Submitted to the Annals of Statistics

ANALYSIS OF SINGULAR SUBSPACES UNDER RANDOM PERTURBATIONS

BY KE WANG[1,a],

[1]*Department of Mathematics, The Hong Kong University of Science and Technology,* [a]*kewang@ust.hk*

We present a comprehensive analysis of singular vector and singular subspace perturbations in the signal-plus-noise matrix model with random Gaussian noise. 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, building on previous results by O'Rourke, Vu, and the author. Our analysis provides fine-grained insights, including ℓ_∞ bounds for singular vectors, $\ell_{2,\infty}$ bounds for singular subspaces, and results for linear and bilinear functions of singular vectors. Additionally, we derive $\ell_{2,\infty}$ bounds on perturbed singular vectors, taking into account the weighting by their corresponding singular values. Finally, we explore practical implications of these results in 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, 49], principal component analysis (PCA) [48], and community detection [66, 67], 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 $r \geq 1$. The singular value decomposition (SVD) of A takes the form $A = UDV^T$, where $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ is a diagonal matrix containing the non-zero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ of A; the columns of the matrices $U = (u_1, \ldots, u_r)$ and $V = (v_1, \ldots, v_r)$ 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 $i \geq 1$. 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 $\widetilde{\sigma}_1 \geq \widetilde{\sigma}_2 \geq \cdots \geq \widetilde{\sigma}_{\min\{N,n\}} \geq 0$, 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 $1 \leq k \leq r$, let us denote

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

MSC2020 subject classifications: Primary 60B20; secondary 62H25, 62H30.

Keywords and phrases: Singular vector perturbation, singular subspace perturbation, low-rank structures, random matrices, spectral clustering, mixture models, submatrix localization.

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 the supplementary material [69].

Denote $\operatorname{diag}(\sigma_i - \widetilde{\sigma}_i) = \operatorname{diag}(\sigma_1 - \widetilde{\sigma}_2, \cdots, \sigma_{\min\{N,n\}} - \widetilde{\sigma}_{\min\{N,n\}})$. 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 [64]).

THEOREM 1.1 (Mirsky). *Let $\widetilde{A} = A + E$ as in (1). Then for any unitarily invariant norm $|||\cdot|||$,*

$$|||\operatorname{diag}(\sigma_i - \widetilde{\sigma}_i)||| \le |||E|||.$$

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 $0 \le \theta_1 \le \cdots \le \theta_k \le \pi/2$. These angles measure the distance between the two subspaces. A detailed definition is given in Section 2.1 below. Denote

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

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 [71]).

THEOREM 1.2 (Wedin [71]). *Let $\widetilde{A} = A + E$ as in (1). If $\hat{\delta}_k := \sigma_k - \widetilde{\sigma}_{k+1} > 0$, then for any unitarily invariant norm $|||\cdot|||$,*

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

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

In the context of a 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 the supplementary material [69], 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 this paper, we aim to develop a stochastic variant of Wedin's theorem under these additional assumptions that the signal matrix A is low-rank and the noise E consists of i.i.d. Gaussian entries. This work builds on the recent developments in [60, 61], and offers several substantial improvements. We extend the classical Davis-Kahan-Wedin theorem for unitarily invariant norms, provide sharper bounds with improved dependence on the signal rank r, and relax several technical assumptions used in earlier analyses. A more detailed comparison with prior works appears in Section 4.2.

There is currently a surging interest in ℓ_∞ analysis (also known as entrywise analysis) of eigenvectors and singular vectors. This dynamic research area focuses on deriving rigorous bounds, such as ℓ_∞ bounds [2, 20, 36, 43, 45, 78] for eigenvectors or singular vectors, and $\ell_{2,\infty}$ bounds for eigenspaces or singular subspaces [1, 4, 27, 32, 53], in relation to perturbed matrix models. These analyses have significant impact across statistics and machine learning applications.

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 $A+E$. We have also established results for linear and bilinear forms of the perturbed singular vectors and subspaces, and investigated the $\ell_{2,\infty}$ bounds on the perturbed singular vectors weighted by their singular values. These new results are presented in Section 2.2. Throughout, we assume that the noise E has i.i.d. Gaussian entries; extensions to sub-Gaussian noise are discussed in Remark 2.15. Section 4 summarizes our main contributions and discusses the optimality of our bounds, with a focus on rank dependence. A high-level overview of the proof strategy appears in Section 6.2, following notation in Section 6.1. Full proofs are deferred to the supplementary material [69].

In Section 5, 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: Section 2 presents our new matrix perturbation results, with Section 2.1 extending Wedin's $\sin\Theta$ theorem to stochastic versions for arbitrary unitarily invariant norms. Section 2.2 provides results on ℓ_∞ and $\ell_{2,\infty}$ norms of singular vectors and subspaces, while Section 3 surveys related literature. Section 4 discusses our main contributions and their optimality. Applications to Gaussian mixture model and submatrix localization are presented in Sections 5.1 and 5.2. Section 6 introduces basic tools and proof strategies, with detailed proofs and extensive numerical simulations provided in the supplementary material [69].

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 $f(n)\le Cg(n)$ 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 $\mathbb{O}^{n\times n}$. We denote by $A\oplus B:=\begin{pmatrix}A&0\\0&B\end{pmatrix}$ the direct sum (block diagonal concatenation) of two matrices A and B.

2. New results on the matrix perturbation bounds.

2.1. *Stochastic Wedin's* $\sin\Theta$ *theorem.* We first generalize the previous results in [60, 61] to an arbitrary unitarily invariant norm $|||\cdot|||$. We start with the concept of principle angles. Let H and W are two subspaces each of dimension k. The principal angles $0 \le \theta_1 \le \cdots \le \theta_k \le \pi/2$ between H and W are defined recursively as follows:

$$\cos(\theta_i) = \max_{h\in H, w\in W} h^T w = h_i^T w_i, \qquad \|h\| = \|w\| = 1$$

subject to the constraint

$$h_i^T h_l = 0, \quad w_i^T w_l = 0 \quad \text{for } l = 1, \ldots, i-1.$$

Denote $\angle(H,W) := \mathrm{diag}(\theta_1, \cdots, \theta_k)$ and $\sin\angle(H,W) := \mathrm{diag}(\sin\theta_1, \cdots, \sin\theta_k)$.

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

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

$P_{U_{k,s}}$ the orthogonal projection onto $U_{k,s}$, and analogously for $V_{k,s}$, $\widetilde{V}_{k,s}$ and $P_{V_{k,s}}$. Denote

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

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 2.1 (Unitarily invariant norms: simplified asymptotic version). *Let A and E be $N \times n$ real matrices, where A is deterministic with rank $r \ge 1$ and the entries of E are i.i.d. $\mathcal{N}(0,\tau^2)$ random variables. Let $|||\cdot|||$ be any normalized, unitarily invariant norm. Consider $1 \le k \le r$ such that $\delta_k \gtrsim \tau r\sqrt{r+\log(N+n)}$. Denote $k_0 = \min\{k, r-k\}$. Then with probability $1-(N+n)^{-C}$ for some $C > 0$,*

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

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

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

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 1.2 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 1.2. 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

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

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 $C(r) \lesssim r^{3/2}$. This improvement is particularly important in statistical applications where $\delta_k \ll \sigma_k$, which frequently occurs in settings such as principal component analysis, matrix completion, and spiked covariance models. In these regimes, the Wedin's bounds become loose or even vacuous because they require the spectral gap to be comparable to the noise level. By contrast, our bounds remain effective by leveraging the signal strength σ_k, providing meaningful and tighter control even when $\delta_k \ll \sigma_k$.

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 2.2. *Under the assumptions of Theorem 2.1, the following holds with probability* $1-(N+n)^{-C}$,

$$|||\sin\angle(U_k, \widetilde{U}_k)||| \lesssim \tau\sqrt{kk_0}\frac{\sqrt{r+\log(N+n)}}{\delta_k} + k\frac{\|E\|}{\sigma_k}. \tag{6}$$

Theorem 2.1 follows immediately from the next general and non-asymptotic result. For notation, let $\sigma_0 := \infty$, $\delta_0 := \infty$, and working with $\tau^{-1}(A+E)$, we assume E has $\mathcal{N}(0,1)$ entries. We define

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

THEOREM 2.3 (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 $r \geq 1$. Let $K > 0$ and $b \geq 2$. Denote $\eta := \frac{11b^2}{(b-1)^2}\sqrt{2(\log 9)r + (K+7)\log(N+n)}$. Assume $(\sqrt{N}+\sqrt{n})^2 \geq 32(K+7)\log(N+n) + 64(\log 9)r$. Consider $1 \leq r_0 \leq r$ such that $\sigma_{r_0} \geq 2b(\sqrt{N}+\sqrt{n}) + 80b\eta r$ and $\delta_{r_0} \geq 75\chi(b)\eta r$. For any $1 \leq k \leq s \leq r_0$, if $\min\{\delta_{k-1}, \delta_s\} \geq 75\chi(b)\eta r$, then*

$$\begin{aligned}|||\sin\angle(U_{k,s}, \widetilde{U}_{k,s})||| \leq& 6\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\sqrt{\min\{s-k+1, r-s+k-1\}}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}} \\ &+ 2\frac{|||P_{U^\perp}EP_{\widetilde{V}_{k,s}} \oplus P_{V^\perp}E^{\mathrm{T}}P_{\widetilde{U}_{k,s}}|||}{\sigma_s}\end{aligned} \tag{7}$$

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

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

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

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

REMARK 2.4. *The parameters K and b are user-specified positive constants. The parameter K controls the probability level: larger K yields a smaller failure probability $(N+n)^{-K}$ at the cost of slightly larger constants. The parameter b governs the trade-off between the required singular value separation (through $\chi(b)$) and the constants in the bound. A larger b corresponds to a stronger signal regime: it reduces $\chi(b)$, thereby relaxing the spectral gap assumption, and can also lead to sharper constants in the final estimates.*

REMARK 2.5. *Throughout the proofs, we work on the event that* $\|E\| \le 2(\sqrt{N}+\sqrt{n})$. *Lemma 6.7 below guarantees this event holds with very high probability. In Theorem 2.3, the parameter* $b \ge 2$ *represents the signal-to-noise ratio, and in the proof, we ensure that* $\sigma_{r_0}/\|E\| \ge b$. *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* $\|E\| \le (1+\epsilon_1)(\sqrt{N}+\sqrt{n})$, *and assume* $b \ge 1+\epsilon_2$ *for* $\epsilon_1, \epsilon_2 > 0$. *By following the same proof, one can arrive at refined constants and bounds.*

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 [60]. 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 2.6 (Singular value bounds: general noise). *Assume* A *has rank* r *and* E *is random. Let* $1 \le k \le r$. *Consider any* $\varepsilon \in (0,1)$.

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

$$\widetilde{\sigma}_k \ge \sigma_k - t. \tag{8}$$

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

$$\widetilde{\sigma}_k \le \sigma_k + 2\sqrt{k}\frac{B^2}{\widetilde{\sigma}_k} + k\frac{B^3}{\widetilde{\sigma}_k^2} + L. \tag{9}$$

THEOREM 2.7 (Singular subspace bounds: general noise). *Assume* A *has rank* r *and* E *is random. Let* $1 \le k \le r$. *For* $\varepsilon > 0$, *assume there exist* $L, B > 0$ *such that* $\|U^{\mathrm{T}} E V\| \le L$ *and* $\|E\| \le B$ *with probability at least* $1-\varepsilon$. *Furthermore, assume* $\delta_k = \sigma_k - \sigma_{k+1} \ge 2L$. *Then for any normalized, unitarily invariant norm* $|||\cdot|||$, *the following holds with probability at least* $1-\varepsilon$,

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

More specifically, for the operator norm,

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

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

To apply Theorems 2.6 and 2.7, it is necessary to obtain effective bounds on $\|E\|$ and $\|U^{\mathrm{T}} E V\|$. In general, matrix concentration inequalities (refer to [65] 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, $U^{\mathrm{T}} E V$ is an $r \times r$ matrix, and $\|U^{\mathrm{T}} E V\|$ typically depends on r. Bounds on $\|U^{\mathrm{T}} E V\|$ can be obtained by applying concentration inequalities. We note that in the symmetric setting, Eldridge et al. [43, Theorem 6] established related eigenvalue upper bounds. Both approaches begin with the min-max characterization, but differ in methodology: we use a direct bilinear form argument, while they leverage signal-aligned subspaces. For random noise matrices, our bounds can provide better control when k is not too large, while their framework offers additional precision through explicit exploitation of spectral gaps when present.

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 $1 \le k \le s \le r$. 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 $\sigma_1 \le n^2$. 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. Theorem 2.10 later will provide a precise treatment of these strong singular values (i.e., without the assumption $\sigma_1 \le n^2$). Denote $\sigma_0 := \infty$ and $\delta_0 := \infty$.

THEOREM 2.8 (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 $r \ge 1$ and the entries of E are i.i.d. $\mathcal{N}(0, \tau^2)$ random variables. Let $1 \le k \le r$ and $\sigma_k \ge 2\|E\|$. Assume that $\sigma_1 \le n^2$.*

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

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

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

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

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 A.5 in [69], 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 A.6 in [69], we establish

$$\|\widetilde{U}_k - U_k O\|_{2,\infty} \le \|\widetilde{U}_k - P_{U_k}\widetilde{U}_k\|_{2,\infty} + \|U_k\|_{2,\infty}\|\sin\angle(U_k, \widetilde{U}_k)\|^2. \tag{10}$$

Therefore, by combining Theorem 2.3 with Theorem 2.8, we obtain the next result.

COROLLARY 2.9. *Under the same assumption as Theorem 2.8, the following holds:*

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

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

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

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

Theorems 2.8 follows as a direct consequence of the next general and non-asymptotic result, which we will prove in [69]. By a scaling, it suffices to assume E has $\mathcal{N}(0,1)$ entries.

THEOREM 2.10. *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 $r \geq 1$. Let $K > 0$ and $b \geq 2$. Denote $\eta := \frac{11b^2}{(b-1)^2}\sqrt{2(\log 9)r + (K+7)\log(N+n)}$. Assume $(\sqrt{N}+\sqrt{n})^2 \geq 32(K+7)\log(N+n) + 64(\log 9)r$. Consider $1 \leq r_0 \leq r$ such that $\sigma_{r_0} \geq 2b(\sqrt{N}+\sqrt{n}) + 80b\eta r$ and $\delta_{r_0} \geq 75\chi(b)\eta r$. For any $1 \leq k \leq s \leq r_0$, if $\min\{\delta_{k-1}, \delta_s\} \geq 75\chi(b)\eta r$, then with probability at least $1 - 40(N+n)^{-K}$,*

$$\begin{aligned}\|\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\|_{2,\infty} \leq\ & 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|U\|_{2,\infty}\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} \\ & + \frac{4\sqrt{2}b^2}{(b-1)^2}\sqrt{\sum_{i\in[\![k,s]\!],\sigma_i\leq n^2}\frac{\gamma^2}{\sigma_i^2} + \sum_{i\in[\![k,s]\!],\sigma_i>n^2}\frac{16n}{\sigma_i^2}},\end{aligned} \tag{11}$$

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 $\sum_{k\leq i\leq s,\sigma_i>n^2}\frac{16n}{\sigma_i^2} < \frac{16}{n^2}$ 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. The complete result follows a format similar to Theorem 2.10. The proofs are provided in [69].

THEOREM 2.11 (Bounds on linear and bilinear forms). *Under the assumptions of Theorem 2.10, 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 $1 - 40(N+n)^{-K}$:*

$$\begin{aligned}\|x^{\mathrm{T}}(\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s})\| \leq\ & 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|x^{\mathrm{T}}U\|\frac{\eta\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} \\ & + \frac{4\sqrt{2}b^2}{(b-1)^2}\sqrt{\sum_{i\in[\![k,s]\!],\sigma_i\leq n^2}\frac{\gamma^2}{\sigma_i^2} + \sum_{i\in[\![k,s]\!],\sigma_i>n^2}\frac{16n}{\sigma_i^2}}\end{aligned}$$

and

$$\begin{aligned}\left|x^{\mathrm{T}}(\widetilde{U}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s})y\right| \leq\ & 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|x^{\mathrm{T}}U\|\frac{\eta\sqrt{\|y\|_0}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}} \\ & + \frac{4\sqrt{2}b^2}{(b-1)^2}\left[\sum_{i\in[\![k,s]\!],\sigma_i\leq n^2}\gamma\frac{|y_i|}{\sigma_i} + \sum_{i\in[\![k,s]\!],\sigma_i>n^2}\frac{4\sqrt{n}}{\sigma_i}\right],\end{aligned} \tag{12}$$

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

REMARK 2.12. *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 2.9, one can leverage the fact provided in Proposition A.6 of [69]:*

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

and combine Theorems 2.3 and 2.11.

A natural extension is to consider entrywise control of the perturbation. By applying (12) with the canonical vectors and using Proposition A.6 of [69], we obtain the following bound under the matrix max-norm (which measures the largest absolute entry): with probability $1-(N+n)^{-C}$,

(13)
$$\|\widetilde{U}_{k,s}-U_{k,s}O\|_{\max}\lesssim\frac{\sqrt{r+\log(N+n)}}{\min\{\delta_{k-1},\delta_s\}}\|U\|_{2,\infty}+\frac{\sqrt{r\log(N+n)}}{\sigma_s}+\frac{\|E\|^2}{\sigma_s^2}\|U_{k,s}\|_{2,\infty}$$

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 the supplementary material [69].

THEOREM 2.13 (Bounds on singular value-adjusted projection perturbation). *Under the assumptions of Theorem 2.10, the following holds with probability at least* $1-40(N+n)^{-K}$*:*

$$\|\widetilde{U}_{k,s}\widetilde{D}_{k,s}-P_{U_{k,s}}\widetilde{U}_{k,s}\widetilde{D}_{k,s}\|_{2,\infty}\le 3\sqrt{2}\frac{(b+1)^2}{(b-1)^2}\|U\|_{2,\infty}\frac{\eta\sigma_k\sqrt{s-k+1}}{\min\{\delta_{k-1},\delta_s\}}\mathbf{1}_{\{s-k+1\neq r\}}$$
$$+\frac{4\sqrt{2}b^2}{(b-1)^2}\sqrt{\gamma^2(s-k+1)+16},$$

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

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, requires more analysis than the unweighted case in Corollary 2.9. Consider the top k-singular subspaces for any $1\le k\le r$. We establish that for some $O\in\mathbb{O}^{k\times k}$:

$$\|\widetilde{U}_k\widetilde{D}_k-U_kO\widetilde{D}_k\|_{2,\infty}\le\|\widetilde{U}_k\widetilde{D}_k-P_{U_k}\widetilde{U}_k\widetilde{D}_k\|_{2,\infty}$$
$$+\|U_k\|_{2,\infty}\|\sin\angle(U_k,\widetilde{U}_k)\|\left(\|E\|+\sigma_{k+1}\|\sin\angle(V_k,\widetilde{V}_k)\|\right). \tag{14}$$

The proof of (14) is provided in [69]. Consequently, the weighted case bounds follow from combining Theorem 2.13 with Theorem 2.3. Of particular interest is the special case $k=r$, which is crucial for our statistical applications. Here, (14) simplifies to:

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

Prior works (e.g., Proposition 3 in [74]) have used an alternative formulation $\|\widetilde{U}_r\widetilde{D}_rO^{\mathrm{T}}-UD\|_{2,\infty}$. In [69] (after the proof of (14)), we analyze their connections.

Combining (15) with Theorems 2.13 and 2.3 yields the following key result for our applications:

COROLLARY 2.14. *Assume the setting of Theorem 2.10, where we set* $r_0=s=r$ *and* $k=1$*, the following holds with probability at least* $1-40(N+n)^{-K}$*:*

$$\min_{O\in\mathbb{O}^{r\times r}}\|\widetilde{U}_r\widetilde{D}_r-UO\widetilde{D}_r\|_{2,\infty}\le\frac{72b^4}{(b-1)^4}r\sqrt{(K+7)\log(N+n)}+2\|U\|_{2,\infty}\frac{\|E\|^2}{\sigma_r}.$$

The proof technique of Theorem 2.13 could be extended to bound the bilinear form $x^{\mathrm{T}}(\widetilde{U}_{k,s}\widetilde{D}_{k,s} - P_{U_{k,s}}\widetilde{U}_{k,s}\widetilde{D}_{k,s})y$ for arbitrary unit vectors x and y, similar to the analysis in Theorem 2.11. However, we do not pursue this generalization in the present work.

REMARK 2.15. *Our theoretical framework, while presented for noise matrices E with i.i.d. centered Gaussian entries, extends naturally to more general settings. The underlying methodology accommodates sub-Gaussian noise matrices through established random matrix theory techniques, particularly isotropic local laws and concentration inequalities. The detailed technical treatment of these extensions, including sub-Gaussian cases and more general noise assumptions, will be addressed in forthcoming work.*

3. Related Work.

This section reviews recent developments in singular vector perturbation theory and their connections to high-dimensional statistics and random matrix theory.

Classical and ℓ_2-Type Subspace Perturbation Results. Foundational results such as the Davis–Kahan and Wedin $\sin\Theta$ theorems [40, 71] provide classical ℓ_2-type perturbation bounds for eigenvectors and singular subspaces. Several deterministic extensions refine or generalize these results. Yu, Wang, and Samworth [75] replace empirical gaps with population gaps in the Frobenius norm setting. Vu and Lei [68] offer a variational form applicable to unaligned subspaces. Cai and Zhang [29] and Luo, Han, and Zhang [56] further generalize to unbalanced dimensions and Schatten-q norm bounds, respectively. Zhang and Zhou [76] give Frobenius norm bounds for closely related matrix pairs, useful for spectral clustering.

In the stochastic regime, Wang [70] analyzes the non-asymptotic distribution of singular vectors under Gaussian noise. Allez and Bouchaud [5] study eigenvector dynamics in symmetric matrices with Brownian motion noise. Benaych-Georges, Enriquez, and Michaïl [15] analyze a perturbative expansion for eigenvector coordinates in symmetric matrices with stochastic noise. Zhong [77] develops a Rayleigh–Schrödinger-type expansion to study angular perturbations of leading eigenvectors in symmetric low-rank plus noise models.

ℓ_∞ and $\ell_{2,\infty}$-Type Subspace Perturbation Results. Recent works have increasingly focused on more refined, entrywise and row-wise perturbation bounds, which offer localized guarantees particularly useful in applications such as submatrix localization and community detection. A foundational contribution in this direction is due to Fan, Wang, and Zhong [45], who derive ℓ_∞ perturbation bounds under incoherence conditions on the low-rank signal matrix. Building on this idea, a series of works [1, 2, 20, 27, 36, 73, 74] develop increasingly sharp ℓ_∞ and $\ell_{2,\infty}$ bounds using leave-one-out techniques tailored to various noise models and matrix structures.

Abbe, Fan, Wang, and Zhong [2] establish $\ell_{2,\infty}$ perturbation bounds for symmetric low-rank plus noise matrices with sub-Gaussian noise. This is further extended by Abbe, Fan, and Wang [1], who provide a comprehensive analysis of $\ell_{2,p}$ norms in the context of hollowed PCA. Chen, Fan, Ma, and Wang [36] generalize the analysis to asymmetric transition matrices, while Lei [53] considers more intricate dependence structures in the noise. Zhong and Boumal [78] obtain ℓ_∞ bounds for the leading eigenvector in the phase synchronization model, using tools from semidefinite programming. Bhardwaj and Vu [20] propose a stochastic analogue of the Davis–Kahan theorem that yields ℓ_∞ bounds for both eigenvectors and singular vectors under general noise. More recently, Yan and Wainwright [74] introduce a novel expansion technique that yields sharp $\ell_{2,\infty}$ bounds and distributional characterizations of the estimation error, refining earlier results by Yan, Chen, and Fan [73]. Agterberg, Lubberts, and Priebe [4] also contribute by analyzing heteroskedastic noise and deriving Berry–Esseen-type results for entrywise estimation of singular vectors.

Several works investigate perturbation of linear and bilinear forms involving singular vectors. Koltchinskii and Xia [52] derive concentration bounds for such forms, later extended to tensor settings by Xia and Zhou [72]. Cape, Tang, and Priebe [32] provide $\ell_{2,\infty}$ bounds for structured matrices, and Eldridge, Belkin, and Wang [43] use a Neumann series expansion to control entrywise deviations. Li, Cai, Poor, and Chen [54] propose de-biased estimators for projected eigenvectors under mild eigengap assumptions. Recent works [3, 38] explore this regime further, especially when eigen-gaps are small or scale with noise.

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

Connections to Random Matrix Theory. Random matrix theory offers asymptotic insights into the spectral behavior of low-rank plus noise models. The BBP phase transition [9] describes a critical threshold for the emergence of outlier eigenvalues and eigenvector localization. Subsequent studies [8, 10, 12–14, 16–19, 23, 33, 34, 42, 44, 62] explore the fine behavior of extreme eigenvalues and eigenvectors across different regimes and ensembles. While these results are typically asymptotic and ensemble-specific, they inform the design and understanding of non-asymptotic perturbation techniques used in this work.

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

4. Discussions.

This section summarizes the key contributions of this work, and examines the lower bounds and the role of rank r in our perturbation bounds.

4.1. *Optimality of our results.*

Our operator norm bound in Theorem 2.1 improves upon prior results, notably [60], in its dependence on the signal rank r. In particular, when $k = 1$, we show that with probability $1 - (N+n)^{-C}$,

$$\sin\angle(u_1, \widetilde{u}_1) \lesssim \frac{\sqrt{r + \log(N+n)}}{\delta_1} + \frac{\|E\|}{\sigma_1}. \tag{16}$$

This bound is conjectured to be near-optimal, up to constant factors. First, the necessity of the second term in (16) is supported by a high-probability lower bound established in [61, Theorem 3], which shows that $\max\{\sin\angle(u_1, \widetilde{u}_1), \sin\angle(v_1, \widetilde{v}_1)\} \gtrsim \frac{\|E\|}{\sigma_1}$. Second, the first term of order $\sqrt{r + \log(N+n)}/\delta_1$ merits further attention. A line of work including [27, 29, 35, 36, 38, 54] establishes minimax lower bounds for singular subspace estimation. In particular, [38, Theorem 3] shows that, under Gaussian noise, any estimator must incur error at least of order $1/\delta_1$. This confirms the necessity of the $1/\delta_1$ scaling in our bound.

To better understand the role of r, we conduct simulations varying r and plot the empirical distribution of $\sin\angle(u_1, \widetilde{u}_1)$. The left panel of Figure 1 shows that the distribution depends on the rank r, while the right panel demonstrates its sensitivity to the matrix dimensions, which controls the noise level $\|E\| \asymp \sqrt{n}$. Although a matching lower bound capturing the full $\sqrt{r}$ dependence is not currently known, we denote the unknown scaling factor as $f(r)$ and empirically investigate its behavior in the supplementary material. These simulations suggest a sublinear but non-negligible scaling with r, with $\sqrt{r}$ emerging as a plausible candidate.

For the ℓ_∞ bound, we observe that

$$\min_{\mathsf{s}\in\{\pm 1\}} \|u_1 - \mathsf{s}\widetilde{u}_1\|_\infty \geq \min_{\mathsf{s}\in\{\pm 1\}} \frac{1}{\sqrt{N}}\|u_1 - \mathsf{s}\widetilde{u}_1\| \geq \frac{1}{\sqrt{N}}\sin\angle(u_1, \widetilde{u}_1).$$

Combining the above discussion, we expect the following lower bound:

$$\min_{\mathsf{s}\in\{\pm 1\}} \|u_1 - \mathsf{s}\widetilde{u}_1\|_\infty \gtrsim \frac{1}{\sqrt{N}}\frac{f(r)}{\delta_1} + \frac{1}{\sigma_1}.$$

On the other hand, under the incoherence assumption $\|U\|_{\max} \lesssim \frac{1}{\sqrt{N}}$, our bound in Corollary 2.9 (with $\tau = 1$) gives with high probability

$$\min_{s\in\{\pm 1\}} \|u_1 - s\widetilde{u}_1\|_\infty \le \frac{1}{\sqrt{N}} \frac{\sqrt{r}\sqrt{r+\log(N+n)}}{\delta_1} + \frac{\sqrt{r\log(N+n)}}{\sigma_1}.$$

The gap in r-dependence between the upper and lower bounds is at most a factor of r (up to logs), suggesting near-optimal rank scaling in our entrywise bound.

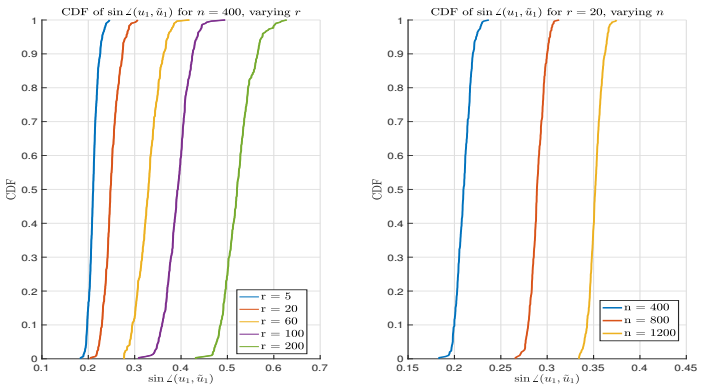

FIG 1. *CDF plots of* $\sin\angle(u_1, \widetilde{u}_1)$ *across 300 trials. The signal matrix* $A \in \mathbb{R}^{n\times n}$ *has rank* r*. The noise matrix* E *has i.i.d. standard Gaussian entries. We set the largest singular value of* A *as* $\sigma_1 = 100$*.* ***Left:*** *We set* $n = 400$*,* $\delta_1 = 20$ *and* $\sigma_2 = \ldots = \sigma_r = 80$*, varying* $r = 5, 20, 60, 100, 200$*.* ***Right:*** *We fix* $r = 20$*,* $\delta_1 = 40$*, and* $\sigma_2 = \ldots = \sigma_r = 60$*, and vary the dimension* $n = 400, 800, 1200$*.*

The supplementary material [69] provides extensive numerical validation of our bounds. While these experiments confirm the tightness of rank dependence in angular and Frobenius norm bounds, they also reveal opportunities for improvement. Specifically, the ℓ_∞ and $\ell_{2,\infty}$ bounds might be refined under incoherence assumptions, and the operator norm bounds could be slightly improved through refined analysis.

4.2. *Contribution of This Work.* This paper establishes near-optimal, non-asymptotic perturbation bounds for singular vectors in the low-rank matrix denoising model, under the assumption that the noise matrix has i.i.d. Gaussian entries. Building on the framework of [61], which introduced the isotropic local law into singular vector analysis, we extend and refine the results in several important directions.

First, we obtain tighter perturbation bounds under milder assumptions by sharpening the decomposition of error terms and enhancing the control of resolvent expansions. Most notably, we eliminate a restrictive condition in [61], which required distinct singular values among $\sigma_k, \cdots, \sigma_s$ to be separated by a distance of order $r^2\sqrt{\log(N+n)}$. This restrictive spectral gap condition, often challenging to verify in practice, is no longer necessary in our theorem through our refined analysis of perturbed singular value locations.

Second, we provide a more detailed investigation of the role of signal rank r, establishing bounds with improved r-dependence. In several key regimes, this dependence is shown to be near-optimal, with our numerical experiments supporting its necessity. Our analysis addresses key limitations of previous works: while [60] provided bounds for general noise without utilizing the isotropic local law, it included an extra additive term that we eliminate here under Gaussian noise. Similarly, [61], though introducing the isotropic local law for Gaussian noise,

focused primarily on operator norm bounds without optimizing rank scaling. The current framework thus achieves sharper results by employing a more delicate decomposition of the perturbation terms and sharper control over resolvent expansions.

Our bounds are formulated under a wide class of norms, including unitarily invariant norms, the $\ell_{2,\infty}$ norm, and a weighted $\ell_{2,\infty}$ norm that captures heterogeneity in the signal structure. This weighted bound is especially powerful in statistical applications, where standard norm-based bounds often fail to achieve optimal practical performance. We illustrate its utility through two well-studied problems–Gaussian mixture model and submatrix localization–in Section 5. These examples demonstrate how our general perturbation bounds yield concise and effective guarantees for important statistical tasks.

Finally, the isotropic local law framework we adopt proves to be both flexible and robust. It accommodates various matrix norms and opens the door to further extensions beyond Gaussian noise.

5. Applications.

5.1. *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

$$X = \mathbb{E}(X) + E. \tag{17}$$

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:

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

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, 55, 76], 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 [55] 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. If the rank r of $\mathbb{E}(X)$ is less than k, then Σ will have $k-r$ zero diagonal entries. 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 [69], 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 holds:

$$\max_{1\le j\le n} \|(\widetilde{U}_k^T X)_j - (U^{\mathrm{T}}\mathbb{E}(X))_j\| < \frac{1}{5}\Delta. \tag{18}$$

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 $\widetilde{\mathbf{z}} \in [k]^n$.
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 [55] and [76]. 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 [55, 67] 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 (18); 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 5.1. *Consider the GMM (17) 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 $(\sqrt{n}+\sqrt{p})^2 \ge 32(L+7)\log(n+p) + 64(\log 9)k$. If*

$$\Delta \ge \max\left\{\frac{40(\sqrt{n}+\sqrt{p})}{\sqrt{c_{\min}}}, 1800k\sqrt{(L+7)\log(n+p)}\right\}, \tag{19}$$

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

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

The proof of Theorem 5.1 is a direct application of Corollary 2.14 and is detailed in [69]. By setting $L = (n+p)/\log(n+p)$, for instance, we achieve an exponential rate of misclassification.

Löffler, Zhang and Zhou [55] 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:

$$\mathbb{E}\mathcal{M}(\mathbf{z}, \widetilde{\mathbf{z}}) \le \exp\left(-(1-o(1))\Delta^2/8\right) + \exp(-0.08n). \tag{20}$$

More recently, Zhang and Zhou [76] have developed another innovative approach to analyze the output $\widetilde{\mathbf{z}}$ and obtained the same asymptotic exponential error rate (20) for the GMM, assuming

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

Additionally, [76, 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 exactly all r singular vectors of $\mathbb{E}(X)$), an exponential error rate is attainable when

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

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 [55] and [76] 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 5.1 aligns with the findings for the sub-Gaussian mixture model in [76]. 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 6.2. Our findings can be extended to scenarios where the entries of E are sub-Gaussian random variables. The extension to sub-Gaussian noise can be achieved through standard random matrix techniques, as discussed in Remark 2.15.

5.2. *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 $\mathcal{Q}$. In particular, 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, 24–26, 28, 37, 39, 41, 46, 47, 51, 57–59, 67] and references therein).

The task of recovering a single submatrix has been intensively explored (see for instance, [26, 28, 37, 47, 58, 67] 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 $m \times n$:

$$X = M + E, \tag{21}$$

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 $1 \leq i \leq k$. 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 $\min_{i,j} |M_{ij}| = \min_{1 \leq l \leq k} |\lambda_l| \gtrsim \sqrt{\log n}$ 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 5.1, we apply the same spectral clustering method to locate the submatrices. Denote $C_0 := [n] \setminus \cup_{i=1}^k C_i$ 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 $0 \leq l \leq k$, it holds that $(U^{\mathrm{T}}M)_i = (U^{\mathrm{T}}M)_j$. For $i, j \in [n]$ from different submatrices, we have that

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

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,

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

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

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

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

Define

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

THEOREM 5.2. *Consider the submatrix localization model (21) with k submatrices. Let $L > 0$ and assume $(\sqrt{n} + \sqrt{p})^2 \geq 32(L+7)\log(n+p) + 64(\log 9)k$. Given that*

$$\Delta_R \geq \max\left\{ \frac{40(\sqrt{m} + \sqrt{n})}{\sqrt{r_{\min}}}, 1800k\sqrt{(L+7)\log(m+n)} \right\},$$

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 $\{\widetilde{R}_i\}_{i=0}^k$.

$$\Delta_C \geq \max\left\{\frac{40(\sqrt{m}+\sqrt{n})}{\sqrt{c_{\min}}}, 1800k\sqrt{(L+7)\log(m+n)}\right\},$$

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

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

The proof of Theorem 5.2 parallels that of Theorem 5.1, and therefore we omit the details.

Previous research on the model (21) 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 $1 \leq i \leq k$ and $\min\{K_R, K_C\} \gtrsim \max\{\sqrt{m}, \sqrt{n}\}$. Given that

$$\lambda \gtrsim \frac{\sqrt{k}}{\min\{\sqrt{K_R}, \sqrt{K_C}\}} + \max\left\{\sqrt{\frac{\log m}{K_C}}, \sqrt{\frac{\log n}{K_R}}\right\} + \frac{\sqrt{m}+\sqrt{n}}{\sqrt{K_R K_C}}, \tag{22}$$

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 $1 \leq i \leq k$, and furthermore $r_0 \gtrsim K_R$ and $c_0 \gtrsim K_C$, our analysis indicates that if

$$\lambda \gtrsim \frac{k\sqrt{\log(m+n)}}{\min\{\sqrt{K_R}, \sqrt{K_C}\}} + \frac{\sqrt{m}+\sqrt{n}}{\min\{K_R, K_C\}}, \tag{23}$$

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 (23) is more stringent than that in (22), 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.

6. Basic tools and proof overview.

6.1. *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 [61].

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 $1 \le j \le r$. 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 $1 \le j \le r$ are their corresponding orthonormal eigenvectors. The spectral decomposition of $\mathcal{A}$ is

$$\mathcal{A} = \mathcal{U}\mathcal{D}\mathcal{U}^{\mathrm{T}}, \tag{24}$$

where $\mathcal{U} := (\mathbf{u}_1, \dots, \mathbf{u}_{2r})$ and $\mathcal{D} := \operatorname{diag}(\lambda_1, \cdots, \lambda_{2r})$. 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 $1 \le j \le \min\{N,n\}$. The eigenvector corresponding to $\widetilde{\lambda}_j$ is denoted by $\widetilde{\mathbf{u}}_j$ 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

$$\Phi(z) := \begin{pmatrix} \frac{1}{\phi_1(z)} I_N & 0 \\ 0 & \frac{1}{\phi_2(z)} I_n \end{pmatrix}, \tag{25}$$

where

$$\phi_1(z) := z - \sum_{t \in [\![N+1, N+n]\!]} G_{tt}(z), \qquad \phi_2(z) := z - \sum_{s \in [\![1,N]\!]} G_{ss}(z). \tag{26}$$

Using the Schur complement, the Green function $G(z)$ of $\mathcal{E}$ is composed of four blocks: the Green function $G_1(z)$ of EE^T, the Green function $G_2(z)$ of E^TE, and $E^TG_1(z)$ and $G_1(z)E$ (with minor modifications). The approximation $\Phi(z)$ is motivated by the quadratic equations satisfied by the Stieltjes transforms of the spectral distributions of EE^{T} and $E^{\mathrm{T}}E$, similar to how the Stieltjes transform of the semi-circle law satisfies a quadratic equation in the symmetric case.

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 (26) as

$$\phi_1(z) = z - \operatorname{tr}\mathcal{I}^{\mathrm{d}}G(z), \qquad \phi_2(z) = z - \operatorname{tr}\mathcal{I}^{\mathrm{u}}G(z). \tag{27}$$

By elementary linear algebra (the proof deferred to [69]), it can be verified that

$$\phi_1(z) = \phi_2(z) - \frac{1}{z}(n-N). \tag{28}$$

From the definition of $\mathcal{U}$ in (24), it is easy to verify that

$$\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U} = \begin{pmatrix} \alpha(z)I_r & \beta(z)I_r \\ \beta(z)I_r & \alpha(z)I_r \end{pmatrix}, \tag{29}$$

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

$$\|\mathcal{U}^{\mathrm{T}}\Phi(z)\mathcal{U}\| = \max\left\{\frac{1}{|\phi_1(z)|}, \frac{1}{|\phi_2(z)|}\right\}. \tag{30}$$

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 [61] and is omitted for brevity.

LEMMA 6.1. *On the event where* $\|E\| \le 2(\sqrt{N}+\sqrt{n})$,

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

and

$$\left(1 - \frac{1}{4b(b-1)}\right)|z| \le |\phi_i(z)| \le \left(1 + \frac{1}{4b(b-1)}\right)|z| \quad \text{for } i = 1,2 \tag{31}$$

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

Consequently, by Lemma 6.1, we obtain

$$\max\{|\operatorname{tr} G(z)|, |\operatorname{tr}\mathcal{I}^u G(z)|, |\operatorname{tr}\mathcal{I}^d G(z)|\} \le (N+n)\|G(z)\| \le \frac{b}{b-1}\frac{N+n}{|z|}. \tag{32}$$

The subsequent isotropic local law is derived using a proof similar to that of [61, Lemma 27]. For completeness, we briefly describe the proof in the supplementary material [69].

LEMMA 6.2. *Let* $K > 0$ *and assume* $(\sqrt{N}+\sqrt{n})^2 \ge 32(K+1)\log(N+n)$. *For any unit vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N+n}$ *and for any* $z \in \mathbb{C}$ *with* $|z| \ge 2b(\sqrt{N}+\sqrt{n})$,

$$\left|\mathbf{x}^{\mathrm{T}}\left(G(z) - \Phi(z)\right)\mathbf{y}\right| \le \frac{5b^2}{(b-1)^2}\frac{\sqrt{(K+1)\log(N+n)}}{|z|^2} \tag{33}$$

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

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

Using the previous lemma and a standard ε-net argument, we obtain the following result that is analogous to [61, Lemma 9]:

LEMMA 6.3. *Under the assumptions of Theorem 2.3, one has*

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

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

Lemma 6.3 improves the rank r-dependence in the bound of [61, Lemma 9]. The proof of Lemma 6.3 is included in [69]. 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 6.5 below).

The following result on the location of perturbed singular values is obtained using Lemma 6.3. Consider the random function

$$\varphi(z) := \phi_1(z)\phi_2(z), \tag{34}$$

where $\phi_1(z)$ and $\phi_2(z)$ are defined in (26). Define the auxiliary functions for $b \ge 2$:

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

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

$$S_\sigma := \left\{w\in\mathbb{C} : |\operatorname{Im}(w)| \le 20\chi(b)\eta r, \sigma - 20\chi(b)\eta r \le \operatorname{Re}(w) \le \chi(b)\sigma + 20\chi(b)\eta r\right\}. \tag{35}$$

THEOREM 6.4 (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 $r\ge 1$. Let $K>0$ and $b\ge 2$. Denote $\eta := \frac{11b^2}{(b-1)^2}\sqrt{2(\log 9)r+(K+7)\log(N+n)}$. Assume $(\sqrt{N}+\sqrt{n})^2 \ge 32(K+7)\log(N+n)+64(\log 9)r$. Let $1\le r_0\le r$ such that $\sigma_{r_0}\ge 2b(\sqrt{N}+\sqrt{n})+80b\eta r$ and $\delta_{r_0}\ge 75\chi(b)\eta r$. Consider any $1\le k\le s\le r_0$ satisfying $\min\{\delta_{k-1},\delta_s\}\ge 75\chi(b)\eta r$. For any $j\in[\![k,s]\!]$, there exists $j_0\in[\![k,s]\!]$ such that $\widetilde{\sigma}_j\in S_{\sigma_{j_0}}$, and*

$$|\varphi(\widetilde{\sigma}_j)-\sigma_{j_0}^2| \le 20\xi(b)\chi(b)\eta r\left(\widetilde{\sigma}_j+\chi(b)\sigma_{j_0}\right) \tag{36}$$

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

Our Theorem 6.4 extends the analysis of Theorem 12 in [61] in an important direction by removing the singular value separation requirement. This improvement stems from a careful refinement of the theoretical framework, detailed in the supplementary material [69].

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 [61], and is omitted.

LEMMA 6.5. *On the event where $\|E\|\le 2(\sqrt{N}+\sqrt{n})$,*

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

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

LEMMA 6.6 (Lemma 13 from [61]). *Under the assumptions of Theorem 2.3,*

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

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 6.7 (Spectral norm bound; see (2.3) from [63]). *Let E be an $N \times n$ matrix whose entries are independent standard Gaussian random variables. Then*

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

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

6.2. *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 $\widetilde{\mathcal{A}} = \mathcal{A} + \mathcal{E}$ in Section 6.1.

At the heart of our analysis is the isotropic local law (Lemma 6.2), 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, 50]. 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 6.4 and achieve the control of $\left\|\mathcal{U}^{\mathrm{T}}\left(G(z) - \Phi(z)\right)\mathcal{U}\right\|$ as given in Lemma 6.3. These instruments have been previously explored in the previous work by O'Rourke, Vu and the author [61]. In this paper, we refine these estimations and ease the conditions in [61]. Furthermore, we deploy these refined tools to derive a variety of new perturbation bounds.

To illustrate the key ideas behind our perturbation bounds, we simply focus on the largest eigenvector $\widetilde{\mathbf{u}}_1$ of $\widetilde{\mathcal{A}}$. While a complete analysis requires considering both $\widetilde{\mathbf{u}}_1$ and $\widetilde{\mathbf{u}}_{r+1}$ (as they jointly involve the largest singular vectors u_1 and v_1), we temporarily consider only $\widetilde{\mathbf{u}}_1$ for clarity. We emphasize that the calculation presented below is simplified to convey the main idea. In the actual proof, we work with all eigenvectors $\mathbf{u}_j$ with $j \neq 1, r+1$ and $j \in [2r]$ simultaneously, leading to a more comprehensive analysis involving the block structure of Φ.

We start with the decomposition

$$\widetilde{\mathbf{u}}_1 = (\mathbf{u}_1\mathbf{u}_1^{\mathrm{T}})\widetilde{\mathbf{u}}_1 + \mathcal{P}_1\widetilde{\mathbf{u}}_1 + \mathcal{Q}\widetilde{\mathbf{u}}_1, \tag{37}$$

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 (37).

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 (37), 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 $\|\mathcal{P}_1\widetilde{\mathbf{u}}_1\| \leq \|\mathcal{U}_1^{\mathrm{T}}\widetilde{\mathbf{u}}_1\|$, 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

$$\mathbf{u}_j^{\mathrm{T}}\widetilde{\mathbf{u}}_1 = \mathbf{u}_j^{\mathrm{T}}\Phi(\widetilde{\lambda}_1)\mathcal{A}\widetilde{\mathbf{u}}_1 + \mathbf{u}_j^{\mathrm{T}}\left(G(\widetilde{\lambda}_1) - \Phi(\widetilde{\lambda}_1)\right)\mathcal{A}\widetilde{\mathbf{u}}_1 \tag{38}$$

Calculations similar to those in (29) indicate that the first term on the right-hand side of (38), $\mathbf{u}_j^{\mathrm{T}}\Phi(\widetilde{\lambda}_1)\mathcal{A}\widetilde{\mathbf{u}}_1$, is exactly $\lambda_j\alpha(\widetilde{\lambda}_1)\mathbf{u}_j^{\mathrm{T}}\widetilde{\mathbf{u}}_1$ (omitting the term containing $\mathbf{u}_{r+j}$). We continue from (38) 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 6.4 to analyze the coefficient $1-\lambda_j\alpha(\widetilde{\lambda}_1)$ that precedes it. Lemma 6.3 is applied to manage the term on the right-hand side.

Next, for the ℓ_∞ analysis, from (37), we obtain

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

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

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

and then applying Lemma 6.3.

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 6.2, 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.

Acknowledgments. 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 [73, 74] to her attention.

Funding. The author was supported by Hong Kong RGC grant GRF 16304222, GRF 16308219 and ECS 26304920.

SUPPLEMENTARY MATERIAL

Supplementary material to "Analysis of singular subspaces under random perturbations"

In this supplementary material [69], we provide the detailed proofs of our main results, as well as the proofs of basic tools related to them and some preliminary materials. Additionally, [69] contains further discussions and numerical simulations that explore the sharpness of our theoretical results.

REFERENCES

[1] ABBE, E., FAN, J. and WANG, K. (2022). An ℓ_p theory of PCA and spectral clustering. *The Annals of Statistics* **50** 2359–2385. https://doi.org/10.1214/22-aos2196

[2] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics* **48** 1452–1474.

[3] AGTERBERG, J. (2023). Distributional Theory and Statistical Inference for Linear Functions of Eigenvectors with Small Eigengaps. *arXiv preprint arXiv:2308.02480*.

[4] AGTERBERG, J., LUBBERTS, Z. and PRIEBE, C. E. (2022). Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Trans. Inform. Theory* **68** 4618–4650. MR4449064

[5] ALLEZ, R. and BOUCHAUD, J.-P. (2014). Eigenvector dynamics under free addition. *Random Matrices Theory Appl.* **3** 1450010, 17. https://doi.org/10.1142/S2010326314500105 MR3256861

[6] ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. *Random Structures & Algorithms* **13** 457–466.

[7] ARIAS-CASTRO, E. and VERZELEN, N. (2014). Community detection in dense random networks. *The Annals of Statistics* **42** 940–969.

[8] BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. H. Poincaré Probab. Statist.* **44** 447–474. https://doi.org/10.1214/07-AIHP118

[9] BAIK, J., AROUS, G. B., PÉCHÉ, S. et al. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability* **33** 1643–1697.

[10] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis* **97** 1382–1408.

[11] BALAKRISHNAN, S., KOLAR, M., RINALDO, A., SINGH, A. and WASSERMAN, L. (2011). Statistical and computational tradeoffs in biclustering. In *NIPS 2011 Workshop on Computational Trade-offs in Statistical Learning* **4**.

[12] BAO, Z., DING, X., WANG, J. and WANG, K. (2022). Statistical inference for principal components of spiked covariance matrices. *The Annals of Statistics* **50** 1144–1169. https://doi.org/10.1214/21-aos2143 MR4404931

[13] BAO, Z., DING, X. and WANG, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *The Annals of Statistics* **49** 370–392.

[14] BAO, Z. and WANG, D. (2021). Eigenvector distribution in the critical regime of BBP transition. *Probability Theory and Related Fields* 1–81.

[15] BENAYCH-GEORGES, F., ENRIQUEZ, N. and MICHAÏL, A. (2021). Eigenvectors of a matrix under random perturbation. *Random Matrices Theory Appl.* **10** 2150023. https://doi.org/10.1142/S2010326321500234 MR4260218

[16] BENAYCH-GEORGES, F., GUIONNET, A. and MAIDA, M. (2011). Fluctuations of the Extreme Eigenvalues of Finite Rank Deformations of Random Matrices. *Electron. J. Probab.* **16** 1621–1662.

[17] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics* **227** 494-521. https://doi.org/10.1016/j.aim.2011.02.007

[18] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis* **111** 120-135. https://doi.org/10.1016/j.jmva.2012.04.019

[19] BENIGNI, L. (2020). Eigenvectors distribution and quantum unique ergodicity for deformed Wigner matrices. *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques* **56** 2822–2867.

[20] BHARDWAJ, A. and VU, V. (2023). Matrix Perturbation: Davis-Kahan in the Infinity Norm. *arXiv preprint arXiv:2304.00328*.

[21] BHATIA, R. (1997). *Matrix analysis*. *Graduate Texts in Mathematics* **169**. Springer-Verlag, New York. https://doi.org/10.1007/978-1-4612-0653-8 MR1477662

[22] BLOEMENDAL, A., ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19** 1-53. https://doi.org/10.1214/ejp.v19-3054 MR3183577

[23] BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2016). On the principal components of sample covariance matrices. *Probability Theory and Related Fields* **164** 459–552. https://doi.org/10.1007/s00440-015-0616-x MR3449395

[24] BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory* 48–166. PMLR.

[25] BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688.

[26] BUTUCEA, C., INGSTER, Y. I. and SUSLINA, I. A. (2015). Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM: Probability and Statistics* **19** 115–134.

[27] CAI, C., LI, G., CHI, Y., POOR, H. V. and CHEN, Y. (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics* **49** 944–967.

[28] CAI, T. T., LIANG, T. and RAKHLIN, A. (2017). Computational and statistical boundaries for submatrix localization in a large noisy matrix. *The Annals of Statistics* **45** 1403–1430.

[29] CAI, T. T. and ZHANG, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics* **46** 60–89.

[30] CANDÈS, E. J. and PLAN, Y. (2010). Matrix Completion With Noise. *Proceedings of the IEEE* **98** 925-936.

[31] CANDÈS, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. https://doi.org/10.1007/s10208-009-9045-5 MR2565240

[32] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics* **47** 2405–2439.

[33] CAPITAINE, M. (2018). Limiting eigenvectors of outliers for Spiked Information-Plus-Noise type matrices. *Séminaire de Probabilités XLIX* 119–164.

[34] CAPITAINE, M. and DONATI-MARTIN, C. (2021). Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of Wigner matrices. *ALEA: Latin American Journal of Probability and Mathematical Statistics* **18** 129–165.

[35] CHEN, Y., CHI, Y., FAN, J. and MA, C. (2021). Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning* **14** 566–806.

[36] CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top-K ranking. *The Annals of Statistics* **47** 2204–2235.

[37] CHEN, Y. and XU, J. (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research* **17** 882–938.

[38] CHENG, C., WEI, Y. and CHEN, Y. (2021). Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *IEEE Transactions on Information Theory* **67** 7380–7419.

[39] DADON, M., HULEIHEL, W. and BENDORY, T. (2024). Detection and Recovery of Hidden Submatrices. *IEEE Transactions on Signal and Information Processing over Networks*.

[40] DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. https://doi.org/10.1137/0707001 MR264450

[41] DEKEL, Y., GUREL-GUREVICH, O. and PERES, Y. (2014). Finding hidden cliques in linear time with high probability. *Combinatorics, Probability and Computing* **23** 29–49.

[42] EL KAROUI, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714.

[43] ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: spectral analysis beyond Davis-Kahan. In *Algorithmic Learning Theory* 321–358. PMLR.

[44] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). Asymptotic theory of eigenvectors for large random matrices. *Journal of the American Statistical Association* **117** 996–1009.

[45] FAN, J., WANG, W. and ZHONG, Y. (2018). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research* **18** 207–207. MR3827095

[46] FEIGE, U. and RON, D. (2010). Finding hidden cliques in linear time. In *Discrete Mathematics and Theoretical Computer Science* 189-204. Discrete Mathematics and Theoretical Computer Science.

[47] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory* **62** 2788–2797.

[48] JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association* **104** 682–693.

[49] KESHAVAN, R., MONTANARI, A. and OH, S. (2009). Matrix completion from noisy entries. *Advances in Neural Information Processing Systems* **22**.

[50] KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1750. https://doi.org/10.1002/cpa.21450 MR3103909

[51] KOLAR, M., BALAKRISHNAN, S., RINALDO, A. and SINGH, A. (2011). Minimax localization of structural information in large noisy matrices. *Advances in Neural Information Processing Systems* **24**.

[52] KOLTCHINSKII, V. and XIA, D. (2016). Perturbation of linear forms of singular vectors under Gaussian noise. In *High dimensional probability VII. Progr. Probab.* **71** 397–423. Springer, [Cham]. https://doi.org/10.1007/978-3-319-40519-3_18 MR3565274

[53] LEI, L. (2019). Unified $\ell_{2\to\infty}$ Eigenspace Perturbation Theory for Symmetric Random Matrices. *arXiv preprint arXiv:1909.04798*.

[54] LI, G., CAI, C., POOR, H. V. and CHEN, Y. (2021). Minimax estimation of linear functions of eigenvectors in the face of small eigen-gaps. *arXiv preprint arXiv:2104.03298*.

[55] LÖFFLER, M., ZHANG, A. Y. and ZHOU, H. H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *The Annals of Statistics* **49** 2506–2530.
[56] LUO, Y., HAN, R. and ZHANG, A. R. (2021). A Schatten-q low-rank matrix perturbation analysis via perturbation projection error bound. *Linear Algebra and its Applications* **630** 225–240.
[57] MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *The Annals of Statistics* 1089–1116.
[58] MCSHERRY, F. (2001). Spectral partitioning of random graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science* 529–537. IEEE.
[59] MONTANARI, A., REICHMAN, D. and ZEITOUNI, O. (2015). 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**.
[60] O'ROURKE, S., VU, V. and WANG, K. (2018). Random perturbation of low rank matrices: improving classical bounds. *Linear Algebra Appl.* **540** 26–59. https://doi.org/10.1016/j.laa.2017.11.014 MR3739989
[61] O'ROURKE, S., VU, V. and WANG, K. (2024). Matrices with Gaussian noise: optimal estimates for singular subspace perturbation. *IEEE Transactions on Information Theory* **70** 1978-2002.
[62] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica* **17** 1617–1642.
[63] RUDELSON, M. and VERSHYNIN, R. (2010). Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III* 1576–1602. Hindustan Book Agency, New Delhi. MR2827856
[64] STEWART, G. W. and SUN, J. G. (1990). *Matrix perturbation theory*. *Computer Science and Scientific Computing*. Academic Press, Inc., Boston, MA. MR1061154
[65] TROPP, J. A. (2015). An Introduction to Matrix Concentration Inequalities. *Found. Trends Mach. Learn.* **8** 1–230. https://doi.org/10.1561/2200000048
[66] VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *The Annals of Statistics* **36** 555–586.
[67] VU, V. (2018). A simple SVD algorithm for finding hidden partitions. *Combinatorics, Probability and Computing* **27** 124–140.
[68] VU, V. Q. and LEI, J. (2013). Minimax sparse principal subspace estimation in high dimensions. *The Annals of Statistics* **41** 2905–2947.
[69] WANG, K. Supplementary material to "Analysis of singular subspaces under random perturbations". *2024*.
[70] WANG, R. (2015). Singular vector perturbation under Gaussian noise. *SIAM J. Matrix Anal. Appl.* **36** 158–177. https://doi.org/10.1137/130938177 MR3310977
[71] WEDIN, P.-. A. (1972). Perturbation bounds in connection with singular value decomposition. *Nordisk Tidskr. Informationsbehandling (BIT)* **12** 99–111. https://doi.org/10.1007/bf01932678 MR309968
[72] XIA, D. and ZHOU, F. (2019). The sup-norm perturbation of HOSVD and low rank tensor denoising. *Journal of Machine Learning Research* **20** 1-42. MR3960915
[73] YAN, Y., CHEN, Y. and FAN, J. (2021). Inference for heteroskedastic PCA with missing data. *arXiv preprint arXiv:2107.12365*.
[74] YAN, Y. and WAINWRIGHT, M. J. (2024). Entrywise Inference for Missing Panel Data: A Simple and Instance-Optimal Approach. *arXiv preprint arXiv:2401.13665*.
[75] YU, Y., WANG, T. and SAMWORTH, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102** 315–323.
[76] ZHANG, A. Y. and ZHOU, H. H. (2022). Leave-one-out singular subspace perturbation analysis for spectral clustering. *arXiv preprint arXiv:2205.14855*.
[77] ZHONG, Y. (2017). Eigenvector Under Random Perturbation: A Nonasymptotic Rayleigh-Schrödinger Theory. *arXiv preprint arXiv:1702.00139*.
[78] ZHONG, Y. and BOUMAL, N. (2018). Near-optimal bounds for phase synchronization. *SIAM J. Optim.* **28** 989–1016. https://doi.org/10.1137/17M1122025 MR3782406