Asymmetry Helps: Eigenvalue and Eigenvector Analyses of Asymmetrically Perturbed Low-Rank Matrices

Yuxin Chen∗ Chen Cheng† Jianqing Fan‡

November 2018; Revised March 2019

Abstract

This paper is concerned with a curious phenomenon reflecting the interplay between asymmetry and the spectral methods. Suppose we are interested in a rank-1 and symmetric matrix \( M^* \in \mathbb{R}^{n \times n} \), yet only a randomly perturbed version \( M \) is observed. The noise matrix \( M - M^* \) is composed of independent and zero-mean entries and is not symmetric. This might arise, for example, when we have two independent samples for each entry of \( M^* \) and arrange them into an asymmetric data matrix \( M \). The aim is to estimate the leading eigenvalue and eigenvector of \( M^* \). We demonstrate that the leading eigenvalue of the data matrix \( M \) can be \( \sqrt{n} \) times more accurate than its leading singular value in eigenvalue estimation. Further, the perturbation of any linear form of the leading eigenvector of \( M \) (e.g. entrywise eigenvector perturbation) is provably well-controlled. We further provide partial theory for the more general rank-\( r \) case; this allows us to accommodate the case when \( M^* \) is rank-1 but asymmetric, by considering eigen-decomposition of the associated rank-2 dilation matrix. The takeaway message is this: arranging the data samples in an asymmetric manner and performing eigen-decomposition (as opposed to SVD) could sometimes be quite beneficial.

Keywords: asymmetric matrices, eigenvalue perturbation, entrywise eigenvector perturbation, linear form of eigenvectors.

1 Introduction

Consider an unknown symmetric and low-rank matrix \( M^* \in \mathbb{R}^{n \times n} \). What we have observed is a corrupted version

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

with \( H \) a noise matrix. A classical problem is concerned with estimating the leading eigenvalues and eigenspace of \( M^* \) given \( M \).

The current paper concentrates on a scenario where the noise matrix \( H \) (and hence \( M \)) is independently random and hence asymmetric. This might arise, for example, when we have available multiple (e.g. two) samples for each entry of \( M^* \) and arrange the samples in an asymmetric fashion. A natural approach that immediately comes to mind is based on singular value decomposition (SVD), which employs the leading singular values (resp. subspace) of \( M \) to approximately estimate the eigenvalues (resp. eigenspace) of \( M^* \). By contrast, a much less popular alternative is based on eigen-decomposition of \( M \), which attempts approximation using the leading eigenvalues and eigenspace of \( M \). Given that eigen-decomposition of an asymmetric matrix is in general not as numerically stable as SVD, conventional wisdom often favors the SVD-based approach, unless certain symmetrization step is implemented prior to eigen-decomposition.

Author names are sorted alphabetically.
*Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA; Email: yuxin.chen@princeton.edu.
†Department of Scientific & Engineering Computing, School of Mathematical Sciences, Peking University, Beijing 100871, China; Email: moriartycc@pku.edu.cn.
‡Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; Email: jqfan@princeton.edu.
A curious empirical phenomenon arises when comparing these two approaches, which largely motivates the research in this paper. Let us generate $M^\star$ as a random rank-1 matrix with leading eigenvalue $\lambda^\star = 1$, and let $H$ be a Gaussian random matrix whose entries are i.i.d. $\mathcal{N}(0, \sigma^2)$ with $\sigma = 1/\sqrt{n\log n}$. Fig. 1(a) compares the empirical accuracy of estimating the 1st eigenvalue of $M^\star$ via the leading eigenvalue (the blue line) and via the leading singular value of $M$ (the red line). As it turns out, eigen-decomposition significantly outperforms SVD in estimating $\lambda^\star$, and the advantage seems increasingly more remarkable as the dimensionality $n$ grows. To facilitate comparison, we include an additional green line in Fig. 1(a), obtained by rescaling the red line by $2.5/\sqrt{n}$. Interestingly, this green line coincides almost perfectly with the blue line, thus suggesting orderwise gain of eigen-decomposition compared to SVD. What is more, this phenomenon does not merely happen under i.i.d. Gaussian noise. Similar numerical behaviors are observed in the problem of matrix completion — as displayed in Fig. 1(b) — even though the components of the equivalent perturbation matrix now are far from i.i.d.

The goal of the current paper is thus to develop a systematic understanding of this interesting phenomenon, that is, why statistical asymmetry empowers eigen-decomposition. Informally, our findings suggest that: when $M^\star$ is rank-1 and $H$ is composed of independent and zero-mean entries,

1. the leading eigenvalue of $M$ could be $O(\sqrt{n})$ times (up to some logarithmic factor) more accurate than the leading singular value of $M$ when estimating the 1st eigenvalue of $M^\star$.\(^\dagger\)

2. the perturbation of the leading eigenvector is well-controlled along an arbitrary deterministic direction. For example, the eigenvector perturbation is well-controlled in any coordinate, indicating that the eigenvector estimation error is spread out across all coordinates.

We will further provide partial theory to accommodate the rank-$r$ case. As an important application, such a theory allows us to estimate the leading singular value and singular vectors of an asymmetric rank-1 matrix via eigen-decomposition (when applied to the associated dilation matrix), which also provably outperforms the vanilla SVD approach.

All in all, when it comes to low-rank matrix estimation, arranging the observed matrix samples in an asymmetric manner and invoking eigen-decomposition properly (as opposed to SVD) could sometimes be quite beneficial.

\(^\dagger\)More precisely, this gain is possible when $\|H\|$ is nearly as large as $\|M^\star\|$ (up to some logarithmic factor).
2 Problem formulation

2.1 Models and assumptions

In this section, we formally introduce our models and assumptions. Consider a symmetric and low-rank matrix $M^\star = [M^\star_{ij}]_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$. Suppose we are given a random copy of $M^\star$ as follows

$$M = M^\star + H,$$

where $H = [H_{ij}]_{1 \leq i,j \leq n}$ is a random noise matrix.

The current paper concentrates on independent and random noise. Specifically, we impose the following assumptions on $H$ throughout this paper.

Assumption 1.

1. (Independent entries) The entries $\{H_{ij}\}_{1 \leq i,j \leq n}$ are independently generated;
2. (Zero mean) $\mathbb{E}[H_{ij}] = 0$ for all $1 \leq i,j \leq n$;
3. (Variance) $\text{Var}(H_{ij}) = \mathbb{E}[H_{ij}^2] \leq \sigma_n^2$ for all $1 \leq i,j \leq n$;
4. (Magnitude) Each $H_{ij}$ $(1 \leq i,j \leq n)$ satisfies either of the following conditions:
   (a) $|H_{ij}| \leq B_n$;
   (b) $H_{ij}$ has a symmetric distribution obeying $\mathbb{P}\{|H_{ij}| > B_n\} \leq c_b n^{-12}$ for some universal constant $c_b > 0$.

Remark 1 (Notational convention). In what follows, the dependency of $\sigma_n$ and $B_n$ on $n$ shall often be suppressed whenever it is clear from the context, so as to simplify notation.

Note that we do not enforce the constraint $H_{ij} = H_{ji}$, and hence $H$ and $M$ are in general asymmetric matrices. In addition, while Condition 4(a) covers the class of bounded random variables, Condition 4(b) allows us to accommodate a large family of heavy-tailed distributions (e.g. sub-exponential distributions).

An immediate consequence of Assumption 1 is the following bound on the spectral norm $\|H\|$ of $H$.

Lemma 1. Under Assumption 1, there exist some universal constants $c_0, C_0 > 0$ such that with probability exceeding $1 - C_0 n^{-10}$,

$$\|H\| \leq c_0 \sqrt{n \sigma^2 \log n} + c_0 B \log n. \quad (3)$$

Proof. This is a standard non-asymptotic result that follows immediately from the matrix Bernstein inequality Tropp (2015) and the union bound (for Assumption 4(b)). We omit the details for conciseness.

The aim is to develop non-asymptotic eigenvalue and eigenvector perturbation bounds under this family of random and asymmetric noise matrices. Our theoretical development is divided into two parts.

• Rank-1 symmetric case. When specialized to the rank-1 case, we assume the eigen-decomposition of $M^\star$ to be

$$M^\star = \lambda^\star u^\star u^\star^\top$$

with $\lambda^\star$ and $u^\star$ being its first eigenvalue and eigenvector, respectively. We also denote by $\lambda$ and $u$ the leading eigenvalue and eigenvector of $M$, respectively. The following quantities are the focal points of this paper (see Section 4):

– Eigenvalue perturbation: $|\lambda - \lambda^\star|$;
– Perturbation of linear forms of eigenvectors: $\min\{|a^\top (u - u^\star)|, |a^\top (u + u^\star)|\}$ for any fixed unit vector $a \in \mathbb{R}^n$;
– Entrywise eigenvector perturbation: $\min\{|u - u^\star|_\infty, |u + u^\star|_\infty\}$.
• **Rank-r symmetric case.** For the general rank-\(r\) case, we let the eigen-decomposition of \(M^*\) be
\[
M^* = U^* \Sigma^* U^{\top}
\]
where \(U^* = [u_1^*, \cdots, u_r^*] \in \mathbb{R}^{n \times r}\) consists of the eigenvectors, and \(\Sigma^* = \text{diag}(\lambda_1^*, \cdots, \lambda_r^*) \in \mathbb{R}^{r \times r}\) is a diagonal matrix with the eigenvalues arranged in descending order, i.e. \(\lambda_1^* \geq \cdots \geq \lambda_r^*\). We let \(\lambda_{\text{max}}^* = \max_{1 \leq j \leq r} |\lambda_j^*|\) and \(\lambda_{\text{min}}^* = \min_{1 \leq j \leq r} |\lambda_j^*|\). In addition, we let the top \(r\) eigenvalues (in magnitude) of \(M\) be \(\lambda_1, \cdots, \lambda_r\) (obeying \(|\lambda_1| \geq \cdots \geq |\lambda_r|\)) and their corresponding normalized eigenvectors be \(u_1, \cdots, u_r\). We will obtain partial eigenvalue perturbation results for this more general case, as detailed in Section 5.

As is well-known, eigen-decomposition can be applied to estimate the singular values and singular vectors of an asymmetric matrix \(M^*\) via the standard dilation trick Tropp (2015). As a consequence, our results are also applicable for singular value and singular vector estimation. See Section 5.2 for details.

Finally, we single out an incoherence parameter that plays an important role in our theory, which captures how well the energy of the eigenvectors is spread out across all entries.

**Definition 1 (Incoherence parameter).** The incoherence parameter of a rank-\(r\) symmetric matrix \(M^*\) with eigen-decomposition \(M^* = U^* \Sigma^* U^{\top}\) is defined to be the smallest quantity \(\mu\) obeying
\[
\|U^*\|_\infty \leq \sqrt{\frac{\mu}{n}},
\]
where \(\|\cdot\|_\infty\) denotes the entrywise \(\ell_\infty\) norm.

**Remark 2.** An alternative definition of the incoherence parameter (Candès and Recht, 2009; Keshavan et al., 2010; Chi et al., 2018) is the smallest quantity \(\mu_0\) satisfying \(\|U^*\|_{2,\infty} \leq \sqrt{\mu_0 r}/n\). This is a weaker assumption than Definition 1, as it only requires the energy of \(U^*\) to be spread out across all of its rows rather than all of its entries. Note that these two incoherent parameters are consistent in the rank-1 case; in the rank-\(r\) case one has \(\mu_0 \leq \mu \leq \mu_0 r\).

### 2.2 Notation

The standard basis vectors for \(\mathbb{R}^n\) are denoted by \(e_1, \cdots, e_n\). For any vector \(z\), we let \(\|z\|_2\) and \(\|z\|_\infty\) denote the \(\ell_2\) norm and the \(\ell_\infty\) norm of \(z\), respectively. For any matrix \(M\), denote by \(\|M\|\), \(\|M\|_F\) and \(\|M\|_\infty\) the spectral norm, the Frobenius norm and the entrywise \(\ell_\infty\) norm (the largest magnitude of all entries) of \(M\), respectively. Let \([n] := \{1, \cdots, n\}\). We use \(A^\top\) and \(A^*\) to denote respectively the transpose and the conjugate transpose of \(A\). In addition, the notation \(f(n) = O(g(n))\) or \(f(n) \lesssim g(n)\) means there is a constant \(c > 0\) such that \(|f(n)| \leq cg(n)|\), \(f(n) \gtrsim g(n)\) means there is a constant \(c > 0\) such that \(|f(n)| \geq cg(n)|\), and \(f(n) \asymp g(n)\) means that there exist constants \(c_1, c_2 > 0\) such that \(c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|\).

### 3 Preliminaries

Before continuing, we gather several preliminary facts that will be useful throughout. The readers familiar with matrix perturbation theory may proceed directly to the main theoretical development in Section 4.

#### 3.1 Perturbation of eigenvalues of asymmetric matrices

We begin with a standard result concerning eigenvalue perturbation of a diagonalizable matrix Bauer and Fike (1960). Note that the matrices under study might be asymmetric.

**Theorem 1 (Bauer-Fike Theorem).** Consider a diagonalizable matrix \(A \in \mathbb{R}^{n \times n}\) with eigen-decomposition \(A = V \Lambda V^{-1}\), where \(V \in \mathbb{C}^{n \times n}\) is a non-singular eigenvector matrix and \(\Lambda\) is diagonal. Let \(\tilde{\lambda}\) be an eigenvalue of \(A + \mathbf{H}\). Then there exists an eigenvalue \(\lambda\) of \(A\) such that
\[
|\lambda - \tilde{\lambda}| \leq \|V\| \|V^{-1}\| \|\mathbf{H}\|.
\]

In addition, if \(A\) is symmetric, then there exists an eigenvalue \(\lambda\) of \(A\) such that
\[
|\lambda - \tilde{\lambda}| \leq \|\mathbf{H}\|.
\]
However, caution needs to be exercised as the Bauer-Fike Theorem does not specify which eigenvalue of \( A \) is close to an eigenvalue of \( A + H \). Encouragingly, in the low-rank case of interest, the Bauer-Fike Theorem together with certain continuity of the spectrum allows one to localize the leading eigenvalues of the perturbed matrix.

**Lemma 2.** Suppose \( M^* \) is a rank-\( r \) symmetric matrix whose top-\( r \) eigenvalues obey \( |\lambda_1^*| \geq \cdots \geq |\lambda_r^*| > 0 \). If \( \|H\| < |\lambda_1^*|/2 \), then the top-\( r \) eigenvalues \( \lambda_1, \ldots, \lambda_r \) of \( M = M^* + H \), sorted by modulus, obey that: for any \( 1 \leq l \leq r \),
\[
|\lambda_l - \lambda_l^*| \leq \|H\| \quad \text{for some} \ 1 \leq j \leq r.
\]
In addition, if \( r = 1 \), then both the leading eigenvalue and the leading eigenvector of \( M \) are real-valued.

This result, which we establish in Appendix A.1, parallels Weyl’s inequality for symmetric matrices. Note, however, that the above bound (9) might be quite loose for specific settings. We will establish much sharper perturbation bounds when \( H \) contains independent random entries (see, e.g. Corollary 1).

### 3.2 Neumann trick and eigenvector perturbation

Next, we introduce a classical result dubbed as the “Neumann trick” Eldridge et al. (2018). This theorem, which is derived based on the Neumann series for a matrix inverse, has been applied to analyze eigenvectors in various settings Erdős et al. (2013); Jain and Netrapalli (2015); Eldridge et al. (2018).

**Theorem 2 (Neumann trick).** Consider the matrices \( M^* \) and \( M \) (see (5) and (2)). Suppose \( \|H\| < |\lambda_l| \) for some \( 1 \leq l \leq n \). Then
\[
u_l = \sum_{j=1}^{r} \frac{\lambda_j^*}{\lambda_l^*} (u_j^* u_l^\top) \left\{ \sum_{s=0}^{\infty} \frac{1}{\lambda_l^*} H^s u_j^* \right\}.
\]

**Proof.** We supply the proof in Appendix A.2 for self-containedness. \( \square \)

**Remark 3.** In particular, if \( M^* \) is a rank-1 matrix and \( \|H\| < \lambda_1 \), then
\[
u_1 = \frac{\lambda_1^*}{\lambda_1} (u_1^* u_1^\top) \left\{ \sum_{s=0}^{\infty} \frac{1}{\lambda_1^*} H^s u_1^* \right\}.
\]

An immediate consequence of the Neumann trick is the following lemma, which asserts that each of the top-\( r \) eigenvectors of \( M \) resides almost within the top-\( r \) eigen-subspace of \( M^* \). The proof is deferred to Appendix A.3.

**Lemma 3.** Suppose \( M^* \) is a rank-\( r \) symmetric matrix with \( r \) non-zero eigenvalues obeying \( 1 = \lambda_{\text{max}}^* = |\lambda_1^*| \geq \cdots \geq |\lambda_r^*| = \lambda_{\text{min}}^* > 0 \) and associated eigenvectors \( u_1^*, \ldots, u_r^* \). Define \( \kappa \triangleq \lambda_{\text{max}}^*/\lambda_{\text{min}}^* \). If \( \|H\| \leq 1/(4\kappa) \), then the top-\( r \) eigenvectors \( u_1, \ldots, u_r \) of \( M = M^* + H \) obey
\[
\sum_{j=1}^{r} |u_j^\top u_l|^2 \geq 1 - \frac{64\kappa^4}{9} \|H\|^2, \quad 1 \leq l \leq r.
\]
In addition, if \( r = 1 \), then one further has
\[
\min\{\|u_1 - u_1^\top\|_2, \|u_1 + u_1^\top\|_2\} \leq \frac{8\sqrt{2}}{3} \|H\|.
\]

### 4 Perturbation analysis for the rank-1 case

#### 4.1 Main results: the rank-1 case

This section presents perturbation analysis results when the truth \( M^* \) is a symmetric rank-1 matrix. The following theorem serves as a master bound that leads to our main findings. Here and throughout, we assume for simplicity that the leading eigenvalue \( \lambda^* \) of \( M^* \) obeys \( \lambda^* = 1 \).
Theorem 3 (Perturbation of linear forms of eigenvectors (rank-1)). Consider a rank-1 symmetric matrix $M^* = u^* u^{\top} \in \mathbb{R}^{n \times n}$ with leading eigenvalue $\lambda^* = 1$ and incoherence parameter $\mu$ (cf. Definition 1). Suppose the noise matrix $H$ obeys Assumption 1, and assume

$$\max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \leq c_1$$

(14)

for some sufficiently small constants $c_1 > 0$. Then for any fixed unit vector $a \in \mathbb{R}^n$, with probability at least $1 - O(n^{-10})$ one has

$$\left| a^\top \left( u - \frac{u^\top u}{\lambda} u^* \right) \right| \leq \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}}.$$  

(15)

In words, Theorem 3 tells us that the approximation $\frac{u^\top v}{\lambda} a^\top u^*$ for a linear function of $u$ is remarkably accurate. Consider, for example, the case where $\sigma^2 \mu \log n \lesssim 1/n$ and $B^2 \mu \log^2 n \lesssim 1$, then the approximation error is at most on the order of $O(1/\sqrt{n})$. Encouragingly, this approximation accuracy holds true for an arbitrary deterministic unit vector $a$. Consequently, one can interpret Theorem 3 as

$$u \approx \frac{1}{\lambda} u^* u^{\top} u = \frac{1}{\lambda} M^* u,$$

(16)

where such an approximation is fairly accurate along any direction. Compared with the identity $u = \frac{1}{\lambda} M u = \frac{1}{\lambda} (M^* + H) u$, our results imply that $H u$ is exceedingly small along any fixed direction. As we shall explain in Section 4.4, this phenomenon usually cannot happen when $H$ is a symmetric random matrix or when one uses the leading singular vector instead, due to the significant bias resulting from symmetry.

**Remark 4.** When both the ground truth $M^*$ and the noise $H$ are symmetric matrices, Abbe et al. (2017) justifies the validity of a different form of approximation

$$u \approx \frac{1}{\lambda^*} M u^* = \frac{1}{\lambda^*} (M^* u^* + H u^*).$$

(17)

This form allows one to exploit the entrywise independence of $M$ (and $H$). By contrast, the approximation (16) highlights the correlation between $u^*$ and $u$, which turns out to be more helpful in controlling the eigenvalue perturbation. In addition, both forms of approximation enable sharp entrywise perturbation bounds for their respective settings, as we shall see shortly.

Theorem 3 has several interesting implications, which we single out as follows.

**Eigenvalue perturbation.** To begin with, Theorem 3 allows us to obtain a much sharper non-asymptotic perturbation bound regarding the leading eigenvalue $\lambda$ of $M$.

**Corollary 1.** Under the assumptions of Theorem 3, with probability at least $1 - O(n^{-10})$ we have

$$|\lambda - \lambda^*| \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}}.$$  

(18)

**Proof.** Recall that $\lambda^* = 1$. Taking $a = u^*$ in Theorem 3, we get

$$\left| u^\top u \right| \left| \frac{\lambda - 1}{\lambda} \right| = \left| u^\top u - u^\top u^* \frac{u^\top u}{\lambda} u \right| \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}}.$$  

(19)

From Lemma 1 and the condition (14), we know $\|H\| < 1/4$, which combines with Lemma 2 and Lemma 3 yields $\lambda \approx |u^\top u| \approx 1$. Substitution into (19) yields

$$|\lambda - 1| \lesssim \left| \frac{\lambda}{u^\top u} \right| \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}} \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}}.$$  

□
In order to facilitate comparison, we denote by $\lambda_{\text{svd}}$ the largest singular value of $M$, and look at $|\lambda_{\text{svd}} - \lambda^*|$. Combining Weyl’s inequality, Lemma 1 and the condition (14), we arrive at

$$|\lambda_{\text{svd}} - \lambda^*| \leq \|H\| \lesssim \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}. \quad (20)$$

When $\mu \approx 1$, this error bound w.r.t. the singular value could be $\sqrt{n}$ times larger than the perturbation bound (18) derived for the leading eigenvalue. This corroborates our motivating experiments in Fig. 1.

As a remark, understanding the bias of $\lambda_{\text{svd}}$ has been the focus of a large line of prior works (e.g. Silverstein (1994); Yin et al. (1988); Pêché (2006); Féral and Pêché (2007); Bryc and Silverstein (2018)). If $H$ contains i.i.d. Gaussian entries with variance $\sigma^2$, it has been shown that $\lambda_{\text{svd}}$ admits the following tighter approximation

$$\lambda^* + \frac{n\sigma^2}{\lambda^*} \approx \lambda_{\text{svd}}. \quad (21)$$

While an approximation of this kind has been found for more general cases beyond i.i.d. noise Bryc and Silverstein (2018), the approximation formula becomes fairly complicated, requires prior knowledge about all variance parameters, and is thus hard to use in practice. In contrast, the vanilla eigenvalue decomposition approach studied here does not rely on any extra parameter, and seems more practically applicable.

**Perturbation of linear forms of eigenvectors.** The master bound in Theorem 3 admits a more convenient form when controlling linear functions of the eigenvectors. The result is this:

**Corollary 2.** Under the same setting of Theorem 3, with probability at least $1 - O(n^{-10})$ we have

$$\min \left\{ |a^T(u - u^*)|, |a^T(u + u^*)| \right\} \lesssim \left( |a^T u^*| + \sqrt{\frac{\mu}{n}} \right) \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}. \quad (22)$$

**Proof.** Without loss of generality, assume that $u^T u \geq 0$. Then one has,

$$|a^T(u - u^*)| \leq |a^T u - a^T u^* | + |a^T u^*| \left| \frac{u^T u}{\lambda} - 1 \right|$$

$$\leq \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}} + |a^T u^*| \left| \frac{u^T u}{\lambda} - 1 \right|,$$

where the last inequality arises from Theorem 3 as well as the definition of $\mu$. In addition, apply Lemma 2 and Lemma 3 to obtain

$$\left| \frac{u^T u}{\lambda} - 1 \right| \leq \frac{u^T u}{\lambda} |1 - \lambda| + |u^T u - 1| \lesssim \|H\| \lesssim \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}.$$

Putting the above bounds together concludes the proof.

The perturbation of linear forms of eigenvectors (or singular vectors) has not yet been well explored even for the symmetric case. One scenario that has been studied is linear forms of singular vectors under i.i.d. Gaussian noise (Koltchinskii and Xia, 2016; Xia, 2016). Our analysis — which is certainly different from Koltchinskii and Xia (2016) as our focus is eigen-decomposition — does not rely on the Gaussiananity assumption, and accommodates a much broader class of random noise. Another work that has looked at linear forms of the leading singular vector is Ma et al. (2017) for phase retrieval and blind deconvolution, although the vector $a$ therein is specific to the problems (i.e. the design vectors) and cannot be made general.

**Entrywise eigenvector perturbation.** A straightforward consequence of Corollary 2 that is worth emphasizing is sharp entrywise control of the leading eigenvector as follows.

**Corollary 3.** Under the same setting of Theorem 3, with probability at least $1 - O(n^{-9})$ we have

$$\min \{ \|u - u^*\|_\infty, \|u + u^*\|_\infty \} \lesssim \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}}. \quad (23)$$
Proof. Recognizing that $\|u - u^*\|_\infty = \max_i |e_i^T(u - e_i^T u^*)|$ and recalling our assumption $|e_i^T u| \leq \sqrt{\mu/n}$, we can invoke Corollary 2 and the union bound to establish this entrywise bound.

We note that: while the $\ell_2$ perturbation (or sin $\Theta$ distance) of eigenvectors or singular vectors has been extensively studied Davis and Kahan (1970); Wedin (1972); Vu (2011); Wang (2015); O’Rourke et al. (2013); Cai and Zhang (2018), the entrywise eigenvector behavior was much less explored. The prior literature contains only a few entrywise eigenvector perturbation analysis results for settings different from ours, e.g. the i.i.d. random matrix case Vu and Wang (2015); O’Rourke et al. (2016), the symmetric low-rank case Fan et al. (2018b); Abbe et al. (2017); Eldridge et al. (2018), and the case with transition matrices for reversible Markov chains (Chen et al., 2017). Our results add another instance to this body of works in providing entrywise eigenvector perturbation bounds. Additionally, the asymptotic behavior of the eigenvalues of asymmetric random matrices has been extensively explored in the physics literature (e.g. (Sommers et al., 1988; Khoruzhenko, 1996; Brezin and Zee, 1998; Chalker and Mehlig, 1998; Feinberg and Zee, 1997; Lytova and Tikhomirov, 2018)). Their focus, however, has largely been to pin down the asymptotic density of the eigenvalues, similar to the semi-circle law in the symmetric case. Nevertheless, a sharp perturbation bound for the leading eigenvalue — particularly for the low-rank case — is beyond their reach.

4.2 Applications
We apply our main results to two concrete matrix estimation problems and examine the effectiveness of these bounds. As before, we assume that $M^*$ is a rank-1 matrix with leading eigenvalue 1 and incoherence parameter $\mu$.

Low-rank matrix estimation from Gaussian noise. Suppose that $H$ is composed of i.i.d. Gaussian random variables $\mathcal{N}(0, \sigma^2)$.\footnote{In this case, one can take $B \propto \sigma \sqrt{\log n}$, which clearly satisfies $B \log n \ll \sqrt{n\sigma^2 \log n}$.} If $\sigma \lesssim \frac{1}{\sqrt{n \log n}}$, applying Corollaries 1-3 reveals that with high probability,

\begin{equation}
|\lambda - \lambda'| \lesssim \sigma \sqrt{\mu \log n} \quad (24a)
\end{equation}

\begin{equation}
\min\{|\|u - u^*\|_\infty, \|u + u^*\|_\infty\} \lesssim \sigma \sqrt{\mu \log n} \quad (24b)
\end{equation}

\begin{equation}
\min\{|a^T(u - u^*), |a^T(u + u^*)|\} \lesssim \left(|a^T u^*| + \sqrt{\frac{\mu}{n}}\right) \sigma \sqrt{n \log n} \quad (24c)
\end{equation}

for any fixed unit vector $a \in \mathbb{R}^n$. We have conducted additional numerical experiments in Fig. 2, which confirm our findings. It is also worth noting that empirically, eigen-decomposition and SVD achieve nearly identical $\ell_2$ and $\ell_\infty$ errors when estimating the leading eigenvector of $M^*$.

![Figure 2: Numerical simulation for rank-1 matrix estimation under i.i.d. Gaussian noise $\mathcal{N}(0, \sigma^2)$, where the rank-1 truth $M^*$ is generated randomly with leading eigenvalue 1. (a): $|\lambda - \lambda'|$ vs. $\sigma$ when $n = 1000$; (b) and (c): $\ell_\infty$ and $\ell_2$ eigenvector estimation errors vs. $n$ with $\sigma = 1/\sqrt{n \log n}$, respectively. The blue (resp. red) lines represent the average errors over 100 independent trials using the eigen-decomposition (resp. SVD) approach.](image-url)
Low-rank matrix completion. Suppose that $M$ is generated using random partial entries of $M^*$ as follows

$$M_{ij} = \begin{cases} \frac{1}{p} M_{ij}^*, & \text{with probability } p, \\ 0, & \text{else,} \end{cases}$$

where $p$ denotes the fraction of the entries of $M^*$ being revealed. It is straightforward to verify that $H = M - M^*$ is zero-mean and obeys $|H_{ij}| \leq \frac{\mu}{np} := B$ and $\text{Var}(H_{ij}) \leq \frac{\mu^2}{pn^2}$. Consequently, if $p \gtrsim \frac{\mu^2 \log n}{n}$, then invoking Corollaries 1-3 yields

$$|\lambda - \lambda^*| \lesssim \frac{1}{\sqrt{n}} \sqrt{\frac{\mu^4 \log n}{pn}}$$

(26a)

$$\min\{\|u - u^*\|_\infty, \|u + u^*\|_\infty\} \lesssim \frac{1}{\sqrt{n}} \sqrt{\frac{\mu^4 \log n}{pn}}$$

(26b)

$$\min\{|a^T (u - u^*)|, |a^T (u + u^*)|\} \lesssim \left( |a^T u^*| + \sqrt{\frac{\mu}{n}} \right) \sqrt{\frac{\mu^2 \log n}{pn}}$$

(26c)

with high probability, where $a \in \mathbb{R}^n$ is any fixed unit vector. Additional numerical simulations have been carried out in Fig. 3 to verify these findings. Empirically, eigen-decomposition outperforms SVD in estimating both the leading eigenvalue and eigenvector of $M^*$.

Figure 3: Numerical simulation for rank-1 matrix completion, where the rank-1 truth $M^*$ is randomly generated with leading eigenvalue 1 and sampling rate is $p = 3 \log n / n$. (a) $|\lambda - \lambda^*|$ vs. $p$ when $n = 1000$; (b) and (c): $\ell_\infty$ and $\ell_2$ eigenvector estimation errors vs. $n$, respectively. The blue (resp. red) lines represent the average errors over 100 independent trials using the eigen-decomposition (resp. SVD) approach.

4.3 Asymmetrization of data samples: two examples

As mentioned earlier, an independent and asymmetric noise matrix arises when we collect two samples for each entry of the matrix of interest and arrange the samples in an asymmetric manner (i.e. placing 1 sample on the upper triangular part and the other on the lower triangular part). Interestingly, our results might be applicable for some cases where we only have 1 sample for each entry. In what follows, we describe two examples of this kind similar to the ones discussed in Section 4.2, but with a symmetric noise matrix. Once again, it is assumed that $M^*$ is a rank-1 matrix with leading eigenvalue 1 and incoherence parameter $\mu$.

Low-rank matrix estimation from Gaussian noise.\footnote{We thank Prof. Zhou Fan for telling us this example.} Suppose that the noise matrix $H$ is a symmetric matrix, and that the upper triangular part of $H$ is composed of i.i.d. Gaussian random variables $\mathcal{N}(0, \sigma^2)$.

When $\sigma$ is known, one can decouple the upper and lower triangular parts of $H$ by adding a skew-symmetric Gaussian matrix $\Delta$. Specifically, our strategy is:

1. For each $1 \leq i \leq j \leq n$, generate $\Delta_{ij} \sim \mathcal{N}(0, \sigma^2)$ independently, and set $\Delta_{ji} = -\Delta_{ij}$;
(2) Compute the leading eigenvalue and eigenvector of \( M + \Delta \).

This is motivated by a simple observation from Gaussianity: \( H + \Delta \) is now an asymmetric matrix whose off-diagonal entries are i.i.d. \( \mathcal{N}(0, 2\sigma^2) \); in fact, it is easy to verify that \( H_{ij} + \Delta_{ij} \) and \( H_{ji} + \Delta_{ji} \) are independent Gaussian random variables. As a result, the orderwise bounds (24) continue to hold if \( \lambda \) and \( u \) are taken to be the leading eigenvalue and eigenvector of \( M + \Delta \), respectively.

While this asymmetrization procedure comes with the price of doubling the noise variance, the eigenvalue perturbation bound may still be significantly smaller — up to a factor of \( O(\sqrt{n}) \) — than the bound for the SVD approach. This is also confirmed in the numerical simulations reported in Fig. 4.

\[ \text{Asymmetrization} \quad \text{Symmetric Perturbation} \]

(a) eigenvalue perturbation vs. \( n \)  \hspace{1cm} (b) eigenvalue perturbation vs. \( \sigma \)  \hspace{1cm} (c) \( \ell_\infty \) eigenvector perturbation

Figure 4: Numerical simulation for rank-1 matrix estimation when \( H \) is symmetric and its upper triangular part consists of i.i.d. Gaussian entries \( \mathcal{N}(0, \sigma^2) \). The truth \( M^* \) is rank-1 and is randomly generated with leading eigenvalue \( \lambda^* = 1 \). (a) \( |\lambda - \lambda^*| \) vs. \( n \) with \( \sigma = 1/\sqrt{n \log n} \); (b) \( |\lambda - \lambda^*| \) vs. \( \sigma \) with \( n = 1000 \); (c) \( \ell_\infty \) eigenvector perturbation vs. \( n \) with \( \sigma = 1/\sqrt{n \log n} \). We compare eigen-decomposition applied to symmetric matrix samples (the red lines) and asymmetrized data (the blue lines). In the green line, the estimate \( \hat{u} \) is obtained by generating 10 independent copies of \( \Delta \) and aggregating the leading eigenvectors of these copies of \( M + \Delta \) (see Remark 5).

\textbf{Remark 5.} While this asymmetrization procedure achieves enhanced eigenvalue estimation accuracy compared to the SVD approach (see Fig. 4(a)(b)), it results in higher eigenvector estimation errors (see Fig. 4(c)). This is perhaps not surprising as we have added extra noise to the observed matrix. One way to mitigate this issue is to (1) generate \( K \) independent copies of \( \Delta \); (2) compute the leading eigenvector of each copy of \( M + \Delta \), denoted by \( \{ u^{(l)} | 1 \leq l \leq K \} \); and (3) aggregate these eigenvectors, namely, compute the leading eigenvector of \( \frac{1}{K} \sum_{l=1}^{K} u^{(l)} u^{(l)\top} \). As can be seen in the green line of Fig. 4(c), this allows us to mitigate the effect of the extra noise component.

\textbf{Low-rank matrix completion.} Suppose that each entry \( M^*_{ij} \) (\( i \geq j \)) is observed independently with probability \( p \). This is different from the settings in Section 4.2, as we do not have additional samples for \( M^*_{ji} \) (\( i \geq j \)). In order to arrange the data samples in an asymmetric and independent fashion, we employ a simple resampling technique to decouple the statistical dependency between \( M_{ij} \) and \( M_{ji} \):

(1) Define \( M^\text{sym} \) such that

\[ M^\text{sym}_{ij} = \begin{cases} \frac{1}{p} M^*_ij, & \text{if } M^*_ij \text{ is observed;} \\ 0, & \text{else.} \end{cases} \]

(2) Define \( p^\text{asym} \) so that \( p = 1 - (1 - p^\text{asym})^2 \) (i.e. \( p^\text{asym} = \frac{p}{1 + \sqrt{1 - p}} \)). For any pair \( i > j \), set

\[ (M^\text{sym}_{ij}, M^\text{asym}_{ji}) = \begin{cases} \frac{p}{p^\text{asym}} (M^\text{sym}_{ij}, 0) & \text{with probability } \frac{1 - p^\text{asym}}{2 - p^\text{asym}}, \\ \frac{p}{p^\text{asym}} (0, M^\text{sym}_{ij}) & \text{with probability } \frac{1 - p^\text{asym}}{2 - p^\text{asym}}, \\ \frac{p}{p^\text{asym}} (M^\text{sym}_{ij}, M^\text{sym}_{ij}) & \text{with probability } \frac{p^\text{asym}}{2 - p^\text{asym}}. \end{cases} \quad (27) \]

For any \( i \), set

\[ M^\text{asym}_{ii} = \begin{cases} \frac{p}{p^\text{asym}} M^\text{sym}_{ii} & \text{with probability } \frac{p^\text{asym}}{p}, \\ 0 & \text{else.} \end{cases} \]
(3) Compute the leading eigenvalue and eigenvector of \( M_{\text{asym}} \).

As can be easily verified, this scheme is equivalent to saying that (i) with probability \( p \), either \( M_{ij}^{\text{asym}} \) or \( M_{ij}^{\text{sym}} \) is taken to be a rescaled version of \( M_{ij} \); (ii) for any \( i \neq j \), the entries \( \{M_{ij}^{\text{asym}}\} \) are independently drawn. Since \( p_{\text{asym}} < p \), our results (26) in Section 4.2 remain valid, as long as \( \lambda \) and \( u \) are set to be the leading eigenvalue and eigenvector of \( M_{\text{asym}} \), respectively. Numerical simulations have been carried out in Fig. 5 to verify the effectiveness of this scheme.

![Fig. 5](image)

(a) eigenvalue perturbation vs. \( n \) 
(b) eigenvalue perturbation vs. \( p \) 
(c) \( \ell_\infty \) eigenvector perturbation

Figure 5: Numerical experiments for rank-1 matrix completion, where the rank-1 truth \( M^* \) is randomly generated with leading eigenvalue 1. (a) \(|\lambda - \lambda^*|\) vs. \( n \) with \( p = 3 \log n / n \); (b) \(|\lambda - \lambda^*|\) vs. \( p \) with \( n = 1000 \); (c) \( \ell_\infty \) perturbation error vs. \( n \) with \( p = 3 \log n / n \). We compare eigen-decomposition applied to symmetric matrix samples and asymmetrized data. The blue (resp. red) lines represent the average errors over 100 independent trials using the eigen-decomposition (resp. SVD) approach.

### 4.4 Why asymmetry helps?

We take a moment to develop some intuition underlying Theorem 3. The key ingredient is the Neumann trick stated in Theorem 2. Specifically, in the rank-1 case we can expand

\[
u = \frac{1}{\lambda} (u^\top u) \sum_{s=0}^{\infty} \frac{1}{\lambda^s} H^s u^*.
\]

A little algebra yields

\[
|a^\top (u - \frac{u^\top u}{\lambda} u)| = \left| \frac{u^\top u}{\lambda} \sum_{s=1}^{\infty} \frac{a^\top H^s u^*}{\lambda^s} \right| \leq \frac{|u^\top u|}{\lambda} \sum_{s=1}^{\infty} \left| \frac{a^\top H^s u^*}{\lambda^s} \right| \leq \sum_{s=1}^{\infty} \frac{|a^\top H^s u^*|}{\lambda^s}.
\]

(28)

where the last inequality holds since (i) \(|u^\top u| \leq 1\), and (ii) \( \lambda \) is real-valued and obeys \( \lambda \approx 1 \) if \( \|H\| \ll 1 \) (in view of Lemma 2). As a result, the perturbation can be well-controlled as long as \( |a^\top H^s u^*| \) is small for every \( s \geq 1 \).

As it turns out, \( a^\top H^s u^* \) might be much better controlled when \( H \) is random and asymmetric, in comparison to the case where \( H \) is random and symmetric. To illustrate this point, it is perhaps the easiest to inspect the second-order term.

- **Asymmetric case**: when \( H \) is composed of independent zero-mean entries each with variance \( \sigma_n^2 \), one has

\[
E[\sigma_n^2 |a^\top H^2 u^*|] = a^\top E[H^2] u^* = a^\top (\sigma^2 I) u^* = \sigma^2 a^\top u^*.
\]

- **Symmetric case**: when \( H \) is symmetric and its upper triangular part consists of independent zero-mean entries with variance \( \sigma_n^2 \), it holds that

\[
E[\sigma_n^2 |a^\top H^2 u^*|] = a^\top E[H^2] u^* = a^\top (n\sigma^2 I) u^* = n\sigma^2 a^\top u^*.
\]

11
In words, the term $a^\top H^2u^*$ in the symmetric case might have a significantly larger bias compared to the asymmetric case. This bias effect is substantial when $a^\top u$ is large (e.g. when $a = u^*$), which plays a crucial role in determining the size of eigenvalue perturbation.

The SVD-based approach can be interpreted in a similar manner. Specifically, we recognize that the leading singular value (resp. left singular vector) can be computed via the leading eigenvalue (resp. eigenvector) of the symmetric matrix $MM^\top$. Given that $MM^\top - M^*M^*^\top$ is also symmetric, the aforementioned bias issue arises as well. This explains why eigen-decomposition might have an advantage over SVD when dealing with asymmetric matrices.

Finally, we remark that the aforementioned bias issue becomes less severe as $\|H\|$ decreases. For example, when $\|H\|$ is exceedingly small, the only dominant term on the right-hand side of (28) is $a^\top Hu$, with all higher-order terms being vanishingly small. In this case, $\mathbb{E}[a^\top Hu] = 0$ for both symmetric and asymmetric zero-mean noise matrices. As a consequence, the advantage of eigen-decomposition becomes negligible when dealing with nearly-zero noise. This observation is also confirmed in the numerical experiments reported in Fig. 2(a) and Fig. 3(a), where the two approaches achieve similar eigenvalue estimation accuracy when $\sigma \to 0$ (resp. $p \to 1$) in matrix estimation under Gaussian noise (resp. matrix completion). In fact, the case with very small $\|H\|$ has been studied in the literature O’Rourke et al. (2013); Vu (2011); Eldridge et al. (2018). For example, it was shown in O’Rourke et al. (2013) that when $\|H\| \lesssim \frac{1}{\sqrt{n}}\|M^*\|$, the singular value perturbation is also $\sqrt{n}$ times smaller than the bound predicted by Weyl’s theorem; similar improvement can be observed w.r.t. eigenvalue perturbation when $H$ is symmetric (cf. (Eldridge et al., 2018, Theorem 6)). In comparison, our eigenvalue perturbation results achieve this gain even when $\|H\|$ is nearly as large as $\|M^*\|$ (up to some logarithmic factor).

4.5 Proof outline of Theorem 3

This subsection outlines the main steps for establishing Theorem 3. Throughout this paper, all the proofs are provided for the case when Conditions 1-3, 4(a) in Assumption 1 are valid. Otherwise, if Condition 4(b) is valid, then we can invoke the union bound to show that

$$M = M^* + \tilde{H}$$

with probability exceeding $1 - O(n^{-10})$, where $\tilde{H}_{ij} \triangleq H_{ij} 1_{\{|H_{ij}| \leq B\}}$ is the truncated noise and has magnitude bounded by $B$. Since $H_{ij}$ has symmetric distribution, it is seen that $\mathbb{E}[\tilde{H}_{ij}] = 0$ and $\text{Var}(\tilde{H}_{ij}) \leq \sigma^2$, which coincides with the case obeying Conditions 1-3, 4(a) in Assumption 1.

As already mentioned in Section 4.4, everything boils down to controlling $|a^\top H^*u^*|$ for $s \geq 1$. This is accomplished via the following lemma.

**Lemma 4 (Bounding higher-order terms).** Consider any fixed unit vector $a \in \mathbb{R}^n$ and any positive integers $s, k$ satisfying $Bsk \leq 2$ and $n\sigma^2 sk \leq 2$. Under the assumptions of Theorem 3,

$$\left|\mathbb{E} \left[ \left( a^\top H^*u^* \right)^k \right] \right| \leq \frac{s^k}{2} \max \left\{ (Bsk)^s, (2n\sigma^2 sk)^{s/2} \right\} \left( \frac{\mu}{n} \right)^{k/2}.$$  

**Proof.** The proof of Lemma 4 is combinatorial in nature, which we defer to Appendix B. \qed

**Remark 6.** A similar result in (Tao, 2013, Lemma 2.3) has studied the bilinear forms of the high order terms of an i.i.d. random matrix, with a few distinctions. First of all, Tao (2013) assumes that each entry of the noise matrix is i.i.d. and has finite fourth moment (if the noise variance is rescaled to be 1); these assumptions break in examples like matrix completion. Moreover, Tao (2013) focuses on the case with $k = 2$, and does not lead to high-probability bounds (which are crucial for, e.g. entrywise error control).

Using Markov’s inequality and the union bound, we can translate Lemma 4 into a high probability bound as follows.

**Corollary 4.** Under the assumptions of Lemma 4, there exists some universal constant $c_2 > 0$ such that

$$|a^\top H^*u^*| \leq \left( c_2 \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \right)^s \sqrt{\frac{\mu}{n}}, \quad \forall s \leq 20 \log n$$

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

12
Proof. See Appendix E.

In addition, in view of Lemma 1 and the condition (14), one has
\[ \|H\| \lesssim \max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} < 1/10 \] (31)
with probability \( 1 - O(n^{-10}) \), which together with Lemma 2 implies \( \lambda \geq 3\|H\| \). This further leads to
\[
\sum_{s:s \geq 20 \log n} \left( \frac{\|H\|}{\lambda} \right)^s \leq \frac{\|H\|}{\lambda} \sum_{s:s \geq 20 \log n - 1} \left( \frac{\|H\|}{\lambda} \right)^s \leq \frac{\|H\|}{\lambda} \sum_{s:s \geq 20 \log n - 1} \frac{1}{3^s} \\
\lesssim \max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \cdot n^{-10}.
\]
Putting the above bounds together and using the fact that \( \lambda \) is real-valued and \( \lambda \geq 1/2 \) (cf. Lemma 2), we have
\[
\left| a^\top \left( u - \frac{u^\top u}{\lambda} u^* \right) \right| = \left| \frac{u^\top u}{\lambda} \sum_{s=1}^{+\infty} \frac{a^\top H^s u^*}{\lambda^s} \right| \\
\lesssim \sum_{s=1}^{20 \log n} \frac{1}{\lambda^s} \left| a^\top H^s u^* \right| + \sum_{s=20 \log n}^{+\infty} \left( \frac{\|H\|}{\lambda} \right)^s \\
\leq \left( \frac{\mu}{n} \right)^{20 \log n} \sum_{s=1}^{20 \log n} \left( 2c_2 \max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \right)^s + \frac{\max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}}{n^{10}} \\
\lesssim \max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \cdot \sqrt{\frac{\mu}{n}},
\]
as long as \( \max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \) is sufficiently small. Here, the last line also uses the fact that \( \mu \geq 1 \) (and hence \( \sqrt{\mu/n} \gg n^{-10} \)). This finishes the proof.

5 Extension: perturbation analysis for the rank-\( r \) case

5.1 Eigenvalue perturbation for the rank-\( r \) case

The eigenvalue perturbation analysis in Section 4 can be extended to accommodate the case where \( M^* \) is symmetric and rank-\( r \), as detailed in this section. As before, assume that the \( r \) non-zero eigenvalues of \( M^* \) obey \( \lambda_{\max}^* = |\lambda_1^*| \geq \cdots \geq |\lambda_r^*| = \lambda_{\min}^* \). We start again with a master bound. Here and throughout, it is assumed that
\[ \lambda_{\max}^* = 1. \]

**Theorem 4 (Perturbation of linear forms of eigenvectors (rank-\( r \))).** Consider a rank-\( r \) symmetric matrix \( M^* \in \mathbb{R}^{n \times n} \) with leading eigenvalue \( \lambda_{\max}^* = 1 \) and incoherence parameter \( \mu \). Define \( \kappa \triangleq \lambda_{\max}^*/\lambda_{\min}^* \).

Suppose that
\[
\max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \leq c_1/\kappa
\]
(32)
for some sufficiently small constants \( c_1 > 0 \). Then for any fixed unit vector \( a \in \mathbb{R}^n \) and any \( 1 \leq l \leq r \), with probability at least \( 1 - O(n^{-10}) \) one has
\[
\left| a^\top \left( u_l - \sum_{j=1}^{r} \frac{\lambda_j u_j^\top u_j}{\lambda_l} u_j^* \right) \right| \lesssim \max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \cdot \frac{1}{|\lambda_l|} \sqrt{\frac{\mu r}{n}} \\
\lesssim \max\left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \cdot \kappa \sqrt{\frac{\mu r}{n}}.
\]
This result allows us to control the perturbation of the linear form of eigenvectors. The perturbation upper bound grows as either the rank $r$ or the condition number $\kappa$ increases.

One of the most important consequences of Theorem 4 is a refinement of the Bauer-Fike theorem concerning eigenvalue perturbations as follows.

**Corollary 5.** Consider the $l$th ($1 \leq l \leq r$) eigenvalue $\lambda_l$ of $M$. Under the assumptions of Theorem 4, with probability at least $1 - O(n^{-10})$, there exists $1 \leq j \leq r$ such that

$$|\lambda_l - \lambda_j^\star| \leq \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} r \sqrt{\frac{\mu}{n}},$$

provided that

$$\max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \leq c_1 / \kappa^2$$

for some sufficiently small constant $c_1 > 0$.

**Proof.** See Appendix F.

In comparison, the Bauer-Fike theorem (Lemma 2) together with Lemma 1 gives a perturbation bound

$$|\lambda_l - \lambda_j^\star| \leq \|H\| \leq \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}$$

for some $1 \leq j \leq r$. (37)

For the low-rank case where $r \ll \sqrt{n}$, the eigenvalue perturbation bound derived in Corollary 5 can be much sharper than the Bauer-Fike theorem.

Another result that comes from Theorem 4 is the following bound that concerns linear forms of the eigen-subspace.

**Corollary 6.** Under the same setting of Theorem 4, with probability $1 - O(n^{-9})$ we have

$$\|a^\top U\|_2 \lesssim \kappa \sqrt{r} \|a^\top U^\star\|_2 + \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \kappa r \sqrt{\frac{\mu}{n}}.$$

**Proof.** See Appendix G.

Consequently, by taking $a = e_i$ ($1 \leq i \leq n$) in Corollary 6, we arrive at the following statement regarding the alternative definition of the incoherence of the eigenvector matrix $U$ (see Remark 2).

**Corollary 7.** Under the same setting of Theorem 4, with probability $1 - O(n^{-8})$ we have

$$\|U\|_{2,\infty} \leq \kappa r \sqrt{\frac{\mu}{n}}.$$  

**Proof.** Given that $\|U\|_{2,\infty} = \max_{1 \leq i \leq n} \|e_i^\top U\|_2$ and recalling our assumption implies $\|U^\star\|_{2,\infty} \leq \sqrt{\mu r / n}$, we can invoke Corollary 6 and the union bound to derive the advertised entrywise bounds.

**Remark 7.** The eigenvector matrix is often employed to form a reasonably good initial guess for several nonconvex statistical estimation problems Keshavan et al. (2010), and the above kind of incoherence property is crucial in guaranteeing fast convergence of the subsequent nonconvex iterative refinement procedures Ma et al. (2017).

Unfortunately, these results fall short of providing simple perturbation bounds for the eigenvectors; in other words, the above-mentioned bounds do not imply the size of the difference between $U$ and $U^\star$. The challenge arises in part due to the lack of orthonormality of the eigenvectors when dealing with asymmetric matrices. Analyzing the eigenspace perturbation for the general rank-$r$ case will likely require new analysis techniques, which we leave for future work. There is, however, some special case in which we can develop eigenvector perturbation theory, as detailed in the next subsection.
5.2 Application: spectral estimation when $M^*$ is asymmetric and rank-1

In some scenarios, the above general rank results allow us to improve spectral estimation when $M^*$ is asymmetric. Consider the case where $M^* = u^* v^{\dagger}$ is an asymmetric rank-1 matrix with leading singular value $\sigma^* = 1$. Suppose that we observe two independent noisy copies of $M^*$, namely,

$$ M_1 = M^* + H_1, \quad M_2 = M^* + H_2, $$

where $H_1$ and $H_2$ are independent noise matrices. The goal is to estimate the singular value and singular vectors of $M^*$ from $M_1$ and $M_2$.

We attempt estimation via the standard dilation trick Tao (2012). This consists of embedding the matrices of interest within a larger block matrix

$$ M^*_\text{dilation} \triangleq \begin{bmatrix} 0 & M^* \\ M^{\dagger}\top & 0 \end{bmatrix}, \quad M_\text{dilation} \triangleq \begin{bmatrix} 0 & M_1 \\ M_2^{\dagger} & 0 \end{bmatrix}. $$

Here, we place $M_1$ and $M_2$ in two different subblocks, in order to “asymmetrize” the dilation matrix. The rationale is that $M^*_\text{dilation}$ is a rank-2 symmetric matrix with exactly two nonzero eigenvalues

$$ \lambda_1(M^*_\text{dilation}) = \sigma^* = 1 \quad \text{and} \quad \lambda_2(M^*_\text{dilation}) = -\sigma^* = -1, $$

whose corresponding eigenvectors are given by

$$ \frac{1}{\sqrt{2}} \begin{pmatrix} u^* \\ v^* \end{pmatrix} \quad \text{and} \quad \frac{1}{\sqrt{2}} \begin{pmatrix} u^* \\ -v^* \end{pmatrix}, $$

respectively. This motivates us to perform eigen-decomposition of $M_\text{dilation}$, and use the top-2 eigenvalues and eigenvectors to estimate $\sigma^*$, $u^*$ and $v^*$, respectively.

Eigenvalue perturbation analysis. As an immediate consequence of Corollary 5, the two leading eigenvalues of $M_\text{dilation}$ provide fairly accurate estimates of the leading singular value $\sigma^*$ of $M^*$, as stated below.

**Corollary 8.** Assume that $M^* \in \mathbb{R}^{n_1 \times n_2}$ is a rank-1 matrix with leading singular value $\sigma^*$ and incoherence parameter $\mu$. Suppose that $\lambda_1^{\dagger} \geq \lambda_2^{\dagger}$ are the two leading eigenvalues of $M^*_\text{dilation}$ (cf. (41)), and that $H_1$ and $H_2$ are independent and satisfy Assumption 1. Define $n \triangleq n_1 + n_2$. With probability at least $1 - O(n^{-10})$, one has

$$ \max \{ |\lambda_1^{\dagger} - \sigma^*|, |\lambda_2^{\dagger} + \sigma^*| \} \lesssim \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}}, $$

provided that

$$ \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \leq c_1 $$

for some sufficiently small constant $c_1 > 0$.

**Proof.** To begin with, it follows from Corollary 5 that both $\lambda_1^{\dagger}$ and $\lambda_2^{\dagger}$ are close to either $\sigma^*$ or $-\sigma^*$. Repeating similar arguments as in the proof of Lemma 2 (which we omit here), we can immediately show the separation between these two eigenvalues, namely, $\lambda_1^{\dagger}$ (resp. $\lambda_2^{\dagger}$) is close to $\sigma^*$ (resp. $-\sigma^*$).

Eigenvector perturbation analysis. We then move on to studying the eigenvector perturbation bounds. Specifically, denote by $u_1^{\dagger}$ and $u_2^{\dagger}$ the eigenvectors of $M^*_\text{dilation}$ associated with its two leading eigenvalues $\lambda_1^{\dagger}$ and $\lambda_2^{\dagger}$, respectively. Without loss of generality, we assume that $\lambda_1^{\dagger} \geq \lambda_2^{\dagger}$. If we write

$$ u_1^{\dagger} = \left( u_{1,1}^{\dagger}, u_{1,2}^{\dagger} \right) \quad \text{with} \quad u_{1,1}^{\dagger} \in \mathbb{R}^{n_1}, u_{1,2}^{\dagger} \in \mathbb{R}^{n_2}, $$

then

$$ \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \leq c_1 $$

for some sufficiently small constant $c_1 > 0$.
then we can employ \( u^\text{dilation}_{1,1} \) and \( u^\text{dilation}_{1,2} \) to estimate \( u^* \) and \( v^* \) after proper normalization, namely,

\[
\begin{align*}
    u & \triangleq \frac{u^\text{dilation}_{1,1}}{\|u^\text{dilation}_{1,1}\|_2}, \\
    v & \triangleq \frac{u^\text{dilation}_{1,2}}{\|u^\text{dilation}_{1,2}\|_2}.
\end{align*}
\]

(44)

The following theorem develops error bounds for both \( u \) and \( v \), which we establish in Appendix H. Here, we denote \( \min \|x + y\|_2 = \min \{\|x - y\|_2, \|x + y\|_2\} \) and \( \min \|x + y\|_\infty = \min \{\|x - y\|_\infty, \|x + y\|_\infty\} \).

**Theorem 5.** Suppose \( M^* = u^* v^* \top \in \mathbb{R}^{n_1 \times n_2} \) is a rank-1 matrix with leading singular value \( \sigma^* = 1 \) and incoherence parameter \( \mu \), where \( \|u^*\|_2 = \|v^*\|_2 = 1 \). Define \( n = n_1 + n_2 \), and fix any unit vectors \( a \in \mathbb{R}^{n_1} \) and \( b \in \mathbb{R}^{n_2} \). Then with probability at least \( 1 - O(n^{-10}) \), the estimates \( u \) and \( v \) (cf. (44)) obey

\[
\begin{align*}
    \max \left\{ \min \|u \pm u^*\|_2, \min \|v \pm v^*\|_2 \right\} & \lesssim \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}, \\
    \min \{\|a^\top (u - u^*)\|, \|a^\top (u + u^*)\|\} & \lesssim \left(\|a^\top u^*\| + \sqrt{\frac{\mu}{n}}\right) \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}, \\
    \min \{\|b^\top (v - v^*)\|, \|b^\top (v + v^*)\|\} & \lesssim \left(\|b^\top v^*\| + \sqrt{\frac{\mu}{n}}\right) \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\},
\end{align*}
\]

provided that

\[
\max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \leq c_1
\]

for some sufficiently small constant \( c_1 > 0 \).

Similar to the symmetric rank-1 case, the estimation errors of the estimates \( u \) and \( v \) are well-controlled in any deterministic direction (e.g. the entrywise errors are well-controlled). This allows us to complete the theory for the case when \( M^* \) is a real-valued and rank-1 matrix.

Figure 6: Numerical experiments for rank-1 matrix completion, where \( M^* = u^* v^* \top \in \mathbb{R}^{n_1 \times n_2} \) is randomly generated with leading singular value \( \sigma^* = 1 \). Let \( n = n_1 = 2n_2 \). Each entry is observed independently with probability \( p \). (a) \( |\sigma - \sigma'| \) vs. \( n \) with \( p = 3 \log n / n \); (b) \( |\sigma - \sigma'| \) vs. \( p \) with \( n = 1000 \); (c) \( \ell_\infty \) eigenvector estimation error vs. \( n \) with \( p = 3 \log n / n \). The blue (resp. red) lines represent the average errors over 100 independent trials using using the eigen-decomposition (resp. SVD) approach applied to \( M_{\text{dilation}} \) (resp. \( M \)).

Finally, we conduct numerical experiments for matrix completion when \( M^* \) is a rank-1 and asymmetric matrix in Fig. 6. Here, we suppose that at most 1 sample is observed for each entry, and we estimate the singular value and singular vectors of \( M^* \) via the above-mentioned dilation trick, coupled with the asymmetrization procedure discussed in Section 4.3. The numerical performance confirms that the proposed technique outperforms SVD in spectral estimation.
5.3 Proof of Theorem 4

To begin with, Lemma 2 implies that for all $1 \leq l \leq r$,
\[ |\lambda_l| \geq |\lambda^*_l| - \|H\| > 1/(2\kappa) > \|H\| \]  
(47)
as long as $\|H\| < 1/(2\kappa)$. In view of the Neumann trick (Theorem 2), we can derive
\[
\left| a^\top u_l - \sum_{j=1}^r \frac{\lambda_j^* u_j^\top u_l}{\lambda_l} a^\top u_j \right| = \left| \sum_{j=1}^r \frac{\lambda_j^*}{\lambda_l} (u_j^\top u_l) \{ \sum_{s=1}^\infty \frac{1}{\lambda_l^s} a^\top H^s u_j \} \right|
\leq \left( \sum_{j=1}^r \frac{\lambda_j^*}{\lambda_l} |u_j^\top u_l| \right) \left\{ \max_{1 \leq j \leq r} \sum_{s=1}^\infty \frac{1}{|\lambda_l|^s} |a^\top H^s u_j| \right\}
\leq \sqrt{r} \cdot \frac{1}{|\lambda_l|} \left\{ \max_{1 \leq j \leq r} \sum_{s=1}^\infty \frac{1}{|\lambda_l|^s} |a^\top H^s u_j^*| \right\},
\]  
(48)
where the third line follows since $\sum_{j=1}^r |u_j^\top u_l|^2 \leq \|u_l\|_2^2 = 1$, and the last inequality makes use of (47). Apply Corollary 4 to reach
\[
(48) \leq \sqrt{r} \sum_{s=1}^\infty \left( 2c_2 \kappa \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \right)^s \sqrt{\frac{\mu}{n}}
\leq \frac{1}{|\lambda_l|} \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}}
\leq \kappa \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}},
\]
with the proviso that $|\lambda_l| > 1/(2\kappa)$ and $\max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \leq c_1/\kappa$ for some sufficiently small constant $c_1 > 0$. The condition $|\lambda_l| > 1/(2\kappa)$ follows immediately by combining Lemma 2, Lemma 1 and the condition (32).

6 Discussions

In this paper, we demonstrate the remarkable advantage of eigen-decomposition over SVD in the presence of asymmetric noise matrices. This is in stark contrast to conventional wisdom, which is generally not in favor of eigen-decomposition for asymmetric matrices. Our results only reflect the tip of an iceberg, and there are many outstanding issues left unanswered. We conclude the paper with a few future directions.

**Sharper eigenvalue perturbation bounds for the rank-r case.** Our current results in Section 5 provide an eigenvalue perturbation bound on the order of $r/\sqrt{n}$, assuming the truth is rank-$r$. However, numerical experiments suggest that the dependency on $r$ might be improvable. It would be interesting to see whether further theoretical refinement is possible, e.g. whether it is possible to improve it to $O(\sqrt{r/n})$.

**Eigenvector perturbation bounds for the rank-r case.** As mentioned before, the current theory falls short of providing eigenvector perturbation bounds for the general rank-$r$ case. The main difficulty lies in the lack of orthogonality of the eigenvectors of the observed matrix $M$. Nevertheless, when the size of the noise is not too large, it is possible to establish certain near-orthogonality of the eigenvectors, which might in turn lead to sharp control of eigenvector perturbation.

**Nearly-symmetric matrices.** So far our theory focuses on the case where the ground truth $M^*$ is symmetric. One might expect the perturbation analysis to extend to more general case. For example, what happens if the truth $M^*$ is nearly (but not exactly) symmetric?
Figure 7: Numerical experiments for the spiked covariance model, where the sample vectors are zero-mean Gaussian vectors with covariance matrix $\Sigma^\ast$ (cf. (49)). We plot $|\lambda - \lambda^\ast|$ vs. $n$ with $d = n/10$. The blue (resp. red) lines represent the average errors over 100 independent trials when $\lambda$ is the leading eigenvalue of $\hat{\Sigma}_{\text{asy}}$ (resp. $\hat{\Sigma}$).

**Beyond independent noise.** The current theoretical development relies heavily on the assumption that the noise matrix $H$ contains independent random entries. There is no shortage of examples where the noise matrix is asymmetric but is not composed of independent entries. For instance, in blind deconvolution Li et al. (2018), the noise matrix is a sum of independent asymmetric matrices. Can we develop eigenvalue perturbation theory for this class of noise?

**Statistical inference of eigenvalues and eigenvectors.** In various applications like network analysis and inference, one might be interested in determining the (asymptotic) eigenvalue and eigenvector distributions of a random data matrix, in order to produce valid confidence intervals Johnstone (2001); Bai et al. (2008); Fan et al. (2018a); Cai et al. (2017); Cape et al. (2018); Xia (2018); Bao et al. (2018). Can we use the current framework to characterize the distributions of the leading eigenvalues as well as certain linear forms of the eigenvectors of $M$ when the noise matrix is non-symmetric?

**Asymmetrization for other applications.** Given the abundant applications of spectral estimation, our findings are likely to be useful for other matrix eigenvalue problems and might extend to the tensor case Zhang and Xia (2018). Here, we conclude the paper with an example in covariance estimation (Baik and Silverstein, 2006; Fan et al., 2018b). Imagine that we observe a collection of $n$ independent Gaussian vectors $X_1, \cdots, X_n \in \mathbb{R}^d$, which have mean zero and covariance matrix

$$
\Sigma^\ast = vv^\top + I_d
$$

(49)

with $v$ being a unit vector. This falls under the category of the spiked covariance model (Johnstone and Lu, 2009; Wang and Fan, 2017). One strategy to estimate the spectral norm $\lambda^\ast = 2$ of $\Sigma^\ast$ is to look at the spectrum of the sample covariance matrix $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n X_iX_i^\top$. Motivated by the results of this paper, we propose an alternative strategy by looking at the following asymmetrized sample covariance matrix

$$
\hat{\Sigma}_{\text{asy}} = \frac{2}{n} \left( \sum_{i=1}^{n/2} \text{Upper} \left( X_iX_i^\top \right) + \sum_{i=n/2+1}^{n} \text{Lower} \left( X_iX_i^\top \right) \right),
$$

(50)

where Upper$(\cdot)$ (resp. Lower$(\cdot)$) extracts out the upper (resp. lower) triangular part of the matrix, including (resp. excluding) the diagonal entries. As can be seen from Fig. 7, the largest eigenvalue of the asymmetrized $\hat{\Sigma}_{\text{asy}}$ is much closer to the true spectral norm of $\Sigma^\ast$, compared to the largest singular value of the sample covariance matrix $\hat{\Sigma}$. We leave the theoretical understanding of such findings to future investigation.

**Acknowledgment**

Y. Chen is supported in part by the AFOSR YIP award FA9550-19-1-0030, by the ARO grant W911NF-18-1-0303, and by the Princeton SEAS innovation award. J. Fan is supported in part by the NSF grants
DMS-1662139 and DMS-1712591 and NIH grant 2R01-GM072611-13. C. Cheng is supported by the Elite Undergraduate Training Program of School of Mathematical Sciences in Peking University. We thank Cong Ma for helpful discussions, and Zhou Fan for telling us an example of asymmetrizing the Gaussian matrix.

A Proofs for preliminary facts

A.1 Proof of Lemma 2

Without loss of generality, suppose the leading eigenvalue of \( M^* \) is 1.

(1) We start by proving the rank-1 case, in which the eigenvalues of \( M^* \) are either 0 or 1. Theorem 1 immediately implies that

\[
\Lambda(M) \subseteq B(1, \|H\|) \cup B(0, \|H\|),
\]

where \( \Lambda(M) \) is the set of eigenvalues of \( M \), and \( B(z, r) \triangleq \{ x \in \mathbb{C} : |x - z| \leq r \} \) denotes a disk of radius \( r \) centered at \( z \).

To establish the lemma, our aim is to show that there is exactly one eigenvalue of \( M \) — denoted by \( \lambda \) — lying in the disk \( B(1, \|H\|) \), and it has multiplicity 1. If this is true, then the assumption \( \|H\| < 1/2 \) indicates that \( |\lambda| \geq 1 - \|H\| > \|H\| \geq z \) for any \( z \in B(0, \|H\|) \), and hence \( \lambda \) must be the leading eigenvalue. Furthermore, since \( M \) is a real-valued matrix, both \( \lambda \) and its complex conjugate \( \bar{\lambda} \) are eigenvalues of \( M \). However, if \( \lambda \neq \bar{\lambda} \), then both of these two eigenvalues fall within \( B(1, \|H\|) \), resulting in contradiction. As a result, one necessarily has \( \lambda = \bar{\lambda} \) and both of them are real-valued. A similar argument demonstrates that the eigenvector \( u \) associated with \( \lambda \) is also real-valued.

We then justify the existence and uniqueness of an eigenvalue in \( B(1, \|H\|) \). Denote \( \Lambda(M) = \{ \lambda_1, \cdots, \lambda_n \} \), and define a set of auxiliary matrices

\[
M_t = M^* + tH, \quad 0 \leq t \leq 1.
\]

Similar to (51), one has

\[
\Lambda(M_t) \subseteq B(1, t \|H\|) \cup B(0, t \|H\|).
\]

Recognizing that the set of eigenvalues of \( M_t \) depends continuously on \( t \) (e.g. (Embree and Trefethen, 2001, Theorem 6)), we can write

\[
\Lambda(M_t) = \{ \lambda_1(t), \lambda_2(t), \cdots, \lambda_n(t) \},
\]

with each \( \lambda_j(t), 1 \leq j \leq n \) being a continuous function in \( t \). Meanwhile, as long as \( \|H\| < 1/2 \) and \( 0 \leq t \leq 1 \), the two disks \( B(1, t \|H\|) \) and \( B(0, t \|H\|) \) are always disjoint. Thus, the continuity of the spectrum (w.r.t. \( t \)) requires \( \lambda_j(t) \) to always stay within the same disk where \( \lambda_j(0) \in \{0, 1\} \) lies, namely,

\[
\lambda_j(t) \in B(\lambda_j(0), t \|H\|).
\]

Given that \( M^* \) (or \( M_0 \)) has \( n - 1 \) eigenvalues equal to 0 and one eigenvalue equal to 1, we establish the lemma for the rank-1 case.

(2) We now turn to the rank-\( r \) case. Repeating the above argument for the rank-1 case, we can immediately show that: if \( \|H\| < \lambda^*_r/2 \), then (i) there are exactly \( n - r \) eigenvalues lying within \( B(0, \|H\|) \); (ii) all other eigenvalues lie within \( \cup_{1 \leq j \leq r} B(\lambda^*_j, \|H\|) \), which are exactly the top-\( r \) leading eigenvalues of \( M \). This concludes the proof.

A.2 Proof of Theorem 2

From the definition of eigenvectors,

\[
(M^* + H) u_l = \lambda_l u_l, \quad \text{or equivalently}, \quad \frac{1}{\lambda_l} M^* u_l = \left( I - \frac{1}{\lambda_l} H \right) u_l.
\]

When \( \|H\|_2 < |\lambda_l| \), one can invert \( I - \frac{1}{\lambda_l} H \) and use the assumption (5) to reach

\[
u_l = \left( I - \frac{1}{\lambda_l} H \right)^{-1} \frac{1}{\lambda_l} M^* u_l
\]
and hence
\[ \sum \text{rank-1 case reads} \]
\[ = \frac{1}{\lambda_i} \left( I - \frac{1}{\lambda_i} H \right)^{-1} \left( \sum_{j=1}^{r} \lambda_j^* u_j^* u_j^{\top} \right) u_i \]
\[ = \sum_{j=1}^{r} \frac{\lambda_j^*}{\lambda_i} (u_j^* u_i) \left( I - \frac{1}{\lambda_i} H \right)^{-1} u_j^* , \]
where the last line follows by rearranging terms. Finally, replacing \((I - \frac{1}{\lambda} H)^{-1}\) with the Neumann series \(\sum_{s=0}^{\infty} \frac{1}{\lambda^s} H^s\), we establish the theorem.

### A.3 Proof of Lemma 3

1. We start with the rank-1 case. Towards this, we resort to the Neumann trick in Theorem 2, which in the rank-1 case reads
\[ u = \frac{1}{\lambda} \left( u^\top u \right) \sum_{s=0}^{\infty} \left( \frac{1}{\lambda} H \right)^s u^* . \]  
(55)
From Lemma 2, we know that \(\lambda\) is real-valued and that \(\lambda > 1 - \|H\| \geq 3/4 \geq \|H\|\) under our assumption. This together with (55) yields
\[ \left\| u - \frac{u^\top u}{\lambda} u^* \right\|_2 \leq \frac{1}{\lambda} \sum_{s=0}^{\infty} \left\| \frac{1}{\lambda} H \right\|^s \left\| u^* \right\|_2 = \frac{1}{\lambda} \sum_{s=0}^{\infty} \left\| \frac{1}{\lambda} H \right\|^s \geq \frac{\|H\|}{\lambda (\lambda - \|H\|)} \leq \frac{8}{3} \|H\| , \]  
where the last inequality holds since \(\lambda \geq 3/4\) and \(\lambda - \|H\| \geq 1 - 2\|H\| \geq 1/2\).

Next, by decomposing \(u\) into two orthogonal components \(u = (u^\top u) u^* + (u - (u^\top u) u^*)\), we obtain
\[ |u^\top u| = \left\| (u^\top u) u^* \right\|_2 = \sqrt{1 - \left\| u - (u^\top u) u^* \right\|_2^2} \]
\[ \geq 1 - \left\| u - (u^\top u) u^* \right\|_2 \]
\[ \geq 1 - \left\| u - \frac{u^\top u}{\lambda} u^* \right\|_2 \]
\[ \geq 1 - \frac{64}{9} \|H\|^2 . \]  
(57)
(58)
The inequality (57) holds since \((u^\top u) u^*\) is orthogonal projection of \(u\) onto the subspace spanned by \(u^*\), and hence \(\|u - (u^\top u) u^*\|_2 \leq \|u - \frac{1}{\lambda} (u^\top u) u^*\|_2\).

Finally, (58) together with the fact that \(u\) is real-valued (cf. Lemma 2) leads to the advertised bound:
\[ \min\{\|u - u^*\|_2, \|u + u^*\|_2\} = \sqrt{\|u\|_2^2 + \|u^*\|_2^2 - 2|u^\top u|} \leq \frac{8\sqrt{2}}{3} \|H\| . \]

2. For the rank-\(r\) case, it is seen that for any \(1 \leq l \leq r\),
\[ \sum_{j=1}^{r} |u_j^* u_l|^2 = \left\| \sum_{j=1}^{r} (u_j^\top u_l) u_j^* \right\|_2^2 = 1 - \left\| u_l - \sum_{j=1}^{r} (u_j^\top u_l) u_j^* \right\|_2^2 \]
\[ \geq 1 - \left\| u_l - \sum_{j=1}^{r} \frac{\lambda_j}{\lambda_1} u_j^* u_l \right\|_2^2 \]  
(59)
where the inequality arises since \(\sum_{j=1}^{r} (u_j^\top u_l) u_j^*\) is the Euclidean projection of \(u_l\) onto the span of \(\{u_1^*, \cdots, u_r^*\}\). In addition, observe that
\[ \left\| \sum_{j=1}^{r} \lambda_j (u_j^* u_l) u_j^* \right\|_2 \leq \sqrt{\sum_{j} (\lambda_j)^2 |u_j^* u_l|^2} \leq \lambda_{\max} \sqrt{\sum_{j} |u_j^* u_l|^2} = \lambda_{\max} \|u_l\|_2 = 1. \]
over the vertex set.

Throughout this section, we will use

We shall often think of this sum-product graphically by viewing each index pair

To establish this lemma, we exploit entrywise independence of

as claimed.

The last two lines follow since, when \( \|H\| < 1/(4\kappa) \), one can invoke Lemma 2 to show (i) \( |\lambda_t| > |\lambda_{\min}| - \|H\| \geq 3/(4\kappa) \) and (ii) \( |\lambda_t| - \|H\| \geq |\lambda_{\min}| - 2\|H\| \geq 1/(2\kappa) \). Substitution into (59) yields

\[
(59) \geq 1 - \frac{64\kappa^2}{9} \|H\|^2
\]
as claimed.

**B Proof of Lemma 4**

To establish this lemma, we exploit entrywise independence of \( H \) and develop a combinatorial trick.

To begin with, we expand the quantity of interest as

\[
(a^\top H^* u^*)^k = \sum_{1 \leq t, 0 \leq i, 1 \leq b \leq k} \prod_{t=1}^{k} a_{i_{t-1}^{(b)}} \left( \prod_{t=1}^{s} H_{i_{t-1}^{(b)}, i_{t}^{(b)}} \right) u_{i_{t}^{(b)}}.
\]

Throughout this section, we will use

\[
\mathcal{I} := \left\{ i_{t}^{(b)} \mid 0 \leq t \leq s, 1 \leq b \leq k \right\} \in [n]^{(s+1)k}
\]
to denote such a collection of \((s+1)k\) indices. Thus, one can write

\[
E \left[ (a^\top H^* u^*)^k \right] = \sum_{\mathcal{I} \in [n]^{(s+1)k}} E \left[ \prod_{t=1}^{k} a_{i_{t-1}^{(b)}} \left( \prod_{t=1}^{s} H_{i_{t-1}^{(b)}, i_{t}^{(b)}} \right) u_{i_{t}^{(b)}} \right].
\]

We shall often think of this sum-product graphically by viewing each index pair \((i_{t-1}^{(b)}, i_t^{(b)})\) as a directed edge over the vertex set \([n]^{(s+1)k}\). As a result, this gives us a set of \(sk\) edges in total \(\{e_{t}^{(b)} \mid 1 \leq t \leq s, 1 \leq b \leq k\}\), where \(e_{t}^{(b)}\) represents \((i_{t-1}^{(b)}, i_t^{(b)})\). See Fig. 8 for an illustration. In what follows, two edges \(e_{t_1}^{(b_1)}\) and \(e_{t_2}^{(b_2)}\) are said to be equivalent in \(\mathcal{I}\), denoted by

\[
e_{t_1}^{(b_1)} = e_{t_2}^{(b_2)},
\]

if the values of the corresponding vertices are identical (namely, \(i_{t_1-1}^{(b_1)} = i_{t_2}^{(b_2)}\) and \(i_{t_1}^{(b_1)} = i_{t_2}^{(b_2)}\)).

When \((s+1)k \ll n\), most of the summands in the above expansion vanish. In fact, for any summand associated with a given \(\mathcal{I}\): as long as there exists a distinct edge \((i_{t-1}^{(b*)}, i_t^{(b*)})\) (i.e. not equal to any other edge associated with \(\mathcal{I}\)), then the contribution of this summand is zero, namely,

\[
E \left[ \prod_{b=1}^{k} a_{i_{t-1}^{(b)}} \left( \prod_{t=1}^{s} H_{i_{t-1}^{(b)}, i_{t}^{(b)}} \right) u_{i_{t}^{(b)}} \right] = \left( \prod_{b=1}^{k} a_{i_{t-1}^{(b)} u_{i_{t}^{(b)}}} \right) E \left[ H_{i_{t-1}^{(b*)}, i_{t}^{(b*)}} \right] E \left[ \prod_{(t,b) \in [s] \times [k], (t,b) \neq (t*,b*)} H_{i_{t-1}^{(b)}, i_{t}^{(b)}} \right]
\]
For any fixed unit vector $e$, non-empty subsets in the partition associated with $\{e_t \mid 1 \leq t \leq s, 1 \leq b \leq k\}$ when $s = 4$ and $k = 2$. (Right) An illustration of a type (or a partition of $sk$ edges), where the edges of the same color belong to the same subset of the partition.

\[
= 0. \tag{61}
\]

As a result, for any term with non-zero contribution, every edge $(i_{t-1}, i_t)$ must appear at least twice.

To enable simple yet effective upper bounds on the non-vanishing terms, we group those terms of the same “type” and look at each “type” separately. Specifically:

- We represent a collection of edges $\{(i_{t-1}, i_t) \mid 1 \leq t \leq s, 1 \leq b \leq k\}$ as $\{e_t \mid 1 \leq t \leq s, 1 \leq b \leq k\}$. See Fig. 8 for an illustration.

- Each type encodes a set of constraints across edges, namely, which edges correspond to the same pair of vertices. To be precise, we define each type to be a partition of $\{e_t \mid 1 \leq t \leq s, 1 \leq b \leq k\}$ into a disjoint union of subsets, so that all edges falling within the same subset are equivalent. For instance, when $s = k = 2$, one possible type is $\{\{e_1^{(1)}, e_2^{(2)}\}, \{e_2^{(1)}, e_1^{(2)}\}\}$, which encodes the constraints $e_1^{(1)} = e_2^{(2)}$ and $e_2^{(1)} = e_1^{(2)}$.

- For each index collection $\mathcal{I}$ (defined in (60)), we write $\text{type}(\mathcal{I}) = \mathcal{T}$ if the associated edge set of $\mathcal{I}$ satisfies the constraints encoded by a type $\mathcal{T}$.

- Define the set $\Gamma^+$ as

$$
\Gamma^+ : \text{the set of all types s.t. each subset of the associated partition has cardinality at least 2. \tag{62}}
$$

With this grouping strategy and (61) in mind, we can derive

\[
\left| \mathbb{E} \left[ (a^\top H u^* )^k \right] \right| = \sum_{\mathcal{I} \in [n]^{s+1}k} \mathbb{E} \left[ \prod_{b=1}^k a_{i_b} \left( \prod_{t=1}^s H_{i_{t-1}, i_t}^{(b)} \right) u_{i_b}^* \right] = \sum_{\mathcal{T} \in \Gamma^+} \sum_{\mathcal{I} \in [n]^{(s+1)k} : \text{type}(\mathcal{I}) = \mathcal{T}} \mathbb{E} \left[ \prod_{b=1}^k a_{i_b} \left( \prod_{t=1}^s H_{i_{t-1}, i_t}^{(b)} \right) u_{i_b}^* \right], \tag{63}
\]

where the last line follows since all types outside $\Gamma^+$ have vanishing contributions (cf. (61)). To bound the right-hand side of (63), we need the following lemma. Here and throughout, $|\mathcal{T}|$ represents the number of non-empty subsets in the partition associated with $\mathcal{T}$.

**Lemma 5.** For any fixed unit vector $a \in \mathbb{R}^{n \times 1}$ and any type $\mathcal{T} \in \Gamma^+$, one has

\[
\left| \sum_{\mathcal{I} \in [n]^{(s+1)k} : \text{type}(\mathcal{I}) = \mathcal{T}} \mathbb{E} \left[ \prod_{b=1}^k a_{i_b} \left( \prod_{t=1}^s H_{i_{t-1}, i_t}^{(b)} \right) u_{i_b}^* \right] \right| \leq B^k \left( \frac{\sigma^2}{B^2} \right)^{|\mathcal{T}|} \left( \sqrt{\frac{\mu}{n}} \right)^k n^{|\mathcal{T}|}. \tag{64}
\]
Armed with this lemma, we can further obtain
\[
(63) \leq \sum_{T \in \Gamma^+} B^{sk} \left( \frac{\sigma^2}{B^2} \right)^{|T|} \left( \sqrt{\frac{\mu}{n}} \right)^k = \sum_{T \in \Gamma^+} B^{sk-2|T|} \left( \frac{\mu}{n} \right)^k
\leq \sum_{T \in \Gamma^+} \sum_{l=1}^{sk/2} \mathbb{1}_{|T|=l} B^{sk-2|T|} \left( \frac{n\sigma^2}{2} \right)^{l} \left( \sqrt{\frac{\mu}{n}} \right)^k
= \sum_{l=1}^{sk/2} \left\{ \sum_{T \in \Gamma^+} \mathbb{1}_{|T|=l} \right\} B^{sk-2l} \left( \frac{n\sigma^2}{2} \right)^{l} \left( \sqrt{\frac{\mu}{n}} \right)^k,
\]
where we have grouped the types based on their cardinality. The last inequality results from $|T| \leq sk/2$ in Lemma 8. The following lemma bounds the number of distinct types having the same cardinality.

Lemma 6. \( \sum_{T \in \Gamma^+} \mathbb{1}_{|T|=l} \leq (\frac{sk}{l}) l^{sk-l} \leq 2^{sk} l^{sk-l} \).

Proof. The quantity of interest is the number of ways to partition \( sk \) edges into \( l \) disjoint subsets, where each subset contains at least 2 edges. To bound this quantity, we first pick \( l \) edges and assign each of them to a distinct subset; there are \( \binom{sk}{l} \) different ways to achieve it. We still have \( sk-l \) edges left, and the number of ways to assign them to \( l \) subsets is clearly upper bounded by \( l^{sk-l} \). This concludes the proof.

Consequently, Lemma 6 yields
\[
(65) \leq 2^{sk} \sum_{l=1}^{sk/2} (Bl)^{sk-2l} \left( \frac{n\sigma^2}{2} \right)^{l} \left( \sqrt{\frac{\mu}{n}} \right)^k
\leq (i) 2^{sk} \sum_{l=1}^{sk/2} \left( \frac{Bsk}{2} \right)^{sk-2l} \left( \frac{n\sigma^2}{2} \right)^{sk/2} \left( \sqrt{\frac{\mu}{n}} \right)^k
\leq (ii) 2^{sk} \sum_{l=1}^{sk/2} \max \left\{ \left( \frac{Bsk}{2} \right)^{sk}, \left( \frac{n\sigma^2}{2} \right)^{sk/2} \right\} \left( \sqrt{\frac{\mu}{n}} \right)^k
= \frac{sk}{2} \max \left\{ (Bsk)^{sk}, (2n\sigma^2 sk)^{sk/2} \right\} \left( \sqrt{\frac{\mu}{n}} \right)^k,
\]
where (i) uses the condition \( l \leq sk/2 \), and (ii) relies on the fact that \( a^{sk-2l}b^l = a^{sk-2l}(\sqrt{b})^{2l} \leq \max \{a^{sk}, (\sqrt{b})^{sk} \} \) for any \( a, b > 0 \).

Taken collectively, the preceding bounds conclude the proof of Lemma 4, provided that Lemma 5 can be established.

C Proof of Lemma 5

To begin with,
\[
\left| \mathbb{E} \left[ \prod_{b=1}^{k} a_{i_b} \left( \prod_{t=1}^{s} H_{i-t_{i_t}^{(b)}} \right) u_{i_{s_b}}^{*} \right] \right| \leq \mathbb{E} \left[ \prod_{b=1}^{k} a_{i_b} \left( \prod_{t=1}^{s} H_{i_{t_{i_t}^{(b)}}} \right) |u_{i_{s_b}}^{*}| \right]
\leq \left( \sqrt{\frac{\mu}{n}} \right)^k \left( \prod_{b=1}^{k} a_{i_b} \right) \mathbb{E} \left[ \prod_{b=1}^{k} \prod_{t=1}^{s} H_{i_{t_{i_t}^{(b)}}} \right],
\]
where the last inequality holds since each entry of \( u^{*} \) is bounded in magnitude by \( \sqrt{\mu/n} \) (see Definition 1).

According to the definition of \( \Gamma^+ \) (see (62)), for any type \( T \in \Gamma^+ \), each edge \( (i_{t_{i_t}^{(b)}}, i_{t_{i_t}^{(b)}}) \) is repeated at least twice. The total number of distinct edges is exactly \( |T| \). For notational simplicity, suppose the distinct
edges are $e_1, \ldots, e_{|T|}$, where $e_i$ is repeated by $l_i \geq 2$ times. Then we can write
\[
E \left[ \prod_{b=1}^k \prod_{t=1}^s |H_{i_t^{(b)}}| \right] = \prod_{i=1}^{|T|} E \left[ |H_{e_i}| \right] \leq \prod_{i=1}^{|T|} \sigma^2 B^{l_i-2} = \left( \frac{\sigma^2}{B^2} \right)^{|T|} B \sum_{l_i} = \left( \frac{\sigma^2}{B^2} \right)^{|T|} B^{|k|},
\]
where the inequality follows since
\[
E \left[ |H_{ij}| \right] \leq B^{l_i-2} E \left[ |H_{ij}|^2 \right] \leq \sigma^2 B^{l_i-2}, \quad l \geq 2.
\]
Substitution into (66) yields
\[
\left| \sum_{I \in [n]^{(s+1)k} : \text{type}(I) = T} E \left[ \prod_{b=1}^k a_{i_t^{(b)}} \left( \prod_{t=1}^s H_{i_t^{(b)}} \right) u_{i_t^{(b)}}^\ast \right] \right| \leq \left( \frac{\mu}{n} \right)^k \left( \frac{\sigma^2}{B^2} \right)^{|T|} B^{|k|} \sum_{I \in [n]^{(s+1)k} : \text{type}(I) = T} \prod_{b=1}^k |a_{i_t^{(b)}}|,
\]
(67)
It then boils down to controlling $\sum_{I \in [n]^{(s+1)k} : \text{type}(I) = T} \prod_{b=1}^k |a_{i_t^{(b)}}|$. This is achieved by the following lemma, which we establish in Appendix D.

**Lemma 7.** For any type $T \in \Gamma^+$ and any unit vector $a \in \mathbb{R}^n$, one has
\[
\sum_{I \in [n]^{(s+1)k} : \text{type}(I) = T} \prod_{b=1}^k |a_{i_t^{(b)}}| \leq n^{|T|}.
\]
(68)
This lemma together with (67) concludes the proof.

### D Proof of Lemma 7

#### D.1 Graphical tools

The proof of Lemma 7 is combinatorial in nature. Before proceeding, we introduce a few graphical notions that will be useful.

To begin with, recall that the index collection we have used so far is
\[
I := \left\{ i_t^{(b)} : 0 \leq t \leq s, 1 \leq b \leq k \right\} \subset [n]^{(s+1)k},
\]
(69)
which involves $(s+1)k$ vertices. As it turns out, it is helpful to further augment it into $2(s+1)k$ vertices via simple duplication. Specifically, introduce the following set of $2(s+1)k$ vertices
\[
\mathcal{V} = \left\{ q_t^{(b)} : 0 \leq t \leq s, 1 \leq b \leq k \right\} \cup \left\{ p_t^{(b)} : 1 \leq t \leq s + 1, 1 \leq b \leq k \right\}.
\]
(70)
Throughout this paper, we shall
- view $\{q_t^{(b)} : 0 \leq t \leq s, 1 \leq b \leq k\}$ as a copy of $\{i_t^{(b)} : 0 \leq t \leq s, 1 \leq b \leq k\}$;
- view $\{p_t^{(b)} : 1 \leq t \leq s + 1, 1 \leq b \leq k\}$ as another copy of $\{i_t^{(b)} : 1 \leq t \leq s + 1, 1 \leq b \leq k\}$.

The main incentive is that it allows us to reparametrize
\[
\prod_{b=1}^k a_{q_t^{(b)}} \left( \prod_{t=1}^s H_{i_t^{(b)}} \right) u_{i_t^{(b)}}^\ast \rightarrow \prod_{b=1}^k a_{q_t^{(b)}} \left( \prod_{t=1}^s H_{i_t^{(b)}} \right) u_{i_t^{(b)}}^\ast \prod_{b=1}^k a_{p_t^{(b)}} \left( \prod_{t=1}^s H_{i_t^{(b)}} \right) u_{i_t^{(b)}}^\ast
\]
(71)
Recall that we have categorized $I$ into different “types”. Given that $\mathcal{V}$ is simply a redundant representation of $I$, we can also associate a type with each $\mathcal{V}$ by enforcing proper constraints. These constraints can be encoded via the following graphical notion.
Figure 9: An example of the induced graph, where \( s = 4, k = 2 \). The left plot shows the partition, while the right plot displays the induced graph and the vertex set \( Q_0 \) (the yellow region).

**Definition 2 (Induced graph).** For any given type \( T \), we define the induced graph \( G(T) \) with the vertex set \( V \) and the (undirected) edge set \( E(T) \) as \( E(T) = E_1 \cup E_2(T) \), where

\[
E_1 \triangleq \left\{ (p_t^{(b)}, q_{t-1}^{(b)}) \mid 1 \leq t \leq s + 1, 1 \leq b \leq k \right\};
\]

\[
E_2(T) \triangleq \left\{ (p_{t_1}^{(b_1)}, p_{t_2}^{(b_2)}), (q_{t_1}^{(b_1)}, q_{t_2}^{(b_2)}) \mid \forall t_1, b_1, t_2, b_2 : (i_{t_1-1}^{(b_1)}, i_{t_1}^{(b_1)}) = (i_{t_2-1}^{(b_2)}, i_{t_2}^{(b_2)}) \right\}.
\]

(72)

In words,

1. we connect the vertices \( p_t^{(b)} \) (which is a copy of \( i_{t}^{(b)} \)) and \( q_{t-1}^{(b)} \) (which represents \( i_{t-1}^{(r)} \)) by an edge;

2. whenever two edges \( (i_{t_1-1}^{(b_1)}, i_{t_1}^{(b_1)}) \) and \( (i_{t_2-1}^{(b_2)}, i_{t_2}^{(b_2)}) \) are equivalent in \( I \) (or \( T \)), we draw two edges connecting the corresponding vertices in the induced graph.

As an illustration, Fig. 9 displays the induced graph for a simple example with \( s = 4 \) and \( k = 2 \), where

\[
T = \left\{ \{(i_0^{(1)}, i_1^{(1)}), (i_1^{(1)}, i_2^{(1)}), (i_1^{(1)}, i_3^{(1)}), (i_3^{(1)}, i_4^{(1)})\}, \{(i_1^{(2)}, i_2^{(2)}), (i_2^{(2)}, i_3^{(2)}), (i_2^{(2)}, i_3^{(2)})\}, \{(i_3^{(2)}, i_4^{(2)})\}, \{(i_0^{(2)}, i_1^{(2)}), (i_1^{(2)}, i_2^{(2)}), \ldots \right\}.
\]

Here, the 45-degree lines correspond to the edges in \( E_1 \), and the remaining edges come from \( E_2(T) \). Throughout the rest of the paper, we will abuse the notation and write type(\( V \)) = \( T \) if \( V \) is induced by an index set of type \( T \).

One useful feature of the above induced graph is that: each edge subset in the partition associated with \( T \) corresponds to a connected component in \( G(T) \). For our purpose, it is convenient to divide all connected components of \( G(T) \) into two classes. To this end, we first define \( Q_0 \triangleq \{ q_0^{(b)} \mid 1 \leq b \leq k \} \) (as illustrated in Fig. 9).

- **Class 1**: a connected component \( C \) belongs to Class 1 if \( |C \cap Q_0| \leq 1 \);

- **Class 2**: a connected component \( C \) belongs to Class 2 if \( |C \cap Q_0| \geq 2 \).

We denote by \( m_1(T) \) (resp. \( m_2(T) \)) the total number of Class 1 (resp. 2) connected components in \( G(T) \).

**D.2 Proof of Lemma 7**

Making use of the augmented set \( V \) (defined in (70)) and the way of reparametrization (71), we can write

\[
\sum_{I \in [n]^{s+1}: \text{type}(I) = T} \mathbb{E} \left[ \prod_{b=1}^{k} a_{i_b^{(b)}} \right] \leq \sum_{V \in [n]^{2s+2}: \text{type}(V) = T} \mathbb{E} \left[ \prod_{b=1}^{k} a_{q_b^{(b)}} \right].
\]

(73)
As we shall see, the benefit of this bound is to allow us sum over the connected components of $G(T)$ in a separate manner, since all indices in the same connected component must be identical. Specifically,

- Let $X_1, X_2, \cdots$ (resp. $Y_1, Y_2, \cdots$) denote the collection of Class 1 (resp. Class 2) connected components in $G(T)$;
- Denote by $x_i$ (resp. $y_i$) the value assigned to all indices in $X_i$ (resp. $Y_i$);

With these notations in mind, we can decompose

$$
\frac{\sum_{s_1=1}^{n} \sum_{s_2=1}^{n} \cdots \sum_{s_k=1}^{n} E \left[ \sum_{x_1=1}^{n} \left| X_1 \cap Q_1 \right| \cdot \sum_{x_2=1}^{n} \left| X_2 \cap Q_2 \right| \cdots \sum_{y_1=1}^{n} \left| Y_1 \cap Q_1 \right| \cdot \sum_{y_2=1}^{n} \left| Y_2 \cap Q_2 \right| \cdots \right]}
$$

where $m_1(T)$ is the total number of Class 1 connected components in the induced graph $G(T)$. Here, (i) comes from the definitions of Class 1 and Class 2 connected components, and (ii) follows since $\|a\|_2^2 = 1$.

We can thus finish the proof by observing that $m_1(T) \leq |T|$, as claimed in the following lemma.

**Lemma 8.** For any type $T \in \Gamma^+$, one has $m_1(T) \leq |T| \leq sk/2$.

### D.3 Proof of Lemma 8

Introduce two disjoint vertex sets:

$$Q_{\lambda_0} \triangleq \{ q^{(b)}_t \mid 1 \leq t \leq s, 1 \leq b \leq k \} \subseteq V;$$

$$Q_0 \triangleq \{ q^{(b)}_0 \mid 1 \leq b \leq k \} \subseteq V. \tag{74}$$

Clearly, one can define a partition of $Q_{\lambda_0} —$ denoted by $T_{Q_{\lambda_0}}$ — induced by $T$. Specifically, we say that $q^{(b_1)}_{t_1}$ and $q^{(b_2)}_{t_2}$ belong to the same connected subgraph of $Q_{\lambda_0}$ if $(t_1 - 1, b_1) = (t_2 - 1, b_2)$ in $T$. Clearly, $|\Gamma| = |T_{Q_{\lambda_0}}|$.

In addition, for any connected subgraph $C_q$ in $Q_{\lambda_0}$, we denote by $C$ the corresponding connected component in $G(T)$. We can thus define a mapping $\psi$ that maps $C_q$ to $C$, whose domain is the set of all connected subgraphs in $Q_{\lambda_0}$. We claim that the collection of Class 1 connected components — denoted by $\{X_j\} —$ obeys

$$\{X_j \mid 1 \leq j \leq m_1(T)\} \subseteq \text{Im}\psi, \tag{75}$$

where $\text{Im}\psi$ denotes the image of $\psi$, and $m_1(T)$ is the total number of Class 1 connected components. If this holds true, then

$$|\Gamma| = |T_{Q_{\lambda_0}}| \geq |\text{Im}\psi| \geq |\{X_j \mid 1 \leq j \leq m_1(T)\}| = m_1(T).$$

To justify the claim (76), it suffices to show that $X_j \cap Q_{\lambda_0} \neq \emptyset$. Given that $T \in \Gamma^+$ (so that each subset of the partition associated with $T$ has cardinality at least 2) and that $X_j$ is defined over the induced graph $G(T)$ (which is a redundant representation of the original index collection), one must have

$$\max \left\{ \left| X_j \cap \left\{ p^{(b)}_t \mid 1 \leq t \leq s, 1 \leq b \leq k \right\} \right|, \left| X_j \cap \left\{ q^{(b)}_t \mid 0 \leq t \leq s, 1 \leq b \leq k \right\} \right| \right\} \geq 2. \tag{77}$$

Further, from the construction of the induced graph, it is easily seen that

$$|X_j \cap \left\{ p^{(b)}_t \mid 1 \leq t \leq s, 1 \leq b \leq k \right\}| = |X_j \cap \left\{ q^{(b)}_t \mid 0 \leq t \leq s, 1 \leq b \leq k \right\}|,$$
thus resulting in
\[ |X_j \cap \left\{ d_t^{(b)} \mid 0 \leq t \leq s, 1 \leq b \leq k \right\} | = |X_j \cap (Q_0 \cup Q_0^c) | \geq 2. \]

However, by definition of Class 1 connected component, we have \( |X_j \cap Q_0 | \leq 1 \), implying that \( X_j \cap Q_0^c \neq \emptyset \). This establishes the claim (76).

Finally, when \( T \in \Gamma^+ \), each connected subgraph of \( T_{Q_0^c} \) also contains at least two nodes. As a consequence,
\[ |T| = |T_{Q_0^c}| \leq |Q_0^c|/2 = sk/2. \]

E Proof of Corollary 4

In the sequel, we assume that 20 log \( n \) is an integer to avoid the clumsy notation \( [20 \log n] \). But it is straightforward to extend it to the case where 20 log \( n \) is not an integer.

It follows from Markov’s inequality that for any even integer \( k \),
\[ \mathbb{P} \left( |a^\top H u^*| \geq \tau \right) \leq \mathbb{E} \left[ \left| a^\top H u^* \right|^k \right] / \tau^k. \]

This combined with Lemma 4 gives
\[ \mathbb{P} \left( |a^\top H u^*| \leq \left( c_2 \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \right)^s \sqrt{\mu/n} \right) \leq \frac{\mathbb{E} \left[ \left( a^\top H u^* \right)^k \right]}{\left( c_2 \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \right)^s \sqrt{\mu/n}} \leq \frac{\sqrt{sk}}{2} \max \left\{ \left( \frac{sk}{c_2 \log n} \right)^s, \left( \frac{2sk}{c_2^2 \log n} \right)^{sk/2} \right\}. \]

For any \( s \leq 20 \log n \), choose \( k \) such that \( sk \in [20 \log n, 40 \log n] \). It is straightforward to show — using the union bound — that with probability at least \( 1 - O(n^{-10}) \),
\[ |a^\top H u^*| \leq \left( c_2 \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \right)^s \sqrt{\mu/n}, \quad \forall s \leq 20 \log n, \]
as long as \( c_2 > 0 \) is some sufficiently large constant.

F Proof of Corollary 5

Taking \( a = u^*_i \) for some \( 1 \leq i \leq r \) in (33), we see that with high probability,
\[ \left| u_i^\top \left( u_i - \sum_{j=1}^r \frac{\lambda_j^* u_j^*}{\lambda_j} u_j^\top u_j^* \right) \right| \preceq \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \frac{1}{|\lambda_j|} \sqrt{\frac{\mu r}{n}}. \]

In addition, since \( u_1^*, \ldots, u_r^* \) are orthogonal to each other, we have that: for any \( 1 \leq i \leq r \),
\[ \left| u_i^\top \left( u_i - \sum_{j=1}^r \frac{\lambda_j^* u_j^*}{\lambda_j} u_j^\top u_j^* \right) \right| = \left| u_i^\top u_i - \frac{\lambda_i^* u_i^*}{\lambda_i} \right| = \left| u_i^\top u_i \right| \frac{\lambda_i - \lambda_i^*}{\lambda_i} \geq \left| u_i^\top u_i \right| \min_{1 \leq j \leq r} \left| \frac{\lambda_i - \lambda_j^*}{\lambda_i} \right|. \]

Therefore, combining the above two bounds yields
\[ \min_{1 \leq j \leq r} \left| \frac{\lambda_i - \lambda_j^*}{\lambda_i} \right| \max_{1 \leq i \leq r} \left| u_i^\top u_i \right| \preceq \max \left\{ B \log n, \sqrt{n \sigma^2 \log n} \right\} \frac{1}{|\lambda_j|} \sqrt{\frac{\mu r}{n}}. \tag{78} \]

Finally, it comes from Lemma 3 that if \( \|H\| \ll 1/k^2 \), then \( \sum_{1 \leq i \leq r} \left| u_i^\top u_i \right|^2 \geq 1 \), and hence
\[ \max_{1 \leq i \leq r} \left| u_i^\top u_i \right| \geq \left( \frac{1}{r} \sum_{1 \leq i \leq r} \left| u_i^\top u_i \right|^2 \right)^{1/2} \geq 1/\sqrt{r}. \tag{79} \]

This combined with (78) establishes the claim, as long as the spectral norm condition on \( \|H\| \) can be guaranteed. In view of Lemma 1, we have \( \|H\| \ll 1/k^2 \) under the condition (36), thus concluding the proof.
G  Proof of Corollary 6

Since $\|a\|_2 = 1$, it follows immediately from (34) that

$$
\|a^T U\|_2 = \sqrt{\sum_{l=1}^r |a^T u_l|^2} \lesssim \sqrt{\sum_{l=1}^r \left( \sum_{j=1}^r \frac{\lambda_j^* u_j^T u_l}{\lambda_l} \right)^2} + \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \kappa \sqrt{\frac{\mu}{n}}. 
$$

Further, we have

$$
(80) \lesssim \sum_{l=1}^r \left( \sum_{j=1}^r \frac{\lambda_j^* u_j^T u_l}{\lambda_l} \right)^2 + \sum_{l=1}^r \left( \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \kappa r \sqrt{\frac{\mu}{n}} \right)^2 
$$

$$
\lesssim \sum_{l=1}^r \left( \sum_{j=1}^r \left| a^T u_j^* \right|^2 \right) \left( \sum_{j=1}^r \left| a^T u_j^* \right|^2 \right) + \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \kappa r \sqrt{\frac{\mu}{n}} 
$$

$$
= \kappa \sqrt{\|a^T U^*\|_2} + \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \kappa r \sqrt{\frac{\mu}{n}}, 
$$

where (i) and (ii) make use of the Minkowski and the Cauchy-Schwarz inequalities, respectively, and (iii) result from the facts $|\lambda_j^*/\lambda_l| \lesssim \kappa$ (which follows from (37) and (36)) and $\sum_{j=1}^r \left| u_j^*\right|^2 \leq 1$. This concludes the proof.

H  Proof of Theorem 5

To simplify presentation, we introduce the following notation throughout this section:

$$
\bar{u}_1 = \frac{1}{\sqrt{2}} \left( \begin{array}{c} u^* \\ v^* \end{array} \right), \quad \bar{u}_2 = \frac{1}{\sqrt{2}} \left( \begin{array}{c} u^* \\ -v^* \end{array} \right), \quad \bar{u}_1 = u_1^{\text{dilation}}, \quad \bar{u}_2 = u_2^{\text{dilation}};
$$

$$
\bar{\lambda}_1^* = \lambda_1 (M_{\text{dilation}}^*), \quad \bar{\lambda}_2^* = \lambda_2 (M_{\text{dilation}}^*), \quad \bar{\lambda}_1 = \lambda_1^{\text{dilation}}, \quad \bar{\lambda}_2 = \lambda_2^{\text{dilation}}.
$$

In addition, we denote $\bar{u}_{1,1} = u_{1,1}^{\text{dilation}}$ and $\bar{u}_{1,2} = u_{1,2}^{\text{dilation}}$. Recall that we assume $\bar{\lambda}_1 \geq \bar{\lambda}_2$. We also denote $\min\{\|x \pm y\|_2\} = \min\{\|x - y\|_2, \|x + y\|_2\}$.

H.1 $\ell_2$ eigenvector perturbation bounds

The $\ell_2$ perturbation bound results from near orthogonality between $\bar{u}_1$ and $\bar{u}_2^*$. By symmetry, it suffices to establish the result for $u$.

To begin with, applying Theorem 4 on $M_{\text{dilation}}$ and taking $a = \bar{u}_2^*$, we derive

$$
\left| \bar{u}_2^* \bar{u}_1 \left( 1 - \frac{\bar{\lambda}_2^*}{\lambda_1} \right) \right| = \left| \bar{u}_2^T \left( \bar{u}_1 - \frac{\bar{\lambda}_1^* \bar{u}_1^*}{\lambda_1} \bar{u}_1 - \frac{\bar{\lambda}_2^* \bar{u}_2^*}{\lambda_1} \bar{u}_2 \right) \right| \lesssim \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}},
$$
where the identity arises since $\widetilde{u}_2^* \widetilde{u}_1^* = 0$. Given that $\widetilde{\lambda}_2 = -1$ and $\widetilde{\lambda}_1 > 0$ (see Corollary 8), we have $1 - \widetilde{\lambda}_2^2 / \widetilde{\lambda}_1 > 1$. The near orthogonality property can then be described as follows

$$\left| \widetilde{u}_2^* \widetilde{u}_1 \right| \leq \left| \widetilde{u}_2^* \widetilde{u}_1 (1 - \widetilde{\lambda}_2^2 / \widetilde{\lambda}_1) \right| \lesssim \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\} \sqrt{\frac{\mu}{n}}. \quad (82)$$

This combined with Lemma 3 yields

$$\min \{ \| \widetilde{u}_1 \pm \widetilde{u}_1^* \|_2 \} = \sqrt{\frac{\| \widetilde{u}_1 \|_2^2 + \| \widetilde{u}_1 \|_2^2 - 2 | \widetilde{u}_1^* \widetilde{u}_1 |}{2(1 - | \widetilde{u}_1^* \widetilde{u}_1 |)}} \leq \sqrt{\frac{2(1 - | \widetilde{u}_1^* \widetilde{u}_1 |)}{2(1 - | \widetilde{u}_1^* \widetilde{u}_1 |)}} \leq 2 \left(1 - | \widetilde{u}_1^* \widetilde{u}_1 | \right) \leq 2 \left(1 - | \widetilde{u}_1^* \widetilde{u}_1 | \right) \leq 2 \left(1 - \sqrt{\frac{\| H_{\text{dilation}} \|^2 + 2 | \widetilde{u}_2^* \widetilde{u}_1 |^2}{\| \widetilde{u}_2^* \widetilde{u}_1 \|}} \right) \leq \| H_{\text{dilation}} \| + | \widetilde{u}_2^* \widetilde{u}_1 | \leq \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}, \quad (83)$$

where (i) results from Lemma 3 and the fact that the condition number of $M_{\text{dilation}}^*$ is 1, (ii) follows since $\sqrt{x^2 + y^2} \leq x + y$ for all $x, y \geq 0$, and (iii) arises from (82), Lemma 1, the identity $\| H_{\text{dilation}} \| = \max \{ \| H_1 \|, \| H_2 \| \}$, and the fact $\sqrt{\mu/n} \leq 1$.

It then boils down to showing that $\min \{ \| u \pm u^* \|_2 \} \lesssim \min \{ \| \widetilde{u}_1 \pm \widetilde{u}_1^* \|_2 \}$. To this end, we see that the estimate (44) satisfies

$$\| u - u^* \|_2 = \left\| \left( \frac{\widetilde{u}_{1,1}}{\| \widetilde{u}_{1,1} \|_2} - \sqrt{2} \widetilde{u}_{1,1} \right) + \left( \sqrt{2} \widetilde{u}_{1,1} - u^* \right) \right\|_2 \leq \left\| \frac{\widetilde{u}_{1,1}}{\| \widetilde{u}_{1,1} \|_2} - \frac{1}{\sqrt{2}} \right\|_2 + \frac{1}{\sqrt{2}} \left\| \widetilde{u}_{1,1} - u^* \right\|_2 \leq \sqrt{2} \left( \frac{1}{\| \widetilde{u}_{1,1} \|_2} + 1 \right) \left\| \widetilde{u}_{1,1} - u^* \right\|_2 \leq \sqrt{2} \left( 1 + \sqrt{\frac{\| \widetilde{u}_{1,1} - \widetilde{u}_1^* \|_2}{\| \widetilde{u}_{1,1} - \widetilde{u}_1^* \|_2}} \right) \| \widetilde{u}_1 - \widetilde{u}_1^* \|_2. \quad (84)$$

Here, (84) makes use of the triangle inequality, and (85) follows since

$$\frac{1}{\| \widetilde{u}_{1,1} \|_2} \leq \| u^* \|_2 / \sqrt{2} - \| \widetilde{u}_{1,1} - u^* / \sqrt{2} \|_2 = \frac{\sqrt{2}}{1 - \| \sqrt{2} \widetilde{u}_{1,1} - u^* \|_2} \leq \frac{\sqrt{2}}{1 - \| \widetilde{u}_{1,1} - u^* \|_2},$$

where the last inequality relies on the fact that $\| \widetilde{u}_{1,1} - u^* / \sqrt{2} \|_2 \leq \| \widetilde{u}_1 - \widetilde{u}_1^* \|_2$ (the first $n_1$ coordinates). Similarly, one can derive the above bounds for $\| u + u^* \|_2$ as well. Therefore, we are left with

$$\min \{ \| u \pm u^* \|_2 \} \leq \sqrt{2} \left( 1 + \frac{\sqrt{2}}{1 - \sqrt{2} \min \{ \| \widetilde{u}_1 \pm \widetilde{u}_1^* \|_2 \}} \right) \min \{ \| \widetilde{u}_1 \pm \widetilde{u}_1^* \|_2 \} \leq \min \{ \| \widetilde{u}_1 \pm \widetilde{u}_1^* \|_2 \} \leq \max \left\{ B \log n, \sqrt{n\sigma^2 \log n} \right\}. \quad (86)$$

This concludes the proof.
H.2 Perturbation bounds for linear forms of eigenvectors

Fix any unit vector \( a \in \mathbb{R}^{n_1} \). Considering the linear transformation by the vector \((a, 0)\), we can apply Theorem 4 to \( M_{\text{dilation}} \) and get

\[
\left| \begin{pmatrix} a \\ 0 \end{pmatrix}^\top \left( \frac{\lambda_1 \tilde{u}_1^\top \tilde{u}_1}{\lambda_1} \tilde{u}_1 - \frac{\lambda_2 \tilde{u}_2^\top \tilde{u}_1}{\lambda_1} \tilde{u}_2^* \right) \right| \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2} \log n \right\} \sqrt{\frac{1}{n}}. \tag{89}
\]

Given that the first \( n_1 \) coordinates of \( \tilde{u}_1^* \) and \( \tilde{u}_2^* \) are both \( u^*/\sqrt{2} \), we can invoke Corollary 8 (i.e. \( \tilde{\lambda}_1 \approx 1 \)) to obtain

\[
\left| a^\top \left( \tilde{u}_{1,1} - \frac{\tilde{u}_1^\top \tilde{u}_1 u^*}{\sqrt{2} \lambda_1} \right) \right| \lesssim \left| \frac{\tilde{u}_2^\top \tilde{u}_1 u^*}{\sqrt{2} \lambda_1} \right| + \max \left\{ B \log n, \sqrt{n \sigma^2} \log n \right\} \sqrt{\frac{1}{n}}. \tag{90}
\]

In view of (82) and the fact \( |a^\top u^*| \leq 1 \), one has

\[
\left| a^\top \left( \tilde{u}_{1,1} - \frac{\tilde{u}_1^\top \tilde{u}_1 u^*}{\sqrt{2} \lambda_1} \right) \right| \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2} \log n \right\} \sqrt{\frac{1}{n}}. \tag{91}
\]

Recall that \( \tilde{\lambda}_1 > 0 \) (cf. Corollary 8). If we further have

\[
\left| \frac{\tilde{u}_1^\top \tilde{u}_1}{\sqrt{2} \lambda_1} - 1 \right| \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2} \log n \right\}, \tag{92}
\]

then we can use the triangle inequality to reach

\[
\min \{ |a^\top (u \pm u^*)| \} \leq \frac{1}{\|u_{1,1}\|_2} \left| a^\top \left( \tilde{u}_{1,1} - \frac{\tilde{u}_1^\top \tilde{u}_1 u^*}{\sqrt{2} \lambda_1} \right) \right| + \left| \frac{\tilde{u}_1^\top \tilde{u}_1}{\sqrt{2} \lambda_1} - 1 \right| |a^\top u^*| \lesssim \left( \sqrt{\frac{1}{n}} + |a^\top u^*| \right) \max \left\{ B \log n, \sqrt{n \sigma^2} \log n \right\} \tag{93}
\]

as claimed. It then remains to justify (92). Towards this, it suffices to combine a series of consequences from (83), Corollary 8, and (88), namely,

\[
1 - \|\tilde{u}_1^\top \tilde{u}_1\| = \frac{1}{2} \min \{ \|u \pm u^*\|_2^2 \} \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2} \log n \right\}, \tag{94}
\]

\[
|\tilde{\lambda}_1 - \lambda_1| = |\tilde{\lambda}_1 - 1| \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2} \log n \right\} \sqrt{\frac{1}{n}}, \tag{95}
\]

\[
\|\tilde{u}_{1,1}\|_2 - \frac{1}{\sqrt{2}} \leq \min \left\{ \|u_{1,1} \pm u^*\|_2^2 \right\} \lesssim \min \{ \|\tilde{u}_1 \pm \tilde{u}_1^*\|_2 \} \lesssim \max \left\{ B \log n, \sqrt{n \sigma^2} \log n \right\}. \tag{96}
\]

The proof for the bounds on \( v \) is similar and is thus omitted.

H.3 Entrywise eigenvector perturbation bounds

Recognizing that \( \|u - u^*\|_\infty = \max_i |e_i^\top u - e_i^\top u^*| \) and using the incoherence \( |e_i^\top u| \leq \sqrt{\mu/n} \), we can prove this claim directly by invoking the results established in Appendix H.2 and taking the union bound.

References


30


