetilde{\sigma}_k \geqslant \min_{\substack{(x,y)\in U_k\times V_k\\ \|x\|=\|y\|=1}} x^T(A+E)y \geqslant \sigma_k - \max_{\substack{(x,y)\in U_k\times V_k\\ \|x\|=\|y\|=1}} |x^T E y|.$$

Note that

$$\|U_k^T E V_k\| = \max_{\substack{(x,y)\in U_k\times V_k\\ \|x\|=\|y\|=1}} |x^T E y|.$$

Our assumption immediately yields that $\widetilde{\sigma}_k \geqslant \sigma_k - t$ with probability at least $1-\varepsilon$.

For the upper bound (10) of $\widetilde{\sigma}_k$, by (80),

$$\sigma_k \geqslant \widetilde{\sigma}_k - \max_{\substack{(x,y)\in \widetilde{U}_k \times \widetilde{V}_k \\ \|x\|=\|y\|=1}} |x^T E y|. \tag{81}$$

It is enough to bound the second term on the right side. For any unit vectors $x \in \widetilde{U}_k$ and $y \in \widetilde{V}_k$, we decompose $x = P_{U^\perp}x + UU^T x$ and $y = P_{V^\perp}y + VV^T\widetilde{y}$. It follows from triangle inequality that

$$|x^T Ey| \leqslant \|P_{U^\perp}x\| \cdot \|P_{V^\perp}y\| \cdot \|E\| + (\|P_{U^\perp}x\| + \|P_{V^\perp}y\|)\|E\| + \|U^T EV\|.$$

To bound the term $\|P_{U^\perp}x\|$, first notice by Cauchy-Schwartz inequality,

$$\max_{x\in \widetilde{U}_k, \|x\|=1} \|P_{U^\perp}x\| \leqslant \sqrt{k} \max_{1\leqslant s\leqslant k} \|P_{U^\perp}\widetilde{u}_s\|.$$

Next, multiplying $P_{U^\perp}$ on the left side of the equation $(A+E)\widetilde{v}_s = \widetilde{\sigma}_s\widetilde{u}_s$, we get $P_{U^\perp}E\widetilde{v}_s = \widetilde{\sigma}_s P_{U^\perp}\widetilde{u}_s$. It follows that

$$\max_{1\leqslant s\leqslant k} \|P_{U^\perp}\widetilde{u}_s\| \leqslant \|E\|/\widetilde{\sigma}_k \tag{82}$$

and thus

$$\max_{x\in \widetilde{U}_k, \|x\|=1} \|P_{U^\perp}x\| \leqslant \sqrt{k}\|E\|/\widetilde{\sigma}_k.$$

The same calculation leads to

$$\max_{y\in \widetilde{V}_k, \|y\|=1} \|P_{V^\perp}y\| \leqslant \sqrt{k}\|E\|/\widetilde{\sigma}_k.$$

Therefore,

$$\max_{\substack{(x,y)\in \widetilde{U}_k \times \widetilde{V}_k \\ \|x\|=\|y\|=1}} |x^T Ey| \leqslant 2\sqrt{k}\frac{\|E\|^2}{\widetilde{\sigma}_k} + k\frac{\|E\|^3}{\widetilde{\sigma}_k^2} + \|U^T EV\|. \tag{83}$$

Since $\|U^T EV\| \leqslant L$ and $\|E\| \leqslant B$ with probability at least $1-\varepsilon$, by (81) and (83), we obtain, with probability at least $1-\varepsilon$

$$\widetilde{\sigma}_k \leqslant \sigma_k + 2\sqrt{k}\frac{B^2}{\widetilde{\sigma}_k} + k\frac{B^3}{\widetilde{\sigma}_k^2} + L.$$

7.2. **Proof of Theorem 9.** In the proof, we work on the event

$$\Omega := \{\|E\| \leqslant B \text{ and } \|U^T EV\| \leqslant L\}.$$

By the supposition, Ω holds with probability at least $1-\varepsilon$. Observe that $\|U_k^T EV_k\| \leqslant \|U^T EV\| \leqslant L$. Using (9), the lower bound for $\widetilde{\sigma}_k$, together with $\delta_k \geqslant 2L$, we have

$$\widetilde{\sigma}_k \geqslant \sigma_k - L \geqslant \sigma_k/2 > 0$$

and

$$\widetilde{\sigma}_k - \sigma_{k+1} = \widetilde{\sigma}_k - \sigma_k + \delta_k \geqslant \delta_k - L \geqslant \delta_k/2.$$

From (25) and triangle inequality, we see

$$\|\!| \sin\angle(U_k, \widetilde{U}_k)|\!\| = \|\!|P_{U_k^\perp}P_{\widetilde{U}_k}|\!\| \leqslant \|\!|P_{U^\perp}P_{\widetilde{U}_k}|\!\| + \mathbf{1}_{\{k<r\}}\|\!|P_{U_{k+1,r}}P_{\widetilde{U}_k}|\!\|. \tag{84}$$

We first bound the first term on the right-hand side of (84). Suppose $P_{U^\perp} = U_0{U_0}^{\mathrm{T}}$ where the columns of U_0 are an orthonormal basis of the subspace $U^\perp$. Then

$$\|\!|P_{U^\perp}P_{\widetilde{U}_k}|\!\| = \|\!|U_0^{\mathrm{T}}\widetilde{U}_k|\!\|.$$

Multiplying U_0^{T} on the both sides of $(A+E)\widetilde{V}_k = \widetilde{U}_k\widetilde{D}_k$, we see $U_0^{\mathrm{T}}E\widetilde{V}_k = U_0^{\mathrm{T}}\widetilde{U}_k\widetilde{D}_k$ and hence, $U_0^{\mathrm{T}}\widetilde{U}_k = U_0^{\mathrm{T}}E\widetilde{V}_k\widetilde{D}_k^{-1}$. It follows that

$$\|\!|P_{U^\perp}P_{\widetilde{U}_k}\|\!| = \|\!|U_0^{\mathrm{T}}E\widetilde{V}_k\widetilde{D}_k^{-1}\|\!| \leqslant \|\!|U_0^{\mathrm{T}}E\widetilde{V}_k\|\!|\|\widetilde{D}_k^{-1}\| = \frac{\|\!|P_{U^\perp}EP_{\widetilde{V}_k}\|\!|}{\widetilde{\sigma}_k}. \tag{85}$$

In particular, for the operator norm, we have

$$\|P_{U^\perp}P_{\widetilde{U}_k}\| \leqslant \frac{\|E\|}{\widetilde{\sigma}_k}. \tag{86}$$

We proceed to bound the second term on the right-hand side of (84). It suffices to consider $k < r$. Observe that

$$\begin{aligned}\|\!|P_{U_{k+1,r}}P_{\widetilde{U}_k}\|\!| = \|\!|{U_{k+1,r}}^{\mathrm{T}}\widetilde{U}_k\|\!| &\leqslant \|{U_{k+1,r}}^{\mathrm{T}}\widetilde{U}_k\|_* \\ &\leqslant \sqrt{\operatorname{rank}({U_{k+1,r}}^{\mathrm{T}}\widetilde{U}_k)}\|{U_{k+1,r}}^{\mathrm{T}}\widetilde{U}_k\|_F \\ &\leqslant \sqrt{\min\{k, r-k\}}\sqrt{\sum_{i=1}^k \|{U_{k+1,r}}^{\mathrm{T}}\widetilde{u}_i\|^2}.\end{aligned}$$

Specially, for the operator norm, we have

$$\|P_{U_{k+1,r}}P_{\widetilde{U}_k}\| \leqslant \|{U_{k+1,r}}^{\mathrm{T}}\widetilde{U}_k\|_F = \sqrt{\sum_{i=1}^k \|{U_{k+1,r}}^{\mathrm{T}}\widetilde{u}_i\|^2}.$$

It remains to bound $\|U_{k+1,r}^{\mathrm{T}}\widetilde{u}_i\|$ for $1 \leqslant i \leqslant k$. Multiplying $U_{k+1,r}^{\mathrm{T}}$ on both sides of $(A+E)\widetilde{v}_i = \widetilde{\sigma}_i\widetilde{u}_i$, we get

$$D_{k+1,r}V_{k+1,r}^{\mathrm{T}}\widetilde{v}_i + U_{k+1,r}^{\mathrm{T}}E\widetilde{v}_i = \widetilde{\sigma}_iU_{k+1,r}^{\mathrm{T}}\widetilde{u}_i,$$

which yields

$$\widetilde{\sigma}_iD_{k+1,r}V_{k+1,r}^{\mathrm{T}}\widetilde{v}_i = \widetilde{\sigma}_i^2U_{k+1,r}^{\mathrm{T}}\widetilde{u}_i - \widetilde{\sigma}_iU_{k+1,r}^{\mathrm{T}}E\widetilde{v}_i. \tag{87}$$

Similarly, multiplying $V_{k+1,r}^{\mathrm{T}}$ on both sides of $(A^{\mathrm{T}} + E^{\mathrm{T}})\widetilde{u}_i = \widetilde{\sigma}_i\widetilde{v}_i$, we also get

$$D_{k+1,r}U_{k+1,r}^{\mathrm{T}}\widetilde{u}_i + V_{k+1,r}^{\mathrm{T}}E^{\mathrm{T}}\widetilde{U}_i = \widetilde{\sigma}_iV_{k+1,r}^{\mathrm{T}}\widetilde{v}_i,$$

which implies

$$\widetilde{\sigma}_iD_{k+1,r}V_{k+1,r}^{\mathrm{T}}\widetilde{v}_i = D_{k+1,r}^2U_{k+1,r}^{\mathrm{T}}\widetilde{u}_i + D_{k+1,r}V_{k+1,r}^{\mathrm{T}}E^{\mathrm{T}}\widetilde{u}_i. \tag{88}$$

Combining (87) and (88), one has

$$(\widetilde{\sigma}_i^2I - D_{k+1,r}^2)U_{k+1,r}^{\mathrm{T}}\widetilde{u}_i = \widetilde{\sigma}_iU_{k+1,r}^{\mathrm{T}}E\widetilde{v}_i + D_{k+1,r}V_{k+1,r}^{\mathrm{T}}E^{\mathrm{T}}\widetilde{u}_i.$$

As a result, by noting $\widetilde{\sigma}_i^2I - D_{k+1,r}^2 = \operatorname{diag}(\widetilde{\sigma}_i^2 - \sigma_{k+1}^2, \cdots, \widetilde{\sigma}_i^2 - \sigma_r^2)$, we obtain the following bound

$$\begin{aligned}\|U_{k+1,r}^{\mathrm{T}}\widetilde{u}_i\| &\leqslant \frac{\widetilde{\sigma}_i\|U_{k+1,r}^{\mathrm{T}}E\widetilde{v}_i\| + \sigma_{k+1}\|V_{k+1,r}^{\mathrm{T}}E^{\mathrm{T}}\widetilde{u}_i\|}{\widetilde{\sigma}_i^2 - \sigma_{k+1}^2} \\ &\leqslant \frac{\max\{\|U_{k+1,r}^{\mathrm{T}}E\widetilde{v}_i\|, \|V_{k+1,r}^{\mathrm{T}}E^{\mathrm{T}}\widetilde{u}_i\|\}}{\widetilde{\sigma}_k - \sigma_{k+1}}.\end{aligned} \tag{89}$$

Now we turn to bound the numerator of the above expression. Decompose $\widetilde{u}_i = P_U\widetilde{u}_i + P_{U^\perp}\widetilde{u}_i$. Then

$$V_{k+1,r}^{\mathrm{T}}E^{\mathrm{T}}\widetilde{u}_i = V_{k+1,r}^{\mathrm{T}}E^{\mathrm{T}}UU^{\mathrm{T}}\widetilde{u}_i + V_{k+1,r}^{\mathrm{T}}E^{\mathrm{T}}P_{U^\perp}\widetilde{u}_i$$

and

$$\begin{aligned}\|V_{k+1,r}^{\mathrm{T}} E^{\mathrm{T}} \widetilde{u}_i\| &\leqslant \|V_{k+1,r}^{\mathrm{T}} E^{\mathrm{T}} U\| + \|E^{\mathrm{T}}\| \cdot \|P_{U^\perp} \widetilde{u}_i\| \\ &\leqslant \|U^{\mathrm{T}} E V\| + \frac{\|E\|^2}{\widetilde{\sigma}_k}.\end{aligned}$$

The last inequality above follows from $\|P_{U^\perp} \widetilde{u}_i\| \leqslant \|P_{U^\perp} P_{\widetilde{U}_k}\|$ and by applying (86). Likewise, we also have

$$\|U_{k+1,r}^{\mathrm{T}} E \widetilde{v}_i\| \leqslant \|U^{\mathrm{T}} E V\| + \frac{\|E\|^2}{\widetilde{\sigma}_k}.$$

Continuing from (89), we see

$$\begin{aligned}\|U_{k+1,r}^{\mathrm{T}} \widetilde{u}_i\| &\leqslant \frac{\|U^{\mathrm{T}} E V\| + \|E\|^2/\widetilde{\sigma}_k}{\widetilde{\sigma}_k - \sigma_{k+1}} \\ &\leqslant 2\frac{\|U^{\mathrm{T}} E V\|}{\delta_k} + 4\frac{\|E\|^2}{\delta_k \sigma_k}\end{aligned}$$

by plugging in $\widetilde{\sigma}_k - \sigma_{k+1} \geqslant \delta_k/2$ and $\widetilde{\sigma}_k \geqslant \sigma_k/2$. Consequently,

$$\begin{aligned}\|\!| P_{U_{k+1,r}} P_{\widetilde{U}_k} |\!\| &\leqslant \sqrt{\min\{k, r-k\}} \sqrt{\sum_{i=1}^{k} \|U_{k+1,r}{}^{\mathrm{T}} \widetilde{u}_i\|^2} \\ &\leqslant 2\sqrt{k \min\{k, r-k\}} \left(\frac{\|U^{\mathrm{T}} E V\|}{\delta_k} + 2\frac{\|E\|^2}{\delta_k \sigma_k} \right).\end{aligned}$$

In particular, for the operator norm,

$$\|P_{U_{k+1,r}} P_{\widetilde{U}_k}\| \leqslant \sqrt{\sum_{i=1}^{k} \|U_{k+1,r}{}^{\mathrm{T}} \widetilde{u}_i\|^2} \leqslant 2\sqrt{k} \left(\frac{\|U^{\mathrm{T}} E V\|}{\delta_k} + 2\frac{\|E\|^2}{\delta_k \sigma_k} \right).$$

Combining the above estimates with (85), and considering (84), we ultimately arrive at

$$\begin{aligned}\|\!| \sin\angle(U_k, \widetilde{U}_k) |\!\| &\leqslant 2\sqrt{k \min\{k, r-k\}} \left(\frac{\|U^{\mathrm{T}} E V\|}{\delta_k} + 2\frac{\|E\|^2}{\delta_k \sigma_k} \right) + \frac{\|\!| P_{U^\perp} E P_{\widetilde{V}_k} |\!\|}{\widetilde{\sigma}_k} \\ &\leqslant 2\sqrt{k \min\{k, r-k\}} \left(\frac{\|U^{\mathrm{T}} E V\|}{\delta_k} + 2\frac{\|E\|^2}{\delta_k \sigma_k} \right) + 2\frac{k\|E\|}{\sigma_k}.\end{aligned}$$

More specifically, for the operator norm, we have

$$\|\sin\angle(U_k, \widetilde{U}_k)\| \leqslant 2\sqrt{k} \left(\frac{\|U^{\mathrm{T}} E V\|}{\delta_k} + 2\frac{\|E\|^2}{\delta_k \sigma_k} \right) \mathbf{1}_{\{k<r\}} + 2\frac{\|E\|}{\sigma_k}.$$

By applying the result to A^{T} and $A^{\mathrm{T}} + E^{\mathrm{T}}$, we observe that the same bounds also hold for $\sin\angle(V_k, \widetilde{V}_k)$. This concludes the proof.

7.3. **Proof of Theorem 10.** We first bound $\|U^{\mathrm{T}} E V\|$. By a standard ϵ-net argument (for example, Exercise 4.4.3. from [70]), for any $\frac{1}{4}$-nets $\mathcal{N}_1$ and $\mathcal{N}_2$ of the unit sphere in $\mathbb{R}^r$, we have

$$\|U^{\mathrm{T}} E V\| \leqslant 2 \sup_{x \in \mathcal{N}_1, y \in \mathcal{N}_2} x^{\mathrm{T}} U^{\mathrm{T}} E V y.$$

Besides, $|\mathcal{N}_i| \leqslant 9^r$ for $i = 1, 2$ (see [70, Corollary 4.2.13.]). Note that for every $x \in \mathcal{N}_1$ and $y \in \mathcal{N}_2$, by Cauchy-Schwartz inequality,

$$|x^{\mathrm{T}} U^{\mathrm{T}} E V y| = \left| \sum_{i,j=1}^{r} x_i y_j u_i^{\mathrm{T}} E v_j \right| \leqslant \sum_{i,j=1}^{r} |x_i y_j| \max_{1 \leqslant i,j \leqslant r} \left| u_i^{\mathrm{T}} E v_j \right| \leqslant r \max_{1 \leqslant i,j \leqslant r} |u_i^{\mathrm{T}} E v_j|.$$

Hence, for $t > 0$,

$$\begin{aligned} \mathbb{P}(\|U^{\mathrm{T}} E V\| \geqslant t) &\leqslant \mathbb{P}\left(\sup_{x \in \mathcal{N}_1, y \in \mathcal{N}_2} x^{\mathrm{T}} U^{\mathrm{T}} E V y \geqslant \frac{t}{2} \right) \\ &\leqslant 9^{2r} \mathbb{P}\left(\max_{1 \leqslant i,j \leqslant r} |u_i^{\mathrm{T}} E v_j| \geqslant \frac{t}{2} \frac{1}{r} \right) \\ &\leqslant r^2 9^{2r} f\left(\frac{t}{2r} \right). \end{aligned}$$

Likewise, we also get that for $t_0 > 0$,

$$\mathbb{P}(\|U_k^{\mathrm{T}} E V_k\| \leqslant t_0) \geqslant 1 - k^2 9^{2k} f\left(\frac{t_0}{2k} \right).$$

Conditioning on $\{\|U_k^{\mathrm{T}} E V_k\| \leqslant t_0 = \delta_k/2\}$, we can apply (9) from Theorem 8 to obtain

$$\widetilde{\sigma}_k - \sigma_{k+1} \geqslant \sigma_k - \sigma_{k+1} - t_0 = \delta_k - t_0 = \delta_k/2.$$

Hence, the event

$$\Omega_0 := \{\|U^{\mathrm{T}} E V\| \leqslant t\} \cap \{\widetilde{\sigma}_k - \sigma_{k+1} \geqslant \delta_k/2\}$$

holds with probability at least

$$1 - r^2 9^{2r} f\left(\frac{t}{2r} \right) - k^2 9^{2k} f\left(\frac{\delta_k}{4k} \right).$$

Going through the same lines of proof of Theorem 9 on the event Ω_0, we prove the theorem.

8. Proof of Theorem 18

We consider the model (16) and rewrite

$$\mathbb{E}(X) = (\theta_{z_1}, \cdots, \theta_{z_n}) = (\theta_1, \cdots, \theta_k) Z^{\mathrm{T}},$$

where $Z \in \{0,1\}^{n \times k}$ with entry $Z_{ij} = 1$ if $z_i = j$ and $Z_{ij} = 0$ otherwise. It is clear that the information regarding the cluster labels $\mathbf{z}$ is entirely encoded within Z. Additionally, let $D = \mathrm{diag}(d_1, \cdots, d_k)$ where d_i represents the cluster size associated with center θ_i. Consequently, the matrix $ZD^{-1/2}$ has orthonormal columns.

Given that θ_i's could be colinear, the rank of $\mathbb{E}(X)$ or $(\theta_1, \cdots, \theta_k)$, denoted as r, could be smaller than the number of clusters k. Consider the SVD of

$$(\theta_1, \cdots, \theta_k) D^{1/2} = U \Lambda W^{\mathrm{T}},$$

where Λ is a $k \times k$ diagonal matrix with rank r and W is a $k \times k$ orthogonal matrix. Observe that if we denote the SVD of $\mathbb{E}(X)$ as $\mathbb{E}(X) = U \Sigma V^{\mathrm{T}}$ with $U \in \mathbb{R}^{p \times k}$ and $V \in \mathbb{R}^{n \times k}$, then the following relationship emerges:

$$\mathbb{E}(X) = (\theta_1, \cdots, \theta_k) D^{1/2} (Z D^{-1/2})^{\mathrm{T}} = U \Lambda (Z D^{-1/2} W)^{\mathrm{T}} = U \Sigma V^{\mathrm{T}}.$$

Therefore, $\Sigma = \Lambda = \text{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0)$ and $V = ZD^{-1/2}W$. Note that the choice of U, V is not unique and we can only decide U or V up to an orthogonal transformation.

Next, we show that the geometric relationship among the columns of $\mathbb{E}(X)$ is preserved among the columns of $U^{\mathrm{T}}\mathbb{E}(X)$. Consider the SVD of $\mathbb{E}(X) = U\Sigma V^{\mathrm{T}}$, where each column θ_j of $\mathbb{E}(X)$ can be expressed as $\theta_j = U\Sigma(V^{\mathrm{T}})_j$ by denoting $(V^{\mathrm{T}})_j$ as the column of V^{T}. Let $(U^{\mathrm{T}}\mathbb{E}(X))_j$ represent the columns of $U^{\mathrm{T}}\mathbb{E}(X) = (U^{\mathrm{T}}\theta_{z_1}, \cdots, U^{\mathrm{T}}\theta_{z_n})$. For any two columns θ_i and θ_j of $\mathbb{E}(X)$, we have

$$\|\theta_i - \theta_j\|^2 = (\theta_i - \theta_j)^{\mathrm{T}}(\theta_i - \theta_j) = \left((V^{\mathrm{T}})_i - (V^{\mathrm{T}})_j\right)^{\mathrm{T}} \Sigma^2 \left((V^{\mathrm{T}})_i - (V^{\mathrm{T}})_j\right)$$

Moreover, their corresponding columns $(U^{\mathrm{T}}\mathbb{E}(X))_i$ and $(U^{\mathrm{T}}\mathbb{E}(X))_j$ of $U^{\mathrm{T}}\mathbb{E}(X)$ satisfy

$$\begin{aligned}\|(U^{\mathrm{T}}\mathbb{E}(X))_i - (U^{\mathrm{T}}\mathbb{E}(X))_j\|^2 &= \|U^{\mathrm{T}}\theta_i - U^{\mathrm{T}}\theta_j\|^2 = (\theta_i - \theta_j)^{\mathrm{T}}UU^{\mathrm{T}}(\theta_i - \theta_j) \\ &= \left((V^{\mathrm{T}})_i - (V^{\mathrm{T}})_j\right)^{\mathrm{T}} \Sigma U^{\mathrm{T}}UU^{\mathrm{T}}U\Sigma \left((V^{\mathrm{T}})_i - (V^{\mathrm{T}})_j\right) \\ &= \|\theta_i - \theta_j\|^2.\end{aligned}$$

It follows that

$$\|(U^{\mathrm{T}}\mathbb{E}(X))_i - (U^{\mathrm{T}}\mathbb{E}(X))_j\| = \|\theta_i - \theta_j\|.$$

Therefore, if $i, j \in [n]$ belong to the same cluster, then $\|(U^{\mathrm{T}}\mathbb{E}(X))_i - (U^{\mathrm{T}}\mathbb{E}(X))_j\| = 0$. On the other hand, if $i, j \in [n]$ belong to the distinct clusters, then $\|(U^{\mathrm{T}}\mathbb{E}(X))_i - (U^{\mathrm{T}}\mathbb{E}(X))_j\| \geqslant \Delta$.

The main step of the proof, as explained in Section 4.1, is to prove (17). Specifically, we aim to show that

$$\max_{1 \leqslant j \leqslant n} \|(\widetilde{U}_k^T X)_j - (U^{\mathrm{T}}\mathbb{E}(X))_j\| < \frac{1}{5}\Delta \tag{90}$$

holds with high probability.

Recall that throughout the paper, we always assume $\|E\| \leqslant 2(\sqrt{n} + \sqrt{p})$. We start with the decompositions

$$U^{\mathrm{T}}\mathbb{E}(X) = \Lambda V^{\mathrm{T}} = \begin{pmatrix} \Lambda_r V_r^{\mathrm{T}} \\ 0 \end{pmatrix}$$

and

$$\widetilde{U}_r^T X = \widetilde{\Lambda}_k \widetilde{V}_k^{\mathrm{T}} = \begin{pmatrix} \widetilde{\Lambda}_r \widetilde{V}_r^{\mathrm{T}} \\ \widetilde{\Lambda}_{[\![r+1,k]\!]} \widetilde{V}_{[\![r+1,k]\!]}^{\mathrm{T}} \end{pmatrix}.$$

Observe that

$$\|\widetilde{\Lambda}_{[\![r+1,k]\!]} \widetilde{V}_{[\![r+1,k]\!]}^{\mathrm{T}}\| \leqslant \widetilde{\sigma}_{r+1} \leqslant \|E\| \leqslant \frac{1}{20}\Delta$$

by Weyl's inequality and $\Delta \geqslant \sigma_r \geqslant 20\|E\|$ from the supposition. Hence,

$$\begin{aligned}&\max_{1 \leqslant j \leqslant n} \|(\widetilde{U}_k^T X)_j - (U^{\mathrm{T}}\mathbb{E}(X))_j\| \\ &\leqslant \max_{1 \leqslant j \leqslant n} \left(\|(\widetilde{\Lambda}_r \widetilde{V}_r^{\mathrm{T}})_j - (\Lambda_r V_r^{\mathrm{T}})_j\| + \|(\widetilde{\Lambda}_{[\![r+1,k]\!]} \widetilde{V}_{[\![r+1,k]\!]}^{\mathrm{T}})_j\|\right) \\ &\leqslant \max_{1 \leqslant j \leqslant n} \|(\widetilde{\Lambda}_r \widetilde{V}_r^{\mathrm{T}})_j - (\Lambda_r V_r^{\mathrm{T}})_j\| + \frac{1}{20}\Delta \\ &= \max_{1 \leqslant j \leqslant n} \|e_j^{\mathrm{T}} \widetilde{V}_r \widetilde{\Lambda}_r - e_j^{\mathrm{T}} V_r \Lambda_r\| + \frac{1}{20}\Delta = \|\widetilde{V}_r \widetilde{\Lambda}_r - V_r \Lambda_r\|_{2,\infty} + \frac{1}{20}\Delta\end{aligned}$$

$$\leqslant \|\widetilde{V}_r\widetilde{\Lambda}_r - V_r\widetilde{\Lambda}_r\|_{2,\infty} + \|V_r(\widetilde{\Lambda}_r - \Lambda_r)\|_{2,\infty} + \frac{1}{20}\Delta. \tag{91}$$

Due to non-uniqueness of the choice of V in the SVD of $\mathbb{E}(X)$, we choose a specified V_r such that the conclusion of Corollary 17 holds (with $b = 20$): that is, with probability at least $1 - 40(N+n)^{-L}$,

$$\|\widetilde{V}_r\widetilde{\Lambda}_r - V_r\widetilde{\Lambda}_r\|_{2,\infty} \leqslant 45r\sqrt{(L+7)\log(n+p)}(1 + \|V_r\|_{2,\infty}) + 8\|V_r\|_{2,\infty}\frac{(\sqrt{n}+\sqrt{p})^2}{\sigma_r}.$$

Note that

$$\|V_r\|_{2,\infty} \leqslant \|V\|_{2,\infty} = \max_i \|e_i^{\mathrm{T}}V\| = \max_i \sqrt{\frac{1}{d_{z_i}}\sum_{j=1}^{k} W_{z_i,j}^2} \leqslant \frac{1}{\sqrt{c_{\min}}} \leqslant 1.$$

Continuing from (91), we further obtain

$$\begin{aligned}
&\max_{1\leqslant j\leqslant n} \|(\widetilde{U}_k^T X)_j - (U^{\mathrm{T}}\mathbb{E}(X))_j\| \\
&\leqslant \|\widetilde{V}_r\widetilde{\Lambda}_r - V_r\widetilde{\Lambda}_r\|_{2,\infty} + \|V_r\|_{2,\infty}\|\widetilde{\Lambda}_r - \Lambda_r\| + \frac{1}{20}\Delta \\
&\leqslant 90k\sqrt{(L+7)\log(n+p)} + \frac{8(\sqrt{n}+\sqrt{p})^2}{\sqrt{c_{\min}}\sigma_r} + \frac{2(\sqrt{n}+\sqrt{p})}{\sqrt{c_{\min}}} + \frac{1}{20}\Delta \\
&\leqslant \frac{1}{20}\Delta + \frac{1}{200}\Delta + \frac{1}{20}\Delta + \frac{1}{20}\Delta < \frac{1}{5}\Delta
\end{aligned}$$

by Weyl's inequality $\|\widetilde{\Lambda}_r - \Lambda_r\| \leqslant \|E\|$ and the suppositions that $\sigma_r \geqslant 40(\sqrt{n}+\sqrt{p})$ and

$$\Delta \geqslant \max\left\{\frac{40(\sqrt{n}+\sqrt{p})}{\sqrt{c_{\min}}}, 1800k\sqrt{(L+7)\log(n+p)}\right\}.$$

This concludes the proof.

Appendix A. Proofs of (26), (27), Proposition 25 and (15)

A.1. **Proof of** (26)**.** It follows from (23) that for the Schatten p-norm with $p \geqslant 2$,

$$\begin{aligned}
\min_{O\in\mathbb{O}^{r\times r}} \|UO - V\|_p^2 &= \min_{O\in\mathbb{O}^{r\times r}} \|(UO-V)^{\mathrm{T}}(UO-V)\|_{p/2} \\
&= \min_{O\in\mathbb{O}^{r\times r}} \|2I_r - O^{\mathrm{T}}U^{\mathrm{T}}V - V^{\mathrm{T}}UO\|_{p/2}.
\end{aligned} \tag{92}$$

Note that the SVD of $U^{\mathrm{T}}V$ can be written as $O_1^{\mathrm{T}}\cos\Theta O_2$ where O_1, O_2 are orthogonal matrices and $\cos\Theta := \cos\angle(U,V) = \operatorname{diag}(\cos\theta_1, \cdots, \cos\theta_r)$. Continuing from (92), by the definition of unitarily invariant norms, we further have

$$\begin{aligned}
(92) &= \min_{O\in\mathbb{O}^{r\times r}} \|2I_r - \cos\Theta \cdot O_2 O^{\mathrm{T}} O_1^{\mathrm{T}} - O_1 O O_2^{\mathrm{T}} \cos\Theta\|_{p/2} \\
&\leqslant \|2I_r - 2\cos\Theta\|_{p/2} = 2\left(\sum_{i=1}^{r}(1-\cos\theta_i)^{p/2}\right)^{2/p} \\
&\leqslant 2\left(\sum_{i=1}^{r}\sin^p\theta_i\right)^{2/p} = 2\|\sin\Theta\|_p^2,
\end{aligned}$$

where we denote $\sin\Theta := \sin\angle(U,V)$. In the first inequality, we choose $O = O_1^{\mathrm{T}}O_2$. The second inequality follows from $1 - \cos\theta_i \leqslant 1 - \cos^2\theta_i = \sin^2\theta_i$.

For the lower bound, we still consider

$$\begin{aligned}(92) &= \min_{O\in\mathbb{O}^{r\times r}} \|2I_r - \cos\Theta\cdot O_2O^{\mathrm{T}}O_1^{\mathrm{T}} - O_1OO_2^{\mathrm{T}}\cos\Theta\|_{p/2} \\ &= \min_{Y\in\mathbb{O}^{r\times r}} \|2I_r - \cos\Theta\cdot Y - Y^{\mathrm{T}}\cos\Theta\|_{p/2} := \min_{Y\in\mathbb{O}^{r\times r}} \|B_Y\|_{p/2}.\end{aligned}$$

Observe that B_Y is positive semidefinite. To see this, let x be an arbitrary unit vector in $\mathbb{R}^r$ and we have

$$x^{\mathrm{T}}B_Yx = 2 - 2x^{\mathrm{T}}\cos\Theta\cdot Yx \geqslant 2(1-|x^{\mathrm{T}}\cos\Theta\cdot Yx|) \geqslant 2(1-\|\cos\Theta\|)\geqslant 0.$$

Denote $p' = p/2$ for brevity. We use the following variational formula for the Schatten norms of positive semidefinite matrices:

$$\|B_Y\|_{p'} = \max_{\|X\|_{q'}\leqslant 1} \mathrm{tr}(B_YX), \tag{93}$$

where $\|X\|_{q'}$ is the Schatten-q' norm of $X\in\mathbb{R}^{r\times r}$ and $\frac{1}{p'}+\frac{1}{q'}=1$. To prove (93), by the Hölder's inequality for Schatten norms,

$$\max_{\|X\|_{q'}\leqslant 1} \mathrm{tr}(B_YX) \leqslant \|B_Y\|_{p'} \max_{\|X\|_{q'}\leqslant 1}\|X\|_{q'} \leqslant \|B_Y\|_{p'}.$$

On the other hand, taking $X = B_Y^{p'-1}/\|B_Y^{p'-1}\|_{q'}$,

$$\max_{\|X\|_{q'}\leqslant 1} \mathrm{tr}(B_YX) \geqslant \mathrm{tr}(B_Y^{p'})/\|B_Y^{p'-1}\|_{q'} = \|B_Y\|_{p'}^{p'}/\|B_Y\|_{p'}^{p'-1} = \|B_Y\|_{p'},$$

where we used $\mathrm{tr}(B_Y^{p'}) = \|B_Y\|_{p'}^{p'}$ since B_Y is positive semidefinite and $\|B_Y^{p'-1}\|_{q'} = \|B_Y\|_{p'}^{p'-1}$ due to $\frac{1}{p'}+\frac{1}{q'}=1$. This proves (93).

Set $S=\sin^2\Theta$ for simplicity and let

$$X = \frac{S^{p'-1}}{\|S^{p'-1}\|_{q'}} = \frac{S^{p'-1}}{\|S\|_{p'}^{p'-1}}.$$

We continue from (93):

$$\begin{aligned}\|B_Y\|_{p'} &\geqslant \mathrm{tr}\left(2X - \cos\Theta\cdot YX - Y^{\mathrm{T}}\cos\Theta X\right) = 2\,\mathrm{tr}(X - X\cos\Theta\cdot Y) \\ &\geqslant 2\left(\mathrm{tr}(X) - \|X\cos\Theta\|_*\right),\end{aligned}$$

where we applied the Hölder's inequality $|\mathrm{tr}(X\cos\Theta\cdot Y)| \leqslant \|Y\|\cdot\|X\cos\Theta\|_* = \|X\cos\Theta\|_*$. Plugging in X and S, we get

$$\begin{aligned}\|B_Y\|_{p'} &\geqslant \frac{2}{\|S\|_{p'}^{p'-1}}\left(\mathrm{tr}(S^{p'-1}) - \|S^{p'-1}\cos\Theta\|_*\right) \\ &= \frac{2}{\|\sin^2\Theta\|_{p'}^{p'-1}}\left(\sum_{i=1}^r (\sin\theta_i)^{2(p'-1)}(1-\cos\theta_i)\right).\end{aligned}$$

Note that $\|\sin^2\Theta\|_{p'}^{p'-1} = \|\sin\Theta\|_{2p'}^{2(p'-1)}$ and

$$1-\cos\theta_i = 2\sin^2(\theta_i/2) = \frac{\sin^2\theta_i}{2\cos^2(\theta_i/2)} \geqslant \frac{1}{2}\sin^2\theta_i.$$

We further obtain

$$\|B_Y\|_{p'} \geqslant \frac{1}{\|\sin\Theta\|_{2p'}^{2(p'-1)}}\sum_{i=1}^r(\sin\theta_i)^{2p'} = \frac{\|\sin\Theta\|_{2p'}^{2p'}}{\|\sin\Theta\|_{2p'}^{2(p'-1)}} = \|\sin\Theta\|_{2p'}^2 = \|\sin\Theta\|_p^2.$$

Therefore,

$$(92) = \min_{Y \in \mathbb{O}^{r\times r}} \|B_Y\|_{p/2} \geqslant \|\sin\Theta\|_p^2.$$

This completes the proof of (26).

A.2. **Proof of** (27)**.** For simplicity, denote $\cos\Theta = \cos\angle(U,V)$ and $\sin\Theta = \sin\angle(U,V)$. Note that for any orthogonal matrices $Y, Z \in \mathbb{R}^{r\times r}$,

$$|||UYZ^{\mathrm{T}} - V||| = |||UY - VZ|||.$$

We use [21, Theorem VII.1.8]: there exist $r\times r$ orthogonal matrices Y, Z and $n\times n$ orthogonal matrix Q such that if $2r \leqslant n$, then

$$QUY = \begin{pmatrix} I_r \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad QUZ = \begin{pmatrix} \cos\Theta \\ \sin\Theta \\ 0 \end{pmatrix}.$$

Hence,

$$\min_{O\in\mathbb{O}^{r\times r}} |||UO - V||| \leqslant |||UYZ^{\mathrm{T}} - V||| = |||QUY - QVZ||| = \left|\left|\left| \begin{pmatrix} I_r - \cos\Theta \\ -\sin\Theta \end{pmatrix} \right|\right|\right|.$$

If $2r > n$, then

$$QUY = \begin{pmatrix} I_{n-r} & 0 \\ 0 & I_{2r-n} \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad QUZ = \begin{pmatrix} \cos\Theta_1 & 0 \\ 0 & I_{2r-n} \\ \sin\Theta_1 & 0 \end{pmatrix},$$

where Θ_1 is a diagonal matrix composed of the largest $n-r$ diagonal entries of Θ (note that the remaining diagonal entries of Θ are all zero). Therefore, by unitary equivalent, we still have

$$\min_{O\in\mathbb{O}^{r\times r}} |||UO - V||| \leqslant |||UYZ^{\mathrm{T}} - V||| = |||QUY - QVZ||| = \left|\left|\left| \begin{pmatrix} I_r - \cos\Theta \\ -\sin\Theta \end{pmatrix} \right|\right|\right|. \quad (94)$$

Note that the matrix on the right-hand side of (94) has singular values

$$\sqrt{(1-\cos\theta_i)^2 + \sin^2\theta_i} = 2\sin\left(\frac{\theta_i}{2}\right)$$

for $i = 1, \cdots, r$. Then by Theorem 21,

$$\min_{O\in\mathbb{O}^{r\times r}} |||UO - V||| \leqslant f(2\sin(\theta_1/2), \cdots, 2\sin(\theta_r/2))$$

for the symmetric gauge function f associated with the norm. Combining the above fact with the inequality $\sin(\theta/2) = \frac{1}{2}\frac{\sin\theta}{\cos(\theta/2)} \leqslant \frac{\sin\theta}{\sqrt{2}}$ for $\theta \in [0, \pi/2]$ and Theorem 23, we get the bound

$$\min_{O\in\mathbb{O}^{r\times r}} |||UO - V||| \leqslant f(\sqrt{2}\sin(\theta_1), \cdots, \sqrt{2}\sin(\theta_r)) = |||\sqrt{2}\sin\angle(U,V)|||.$$

A.3. **Proof of Proposition 25.** For any orthogonal matrix O, we first have

$$\begin{aligned}\|x^{\mathrm{T}}(V-UO)\| &\leqslant \|x^{\mathrm{T}}(V-P_UV)\| + \|x^{\mathrm{T}}(UU^{\mathrm{T}}V-UO)\| \\ &= \|x^{\mathrm{T}}(V-P_UV)\| + \|x^{\mathrm{T}}U(U^{\mathrm{T}}V-O)\| \\ &\leqslant \|x^{\mathrm{T}}(V-P_UV)\| + \|x^{\mathrm{T}}U\|\|U^{\mathrm{T}}V-O\|.\end{aligned}$$

It remains to estimate $\|U^{\mathrm{T}}V-O\|$. Now consider a specific orthogonal matrix $O = O_1O_2^{\mathrm{T}}$, where as per (28), we have $U^{\mathrm{T}}V = O_1\cos\angle(U,V)O_2^{\mathrm{T}}$. Hence,

$$\begin{aligned}\|U^{\mathrm{T}}V-O\| &= \|O_1\cos\angle(U,V)O_2^{\mathrm{T}} - O_1O_2^{\mathrm{T}}\| = \|\cos\angle(U,V) - I_r\| \\ &= 1-\cos\theta_r \leqslant 1-\cos^2\theta_r = \sin^2\theta_r = \|\sin\angle(U,V)\|^2.\end{aligned}$$

Putting these estimates together, we arrive at

$$\|x^{\mathrm{T}}(V-UO)\| \leqslant \|x^{\mathrm{T}}(V-P_UV)\| + \|x^{\mathrm{T}}U\|\|\sin\angle(U,V)\|^2.$$

The other inequalities can be proved immediately by noting that

$$\begin{aligned}\left|x^{\mathrm{T}}(V-UO)y\right| &\leqslant \left|x^{\mathrm{T}}(V-P_UV)y\right| + \left|x^{\mathrm{T}}(UU^{\mathrm{T}}V-UO)y\right| \\ &\leqslant \left|x^{\mathrm{T}}(V-P_UV)y\right| + \|x^{\mathrm{T}}U\|\|U^{\mathrm{T}}V-O\|\end{aligned}$$

and

$$\begin{aligned}\|V-UO\|_{2,\infty} &\leqslant \|V-P_UV\|_{2,\infty} + \|UU^{\mathrm{T}}V-UO\|_{2,\infty} \\ &\leqslant \|V-P_UV\|_{2,\infty} + \|U\|_{2,\infty}\|U^{\mathrm{T}}V-O\|.\end{aligned}$$

A.4. **Proof of** (15)**.** As in (28), from the SVD of $U^{\mathrm{T}}\widetilde{U}_r = O_1\cos\angle(U,\widetilde{U}_r)O_2^{\mathrm{T}}$, we choose the orthogonal matrix $O = O_1O_2^{\mathrm{T}}$. For notational simplicity, let us denote $\cos\angle(U,\widetilde{U}_r) = \mathrm{diag}(\cos\theta_1,\cdots,\cos\theta_r) = \cos\Theta$.

Using a similar argument as in the proof of Proposition 25, we obtain

$$\|\widetilde{U}_r\widetilde{D}_r - UO\widetilde{D}_r\|_{2,\infty} \leqslant \|\widetilde{U}_r\widetilde{D}_r - P_U\widetilde{U}_r\widetilde{D}_r\|_{2,\infty} + \|U\|_{2,\infty}\|(U^{\mathrm{T}}\widetilde{U}_r - O)\widetilde{D}_r\|. \quad (95)$$

It suffices to establish a bound for $\|(U^{\mathrm{T}}\widetilde{U}_r - O)\widetilde{D}_r\|$ in the second term on the right-hand side of (95). Such a bound has been previously established in [79, Lemma 15]. To ensure our proof is self-contained, we repeat their proof here and provide the explicit constants.

Let U_0 denote the matrix whose columns are orthonormal and span the complement of the subspace U. We first show that

$$O - U^{\mathrm{T}}\widetilde{U}_r = 2O_1\cdot\sin(\Theta/2)\cdot(\sin\Theta)^{-1}\sin(\Theta/2)\cdot O_3^{\mathrm{T}}U_0^{\mathrm{T}}\widetilde{U}_r \quad (96)$$

for some orthogonal matrix O_3. Here $\sin(\Theta/2) = \mathrm{diag}(\sin(\theta_1/2),\cdots,\sin(\theta_r/2))$ and

$$(\sin\Theta)^{-1}\sin(\Theta/2) = \mathrm{diag}\left(\frac{\sin(\theta_1/2)}{\sin(\theta_1)},\cdots,\frac{\sin(\theta_r/2)}{\sin(\theta_r)}\right).$$

To see (96), since

$$\begin{aligned}(\widetilde{U}_r^{\mathrm{T}}U_0)(\widetilde{U}_r^{\mathrm{T}}U_0)^{\mathrm{T}} = \widetilde{U}_r^{\mathrm{T}}U_0U_0^{\mathrm{T}}\widetilde{U}_r &= \widetilde{U}_r^{\mathrm{T}}(I-UU^{\mathrm{T}})\widetilde{U}_r \\ &= I - O_2\cos^2(\Theta)O_2^{\mathrm{T}} = O_2\sin^2(\Theta)O_2^{\mathrm{T}},\end{aligned}$$

the SVD of $\widetilde{U}_r^{\mathrm{T}}U_0$ is given by

$$\widetilde{U}_r^{\mathrm{T}}U_0 = O_2(\sin\Theta)O_3^{\mathrm{T}} \quad (97)$$

for some orthogonal matrix O_3. Combining (97) with

$$O - U^{\mathrm{T}}\widetilde{U}_r = O_1(I - \cos\Theta)O_2^{\mathrm{T}} = 2O_1 \sin^2(\Theta/2)O_2^{\mathrm{T}},$$

we prove (96) and consequently

$$(O - U^{\mathrm{T}}\widetilde{U}_r)\widetilde{D}_r = 2O_1 \cdot \sin(\Theta/2) \cdot (\sin\Theta)^{-1} \sin(\Theta/2) \cdot O_3^{\mathrm{T}} U_0^{\mathrm{T}} \widetilde{U}_r \widetilde{D}_r \tag{98}$$

To bound $\|(O - U^{\mathrm{T}}\widetilde{U}_r)\widetilde{D}_r\|$, first observe

$$\|\sin(\Theta/2)\| \leqslant \frac{\sqrt{2}}{2}\|\sin\Theta\| \quad \text{and} \quad \|(\sin\Theta)^{-1}\sin(\Theta/2)\| \leqslant \frac{\sqrt{2}}{2}$$

by the facts that $\cos(\theta/2) \geqslant 1/\sqrt{2}$ and $\sin(\theta/2) = \frac{\sin\theta}{2\cos(\theta/2)} \leqslant \frac{\sqrt{2}}{2}\sin\theta$ for $\theta \in [0, \pi/2]$. Then continuing from (98), we have

$$\|(O - U^{\mathrm{T}}\widetilde{U}_r)\widetilde{D}_r\| \leqslant \|\sin\Theta\| \cdot \|U_0^{\mathrm{T}}\widetilde{U}_r\widetilde{D}_r\|.$$

It remains to bound $\|U_0^{\mathrm{T}}\widetilde{U}_r\widetilde{D}_r\|$. Note that $\widetilde{U}_r\widetilde{D}_r = \widetilde{A}\widetilde{V}_r$. Multiplying U_0^{T} on both sides and using $U_0^{\mathrm{T}}A = 0$, we get the desired bound

$$\|U_0^{\mathrm{T}}\widetilde{U}_r\widetilde{D}_r\| = \|U_0^{\mathrm{T}}\widetilde{A}\widetilde{V}_r\| = \|U_0^{\mathrm{T}}(A+E)\widetilde{V}_r\| = \|U_0^{\mathrm{T}}E\widetilde{V}_r\| \leqslant \|E\|.$$

This completes the proof.

Appendix B. Proofs of Lemma 27 and Lemma 28

B.1. **Proof of Lemma 27.** By the rotational invariance of E and definition of $\mathcal{E}$, we observe that for any orthogonal matrices

$$\mathcal{O}_1 = \begin{pmatrix} O_1 & 0 \\ 0 & O_2 \end{pmatrix}, \; \mathcal{O}_2 = \begin{pmatrix} \widehat{O}_1 & 0 \\ 0 & \widehat{O}_2 \end{pmatrix}$$

where $O_1, \widehat{O}_1 \in \mathbb{R}^{N\times N}$ and $O_2, \widehat{O}_2 \in \mathbb{R}^{n\times n}$ are orthogonal matrices,

$$\mathcal{O}_1\mathcal{E}\mathcal{O}_2 \sim \mathcal{E}.$$

Consequently,

$$\mathbf{x}^{\mathrm{T}}\left(G(z) - \Phi(z)\right)\mathbf{y} \sim (\mathcal{O}_1\mathbf{x})^{\mathrm{T}}\left(G(z) - \Phi(z)\right)(\mathcal{O}_2\mathbf{y}).$$

Hence, it suffices to assume $\mathbf{x} = (x_1, 0\cdots, 0, x_{N+1}, 0, \cdots, 0)^{\mathrm{T}}$ with $x_1^2 + x_{N+1}^2 = 1$ and $\mathbf{y} = (y_1, 0\cdots, 0, y_{N+1}, 0, \cdots, 0)^{\mathrm{T}}$ with $y_1^2 + y_{N+1}^2 = 1$. Furthermore,

$$\begin{aligned} &\mathbf{x}^{\mathrm{T}}\left(G(z) - \Phi(z)\right)\mathbf{y} \\ &= x_1y_1(G_{11}(z) - \Phi_{11}(z)) + x_{N+1}y_{N+1}(G_{N+1,N+1}(z) - \Phi_{N+1,N+1}(z)) \\ &\quad + x_1y_{N+1}G_{1,N+1}(z) + x_{N+1}y_1G_{N+1,1}(z). \end{aligned} \tag{99}$$

In order to prove Lemma 27, it suffices to show that for each fixed $k \in [\![1, N+n]\!]$,

$$|G_{kk}(z) - \Phi_{kk}(z)| \leqslant \frac{2b^2}{(b-1)^2}\frac{\sqrt{(K+1)\log(N+n)}}{|z|^2} \tag{100}$$

with probability at least $1 - 4(N+n)^{-(K+1)}$ and for fixed $i \neq j \in [\![1, N+n]\!]$,

$$|G_{ij}(z)| \leqslant 2\sqrt{2}\left(\frac{b}{b-1}\right)^2 \frac{\sqrt{(K+1)\log(N+n)}}{|z|^2} \tag{101}$$

with probability at least $1 - 0.5(N+n)^{-(K+1)}$.

If so, continuing from (99), we find that

$$\begin{aligned}&\left|\mathbf{x}^{\mathrm{T}}\left(G(z)-\Phi(z)\right)\mathbf{y}\right|\\&\leqslant\left(2\left(\frac{b}{b-1}\right)^2+2\sqrt{2}\left(\frac{b}{b-1}\right)^2\right)\frac{\sqrt{(K+1)\log(N+n)}}{|z|^2}\\&\leqslant 5\left(\frac{b}{b-1}\right)^2\frac{\sqrt{(K+1)\log(N+n)}}{|z|^2}\end{aligned}$$

with probability at least $1-9(N+n)^{-(K+1)}$. Here, we use the fact that $|x_1y_1|+|x_{N+1}y_{N+1}|\leqslant 1$ by Cauchy-Schwarz inequality. The proofs of Equation (100) and Equation (101) closely resemble the proof presented in [64, Lemma 28], with only minor cosmetic modifications. For the sake of brevity, we will omit the detailed proofs here.

B.2. **Proof of Lemma 28.** In this section, we prove Lemma 28 using Lemma 27 and a standard ε-net argument.

For convenience, denote $\Delta(z):=G(z)-\Phi(z)$. We first show that for any fixed $z\in\mathbb{C}$ with $|z|\geqslant 2b(\sqrt{N}+\sqrt{n})$,

$$\|\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\|\leqslant 10\left(\frac{b}{b-1}\right)^2\frac{\sqrt{(K+1)\log(N+n)+2(\log 9)r}}{|z|^2}$$

with probability at least $1-9(N+n)^{-(K+1)}$.

Let $\mathcal{N}$ be the 1/4-net of the unit sphere $\mathbb{S}^{2r-1}$. A simple volume argument (see for instance [70, Corollary 4.2.13]) shows $\mathcal{N}$ can be chosen such that $|\mathcal{N}|\leqslant 9^{2r}$. Furthermore, since for any $\mathbf{y}\in\mathbb{S}^{2r-1}$, there exists a $\mathbf{x}\in\mathcal{N}$ such that $\|\mathbf{y}-\mathbf{x}\|\leqslant 1/4$, we have

$$\begin{aligned}\left|\mathbf{y}^{\mathrm{T}}\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\mathbf{y}\right|&\leqslant\left|\mathbf{x}^{\mathrm{T}}\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\mathbf{x}\right|+\left|(\mathbf{y}-\mathbf{x})^{\mathrm{T}}\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\mathbf{x}\right|+\left|\mathbf{y}^{\mathrm{T}}\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}(\mathbf{y}-\mathbf{x})\right|\\&\leqslant\left|\mathbf{x}^{\mathrm{T}}\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\mathbf{x}\right|+\frac{1}{2}\|\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\|.\end{aligned}$$

Therefore, $\|\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\|\leqslant 2\max_{\mathbf{x}\in\mathcal{N}}\left|\mathbf{x}^{\mathrm{T}}\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\mathbf{x}\right|$ and for any $K_1>0$ and $(\sqrt{N}+\sqrt{n})^2\geqslant 32(K_1+1)\log(N+n)$, by the union bound,

$$\begin{aligned}&\mathbb{P}\left(\|\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\|\geqslant 10\left(\frac{b}{b-1}\right)^2\frac{\sqrt{(K_1+1)\log(N+n)}}{|z|^2}\right)\\&\leqslant\mathbb{P}\left(\max_{\mathbf{x}\in\mathcal{N}}\left|\mathbf{x}^{\mathrm{T}}\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\mathbf{x}\right|\geqslant 5\left(\frac{b}{b-1}\right)^2\frac{\sqrt{(K_1+1)\log(N+n)}}{|z|^2}\right)\\&\leqslant 9^{2r+1}(N+n)^{-(K_1+1)},\end{aligned}$$

where in the last inequality, we applied Lemma 27. Now choose $K_1=K+\frac{2\log 9}{\log(N+n)}r$ and assume

$$(\sqrt{N}+\sqrt{n})^2\geqslant 32(K_1+1)\log(N+n)=32(K+1)\log(N+n)+64(\log 9)r.$$

The conclusion becomes that

$$\|\mathcal{U}^{\mathrm{T}}\Delta(z)\mathcal{U}\|\leqslant 10\left(\frac{b}{b-1}\right)^2\frac{\sqrt{(K+1)\log(N+n)+2(\log 9)r}}{|z|^2}$$

with probability at least $1-10(N+n)^{-(K+1)}$.

In particular, for any $z \in \mathsf{D} = \{z \in \mathbb{C} : 2b(\sqrt{N} + \sqrt{n}) \leqslant |z| \leqslant 2n^3\}$,

$$\|\mathcal{U}^T \Delta(z) \mathcal{U}\| \leqslant 10 \left(\frac{b}{b-1}\right)^2 \frac{\sqrt{(K+7)\log(N+n) + 2(\log 9)r}}{|z|^2}$$

with probability at least $1 - 9(N+n)^{-(K+7)}$, as long as

$$(\sqrt{N} + \sqrt{n})^2 \geqslant 32(K+7)\log(N+n) + 64(\log 9)r. \tag{102}$$

Let $\mathcal{N}$ be a 1-net of D. A simple volume argument (see for instance [65, Lemma 3.3]) shows $\mathcal{N}$ can be chosen so that $|\mathcal{N}| \leqslant (1+8n^3)^2 < n^7$. By the union bound,

$$\max_{z \in \mathcal{N}} |z|^2 \left\|\mathcal{U}^{\mathrm{T}} \Delta(z) \mathcal{U}\right\| \leqslant 10 \left(\frac{b}{b-1}\right)^2 \sqrt{(K+7)\log(N+n) + 2(\log 9)r} \tag{103}$$

with probability at least $1 - 9(N+n)^{-K}$. We now wish to extend this bound to all $z \in \mathsf{D}$.

Define the functions

$$f(z) := z^2 \mathcal{U}^{\mathrm{T}} G(z) \mathcal{U}, \qquad g(z) := z^2 \mathcal{U}^{\mathrm{T}} \Phi(z) \mathcal{U}.$$

In order to complete the proof, it suffices to show that f and g are $\frac{3b^2}{(b-1)^2}$-Lipschitz in D. In other words, we want to show that $\|f(z) - f(w)\| \leqslant \frac{3b^2}{(b-1)^2}|z-w|$ and $\|g(z) - g(w)\| \leqslant \frac{3b^2}{(b-1)^2}|z-w|$ for all $z, w \in \mathsf{D}$. Indeed, in view of (103), if $z \in \mathsf{D}$, then there exists $w \in \mathcal{N}$ so that $|z-w| \leqslant 1$, and hence

$$\begin{aligned}
&|z|^2 \left\|\mathcal{U}^{\mathrm{T}} G(z) \mathcal{U} - \mathcal{U}^T \Phi(z) \mathcal{U}\right\| \\
&\leqslant \|f(z) - f(w)\| + \|f(w) - g(w)\| + \|g(w) - g(z)\| \\
&\leqslant \frac{6b^2}{(b-1)^2} + |w|^2 \left\|\mathcal{U}^{\mathrm{T}} G(w) \mathcal{U} - \mathcal{U}^T \Phi(w) \mathcal{U}\right\| \\
&\leqslant \frac{6b^2}{(b-1)^2} + 10 \left(\frac{b}{b-1}\right)^2 \sqrt{(K+7)\log(N+n) + 2(\log 9)r} \\
&< 11 \left(\frac{b}{b-1}\right)^2 \sqrt{(K+7)\log(N+n) + 2(\log 9)r} = \eta,
\end{aligned}$$

where we used the Lipschitz continuity of f and g in the second inequality. In the last inequality, we use (102) to obtain a crude bound $N + n \geqslant 32 \cdot 7 \log(N+n)$ and hence $N + n \geqslant 1600$. This implies $\sqrt{(K+7)\log(N+n) + 2(\log 9)r} \geqslant \sqrt{7\log(1600) + 2(\log 9)} \approx 7.5$.

It remains to show that f and g are $\frac{3b^2}{(b-1)^2}$-Lipschitz in D. Recall that we work on the event where $\|E\| \leqslant 2(\sqrt{N} + \sqrt{n})$ through the proofs. Let $z, w \in \mathsf{D}$, and assume without loss of generality that $|z| \geqslant |w| \geqslant 2b(\sqrt{N} + \sqrt{n})$. Then

$$\begin{aligned}
\|f(z) - f(w)\| &\leqslant \|z^2 \mathcal{U}^{\mathrm{T}} G(z) \mathcal{U} - zw \mathcal{U}^{\mathrm{T}} G(z) \mathcal{U}\| + \|zw \mathcal{U}^{\mathrm{T}} G(z) \mathcal{U} - w^2 \mathcal{U}^{\mathrm{T}} G(z) \mathcal{U}\| \\
&\quad + \|w^2 \mathcal{U}^{\mathrm{T}} G(z) \mathcal{U} - w^2 \mathcal{U}^{\mathrm{T}} G(w) \mathcal{U}\| \\
&\leqslant |z| \|G(z)\| |z-w| + |w| |z-w| \|G(z)\| + |w|^2 |z-w| \|G(z)\| \|G(w)\| \\
&\leqslant \frac{2b}{b-1}|z-w| + \frac{b^2}{(b-1)^2}|z-w| = \frac{b(3b-2)}{(b-1)^2}|z-w| \\
&\leqslant \frac{3b^2}{(b-1)^2}|z-w|,
\end{aligned}$$

where we used the resolvent identity $B^{-1} - C^{-1} = B^{-1}(C - B)C^{-1}$, Lemma 26, and the fact that $\frac{|w|}{|z|} \leqslant 1$. This shows that f is $\frac{3b^2}{(b-1)^2}$-Lipschitz in D.

The proof for g is similar. First, by the triangle inequality, we have

$$\begin{aligned}
&\|g(z) - g(w)\| \\
&\quad \leqslant \|z^2\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U} - zw\mathcal{U}^{\mathrm{T}}\Phi(z)U\| + \|zw\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U} - w^2\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\| \\
&\qquad + \|w^2\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U} - w^2\mathcal{U}^{\mathrm{T}}\Phi(w)\mathcal{U}\| \\
&\quad \leqslant |z||z-w|\|\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\| + |w||z-w|\|\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\| + |w|^2\|\mathcal{U}^{\mathrm{T}}(\Phi(z) - \Phi(w))\mathcal{U}\|.
\end{aligned} \tag{104}$$

Using the explicit expression in (34), we find that

$$\|\mathcal{U}^{\mathrm{T}}(\Phi(z) - \Phi(w))\mathcal{U}\| = \max\left\{\frac{|\phi_1(z) - \phi_1(w)|}{|\phi_1(z)\phi_1(w)|}, \frac{|\phi_2(z) - \phi_2(w)|}{|\phi_2(z)\phi_2(w)|}\right\}.$$

By (32) and the resolvent identity $B^{-1} - C^{-1} = B^{-1}(C - B)C^{-1}$,

$$\begin{aligned}
|\phi_1(z) - \phi_1(w)| &= |z - w - \operatorname{tr}\mathcal{I}^{\mathrm{d}}(G(z) - G(w))| \\
&= |z - w - (z-w)\operatorname{tr}\mathcal{I}^{\mathrm{d}}G(z)G(w)| \\
&\leqslant |z-w|\,(1 + (N+n)\|G(z)\|\|G(w)\|) \\
&\leqslant |z-w|\left(1 + \frac{b^2}{(b-1)^2}\frac{(N+n)}{|z||w|}\right) \\
&\leqslant \left(1 + \frac{1}{4(b-1)^2}\right)|z-w|,
\end{aligned}$$

where we used Lemma 26 and the facts that $\frac{N+n}{|w|} \leqslant \frac{|w|}{4b^2}$ and $\frac{|w|}{|z|} \leqslant 1$. The same upper bound also holds for $|\phi_2(z) - \phi_2(w)|$. Combining these estimates with (36), we have

$$\|\mathcal{U}^{\mathrm{T}}(\Phi(z) - \Phi(w))\mathcal{U}\| \leqslant \frac{1 + 1/4(b-1)^2}{(1 - 1/4b(b-1))^2}\frac{|z-w|}{|z||w|}.$$

Notice that $\|\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\| \leqslant \frac{1}{1-1/4b(b-1)}\frac{1}{|z|}$ for any $|z| \geqslant 2b(\sqrt{N} + \sqrt{n})$, which can be verified using (35) and the bounds in (36). Inserting these bounds into (104) yields that

$$\begin{aligned}
\|g(z) - g(w)\| &\leqslant \frac{2}{1 - 1/4b(b-1)}|z-w| + \frac{1 + 1/4(b-1)^2}{(1 - 1/4b(b-1))^2}|z-w| \\
&\leqslant \frac{4b(12b^3 - 24b^2 + 11b + 2)}{(4b^2 - 4b - 1)^2}|z-w| < \frac{3b^2}{(b-1)^2}|z-w|,
\end{aligned}$$

where the last inequality is check via Mathematica. Hence, g is $\frac{3b^2}{(b-1)^2}$-Lipschitz in D.

Appendix C. Proof of Lemma 52

In this section, we estimate $\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathbf{u}}_i\|$ for each $i \in I = [\![k, s]\!] \cup [\![r+k, r+s]\!]$. Recall the decomposition of $\widetilde{\mathbf{u}}_i$ in (45):

$$\widetilde{\mathbf{u}}_i = \Pi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i + \Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i, \tag{105}$$

where $\Pi(z)$ is a function to be further specified during the proof, and $\Xi(z) = G(z) - \Pi(z)$.

We split the estimation of $\|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\|$ for $i \in I$ into two cases: when $|\lambda_i|$ is large and when $|\lambda_i|$ is relatively small.

We start with the simpler case when $|\lambda_i| > n^2/2$. We choose

$$\Pi(\widetilde{\lambda}_i) := \frac{1}{\widetilde{\lambda}_i} I_{N+n} + \frac{1}{\widetilde{\lambda}_i^2}\mathcal{E}.$$

We work on the event

$$\max_{l \in [\![1,r_0]\!] : \sigma_l > \frac{1}{2}n^2} |\widetilde{\sigma}_l - \sigma_l| \leqslant \eta r.$$

By Lemma 31, this event holds with probability at least $1 - (N+n)^{-1.5r^2(K+4)}$. Hence, $|\widetilde{\lambda}_i| \geqslant |\lambda_i| - \eta r \geqslant 2b(\sqrt{N} + \sqrt{n})$. We apply Lemma 30 to get

$$\left\|\Xi(\widetilde{\lambda}_i)\right\| = \left\|G(\widetilde{\lambda}_i) - \Pi(\widetilde{\lambda}_i)\right\| \leqslant \frac{b}{b-1}\frac{\|\mathcal{E}\|^2}{|\widetilde{\lambda}_i|^3}. \tag{106}$$

Multiplying $\mathcal{U}_J^{\mathrm{T}}$ on both sides of (105), we obtain the following equation:

$$\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i = \mathcal{U}_J^{\mathrm{T}}\Pi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i + \mathcal{U}_J^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i.$$

Plugging in (46) and using the facts $\mathcal{U}_J^{\mathrm{T}}\mathcal{U}_I = 0$ and $\mathcal{U}_J^{\mathrm{T}}\mathcal{U}_J = I$, we further get

$$\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i = \frac{1}{\widetilde{\lambda}_i}\mathcal{D}_J\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i + \frac{1}{\widetilde{\lambda}_i^2}\mathcal{U}_J^{\mathrm{T}}\mathcal{E}\mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i + \mathcal{U}_J^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i,$$

which, by rearranging the terms, is reduced to

$$(\widetilde{\lambda}_i I - \mathcal{D}_J)\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i = \frac{1}{\widetilde{\lambda}_i}\mathcal{U}_J^{\mathrm{T}}\mathcal{E}\mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i + \widetilde{\lambda}_i\mathcal{U}_J^{\mathrm{T}}\Xi(\widetilde{\lambda}_i)\mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i.$$

Hence,

$$\min_{j \in J}|\widetilde{\lambda}_i - \lambda_j| \cdot \|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\| \leqslant \frac{1}{|\widetilde{\lambda}_i|}\|\mathcal{U}^{\mathrm{T}}\mathcal{E}\mathcal{U}\| \cdot \|\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\| + |\widetilde{\lambda}_i|\|\Xi(\widetilde{\lambda}_i)\| \cdot \|\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\|.$$

Note that $\|\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\| \leqslant \|\mathcal{E}\| + |\widetilde{\lambda}_i| \leqslant \frac{b}{b-1}|\widetilde{\lambda}_i|$ as in (62). Inserting (106) into the above inequality, we arrive at

$$\min_{j \in J}|\widetilde{\lambda}_i - \lambda_j| \cdot \|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\| \leqslant \frac{b}{b-1}\|\mathcal{U}^{\mathrm{T}}\mathcal{E}\mathcal{U}\| + \frac{b^2}{(b-1)^2}\frac{\|E\|^2}{|\widetilde{\lambda}_i|}. \tag{107}$$

For the remaining arguments, we work on the event

$$\mathsf{F} := \left\{\|\mathcal{U}^{\mathrm{T}}\mathcal{E}\mathcal{U}\| \leqslant 2\sqrt{r} + \sqrt{2(K+7)\log(N+n)}\right\}.$$

The following lemma is proved in [63, Lemma 18].

Lemma 35. *Let K be an arbitrary positive constant. With probability at least $1 - 2(N+n)^{-K}$, we have*

$$\|\mathcal{U}^{\mathrm{T}}\mathcal{E}\mathcal{U}\| \leqslant 2\sqrt{r} + \sqrt{2K\log(N+n)}.$$

Therefore, the event F holds with probability at least $1 - 2(N+n)^{-(K+7)}$. We continue the estimation of $\|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\|$ from (107). Note that

$$\|\mathcal{U}^{\mathrm{T}}\mathcal{E}\mathcal{U}\| \leqslant 2\sqrt{r} + \sqrt{2(K+7)\log(N+n)} < \eta.$$

Also, $\|E\|^2/|\widetilde{\lambda}_i| \leqslant 4(2\sqrt{n})^2/n^2 < \eta$ where we used the crude bound $|\widetilde{\lambda}_i| \geqslant \frac{1}{4}n^2$ by Weyl's inequality. It follows that

$$\min_{j \in J}|\widetilde{\lambda}_i - \lambda_j| \cdot \|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\| \leqslant \frac{b(2b-1)}{(b-1)^2}\eta. \tag{108}$$

To bound the left-hand side of (108), we first consider $i \in [\![k, s]\!]$. Then

$$\min_{j\in J} |\widetilde{\lambda}_i - \lambda_j| = \min_{j\in[\![1,k-1]\!]\cup[\![s+1,r]\!]} |\widetilde{\sigma}_i - \sigma_j| = \min\{\sigma_{k-1} - \widetilde{\sigma}_i, \widetilde{\sigma}_i - \sigma_{s+1}\}$$

by $|\widetilde{\sigma}_i - \sigma_i| \leqslant \eta r$ and the supposition $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$. Next, applying $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$ again, we get

$$\min_{j\in J} |\widetilde{\lambda}_i - \lambda_j| \geqslant \left(1 - \frac{1}{75\chi(b)}\right) \min\{\delta_{k-1}, \delta_s\}.$$

It follows from (108) that

$$\begin{aligned} \|\mathcal{U}_J^{\mathrm{T}} \widetilde{\mathbf{u}}_i\| &\leqslant \frac{\eta}{\min\{\delta_{k-1}, \delta_s\}} \left(1 - \frac{1}{75\chi(b)}\right)^{-1} \frac{b(2b-1)}{(b-1)^2} \\ &= \frac{75(2b-1)^3 b}{(b-1)^2(296b^2 - 296b + 75)} \frac{\eta}{\min\{\delta_{k-1}, \delta_s\}} \\ &< 3\frac{(b+1)^2}{(b-1)^2} \frac{\eta}{\min\{\delta_{k-1}, \delta_s\}} \end{aligned} \tag{109}$$

for every $i \in [\![k, s]\!]$ satisfying $\lambda_i = \sigma_i \geqslant n^2/2$. The last inequality was checked by Mathematica. Finally, for $i \in [\![r + k, r + s]\!]$ such that $|\lambda_i| \geqslant n^2/2$, analogous arguments yield that the same bound

$$\|\mathcal{U}_J^{\mathrm{T}} \widetilde{\mathbf{u}}_i\| \leqslant 3\frac{(b+1)^2}{(b-1)^2} \frac{\eta}{\min\{\delta_{k-1}, \delta_s\}}. \tag{110}$$

The estimation of $\|\mathcal{U}_J^{\mathrm{T}} \widetilde{\mathbf{u}}_i\|$ when $|\lambda_i|$ is relatively small is more involved. From the previous discussion, it suffices to assume there is a certain $l_0 \in [\![1, r_0]\!]$ for which $\sigma_{l_0} \leqslant n^2/2$. We claim that there exists an index $i_0 \in [\![1, r_0]\!]$ such that $\sigma_j \leqslant n^3$ for $j \geqslant i_0$ and $\sigma_j > n^3$ for $j < i_0$, and

$$\delta_{i_0-1} = \sigma_{i_0-1} - \sigma_{i_0} \geqslant 75\chi(b)\eta r.$$

To determine i_0, we propose a simple iterative algorithm: start with σ_1. If $\sigma_1 \leqslant n^3$, set $i_0 = 1$ and terminate the algorithm, since $\sigma_0 = \infty$ and $\delta_0 = \infty$ by definition. Assume $\sigma_1 > n^3$ and evaluate σ_2. If $\sigma_2 \leqslant n^3 - 75\chi(b)\eta r$, set $i_0 = 2$ and exit. Assume $\sigma_2 > n^3 - 75\chi(b)\eta r$ and evaluate σ_3. We continue this process and terminate the algorithm with $i_0 = k$ unless

$$\sigma_1 > n^3, \sigma_2 > n^3 - 75\chi(b)\eta r, \cdots, \sigma_k > n^3 - 75\chi(b)\eta r. \tag{111}$$

Note that the condition (111) cannot hold for $k = l_0$ because $\sigma_{l_0} \leqslant n^2/2 < n^3 - 75\chi(b)\eta r$, based on the assumption that $(\sqrt{N} + \sqrt{n})^2 \geqslant 32(K + 7)\log(N + n) + 64(\log 9)r$. Therefore, i_0 must satisfy $i_0 \leqslant l_0 - 1$.

We shall fix such an index i_0 throughout the rest of the proof. We now turn our attention to estimating $\|\mathcal{U}_J^{\mathrm{T}} \widetilde{\mathbf{u}}_i\|$ for $i \in [\![k_0, s]\!] \cup [\![r + k_0, r + s]\!]$, where we define

$$k_0 := \max\{k, i_0\}$$

for the sake of notational simplicity. Note that $\min\{\delta_{k_0-1}, \delta_s\} \geqslant 75\chi(b)\eta r$. Furthermore, in this scenario, $|\lambda_i| \leqslant n^3$. We take

$$\Pi(\widetilde{\mathbf{u}}_i) = \Phi(\widetilde{\mathbf{u}}_i).$$

Continuing from (45), we have

$$\widetilde{\mathbf{u}}_i = \Phi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i + \Xi(\widetilde{\lambda}_i)\mathcal{A}\widetilde{\mathbf{u}}_i$$

and multiply on the left by $\mathcal{U}_J^{\mathrm{T}}$ to get

$$\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i = \mathcal{U}_J^{\mathrm{T}}\Phi(\tilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i + \mathcal{U}_J^{\mathrm{T}}\Xi(\tilde{\lambda}_i)\mathcal{A}\tilde{\mathbf{u}}_i. \tag{112}$$

Plugging in (46), we further have

$$\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i = \mathcal{U}_J^{\mathrm{T}}\Phi(\tilde{\lambda}_i)\mathcal{U}_J\mathcal{D}_J\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i + \mathcal{U}_J^{\mathrm{T}}\Xi(\tilde{\lambda}_i)\mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i,$$

where we used $\mathcal{U}_J^{\mathrm{T}}\Phi(\tilde{\lambda}_i)\mathcal{U}_I = 0$. Hence,

$$\left(I_{2(r-s+k-1)} - \mathcal{U}_J^{\mathrm{T}}\Phi(\tilde{\lambda}_i)\mathcal{U}_J\mathcal{D}_J\right)\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i = \mathcal{U}_J^{\mathrm{T}}\Xi(\tilde{\lambda}_i)\mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i. \tag{113}$$

We are now in position to bound $\|\mathcal{U}_J^{\mathrm{T}}\tilde{\mathbf{u}}_i\|$. This can be achieved by obtaining an upper bound for the right-hand side of (113) and estimating the smallest singular value of the matrix

$$I_{2(r-s+k-1)} - \mathcal{U}_J^{\mathrm{T}}\Phi(\tilde{\lambda}_i)\mathcal{U}_J\mathcal{D}_J \tag{114}$$

on the left-hand side of (113). We establish these estimates in the following two steps. Recall that

$$\xi(b) = 1 + \frac{1}{2(b-1)^2} \quad \text{and} \quad \chi(b) = 1 + \frac{1}{4b(b-1)}.$$

For each $k_0 \leqslant i \leqslant s$, by Theorem 29, there exists $k_0 \leqslant l_i \leqslant s$ such that $\tilde{\sigma}_i \in S_{\sigma_{l_i}}$ specified in (40), and

$$|\varphi(\tilde{\sigma}_i) - \sigma_{l_i}^2| \leqslant 20\xi(b)\chi(b)\eta r\left(\tilde{\sigma}_i + \chi(b)\sigma_{l_i}\right) \tag{115}$$

with probability at least $1 - 10(N+n)^{-K}$. Denote this event as E_1. Furthermore, on the event E_1, by Lemma 28, for every $k_0 \leqslant i \leqslant s$,

$$\left\|\mathcal{U}^{\mathrm{T}}\Xi(\tilde{\sigma}_i)\mathcal{U}\right\| \leqslant \frac{\eta}{\tilde{\sigma}_i^2}$$

holds with probability at least $1 - 9(N+n)^{-K}$. Let us denote this event as E_2.

In the remaining proof, we will work on the event $\mathsf{E}_1 \cap \mathsf{E}_2$ which holds with probability at least $1 - 19(N+n)^{-K}$.

Step 1. Upper bound for the right-hand side of (113). We first consider the case when $i \in [\![k_0, s]\!]$ and $\tilde{\lambda}_i = \tilde{\sigma}_i$. Note that $\mathcal{U}_J^{\mathrm{T}}\Xi(\tilde{\sigma}_i)\mathcal{U}$ is a sub-matrix of $\mathcal{U}^{\mathrm{T}}\Xi(\tilde{\sigma}_i)\mathcal{U}$. Thus, using (64) and the fact that the spectral norm of any sub-matrix is bounded by the spectral norm of the full matrix, we deduce that

$$\|\mathcal{U}_J^{\mathrm{T}}\Xi(\tilde{\sigma}_i)\mathcal{U}\cdot\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\| \leqslant \frac{\eta}{\tilde{\sigma}_i^2}\|\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\|.$$

Recall the bound in (62):

$$\|\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\| \leqslant \frac{b}{b-1}\tilde{\sigma}_i \tag{116}$$

Hence,

$$\left\|\mathcal{U}_J^{\mathrm{T}}\Xi(\tilde{\sigma}_i)\mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}\tilde{\mathbf{u}}_i\right\| \leqslant \frac{b}{b-1}\frac{\eta}{\tilde{\sigma}_i}. \tag{117}$$

For the case when $i \in [\![r+k_0, r+s]\!]$, $\tilde{\lambda}_i = -\tilde{\sigma}_{i-r}$. Observe that

$$G(-\tilde{\sigma}_{i-r}) = (-\tilde{\sigma}_{i-r} - \mathcal{E})^{-1} = -(\tilde{\sigma}_{i-r} + \mathcal{E})^{-1} \sim -(\tilde{\sigma}_{i-r} - \mathcal{E})^{-1} = -G(\tilde{\sigma}_{i-r})$$

because the distribution of $\mathcal{E}$ is symmetric. Hence

$$\Phi(-\tilde{\sigma}_{i-r}) \sim -\Phi(\tilde{\sigma}_{i-r})$$

by the definition (30). Repeating the arguments from the previous case, we see that

$$\left\| \mathcal{U}_J^{\mathrm{T}} \Xi(\widetilde{\lambda}_i) \mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}} \widetilde{\mathbf{u}}_i \right\| \leqslant \frac{b}{b-1} \frac{\eta}{\widetilde{\sigma}_{i-r}}. \tag{118}$$

Step 2. Lower bound for the smallest singular value of the matrix (114). In fact, the singular values of the matrix (114) can be calculated explicitly via elementary linear algebra. The following proposition presents a subtle modification of the one found in [63, Proposition 10].

Proposition 36. *For $1 \leqslant r_0 < r$ and $1 \leqslant k \leqslant s \leqslant r_0$, denote the index sets $I := [\![k, s]\!] \cup [\![r+k, r+s]\!]$ and $J := [\![1, 2r]\!] \backslash I$. For any $x \in \mathbb{R}$ satisfying $|x| > \|\mathcal{E}\|$, the singular values of $I_{2(r-s+k-1)} - \mathcal{U}_J^{\mathrm{T}} \Phi(x) \mathcal{U}_J \mathcal{D}_J$ are given by*

$$\left| \sqrt{1 + \beta(x)^2 \sigma_t^2} \pm |\alpha(x)| \sigma_t \right|$$

for $t \in [\![1, k-1]\!] \cup [\![s+1, r]\!]$.

In order to bound the singular values, we first estimate $\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)$, $\phi_1(\widetilde{\sigma}_i)$ and $\phi_2(\widetilde{\sigma}_i)$ for $i \in [\![k_0, s]\!]$. Since $\widetilde{\sigma}_i \in S_{\sigma_{l_i}}$ for some $l_i \in [\![k_0, s]\!]$ where S_σ is defined in (40), we have

$$\widetilde{\sigma}_i \geqslant \sigma_{l_i} - 20\chi(b)\eta r \geqslant \left(1 - \frac{\chi(b)}{4b}\right) \sigma_{l_i} \tag{119}$$

and

$$\widetilde{\sigma}_i \leqslant \chi(b)\sigma_{l_i} + 20\chi(b)\eta r \leqslant \chi(b) \left(1 + \frac{1}{4b}\right) \sigma_{l_i} \tag{120}$$

by the supposition $\sigma_{l_i} \geqslant 2b(\sqrt{N} + \sqrt{n}) + 80b\eta r$.

Observe from (36) that

$$\left(1 - \frac{1}{4b(b-1)}\right) \widetilde{\sigma}_i \leqslant \phi_s(\widetilde{\sigma}_i) \leqslant \chi(b)\widetilde{\sigma}_i \quad \text{for } s = 1, 2. \tag{121}$$

Using these estimates, we crudely bound

$$0 < \alpha(\widetilde{\sigma}_i) = \frac{1}{2}\left(\frac{1}{\phi_1(\widetilde{\sigma}_i)} + \frac{1}{\phi_2(\widetilde{\sigma}_i)}\right) \leqslant \frac{\tau(b)}{\widetilde{\sigma}_i} \quad \text{with } \tau(b) := \left(1 - \frac{1}{4b(b-1)}\right)^{-1}$$

and by (33),

$$\begin{aligned} \beta(\widetilde{\sigma}_i) &= \frac{1}{2}\left(\frac{1}{\phi_1(\widetilde{\sigma}_i)} - \frac{1}{\phi_2(\widetilde{\sigma}_i)}\right) = \frac{\phi_2(\widetilde{\sigma}_i) - \phi_1(\widetilde{\sigma}_i)}{2\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)} \\ &= \frac{n-N}{\widetilde{\sigma}_i} \frac{1}{2\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)} \leqslant \frac{\tau(b)^2}{8b^2 \widetilde{\sigma}_i} \end{aligned}$$

by noting that $\widetilde{\sigma}_i^2 \geqslant (2b(\sqrt{N} + \sqrt{n}))^2 > 4b^2(N+n)$.

We are ready to bound the singular values of $I_{2(r-s+k-1)} - \mathcal{U}_J^{\mathrm{T}} \Phi(\widetilde{\sigma}_i) \mathcal{U}_J \mathcal{D}_J$. We start with the case when $i \in [\![k_0, s]\!]$ and $\widetilde{\lambda}_i = \widetilde{\sigma}_i$. In view of Proposition 36, the goal is to bound

$$\begin{aligned} &\min_{t \in [\![1, k-1]\!] \cup [\![s+1, r]\!]} \left| \sqrt{1 + \beta(\widetilde{\sigma}_i)^2 \sigma_t^2} \pm |\alpha(\widetilde{\sigma}_i)| \sigma_t \right| \\ &= \min_{t \in [\![1, k-1]\!] \cup [\![s+1, r]\!]} \left| \sqrt{1 + \beta(\widetilde{\sigma}_i)^2 \sigma_t^2} - \alpha(\widetilde{\sigma}_i) \sigma_t \right| \end{aligned}$$

$$= \min_{t\in[\![1,k-1]\!]\cup[\![s+1,r]\!]} \left| \frac{1-(\alpha(\widetilde{\sigma}_i)^2-\beta(\widetilde{\sigma}_i)^2)\sigma_t^2}{\sqrt{1+\beta(\widetilde{\sigma}_i)^2\sigma_t^2}+\alpha(\widetilde{\sigma}_i)\sigma_t} \right|$$

$$= \min_{t\in[\![1,k-1]\!]\cup[\![s+1,r]\!]} \frac{\left|1-\frac{\sigma_t^2}{\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)}\right|}{\sqrt{1+\beta(\widetilde{\sigma}_i)^2\sigma_t^2}+\alpha(\widetilde{\sigma}_i)\sigma_t}.$$

The upper bounds of $\alpha(\widetilde{\sigma}_i)$ and $\beta(\widetilde{\sigma}_i)$ obtained above yield that

$$\sqrt{1+\beta(\widetilde{\sigma}_i)^2\sigma_t^2}+\alpha(\widetilde{\sigma}_i)\sigma_t \leqslant \sqrt{1+\frac{\tau(b)^4}{64b^4}\frac{\sigma_t^2}{\widetilde{\sigma}_i^2}}+\tau(b)\frac{\sigma_t}{\widetilde{\sigma}_i} \leqslant 1+\tau(b)\left(1+\frac{\tau(b)}{8b^2}\right)\frac{\sigma_t}{\widetilde{\sigma}_i}$$

for any $t\in[\![1,k-1]\!]\cup[\![s+1,r]\!]$. Hence,

$$\begin{aligned}
&\min_{t\in[\![1,k-1]\!]\cup[\![s+1,r]\!]} \left|\sqrt{1+\beta(\widetilde{\sigma}_i)^2\sigma_t^2}\pm\alpha(\widetilde{\sigma}_i)\sigma_t\right| \\
&\geqslant \min_{t\in[\![1,k-1]\!]\cup[\![s+1,r]\!]} \frac{1}{1+\tau(b)\left(1+\frac{\tau(b)}{8b^2}\right)\frac{\sigma_t}{\widetilde{\sigma}_i}} \left|\frac{\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)-\sigma_t^2}{\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)}\right| \\
&\geqslant \min_{t\in[\![1,k-1]\!]\cup[\![s+1,r]\!]} \frac{1}{1+\tau(b)\left(1+\frac{\tau(b)}{8b^2}\right)\frac{\sigma_t}{\widetilde{\sigma}_i}} \frac{|\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)-\sigma_t^2|}{\chi(b)^2\widetilde{\sigma}_i^2}.
\end{aligned}$$

To continue the estimates, we consider the cases $t\in[\![1,k-1]\!]$ and $t\in[\![s+1,r]\!]$ separately.

First, for any $t\in[\![1,k-1]\!]$, $\sigma_t\geqslant\sigma_i$ and $\sigma_t\geqslant\sigma_{l_i}$ since $i,l_i\in[\![k_0,s]\!]$. By (115),

$$\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)\leqslant\sigma_{l_i}^2+20\xi(b)\chi(b)\eta r(\widetilde{\sigma}_i+\chi(b)\sigma_{l_i}).$$

Thus, we obtain

$$\begin{aligned}
\sigma_t^2-\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i) &\geqslant \sigma_t^2-\sigma_{l_i}^2-20\xi(b)\chi(b)\eta r(\widetilde{\sigma}_i+\chi(b)\sigma_{l_i}) \\
&\geqslant (\sigma_t-\sigma_{l_i})(\sigma_t+\sigma_{l_i})-20\xi(b)\chi(b)^2\eta r\left(\left(1+\frac{1}{4b}\right)\sigma_t+\sigma_{l_i}\right).
\end{aligned} \tag{122}$$

The last inequality is due to

$$\widetilde{\sigma}_i\leqslant\chi(b)\left(1+\frac{1}{4b}\right)\sigma_{l_i}\leqslant\chi(b)\left(1+\frac{1}{4b}\right)\sigma_t$$

from (120) and $\sigma_t\geqslant\sigma_{l_i}$. Since $\sigma_t-\sigma_{l_i}\geqslant\delta_{k-1}\geqslant 75\chi(b)\eta r$, we further get

$$\sigma_t^2-\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)\geqslant\left(1-\frac{4}{15}\xi(b)\chi(b)\left(1+\frac{1}{4b}\right)\right)\delta_{k-1}(\sigma_t+\sigma_{l_i})>0$$

since $1-\frac{4}{15}\xi(b)\chi(b)\left(1+\frac{1}{4b}\right)\geqslant 1-\frac{4}{15}\xi(2)\chi(2)\left(1+\frac{1}{8}\right)\approx 0.49$.

Hence, we further have

$$\begin{aligned}
&\min_{t\in[\![1,k-1]\!]} \frac{1}{1+\tau(b)\left(1+\frac{\tau(b)}{8b^2}\right)\frac{\sigma_t}{\widetilde{\sigma}_i}} \frac{|\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)-\sigma_t^2|}{\chi(b)^2\widetilde{\sigma}_i^2} \\
&= \frac{1}{\chi(b)^2\widetilde{\sigma}_i}\min_{t\in[\![1,k-1]\!]} \frac{\sigma_t^2-\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i)}{\widetilde{\sigma}_i+\tau(b)\left(1+\frac{\tau(b)}{8b^2}\right)\sigma_t} \\
&\geqslant \left(1-\frac{4}{15}\xi(b)\chi(b)\left(1+\frac{1}{4b}\right)\right)\frac{\delta_{k-1}}{\chi(b)^2\widetilde{\sigma}_i}\min_{t\in[\![1,k-1]\!]} \frac{\sigma_t+\sigma_{l_i}}{\widetilde{\sigma}_i+\tau(b)\left(1+\frac{\tau(b)}{8b^2}\right)\sigma_t}
\end{aligned}$$

$$\geqslant \left(1 - \frac{4}{15}\xi(b)\chi(b)\big(1 + \frac{1}{4b}\big)\right) \frac{\delta_{k-1}}{\chi(b)^2\widetilde{\sigma}_i} \min_{t\in[\![1,k-1]\!]} \frac{\sigma_{l_i} + \sigma_t}{\chi(b)\left(1 + \frac{1}{4b}\right)\sigma_{l_i} + \tau(b)\left(1 + \frac{\tau(b)}{8b^2}\right)\sigma_t}.$$

Note that $\chi(b)\left(1 + \frac{1}{4b}\right) \geqslant \tau(b)\left(1 + \frac{\tau(b)}{8b^2}\right)$ for $b \geqslant 2$ (checked via Mathematics). We conclude that

$$\min_{t\in[\![1,k-1]\!]} \left|\sqrt{1 + \beta(\widetilde{\sigma}_i)^2\sigma_t^2} \pm \alpha(\widetilde{\sigma}_i)\sigma_t\right| \geqslant \frac{1 - \frac{4}{15}\xi(b)\chi(b)\big(1 + \frac{1}{4b}\big)}{\chi(b)^3\left(1 + \frac{1}{4b}\right)} \frac{\delta_{k-1}}{\widetilde{\sigma}_i}. \tag{123}$$

Next, for any $t \in [\![s+1, r]\!]$, $\sigma_t/\sigma_{l_i} \leqslant 1$ and by (119), $\sigma_t/\widetilde{\sigma}_i \leqslant \left(1 - \frac{\chi(b)}{4b}\right)^{-1}$. Consequently,

$$\begin{aligned}
&\min_{t\in[\![s+1,r]\!]} \frac{1}{1 + \tau(b)\left(1 + \frac{\tau(b)}{8b^2}\right)\frac{\sigma_t}{\widetilde{\sigma}_i}} \frac{|\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i) - \sigma_t^2|}{\chi(b)^2\widetilde{\sigma}_i^2} \\
&\geqslant \frac{1}{1 + \tau(b)\left(1 + \frac{\tau(b)}{8b^2}\right)\left(1 - \frac{\chi(b)}{4b}\right)^{-1}} \frac{1}{\chi(b)^2} \min_{t\in[\![s+1,r]\!]} \frac{|\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i) - \sigma_t^2|}{\widetilde{\sigma}_i^2}.
\end{aligned} \tag{124}$$

By (115),

$$\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i) \geqslant \sigma_{l_i}^2 - 20\xi(b)\chi(b)\eta r(\widetilde{\sigma}_i + \chi(b)\sigma_{l_i}).$$

Using a similar argument as (122), one has

$$\begin{aligned}
\phi_1(\widetilde{\sigma}_i)\phi_2(\widetilde{\sigma}_i) - \sigma_t^2 &\geqslant \sigma_{l_i}^2 - \sigma_t^2 - 20\xi(b)\chi(b)^2\eta r\left(\big(1 + \frac{1}{4b}\big)\sigma_t + \sigma_{l_i}\right) \\
&\geqslant \left(1 - \frac{4}{15}\xi(b)\chi(b)\big(1 + \frac{1}{4b}\big)\right)\delta_s(\sigma_t + \sigma_{l_i}) > 0
\end{aligned}$$

since $\sigma_{l_i} - \sigma_t \geqslant \delta_s \geqslant 75\chi(b)\eta r$. Continuing from (124), we further get

$$\begin{aligned}
&\min_{t\in[\![s+1,r]\!]} \left|\sqrt{1 + \beta(\widetilde{\sigma}_i)^2\sigma_t^2} \pm \alpha(\widetilde{\sigma}_i)\sigma_t\right| \\
&\geqslant \frac{1 - \frac{4}{15}\xi(b)\chi(b)\big(1 + \frac{1}{4b}\big)}{1 + \tau(b)\left(1 + \frac{\tau(b)}{8b^2}\right)\left(1 - \frac{\chi(b)}{4b}\right)^{-1}} \frac{1}{\chi(b)^2} \frac{\delta_s}{\widetilde{\sigma}_i} \min_{t\in[\![s+1,r]\!]} \frac{\sigma_t + \sigma_{l_i}}{\widetilde{\sigma}_i} \\
&\geqslant \frac{1 - \frac{4}{15}\xi(b)\chi(b)\big(1 + \frac{1}{4b}\big)}{1 + \tau(b)\left(1 + \frac{\tau(b)}{8b^2}\right)\left(1 - \frac{\chi(b)}{4b}\right)^{-1}} \frac{1}{\chi(b)^3\big(1 + \frac{1}{4b}\big)} \frac{\delta_s}{\widetilde{\sigma}_i} := \nu(b)\frac{\delta_s}{\widetilde{\sigma}_i},
\end{aligned}$$

where the last inequality follows from (120).

Comparing (123) and (125), together with the observation

$$\frac{1 - \frac{4}{15}\xi(b)\chi(b)\big(1 + \frac{1}{4b}\big)}{\chi(b)^3\left(1 + \frac{1}{4b}\right)} \geqslant \nu(b)$$

for $b \geqslant 2$ (checked via Mathematica), we conclude that

$$\min_{t\in[\![1,k-1]\!]\cup[\![s+1,r]\!]} \left|\sqrt{1 + \beta(\widetilde{\lambda}_i)^2\sigma_t^2} \pm |\alpha(\widetilde{\lambda}_i)|\sigma_t\right| \geqslant \nu(b)\frac{\min\{\delta_{k-1}, \delta_s\}}{\widetilde{\sigma}_i}. \tag{125}$$

For the case when $i \in [\![r + k_0, r + s]\!]$ and $\widetilde{\lambda}_i = -\widetilde{\sigma}_{i-r}$. Use the observation that $\alpha(\widetilde{\lambda}_i) \sim -\alpha(\widetilde{\sigma}_{i-r})$ and $\beta(\widetilde{\lambda}_i) \sim -\beta(\widetilde{\sigma}_{i-r})$ from the definitions (32). A simple modification of the previous proof shows that

$$\min_{t\in[\![1,k-1]\!]\cup[\![s+1,r]\!]}\left|\sqrt{1+\beta(\widetilde{\lambda}_i)^2\sigma_t^2}\pm|\alpha(\widetilde{\lambda}_i)|\sigma_t\right|\geqslant\nu(b)\frac{\min\{\delta_{k-1},\delta_s\}}{\widetilde{\sigma}_{i-r}}. \tag{126}$$

Step 3. Combining the bounds above. With the estimates deduced in the previous two steps, we are in a position to bound $\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathbf{u}}_i\|$. For $i\in[\![k_0,s]\!]$, plugging (117) and (125) and into (113), we find that

$$\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathbf{u}}_i\|\leqslant\frac{b}{(b-1)\nu(b)}\frac{\eta}{\min\{\delta_{k-1},\delta_s\}}.$$

Finally, for simplicity, we employ the following bound

$$\frac{b}{(b-1)\nu(b)}<3\frac{(b+1)^2}{(b-1)^2}$$

for $b\geqslant 2$ (checked via Mathematica). We arrive at

$$\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathbf{u}}_i\|\leqslant 3\frac{(b+1)^2}{(b-1)^2}\frac{\eta}{\min\{\delta_{k-1},\delta_s\}}.$$

Likewise, for $i\in[\![r+k_0,r+s]\!]$, using (118) and (126), we also get

$$\|\mathcal{U}_J^{\mathrm{T}}\widetilde{\mathbf{u}}_i\|\leqslant 3\frac{(b+1)^2}{(b-1)^2}\frac{\eta}{\min\{\delta_{k-1},\delta_s\}}.$$

This completes the proof.

Appendix D. Proof of Theorem 29

This section is devoted to the proof of Theorem 29. For convenience, denote

$$M:=2b(\sqrt{N}+\sqrt{n}).$$

Note that the assumptions of Theorem 29 guarantees that for any $z\in S_{\sigma_j}$ $(1\leqslant j\leqslant r_0)$, $|z|\geqslant\mathrm{Re}(z)\geqslant\sigma_j-20\chi(b)\eta r\geqslant\sigma_j-20\chi(2)\eta r=\sigma_j-\frac{45}{2}\eta r>M$.

We start with some reduction of the proof. First, note that if $\sigma_j>n^2/2$ for $1\leqslant j\leqslant r_0$, then by Lemma 31, with probability at least $1-(N+n)^{-1.5r^2(K+4)}$,

$$|\widetilde{\sigma}_j-\sigma_j|\leqslant\eta r.$$

Since $\varphi(z)=(z-\mathrm{tr}\,\mathcal{I}^{\mathrm{d}}G(z))(z-\mathrm{tr}\,\mathcal{I}^{\mathrm{u}}G(z))$,

$$\begin{aligned}|\varphi(\widetilde{\sigma}_j)-\sigma_j^2|&=\left|\widetilde{\sigma}_j^2-\sigma_j^2-\widetilde{\sigma}_j\,\mathrm{tr}\,G(\widetilde{\sigma}_j)+(\mathrm{tr}\,\mathcal{I}^{\mathrm{u}}G(\widetilde{\sigma}_j))(\mathrm{tr}\,\mathcal{I}^{\mathrm{d}}G(\widetilde{\sigma}_j))\right|\\&\leqslant\eta r(\widetilde{\sigma}_j+\sigma_j)+\widetilde{\sigma}_j\,|\mathrm{tr}\,G(\widetilde{\sigma}_j)|+|\mathrm{tr}\,\mathcal{I}^{\mathrm{u}}G(\widetilde{\sigma}_j)|\left|\mathrm{tr}\,\mathcal{I}^{\mathrm{d}}G(\widetilde{\sigma}_j)\right|.\end{aligned}$$

Note that by Weyl's inequality, $\widetilde{\sigma}_j\geqslant\sigma_j-\|E\|\geqslant\max\{M+80b\eta r,n^2/2\}-2(\sqrt{N}+\sqrt{n})\geqslant M$ by the suppositions on N,n. Hence, by (37),

$$\max\{|\,\mathrm{tr}\,G(\widetilde{\sigma}_j)|,|\,\mathrm{tr}\,\mathcal{I}^u G(\widetilde{\sigma}_j)|,|\,\mathrm{tr}\,\mathcal{I}^d G(\widetilde{\sigma}_j)|\}\leqslant\frac{b}{b-1}\frac{N+n}{\widetilde{\sigma}_j}\leqslant 2\frac{N+n}{\widetilde{\sigma}_j}.$$

It follows that

$$|\varphi(\widetilde{\sigma}_j)-\sigma_j^2|\leqslant\eta r(\widetilde{\sigma}_j+\sigma_j)+2(N+n)+4\frac{(N+n)^2}{\widetilde{\sigma}_j^2}\leqslant 10\eta r(\widetilde{\sigma}_j+\sigma_j)$$

by the supposition that $\sigma_j>n^2/2$ and the Weyl's inequality. In particular, the conclusion of Theorem 29 holds.

Consequently, it is enough to examine the scenario where there is a certain $l_0 \in [\![1, r_0]\!]$ for which $\sigma_{l_0} \leqslant n^2/2$. We claim that there exists an index $i_0 \in [\![1, r_0]\!]$ such that $\sigma_j \leqslant n^3$ for $j \geqslant i_0$ and $\sigma_j > n^3$ for $j < i_0$, and

$$\delta_{i_0-1} = \sigma_{i_0-1} - \sigma_{i_0} \geqslant 75\chi(b)\eta r.$$

To determine i_0, we propose a simple iterative algorithm: start with σ_1. If $\sigma_1 \leqslant n^3$, set $i_0 = 1$ and terminate the algorithm, since $\sigma_0 = \infty$ and $\delta_0 = \infty$ by definition. Assume $\sigma_1 > n^3$ and evaluate σ_2. If $\sigma_2 \leqslant n^3 - 75\chi(b)\eta r$, set $i_0 = 2$ and exit. Assume $\sigma_2 > n^3 - 75\chi(b)\eta r$ and evaluate σ_3. We continue this process and terminate the algorithm with $i_0 = k$ unless

$$\sigma_1 > n^3, \sigma_2 > n^3 - 75\chi(b)\eta r, \cdots, \sigma_k > n^3 - 75\chi(b)\eta r. \tag{127}$$

Note that the condition (127) cannot hold for $k = l_0$ because $\sigma_{l_0} \leqslant n^2/2 < n^3 - 75\chi(b)\eta r$, based on the assumption that $(\sqrt{N} + \sqrt{n})^2 \geqslant 32(K+7)\log(N+n) + 64(\log 9)r$. Therefore, i_0 must satisfy $i_0 \leqslant l_0 - 1$.

We shall fix such an index i_0 throughout the rest of the proof. The goal is to demonstrate that the following holds with a probability of at least $1 - 10(N+N)^{-K}$: assume any $i_0 \leqslant k \leqslant s \leqslant r_0$ that fulfills $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$. For any $j \in [\![k, s]\!]$, there exists $j_0 \in [\![k, s]\!]$ such that $\widetilde{\sigma}_j \in S_{\sigma_{j_0}}$ and (41) is satisfied.

Before moving forward with the proof, we review several results and introduce necessary notations collected from [64]. The proofs of these results are identical to these in [64], utilizing Lemma 26, and we will not repeat them here.

Lemma 37 (Eigenvalue location criterion, Lemma 21 from [64])**.** *Assume $\mathcal{A}$ has rank $2r$ with the spectral decomposition $\mathcal{A} = \mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}$, where $\mathcal{U}$ is an $(N+n) \times 2r$ matrix satisfying $\mathcal{U}^{\mathrm{T}}\mathcal{U} = I_{2r}$ and $\mathcal{D}$ is a $2r \times 2r$ diagonal matrix with non-zero $\lambda_1, \ldots, \lambda_{2r}$ on the diagonal. Then the eigenvalues of $\mathcal{A} + \mathcal{E}$ outside of $[-\|\mathcal{E}\|, \|\mathcal{E}\|]$ are the zeros of the function*

$$z \mapsto \det(\mathcal{D}^{-1} - \mathcal{U}^{\mathrm{T}}G(z)\mathcal{U}).$$

Moreover, the algebraic multiplicity of each eigenvalue matches the corresponding multiplicity of each zero.

Define the functions

$$f(z) := \det(\mathcal{D}^{-1} - \mathcal{U}^{\mathrm{T}}G(z)\mathcal{U}), \qquad g(z) := \det\left(\mathcal{D}^{-1} - \mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\right),$$

where $\Phi(z)$ is given in (30). Observe that, by Lemma 26, $1/\phi_1(z)$, $1/\phi_2(z)$ and thus $\Phi(z)$ are well-defined for any $|z| > M$. Therefore, f and g are both complex analytic in the region $\{z \in \mathbb{C} : |z| > M\}$. Furthermore, a direct computation using (34) suggests that the zeros of $g(z)$ are the values $z \in \mathbb{C}$ which satisfy the equations $\phi_1(z)\phi_2(z) = \sigma_l^2$.

Recall from (39) and (32) that

$$\varphi(z) = \phi_1(z)\phi_2(z) = (z - \operatorname{tr}\mathcal{I}^{\mathrm{d}}G(z))(z - \operatorname{tr}\mathcal{I}^{\mathrm{u}}G(z)).$$

We use the function

$$\xi(b) = 1 + \frac{1}{2(b-1)^2}. \tag{128}$$

The subsequent lemma establishes a set of properties exhibited by φ within the complex plane as well as on the real axis.

Lemma 38 (Lemma 22 from [64])**.** *The function φ satisfies the following properties.*

(i) *For* $z, w \in \mathbb{C}$ *with* $|z|, |w|, |z+w| \geqslant M$,

$$\left(1 - \frac{1}{2(b-1)^2}\right)|z^2 - w^2| \leqslant |\varphi(z) - \varphi(w)| \leqslant \xi(b)|z^2 - w^2|. \tag{129}$$

(ii) *(Monotone)* φ *is real-valued and strictly increasing on* $[M, \infty)$.

(iii) *(Crude bounds)* $0 < \varphi(z) < z^2$ *for any* $z \in [M, \infty)$.

Fix an index $j \in [\![1, r_0]\!]$. Since $\varphi(M) < M^2$ and $\lim_{z\to\infty} \varphi(z) = \infty$, it follows from the previous lemma that there exists a unique positive real number $z_j > M$ such that $\varphi(z_j) = \sigma_j^2$. Similarly, if $\sigma_l > M$ for $\sigma_l \neq \sigma_j$, then there exists a unique positive real number z_l with $\varphi(z_l) = \sigma_l^2$ so that $z_j > z_l$ if $l > j$ and $z_j < z_l$ if $l < j$.

For the next result, we define the half space

$$H_j := \{z \in \mathbb{C} : \mathrm{Re}(z) \geqslant z_j - 20\chi(b)\eta r\} \quad \text{with } \chi(b) = 1 + \frac{1}{4b(b-1)}.$$

Proposition 39 (Proposition 23 from [64])**.** *Under the assumptions of Theorem 5, for every* $z \in H_j$,

$$|z| \geqslant \sigma_j \geqslant M.$$

In particular,

$$\sigma_j \leqslant z_j \leqslant \chi(b)\sigma_j. \tag{130}$$

Proposition 40 (Proposition 24 from [64])**.** *If* $\sigma_j > \frac{1}{2}n^2$, *then* $|z_j - \sigma_j| \leqslant \frac{3b}{b-1}\frac{1}{n}$.

We now complete the proof of Theorem 29. Let j be a fixed index in $[\![i_0, r_0]\!]$. We will work in the set $H_j \cap S_{\sigma_j}$, where S_{σ_j} is specified in (40). It follows from Corollary 2.14 in [48] that

$$\frac{|f(z) - g(z)|}{|g(z)|} \leqslant (1 + \varepsilon(z))^{2r} - 1, \tag{131}$$

where

$$\varepsilon(z) := \left\| \left(\mathcal{D}^{-1} - \mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\right)^{-1} \right\| \left\| \mathcal{U}^{\mathrm{T}}(G(z) - \Phi(z))\mathcal{U} \right\|.$$

The next result facilitates the comparison of the numbers of zeros of f and g inside a region and will be used repeatedly in the later arguments.

Lemma 41. *For any region* $\mathcal{K} \subset \mathbb{C}$ *with closed contour* $\partial\mathcal{K}$, *if* $\varepsilon(z) \leqslant \frac{0.34}{r}$ *for all* $z \in \partial\mathcal{K}$, *then the number of zeros of* f *inside* $\mathcal{K}$ *is the same as the number of zeros of* g *inside* $\mathcal{K}$.

Proof. Continuing from (131), we find that

$$\frac{|f(z) - g(z)|}{|g(z)|} \leqslant \left(1 + \frac{0.34}{r}\right)^{2r} - 1 \leqslant e^{0.68} - 1 < 1 \tag{132}$$

for each $z \in \partial\mathcal{K}$. Therefore, by Rouché's theorem, we conclude that the numbers of zeros of f and g inside $\mathcal{K}$ are the same. □

In the remaining of the proof, we work on the event

$$\mathcal{F} := \left\{ \max_{z\in \mathsf{D}} |z|^2 \left\| \mathcal{U}^{\mathrm{T}}\left(G(z) - \Phi(z)\right)\mathcal{U} \right\| \leqslant \eta \right\} \cap \left\{ \max_{l \in [\![1, r_0]\!] : \sigma_l > \frac{1}{2}n^2} |\widetilde{\sigma}_l - \sigma_l| \leqslant \eta r \right\}, \tag{133}$$

where $\mathsf{D} = \{z \in \mathbb{C} : 2b(\sqrt{N} + \sqrt{n}) \leqslant |z| \leqslant 2n^3\}$. By Lemma 28 and Lemma 31, the event $\mathcal{F}$ holds with probability at least $1 - 9(N+n)^{-K} - (N+n)^{-1.5r^2(K+4)} > 1 - 10(N+n)^{-K}$.

We first bound $\varepsilon(z)$ for $z \in \mathsf{D}$. By Proposition 11 from [64],

$$\left\| \left(\mathcal{D}^{-1} - \mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\right)^{-1} \right\| = \max_{1 \leqslant l \leqslant r} \frac{\sigma_l}{|\sigma_l^2 - \phi_1\phi_2|} \mathcal{Q}^{1/2},$$

where

$$\mathcal{Q} := |\phi_1\phi_2|^2 + \frac{1}{2}\sigma_l^2(|\phi_1|^2 + |\phi_2|^2) + \frac{1}{2}\sigma_l\left[4|\phi_1\phi_2|^2|\phi_1 + \bar{\phi}_2|^2 + \sigma_l^2(|\phi_1|^2 - |\phi_2|^2)^2\right]^{1/2}.$$

Recall $\chi(b) = 1 + \frac{1}{4b(b-1)}$. Using (36) from Lemma 26, for $z \in \mathsf{D}$, we get $|\phi_i(z)| \leqslant \chi(b)|z|$ for $i = 1, 2$, and

$$\begin{aligned}
\mathcal{Q} &\leqslant \chi(b)^4|z|^4 + \chi(b)^2\sigma_l^2|z|^2 + \chi(b)^2\sigma_l|z|^2\sqrt{\sigma_l^2 + 4\chi(b)^2|z|^2} \\
&\leqslant \chi(b)^4|z|^4 + \chi(b)^2\sigma_l^2|z|^2 + \chi(b)^2\sigma_l|z|^2(\sigma_l + 2\chi(b)|z|) \\
&\leqslant \chi(b)^4|z|^4 + 2\chi(b)^2\sigma_l^2|z|^2 + 2\chi(b)^3\sigma_l|z|^3 \\
&\leqslant \chi(b)^2|z|^2\left(\chi(b)|z| + \sqrt{2}\sigma_l\right)^2,
\end{aligned}$$

and thus

$$\left\| \left(\mathcal{D}^{-1} - \mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\right)^{-1} \right\| \leqslant \chi(b)|z| \max_{1 \leqslant l \leqslant r} \frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|}.$$

Hence, we obtain that on the event $\mathcal{F}$,

$$\varepsilon(z) \leqslant \max_{1 \leqslant l \leqslant r} \chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|} \tag{134}$$

for all $z \in \mathsf{D}$. Note that $S_{\sigma_j} \subset \mathsf{D}$ for all $j \in [\![i_0, r_0]\!]$.

For each $j \in [\![i_0, r_0]\!]$, we take $\mathcal{C}_j$ to be the circle of radius $20\chi(b)\eta r$ centered at z_j and contained in $H_j \cap S_{\sigma_j}$. Note that $\mathcal{C}_j$'s may intersect each other. For any $i_0 \leqslant k \leqslant s \leqslant r_0$ satisfying $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$. Let

$$\mathcal{K}_{k,s} := \cup_{l=k}^{s}\mathcal{C}_l.$$

We now restrict ourselves to values of z contained on $\partial\mathcal{K}_{k,s}$. The goal is to show $\varepsilon(z)$ is small for all $z \in \partial\mathcal{K}_{k,s}$. Continuing from (134), it suffices to show

$$\chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|} \tag{135}$$

is small for all $1 \leqslant l \leqslant r$.

Fix $z \in \partial\mathcal{K}_{k,s}$. Assume $z \in \mathcal{C}_{j_0}$ for some $j_0 \in [\![k, s]\!]$. Then

$$|z - z_{j_0}| = 20\chi(b)\eta r.$$

Note that $\sigma_{j_0} \geqslant 80b\eta r$. Using (130), we have

$$\begin{aligned}
|z| &\leqslant z_{j_0} + 20\chi(b)\eta r \leqslant \chi(b)\left(1 + \frac{1}{4b}\right)\sigma_{j_0}, \\
|z| &\geqslant z_{j_0} - 20\chi(b)\eta r \geqslant \left(1 - \frac{\chi(b)}{4b}\right)\sigma_{j_0}.
\end{aligned} \tag{136}$$

We split the discussion into two cases: $|\sigma_l - \sigma_{j_0}| \leqslant 120\eta r$ and $|\sigma_l - \sigma_{j_0}| > 120\eta r$.

Case 1. For any $l \in [\![1, r]\!]$ satisfying $|\sigma_l - \sigma_{j_0}| \leqslant 120\eta r$, observe that $|z - z_l| \geqslant 20\chi(b)\eta r$. In view of (129), , we have

$$|\sigma_l^2 - \varphi(z)| = |\varphi(z_l) - \varphi(z)| \geqslant \left(1 - \frac{1}{2(b-1)^2}\right)|z_l^2 - z^2| \geqslant \frac{1}{2}|z_l - z||z_l + z|$$
$$\geqslant 10\chi(b)\eta r|z_l + z| \geqslant \left(1 - \frac{\chi(b)}{4b}\right)\sigma_{j_0} + \sigma_l,$$

where, in the last inequality, we used

$$|z_l + z| = |z_l + z_{j_0} + z - z_{j_0}| \geqslant z_l + z_{j_0} - 20\chi(b)\eta r$$
$$\geqslant \sigma_l + \sigma_{j_0} - 20\chi(b)\eta r \geqslant \left(1 - \frac{\chi(b)}{4b}\right)\sigma_{j_0} + \sigma_l$$

by (130) and the supposition $\eta r \leqslant \sigma_{j_0}/80b$. Combining with (136), we estimate (135) as follows:

$$\chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|} \leqslant \frac{1}{r}\frac{1}{10(1 - \frac{\chi(b)}{4b})}\frac{\sigma_l}{\sigma_{j_0}}\frac{\chi(b)^2(1 + \frac{1}{4b})\sigma_{j_0} + \sqrt{2}\sigma_l}{(1 - \frac{\chi(b)}{4b})\sigma_{j_0} + \sigma_l}.$$

Note that $\sigma_l \leqslant \sigma_{j_0} + 120\eta r \leqslant (1 + 120/80b)\sigma_{j_0} \leqslant (7/4)\sigma_{j_0}$ for $b \geqslant 2$. Also, $\chi(b) \leqslant \chi(2) = 9/8$ for $b \geqslant 2$. We further obtain that

$$\chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|} \leqslant \frac{1}{r}\frac{1}{10(1 - 9/64)}\frac{7}{4}\frac{(9/8)^3\sigma_{j_0} + \sqrt{2}\sigma_l}{(1 - 9/64)\sigma_{j_0} + \sigma_l}$$
$$\leqslant \frac{1}{r}\frac{7}{40(1 - 9/64)}\frac{(9/8)^3}{1 - 9/64} < \frac{0.34}{r}.$$

Case 2. For any $l \in [\![1, r]\!]$ satisfying $|\sigma_l - \sigma_{j_0}| > 120\eta r$, we start with

$$|\sigma_l^2 - \varphi(z)| \geqslant |\sigma_l^2 - \sigma_{j_0}^2| - |\sigma_{j_0}^2 - \varphi(z)| = |\sigma_l^2 - \sigma_{j_0}^2| - |\varphi(z_{j_0}) - \varphi(z)|$$
$$\geqslant |\sigma_l - \sigma_{j_0}|(\sigma_l + \sigma_{j_0}) - 20\chi(b)\xi(b)\eta r|z_{j_0} + z|.$$

by (129) and $|z - z_{j_0}| = 20\chi(b)\eta r$. Since

$$|z_{j_0} + z| \leqslant 2z_{j_0} + |z - z_{j_0}| = 2z_{j_0} + 20\chi(b)\eta r$$
$$\leqslant 2\chi(b)\sigma_{j_0} + 20\chi(b)\eta r \leqslant 2\chi(b)\left(1 + 1/8b\right)\sigma_{j_0} \leqslant \frac{34}{16}\chi(b)\sigma_{j_0}$$

and $|\sigma_l - \sigma_{j_0}| > 120\eta r$, we further get

$$|\sigma_l^2 - \varphi(z)| \geqslant |\sigma_l - \sigma_{j_0}|(\sigma_l + \sigma_{j_0}) - 20\chi(b)^2\xi(b)\frac{34}{16}\frac{1}{120}|\sigma_l - \sigma_{j_0}|(\sigma_l + \sigma_{j_0})$$
$$= \left(1 - \frac{17}{48}\chi(b)^2\xi(b)\right)|\sigma_l - \sigma_{j_0}|(\sigma_l + \sigma_{j_0}).$$

Hence, using (136), we get

$$\chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|} \leqslant \frac{\chi(b)\eta}{(1 - \frac{\chi(b)}{4b})(1 - \frac{17}{48}\chi(b)^2\xi(b))}\frac{\sigma_l}{\sigma_{j_0}}\frac{\chi(b)^2(1 + \frac{1}{4b})\sigma_{j_0} + \sqrt{2}\sigma_l}{|\sigma_l - \sigma_{j_0}|(\sigma_l + \sigma_{j_0})}.$$

To continue the estimates, we simply use the fact that $\chi(b), \xi(b)$ are decreasing. Thus $\chi(b) \leqslant \chi(2) = 9/8$ and $\xi(b) = 1 + 1/(2(b-1)^2) \leqslant 3/2$ for $b \geqslant 2$. Hence,

$$\chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|} \leqslant \frac{(9/8)^4}{\frac{55}{64}(1 - \frac{17}{48}(\frac{9}{8})^2\frac{3}{2})}\frac{\sigma_l}{\sigma_{j_0}}\frac{\eta}{|\sigma_{j_0} - \sigma_l|} < 6\frac{\sigma_l}{\sigma_{j_0}}\frac{\eta}{|\sigma_{j_0} - \sigma_l|}.$$

If $\sigma_l \leqslant 2\sigma_{j_0}$, then

$$6\frac{\sigma_l}{\sigma_{j_0}}\frac{\eta}{|\sigma_{j_0}-\sigma_l|} \leqslant 12\frac{\eta}{120\eta r} = \frac{0.1}{r}.$$

If $\sigma_l \geqslant 2\sigma_{j_0}$, then $\sigma_l - \sigma_{j_0} \geqslant 0.5\sigma_l$ and

$$6\frac{\sigma_l}{\sigma_{j_0}}\frac{\eta}{|\sigma_{j_0}-\sigma_l|} \leqslant 12\frac{\eta}{\sigma_{j_0}} \leqslant 12\frac{\eta}{160\eta r} = \frac{0.075}{r}.$$

Thus, we conclude that

$$\varepsilon(z) \leqslant \frac{0.34}{r} \tag{137}$$

for all $z \in \mathcal{K}_{k,s}$. By Lemma 41, the number of zeros of f inside $\mathcal{K}_{k,s}$ is the same as the number of zeros of g inside $\mathcal{K}_{k,s}$.

Since $\min\{\delta_{i_0-1}, \delta_{r_0}\} \geqslant 75\chi(b)\eta r$ by our supposition, we could take $k = i_0$ and $s = r_0$ and thus $\mathcal{K}_{i_0,r_0} = \cup_{l=i_0}^{r_0}\mathcal{C}_l$. Since g has $r_0 - i_0 + 1$ zeros inside $\mathcal{K}_{i_0,r_0}$, it follows that $\mathcal{A}+\mathcal{E}$ has exactly $r_0 - i_0 + 1$ eigenvalues inside $\mathcal{K}_{i_0,r_0}$. More generally, for any $i_0 \leqslant k \leqslant s \leqslant r_0$ satisfying $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$, we conclude that the number of eigenvalues of $\mathcal{A}+\mathcal{E}$ inside $\mathcal{K}_{k,s}$ is $s-k+1$, the same as the number of zeros of g inside $\mathcal{K}_{k,s}$. It remains to show that these eigenvalues are exactly $\widetilde{\sigma}_k, \cdots, \widetilde{\sigma}_s$. If this is the case, then for any $j \in [\![k,s]\!]$, there exists $j_0 \in [\![k,s]\!]$ such that $\widetilde{\sigma}_j \in \mathcal{C}_{j_0}$ and thus

$$|\widetilde{\sigma}_j - z_{j_0}| \leqslant 20\chi(b)\eta r.$$

In particular, $\widetilde{\sigma}_j \in S_{\sigma_{j_0}}$. By $\varphi(z_{j_0}) = \sigma_{j_0}^2$, (129) and (130),

$$|\varphi(\widetilde{\sigma}_j) - \sigma_{j_0}^2| = |\varphi(\widetilde{\sigma}_j) - \varphi(z_{j_0})| \leqslant 20\xi(b)\chi(b)\eta r\left(\widetilde{\sigma}_j + \chi(b)\sigma_{j_0}\right).$$

This will complete the proof.

It remains to prove that for any $i_0 \leqslant k \leqslant s \leqslant r_0$ satisfying $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$, the eigenvalues of $\mathcal{A}+\mathcal{E}$ inside $\cup_{l=k}^{s}\mathcal{C}_l$ are exactly $\widetilde{\sigma}_k, \cdots, \widetilde{\sigma}_s$. We will do so by proving the following claims hold on the event $\mathcal{F}$ (see Figure 1 for an illustration):

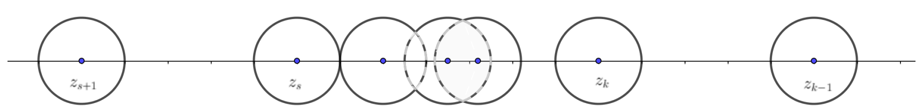

FIGURE 1. Distinct circles $\mathcal{C}_j$ with centers z_j on the real line for $i_0 \leqslant j \leqslant r_0$.

Claim 1. For any $i_0 \leqslant k \leqslant s \leqslant r_0$ satisfying $\min\{\delta_{k-1}, \delta_s\} \geqslant 75\chi(b)\eta r$, $\cup_{l=k}^{s}\mathcal{C}_l$ does not intersect other circles.

Claim 2. $\mathcal{A}+\mathcal{E}$ has exactly $i_0 - 1$ eigenvalues larger than $z_{i_0} + 20\chi(b)\eta r$.

Claim 3. No eigenvalues of $\mathcal{A}+\mathcal{E}$ lie between disjoint circles.

For the moment, let us assume these claims are true. Note that $\widetilde{\sigma}_{i_0}$ has to lie inside one of the $\mathcal{C}_j$'s ($i_0 \leqslant j \leqslant r_0$) because it is the largest eigenvalue of $\mathcal{A}+\mathcal{E}$ that is no larger than $z_{i_0} + 20\chi(b)\eta r$ (due to *Claim 2*) and thus it satisfies $\widetilde{\sigma}_{i_0} > z_{r_0} - 20\chi(b)\eta r$. Since the number of zeros of $g(z)$ located inside $\mathcal{K}_{i_0,r_0} = \cup_{j=i_0}^{r_0}\mathcal{C}_j$, which is $r_0 - i_0 + 1$, is the same as that of $f(z)$ inside $\mathcal{K}_{i_0,r_0}$, we have $\widetilde{\sigma}_{i_0}, \ldots, \widetilde{\sigma}_{r_0}$ lie inside $\mathcal{K}_{i_0,r_0}$. The conclusion follows by *Claim 1*, *Claim 3* and the fact that the number of zeros of $g(z)$ in each $\mathcal{K}_{k,s}$ is the same as that of $f(z)$.

We start with the proof of the *Claim 1*. It suffices to show that if $|\sigma_l - \sigma_j| \geqslant 75\chi(b)\eta r$, then $\mathcal{C}_l$ and $\mathcal{C}_j$ do not intersect. By Lemma 38,

$$|z_l^2 - z_j^2| \geqslant \frac{1}{\xi(b)}|\varphi(z_l) - \varphi(z_j)| = \frac{|\sigma_l^2 - \sigma_j^2|}{\xi(b)} \geqslant \frac{75\chi(b)\eta r}{\xi(b)}(\sigma_l + \sigma_j).$$

Since $|z_l^2 - z_j^2| = (z_l + z_j)|z_l - z_j| \leqslant \chi(b)(\sigma_l + \sigma_j)|z_l - z_j|$ by Proposition 39, we have

$$|z_l - z_j| \geqslant \frac{75}{\xi(b)}\eta r, \tag{138}$$

and thus

$$\mathrm{dist}(\mathcal{C}_j, \mathcal{C}_l) \geqslant |z_l - z_j| - 40\chi(b)\eta r \geqslant \frac{75}{\xi(b)}\eta r - 40\chi(b)\eta r \geqslant \left(\frac{75}{\xi(2)} - 40\chi(2)\right)\eta r > 0.$$

Next, we prove *Claim 2*. We split the proof into two cases: $i_0 = 1$ and $i_0 > 1$. *Case 1:* $i_0 = 1$. We prove that no eigenvalues of $\mathcal{A} + \mathcal{E}$ are larger than $z_1 + 20\chi(b)\eta r$. We now take $\mathcal{C}_0$ to be any circle with radius $20\chi(b)\eta r$ centered at a point $z_0 > z_1 + 20\chi(b)\eta r$ on the real line inside the region $H_1 \cap \hat{S}_{\sigma_1}$ such that $\mathrm{dist}(z_1, \mathcal{C}_0) \geqslant 20\chi(b)\eta r$. Here

$$\begin{aligned}\hat{S}_{\sigma_1} := \{w \in \mathbb{C} :& |\,\mathrm{Im}(w)| \leqslant 20\chi(b)\eta r,\\ & 2b(\sqrt{N} + \sqrt{n}) + 138\eta r \leqslant \mathrm{Re}(w) \leqslant \frac{3}{2}\sigma_1 + 20\chi(b)\eta r\}\end{aligned} \tag{139}$$

is a slight modification of the set S_σ in (40). Note that $\widetilde{\sigma}_1 \in \hat{S}_{\sigma_1}$: the upper bound $\widetilde{\sigma}_1 \leqslant \frac{3}{2}\sigma_1$ follows from the Weyl's inequality and the supposition $\|\mathcal{E}\| \leqslant \frac{1}{b}\sigma_1 \leqslant \frac{1}{2}\sigma_1$; the lower bound is because it is the largest eigenvalue and $\widetilde{\sigma}_1 \geqslant z_j - 20\chi(b)\eta r \geqslant \sigma_j - 20\chi(b)\eta r$ for any $j \in [\![i_0, r_0]\!]$ due to fact that the number of eigenvalues of $\mathcal{A}+\mathcal{E}$ inside $\cup_{l=i_0}^{r_0}\mathcal{C}_l$ is $r_0 - i_0 + 1$. For $z \in \hat{S}_{\sigma_1}$,

$$2b(\sqrt{N} + \sqrt{n}) \leqslant |z| \leqslant 40\chi(b)\eta r + \frac{3}{2}\sigma_2 \leqslant \frac{40\chi(b)}{80b}\sigma_1 + \frac{3}{2}\sigma_1 \leqslant \frac{57}{32}\sigma_1 < 2n^2,$$

hence $z \in \mathsf{D}$ and the conclusion of Lemma 28 holds. In particular, the bound (134) also holds for $z \in \hat{S}_{\sigma_1}$. We show

$$\varepsilon(z) < \frac{1}{3r}$$

for all $z \in \mathcal{C}_0$. The proof is similar to the proof of (137) and we sketch it here. For any $z \in \mathcal{C}_0$, from $|z - z_0| = 20\chi(b)\eta r$ and $z_0 - z_1 > 40\chi(b)\eta r$, we obtain $|z| \leqslant z_0 + 20\chi(b)\eta r$ and

$$|z| \geqslant z_0 - 20\chi(b)\eta r \geqslant z_1 + 20\chi(b)\eta r > \sigma_1 + 20\chi(b)\eta r > \sigma_1.$$

Again, by Lemma 38, we see for any $1 \leqslant l \leqslant r$,

$$\begin{aligned}|\sigma_l^2 - \varphi(z)| = |\varphi(z_l) - \varphi(z)| &\geqslant \frac{1}{2}|z_l^2 - z^2| \\ &\geqslant \frac{1}{2}(z_l + \mathrm{Re}(z))(\mathrm{Re}(z) - z_l) \\ &\geqslant \frac{1}{2}(\sigma_l + z_0 - 20\chi(b)\eta r)(z_0 - z_l - 20\chi(b)\eta r) \\ &\geqslant 10\chi(b)\eta r(\sigma_l + z_0 - 20\chi(b)\eta r). \qquad (140)\end{aligned}$$

Plugging these estimates back into (134), we see

$$\varepsilon(z) \leqslant \max_{1\leqslant l\leqslant r} \chi(b)\frac{\eta\sigma_l}{\sigma_1}\frac{\chi(b)(z_0 + 20\chi(b)\eta r) + \sqrt{2}\sigma_l}{10\chi(b)\eta r(z_0 + \sigma_l - 20\chi(b)\eta r)} < \frac{1}{5r},$$

where we used the bound $\chi(b)(z_0 + 20\chi(b)\eta r) + \sqrt{2}\sigma_l \leqslant 2(z_0 + \sigma_l - 20\chi(b)\eta r)$ in the last inequality.

By Lemma 41, f has the same number of zeros inside $\mathcal{C}_0$ as g. As g has no zeros inside $\mathcal{C}_0$[1], $\mathcal{A} + \mathcal{E}$ has no eigenvalues inside $\mathcal{C}_0$. Since the circle $\mathcal{C}_0$ was arbitrarily chosen inside this region, we conclude that $\mathcal{A} + \mathcal{E}$ has no eigenvalues larger than $z_1 + 20\chi(b)\eta r$.

Case 2: $i_0 > 1$. On the event $\mathcal{F}$, we have

$$\max_{l\in[\![1,r_0]\!];\sigma_l>n^2/2} |\widetilde{\sigma}_l - \sigma_l| \leqslant \eta r. \qquad (141)$$

Note that $\sigma_{i_0-1} > n^3 > n^2$. Combining (141), Proposition 40 and

$$z_{i_0-1} - z_{i_0} \geqslant \frac{75}{\xi(b)}\eta r \geqslant \frac{75}{\xi(2)}\eta r = 50\eta r,$$

which follows from the supposition $\delta_{i_0-1} \geqslant 75\chi(b)\eta r$ and the same argument as (138), we get

$$\widetilde{\sigma}_{i_0-1} \geqslant \sigma_{i_0-1} - \eta r \geqslant z_{i_0-1} - \frac{3b}{b-1}\frac{1}{n} - \eta r \geqslant z_{i_0} + 50\eta r - \frac{6}{n} > z_{i_0} + 20\chi(b)\eta r.$$

Hence, $\mathcal{A} + \mathcal{E}$ has at least $i_0 - 1$ eigenvalues larger than $z_{i_0} + 20\chi(b)\eta r$.

We first consider $\sigma_{i_0} > \frac{1}{2}n^2$. It follows from (141) and Proposition 40 that

$$\widetilde{\sigma}_{i_0} \leqslant \sigma_{i_0} + \eta r \leqslant z_{i_0} + \frac{3b}{b-1}\frac{1}{n} + \eta r \leqslant z_{i_0} + \frac{6}{n} + \eta r < z_{i_0} + 20\chi(b)\eta r.$$

This shows that $\mathcal{A} + \mathcal{E}$ has exactly $i_0 - 1$ eigenvalues larger than $z_{i_0} + 20\chi(b)\eta r$.

Now consider $\sigma_{i_0} \leqslant \frac{1}{2}n^2$. By Weyl's inequality, $\widetilde{\sigma}_{i_0} \leqslant \sigma_{i_0} + \|E\| \leqslant (1 + \frac{1}{b})\sigma_{i_0}$. If $(1 + \frac{1}{b})\sigma_{i_0} \leqslant z_{i_0} + 20\chi(b)\eta r$, the proof is already done. Now we assume $(1 + \frac{1}{b})\sigma_{i_0} > z_{i_0} + 20\chi(b)\eta r$. If $(1 + \frac{1}{b})\sigma_{i_0} - (z_{i_0} + 20\chi(b)\eta r) < 20\eta r$, following (130), we have $\chi(b)\sigma_{i_0} \geqslant z_{i_0} > (1 + \frac{1}{b})\sigma_{i_0} - 20(\chi(b) + 1)\eta r$ and thus $\sigma_{i_0} < 80b\eta r\frac{(b-1)(\chi(b)+1)}{4b-5}$. Note that $\frac{(b-1)(\chi(b)+1)}{4b-5}$ is decreasing for $b \geqslant 2$ and $\frac{(b-1)(\chi(b)+1)}{4b-5} \leqslant 17/24$. Hence, $\sigma_{i_0} < 80b\eta r$ contradicts the supposition that $\sigma_{i_0} \geqslant 80b\eta r$.

It suffices to assume $(1 + \frac{1}{b})\sigma_{i_0} - (z_{i_0} + 20\chi(b)\eta r) \geqslant 20\eta r$, which implies that $z_{i_0} \leqslant (1 + \frac{1}{b})\sigma_{i_0} - 20(\chi(b) + 1)\eta r$. To prove $\widetilde{\sigma}_{i_0} \leqslant z_{i_0} + 20\chi(b)\eta r$, we show that f has no zeros on the interval $(z_{i_0} + 20\chi(b)\eta r, (1 + \frac{1}{b})\sigma_{i_0})$. The proof is similar to the proof of *Case 1* when $i_0 = 1$. We only mention the differences. Define

[1]This follows from Lemma 38 and the fact that $\mathrm{Im}(\varphi(z)) \neq 0$ whenever $\mathrm{Im}\, z \neq 0$ for all $|z| > M$.

$\hat{S}_{\sigma_{i_0}}$ as in (139) and the bound (134) also holds for $z \in \hat{S}_{\sigma_{i_0}}$. The goal is to show $\varepsilon(z) < 1/3r$ for all $z \in \mathcal{C}_0$, where $\mathcal{C}_0$ is any circle with radius $10\eta r$ centered at a point $z_0 \in (z_{i_0} + 20\chi(b)\eta r, (1+\frac{1}{b})\sigma_{i_0})$ inside the region $H_{i_0} \cap \hat{S}_{\sigma_{i_0}}$ such that $\mathrm{dist}(z_0, z_{i_0} + 20\chi(b)\eta r) \geqslant 10\eta r$ and $\mathrm{dist}(z_0, (1+\frac{1}{b})\sigma_{i_0}) \geqslant 10\eta r$. If so, by Lemma 41, f has the same number of zeros inside $\mathcal{C}_0$ as g. Note that g has no zeros inside $\mathcal{C}_0$ since $\mathrm{Im}(\varphi(z)) \neq 0$ whenever $\mathrm{Im}\, z \neq 0$ for all $|z| > M$ and $z_{i_0-1} \geqslant \sigma_{i_0-1} - \frac{3b}{b-1}\frac{1}{n} > n^2 - \frac{3b}{b-1}\frac{1}{n} > \frac{3}{2}\sigma_{i_0} \geqslant (1+\frac{1}{b})\sigma_{i_0}$ by Proposition 40. Since $\mathcal{C}_0$ was arbitrarily chosen, $\mathcal{A} + \mathcal{E}$ has no eigenvalues on $(z_{i_0} + 20\chi(b)\eta r, (1+\frac{1}{b})\sigma_{i_0})$.

It remains to bound $\varepsilon(z)$ from (134). Note that $z_0 - z_{i_0} \geqslant 10\eta r + 20\chi(b)\eta r$ and $(1+\frac{1}{b})\sigma_{i_0} - z_0 \geqslant 10\eta r$. For $z \in \mathcal{C}_0$, from $|z - z_0| = 10\eta r$, we get $|z| \geqslant z_0 - 10\eta r \geqslant z_{i_0} + 20\chi(b)\eta r \geqslant \sigma_{i_0} + 20\chi(b)\eta r > \sigma_{i_0}$ and $|z| \leqslant z_0 + 10\eta r \leqslant (1+\frac{1}{b})\sigma_{i_0}$.

The same arguments as those in *Case 1* yield that

$$\max_{i_0 \leqslant l \leqslant r} \chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|} < \frac{1}{3r}$$

for any $z \in \mathcal{C}_0$. We only need to control

$$\max_{1 \leqslant l \leqslant i_0-1} \chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|}.$$

For any $1 \leqslant l \leqslant i_0 - 1$, using similar computation from (140), we get

$$|\sigma_l^2 - \varphi(z)| \geqslant \frac{1}{2}(\sigma_l + z_0 - 10\eta r)(z_l - z_0 - 10\eta r).$$

Plugging in $z_0 \geqslant z_{i_0} + 10\eta r + 20\chi(b)\eta r \geqslant \sigma_{i_0} + 10\eta r + 20\chi(b)\eta r$, we obtain

$$\sigma_l + z_0 - 20\chi(b)\eta r \geqslant \sigma_l + \sigma_{i_0} + 20\chi(b)\eta r.$$

From $\sigma_l \geqslant n^3$, we see $\sigma_{i_0} \leqslant \frac{1}{2}n^2 < \frac{1}{2}\sigma_l$. This, together with (130) and $z_0 \leqslant (1+\frac{1}{b})\sigma_{i_0} - 10\eta r$, implies that

$$z_l - z_0 - 10\eta r \geqslant \sigma_l - (1+\frac{1}{b})\sigma_{i_0} \geqslant \sigma_l - \frac{1}{2}(1+\frac{1}{b})\sigma_l \geqslant \frac{1}{4}\sigma_l.$$

Hence, $|\sigma_l^2 - \varphi(z)| \geqslant \frac{1}{8}\sigma_l(\sigma_l + \sigma_{i_0} + 20\chi(b)\eta r)$ and

$$\begin{aligned}\max_{1 \leqslant l \leqslant i_0-1} \chi(b)\frac{\eta}{|z|}\frac{\sigma_l(\chi(b)|z| + \sqrt{2}\sigma_l)}{|\sigma_l^2 - \varphi(z)|} &\leqslant \max_{1 \leqslant l \leqslant i_0-1} \chi(b)\frac{\eta\sigma_l}{\sigma_{i_0}}\frac{\chi(b)(1+\frac{1}{b})\sigma_{i_0} + \sqrt{2}\sigma_l}{\frac{1}{8}(\sigma_l + \sigma_{i_0} + 20\chi(b)\eta r)\sigma_l} \\ &< 45\frac{\eta}{\sigma_{i_0}} \leqslant \frac{45\eta}{160\eta r} < \frac{1}{3r}\end{aligned}$$

using the assumption $\sigma_{i_0} \geqslant 80b\eta r \geqslant 160\eta r$ and the bound $\chi(b) \leqslant \chi(2) = 9/8$. Therefore, $\varepsilon(z) < 1/3r$ for all $z \in \mathcal{C}_0$.

The proof of *Claim 3* is similar to the previous argument. Let $\mathcal{C}_{j_1}, \mathcal{C}_{j_2}$ be two disjoint circles for $j_1, j_2 \in [\![i_0, r_0]\!]$. Then $|z_{j_1} - z_{j_2}| > 40\chi(b)\eta r$. Let $\mathsf{d} := \mathrm{dist}(\mathcal{C}_{j_1}, \mathcal{C}_{j_2}) > 0$. We show that $\mathcal{A} + \mathcal{E}$ has no eigenvalues lying on the real line between $\mathcal{C}_{j_1}$ and $\mathcal{C}_{j_2}$. Take any point x on the real line between the two circles so that $\mathcal{C}_x$, the circle centered at x with radius $\mathsf{r} := \frac{1}{10}\min\{d, 20\chi(b)\eta r\}$ (say), is inside the region $H_{j_1} \cap S_{\sigma_{j_1}}$ or $H_{j_2} \cap S_{\sigma_{j_2}}$, where $\mathrm{dist}(x, \mathcal{C}_{j_1}) > \mathsf{r}$ and $\mathrm{dist}(x, \mathcal{C}_{j_2}) > \mathsf{r}$. Then using similar calculations as in the proof of *Claim 2*, it suffices to show that $\varepsilon(z) < 1/3r$. The remaining arguments are similar to those in the proof of *Claim 2*; we omit the details.

Acknowledgement: The author would like to thank Sean O'Rourke and Van Vu for their insightful discussions and invaluable suggestions. The author would also like to thank Yuling Yan for bringing relevant results [78, 79] to her attention.

References

[1] E. Abbe, J. Fan, and K. Wang. An ℓ_p theory of PCA and spectral clustering. *The Annals of Statistics*, 50(4):2359–2385, 2022.
[2] E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics*, 48(3):1452–1474, 2020.
[3] J. Agterberg. Distributional theory and statistical inference for linear functions of eigenvectors with small eigengaps. *arXiv preprint arXiv:2308.02480*, 2023.
[4] J. Agterberg, Z. Lubberts, and C. E. Priebe. Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Trans. Inform. Theory*, 68(7):4618–4650, 2022.
[5] R. Allez and J.-P. Bouchaud. Eigenvector dynamics under free addition. *Random Matrices Theory Appl.*, 3(3):1450010, 17, 2014.
[6] N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.
[7] E. Arias-Castro and N. Verzelen. Community detection in dense random networks. *The Annals of Statistics*, 42(3):940–969, 2014.
[8] Z. Bai and J. Yao. Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. H. Poincaré Probab. Statist.*, 44(3):447–474, 2008.
[9] J. Baik, G. B. Arous, S. Péché, et al. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.
[10] J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.
[11] S. Balakrishnan, M. Kolar, A. Rinaldo, A. Singh, and L. Wasserman. Statistical and computational tradeoffs in biclustering. In *NIPS 2011 Workshop on Computational Trade-offs in Statistical Learning*, volume 4, 2011.
[12] Z. Bao, X. Ding, J. Wang, and K. Wang. Statistical inference for principal components of spiked covariance matrices. *The Annals of Statistics*, 50(2):1144–1169, 2022.
[13] Z. Bao, X. Ding, and K. Wang. Singular vector and singular subspace distribution for the matrix denoising model. *The Annals of Statistics*, 49(1):370–392, 2021.
[14] Z. Bao and D. Wang. Eigenvector distribution in the critical regime of BBP transition. *Probability Theory and Related Fields*, pages 1–81, 2021.
[15] F. Benaych-Georges, N. Enriquez, and A. Michaïl. Eigenvectors of a matrix under random perturbation. *Random Matrices Theory Appl.*, 10(2):2150023, 2021.
[16] F. Benaych-Georges, A. Guionnet, and M. Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.*, 16:1621–1662, 2011.
[17] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.
[18] F. Benaych-Georges and R. R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.
[19] L. Benigni. Eigenvectors distribution and quantum unique ergodicity for deformed wigner matrices. *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, 56(4):2822–2867, 2020.
[20] A. Bhardwaj and V. Vu. Matrix perturbation: Davis-Kahan in the infinity norm. *arXiv preprint arXiv:2304.00328*, 2023.
[21] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.
[22] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19(33):1–53, 2014.
[23] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Probability Theory and Related Fields*, 164(1-2):459–552, 2016.

[24] M. Brennan, G. Bresler, and W. Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.

[25] C. Butucea and Y. I. Ingster. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 19(5B):2652–2688, 2013.

[26] C. Butucea, Y. I. Ingster, and I. A. Suslina. Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM: Probability and Statistics*, 19:115–134, 2015.

[27] C. Cai, G. Li, Y. Chi, H. V. Poor, and Y. Chen. Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics*, 49(2):944–967, 2021.

[28] T. T. Cai, T. Liang, and A. Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. *The Annals of Statistics*, 45(4):1403–1430, 2017.

[29] T. T. Cai and A. Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89, 2018.

[30] E. J. Candès and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98:925–936, 2010.

[31] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[32] J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439, 2019.

[33] M. Capitaine. Limiting eigenvectors of outliers for Spiked Information-Plus-Noise type matrices. *Séminaire de Probabilités XLIX*, pages 119–164, 2018.

[34] M. Capitaine and C. Donati-Martin. Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of wigner matrices. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 18(1):129–165, 2021.

[35] Y. Chen, Y. Chi, J. Fan, and C. Ma. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806, 2021.

[36] Y. Chen, J. Fan, C. Ma, and K. Wang. Spectral method and regularized MLE are both optimal for top-K ranking. *The Annals of Statistics*, 47(4):2204–2235, 2019.

[37] Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research*, 17(1):882–938, 2016.

[38] C. Cheng, Y. Wei, and Y. Chen. Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *IEEE Transactions on Information Theory*, 67(11):7380–7419, 2021.

[39] M. Dadon, W. Huleihel, and T. Bendory. Detection and recovery of hidden submatrices. *IEEE Transactions on Signal and Information Processing over Networks*, 2024.

[40] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970.

[41] Y. Dekel, O. Gurel-Gurevich, and Y. Peres. Finding hidden cliques in linear time with high probability. *Combinatorics, Probability and Computing*, 23(1):29–49, 2014.

[42] N. El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 2007.

[43] J. Eldridge, M. Belkin, and Y. Wang. Unperturbed: spectral analysis beyond Davis-Kahan. In *Algorithmic Learning Theory*, pages 321–358. PMLR, 2018.

[44] J. Fan, Y. Fan, X. Han, and J. Lv. Asymptotic theory of eigenvectors for large random matrices. *Journal of the American Statistical Association*, 117(538):996–1009, 2022.

[45] J. Fan, W. Wang, and Y. Zhong. An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18:207–207, 2018.

[46] U. Feige and D. Ron. Finding hidden cliques in linear time. In *Discrete Mathematics and Theoretical Computer Science*, pages 189–204. Discrete Mathematics and Theoretical Computer Science, 2010.

[47] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.

[48] I. C. F. Ipsen and R. Rehman. Perturbation bounds for determinants and characteristic polynomials. *SIAM J. Matrix Anal. Appl.*, 30(2):762–776, 2008.

[49] I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693, 2009.

[50] R. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. *Advances in Neural Information Processing Systems*, 22, 2009.

[51] A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66(11):1663–1750, 2013.

[52] M. Kolar, S. Balakrishnan, A. Rinaldo, and A. Singh. Minimax localization of structural information in large noisy matrices. *Advances in Neural Information Processing Systems*, 24, 2011.

[53] 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*, 52(4):1976–2013, 2016.

[54] V. Koltchinskii and D. Xia. Perturbation of linear forms of singular vectors under Gaussian noise. In *High dimensional probability VII*, volume 71 of *Progr. Probab.*, pages 397–423. Springer, [Cham], 2016.

[55] L. Lei. Unified $\ell_{2\to\infty}$ eigenspace perturbation theory for symmetric random matrices. *arXiv preprint arXiv:1909.04798*, 2019.

[56] G. Li, C. Cai, H. V. Poor, and Y. Chen. Minimax estimation of linear functions of eigenvectors in the face of small eigen-gaps. *arXiv preprint arXiv:2104.03298*, 2021.

[57] M. Löffler, A. Y. Zhang, and H. H. Zhou. Optimality of spectral clustering in the gaussian mixture model. *The Annals of Statistics*, 49(5):2506–2530, 2021.

[58] Y. Luo, R. Han, and A. R. Zhang. A Schatten-q low-rank matrix perturbation analysis via perturbation projection error bound. *Linear Algebra and its Applications*, 630:225–240, 2021.

[59] Z. Ma and Y. Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, pages 1089–1116, 2015.

[60] X. Mao, P. Sarkar, and D. Chakrabarti. Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association*, 116(536):1928–1940, 2021.

[61] F. McSherry. Spectral partitioning of random graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537. IEEE, 2001.

[62] A. Montanari, D. Reichman, and O. Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. *Advances in Neural Information Processing Systems*, 28, 2015.

[63] S. O'Rourke, V. Vu, and K. Wang. Random perturbation of low rank matrices: improving classical bounds. *Linear Algebra Appl.*, 540:26–59, 2018.

[64] S. O'Rourke, V. Vu, and K. Wang. Matrices with Gaussian noise: optimal estimates for singular subspace perturbation. *IEEE Transactions on Information Theory*, 70(3):1978–2002, 2024.

[65] S. O'Rourke and N. Williams. Pairing between zeros and critical points of random polynomials with independent roots. *Trans. Amer. Math. Soc.*, 371(4):2343–2381, 2019.

[66] D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.

[67] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1576–1602. Hindustan Book Agency, New Delhi, 2010.

[68] G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1990.

[69] J. A. Tropp. An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8(1-2):1–230, May 2015.

[70] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[71] U. Von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.

[72] V. Vu. Singular vectors under random perturbation. *Random Structures & Algorithms*, 39(4):526–538, 2011.

[73] V. Vu. A simple SVD algorithm for finding hidden partitions. *Combinatorics, Probability and Computing*, 27(1):124–140, 2018.

[74] V. Q. Vu and J. Lei. Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics*, 41(6):2905–2947, 2013.

[75] R. Wang. Singular vector perturbation under Gaussian noise. *SIAM J. Matrix Anal. Appl.*, 36(1):158–177, 2015.
[76] P.-A. Wedin. Perturbation bounds in connection with singular value decomposition. *Nordisk Tidskr. Informationsbehandling (BIT)*, 12:99–111, 1972.
[77] D. Xia and F. Zhou. The sup-norm perturbation of HOSVD and low rank tensor denoising. *Journal of Machine Learning Research*, 20(61):1–42, 2019.
[78] Y. Yan, Y. Chen, and J. Fan. Inference for heteroskedastic PCA with missing data. *arXiv preprint arXiv:2107.12365*, 2021.
[79] Y. Yan and M. J. Wainwright. Entrywise inference for causal panel data: A simple and instance-optimal approach. *arXiv preprint arXiv:2401.13665*, 2024.
[80] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.
[81] A. R. Zhang, T. T. Cai, and Y. Wu. Heteroskedastic pca: Algorithm, optimality, and applications. *The Annals of Statistics*, 50(1):53–80, 2022.
[82] A. Y. Zhang and H. H. Zhou. Leave-one-out singular subspace perturbation analysis for spectral clustering. *arXiv preprint arXiv:2205.14855*, 2022.
[83] Y. Zhong. Eigenvector under random perturbation: A nonasymptotic Rayleigh-Schrödinger theory. *arXiv preprint arXiv:1702.00139*, 2017.
[84] Y. Zhong and N. Boumal. Near-optimal bounds for phase synchronization. *SIAM J. Optim.*, 28(2):989–1016, 2018.

Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

E-mail address: kewang@ust.hk