The Annals of Statistics
2021, Vol. 49, No. 1, 435–458
https://doi.org/10.1214/20-AOS1963

ASYMMETRY HELPS: EIGENVALUE AND EIGENVECTOR ANALYSES OF ASYMMETRICALLY PERTURBED LOW-RANK MATRICES[1]

BY YUXIN CHEN[1] CHEN CHENG[2] AND JIANQING FAN[3]

[1]*Department of Electrical Engineering, Princeton University, yuxin.chen@princeton.edu*
[2]*Department of Statistics, Stanford University, chencheng@stanford.edu*
[3]*Department of Operations Research & Financial Engineering, Princeton University, jqfan@princeton.edu*

This paper is concerned with the interplay between statistical asymmetry and spectral methods. Suppose we are interested in estimating a rank-1 and symmetric matrix $\boldsymbol{M}^\star \in \mathbb{R}^{n\times n}$, yet only a randomly perturbed version $\boldsymbol{M}$ is observed. The noise matrix $\boldsymbol{M} - \boldsymbol{M}^\star$ is composed of independent (but not necessarily homoscedastic) entries and is, therefore, not symmetric in general. This might arise if, for example, when we have two independent samples for each entry of $\boldsymbol{M}^\star$ and arrange them in an *asymmetric* fashion. The aim is to estimate the leading eigenvalue and the leading eigenvector of $\boldsymbol{M}^\star$.

We demonstrate that the leading eigenvalue of the data matrix $\boldsymbol{M}$ can be $O(\sqrt{n})$ times more accurate (up to some log factor) than its (unadjusted) leading singular value of $\boldsymbol{M}$ in eigenvalue estimation. Moreover, the eigendecomposition approach is fully adaptive to heteroscedasticity of noise, without the need of any prior knowledge about the noise distributions. In a nutshell, this curious phenomenon arises since the statistical asymmetry automatically mitigates the bias of the eigenvalue approach, thus eliminating the need of careful bias correction. Additionally, we develop appealing nonasymptotic eigenvector perturbation bounds; in particular, we are able to bound the perturbation of any linear function of the leading eigenvector of $\boldsymbol{M}$ (e.g., entrywise eigenvector perturbation). We also provide partial theory for the more general rank-r case. The takeaway message is this: arranging the data samples in an asymmetric manner and performing eigendecomposition could sometimes be quite beneficial.

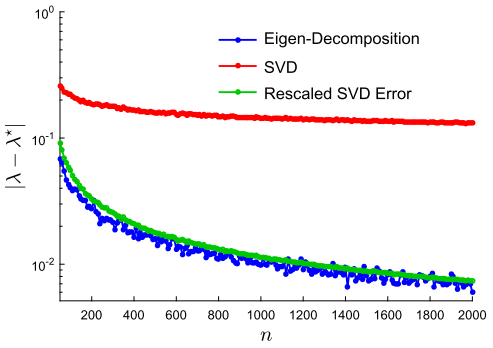
1. Introduction. Consider an unknown symmetric and low-rank matrix $\boldsymbol{M}^\star \in \mathbb{R}^{n\times n}$. What we have observed is a corrupted version

$$\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}, \tag{1}$$

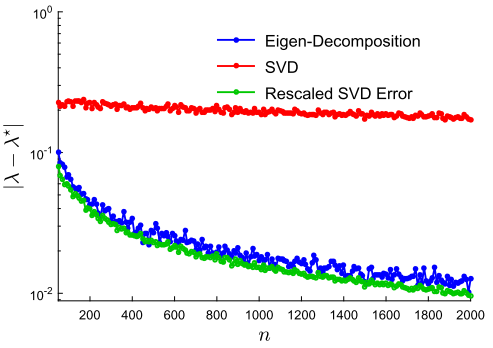
with $\boldsymbol{H}$ denoting a noise matrix. A classical problem is concerned with estimating the leading eigenvalues and eigenspace of $\boldsymbol{M}^\star$ given observation $\boldsymbol{M}$.

The current paper concentrates on a scenario where the noise matrix $\boldsymbol{H}$ (and hence $\boldsymbol{M}$) consists of independently generated random entries and is hence *asymmetric* in general. This might arise, for example, when we have available multiple (e.g., two) samples for each entry of $\boldsymbol{M}^\star$ 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 $\boldsymbol{M}$ to approximately estimate the eigenvalues (resp., eigenspace) of $\boldsymbol{M}^\star$. By contrast, a much less popular alternative is based on eigendecomposition of the asymmetric data matrix $\boldsymbol{M}$, which attempts approximation using the

Received July 2019.
[1]Author names are sorted alphabetically.
MSC2020 subject classifications. Primary 62H25; secondary 62H12.
Key words and phrases. Spectral methods, eigenvalue perturbation, entrywise eigenvector perturbation, linear form of eigenvectors, heteroscedasticity.



(a) eigenvalue perturbation for matrix estimation



(b) eigenvalue perturbation for matrix completion

FIG. 1. *Numerical error* $|\lambda - \lambda^\star|$ *versus the matrix dimension* n, *where* λ *is either the leading eigenvalue* (*the blue line*) *or the leading singular value* (*the red line*) *of* $\boldsymbol{M}$. *Here*, (*a*) *is the case when* $\{H_{ij}\}$ *are i.i.d.* $\mathcal{N}(0, \sigma^2)$ *with* $\sigma = 1/\sqrt{n \log n}$, *and* (*b*) *is the matrix completion case with sampling rate* $p = 3 \log n / n$, *where* $M_{i,j} = \frac{1}{p} M^\star_{i,j}$ *independently with probability* p *and* 0 *otherwise. The results are averaged over* 100 *independent trials. The green lines are obtained by rescaling the corresponding red lines by* $2.5/\sqrt{n}$.

leading eigenvalues and eigenspace of $\boldsymbol{M}$. Given that eigendecomposition 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 eigendecomposition.

When comparing these two approaches numerically, however, a curious phenomenon arises, which largely motivates the research in this paper. Let us generate $\boldsymbol{M}^\star$ as a random rank-1 matrix with leading eigenvalue $\lambda^\star = 1$, and let $\boldsymbol{H}$ be a Gaussian random matrix whose entries are i.i.d. $\mathcal{N}(0, \sigma^2)$ with $\sigma = 1/\sqrt{n \log n}$. Figure 1(a) compares the empirical accuracy of estimating the first eigenvalue of $\boldsymbol{M}^\star$ via the leading eigenvalue (the blue line) and via the leading singular value of $\boldsymbol{M}$ (the red line). As it turns out, eigendecomposition significantly outperforms vanilla 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 Figure 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 eigendecomposition compared to SVD. What is more, this phenomenon does not merely happen under i.i.d. noise. Similar numerical behaviors are observed in the problem of matrix completion—as displayed in Figure 1(b)—even though the components of the equivalent perturbation matrix are apparently far from identically distributed or homoscedastic.

The goal of the current paper is thus to develop a systematic understanding of this phenomenon, that is, why statistical asymmetry empowers eigendecomposition and how to exploit this feature in statistical estimation. Informally, our findings suggest that: when $\boldsymbol{M}^\star$ is rank-1 and $\boldsymbol{H}$ is composed of zero-mean and independent (but not necessarily identically distributed or homoscedastic) entries,

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

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.

[1]More precisely, this gain is possible when $\|\boldsymbol{H}\|$ is nearly as large as $\|\boldsymbol{M}^\star\|$ (up to some logarithmic factor).

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 eigendecomposition of a certain dilation matrix, which also outperforms the vanilla SVD approach.

We would like to immediately remark that: for some scenarios (e.g., the case with i.i.d. Gaussian noise), it is possible to adjust the leading singular value of $\boldsymbol{M}$ to obtain the same accuracy as the leading eigenvalue of $\boldsymbol{M}$. As it turns out, the advantages of the eigendecomposition approach may become more evident in the presence of heteroscedasticity—the case where the noise has location-varying and unknown variance. We shall elaborate on this point in Section 4.1.2.

All in all, when it comes to low-rank matrix estimation, arranging the observed matrix samples in an asymmetric manner and invoking eigendecomposition properly could sometimes be statistically beneficial.

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 $\boldsymbol{M}^\star = [M^\star_{ij}]_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$. Suppose we are given a random copy of $\boldsymbol{M}^\star$ as follows:

$$\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}, \tag{2}$$

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

The current paper concentrates on independent—but not necessarily identically distributed or homoscedastic—noise. Specifically, we impose the following assumptions on $\boldsymbol{H}$ throughout this paper.

ASSUMPTION 1.

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

REMARK 1 (Notational convention). In what follows, the dependency of σ_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 $\boldsymbol{H}$ and $\boldsymbol{M}$ are in general asymmetric matrices. Also, Condition 3 does not require the H_{ij}'s to have equal variance across different locations; in fact, they can be heteroscedastic. 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., subexponential distributions). An immediate consequence of Assumption 1 is the following bound on the spectral norm $\|\boldsymbol{H}\|$ of $\boldsymbol{H}$.

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

$$\|\boldsymbol{H}\| \le c_0 \sigma \sqrt{n \log n} + c_0 B \log n. \tag{3}$$

PROOF. This is a standard *nonasymptotic* 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. □

2.2. *Our goal*. The aim is to develop *nonasymptotic* eigenvalue and eigenvector perturbation bounds under this family of random and asymmetric noise matrices. Our theoretical development is divided into two parts. Below we introduce our goal as well as some notation used throughout.

Rank-1 symmetric case. For the rank-1 case, we assume the eigendecomposition of $\boldsymbol{M}^\star$ to be

$$\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} \tag{4}$$

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

1. *Eigenvalue perturbation*: $|\lambda - \lambda^\star|$;
2. *Perturbation of linear forms of eigenvectors*: $\min\{|\boldsymbol{a}^\top(\boldsymbol{u} - \boldsymbol{u}^\star)|, |\boldsymbol{a}^\top(\boldsymbol{u} + \boldsymbol{u}^\star)|\}$ for any fixed unit vector $\boldsymbol{a} \in \mathbb{R}^n$;
3. *Entrywise eigenvector perturbation*: $\min\{\|\boldsymbol{u} - \boldsymbol{u}^\star\|_\infty, \|\boldsymbol{u} + \boldsymbol{u}^\star\|_\infty\}$.

Rank-r symmetric case. For the general rank-r case, we let the eigendecomposition of $\boldsymbol{M}^\star$ be

$$\boldsymbol{M}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{U}^{\star\top}, \tag{5}$$

where the columns of $\boldsymbol{U}^\star = [\boldsymbol{u}_1^\star, \ldots, \boldsymbol{u}_r^\star] \in \mathbb{R}^{n\times r}$ are the eigenvectors, and $\boldsymbol{\Sigma}^\star = \mathsf{diag}(\lambda_1^\star, \ldots, \lambda_r^\star) \in \mathbb{R}^{r\times r}$ is a diagonal matrix with the eigenvalues arranged in descending order by their magnitude, that is, $|\lambda_1^\star| \geq \cdots \geq |\lambda_r^\star|$. We let $\lambda_{\max}^\star = |\lambda_1^\star|$ and $\lambda_{\min}^\star = |\lambda_r^\star|$. In addition, we let the top-r eigenvalues (in magnitude) of $\boldsymbol{M}$ be $\lambda_1, \ldots, \lambda_r$ (obeying $|\lambda_1| \geq \cdots \geq |\lambda_r|$) and their corresponding normalized eigenvectors be $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$. We will present partial eigenvalue perturbation results for this more general case, as detailed in Section 5.

As is well known, eigendecomposition can be applied to estimate the singular values and singular vectors of an asymmetric matrix $\boldsymbol{M}^\star$ 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.

2.3. *Incoherence conditions*. 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 $\boldsymbol{M}^\star$ with eigendecomposition $\boldsymbol{M}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{U}^{\star\top}$ is defined to be the smallest quantity μ obeying

$$\|\boldsymbol{U}^\star\|_\infty \leq \sqrt{\frac{\mu}{n}}, \tag{6}$$

where $\|\cdot\|_\infty$ denotes the entrywise ℓ_∞ norm.

REMARK 2. An alternative definition of the incoherence parameter (Candès and Recht (2009), Chen et al. (2019a), Chi, Lu and Chen (2019), Keshavan, Montanari and Oh (2010)) is the smallest quantity μ_0 satisfying $\|\boldsymbol{U}^\star\|_{2,\infty} \leq \sqrt{\mu_0 r/n}$. This is a weaker assumption than Definition 1, as it only requires the energy of $\boldsymbol{U}^\star$ 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 $\mu_0 \leq \mu \leq \mu_0 r$.

2.4. *Notation.* The standard basis vectors in $\mathbb{R}^n$ are denoted by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. For any vector $\boldsymbol{z}$, we let $\|\boldsymbol{z}\|_2$ and $\|\boldsymbol{z}\|_\infty$ denote the ℓ_2 norm and the ℓ_∞ norm of $\boldsymbol{z}$, respectively. For any matrix $\boldsymbol{M}$, denote by $\|\boldsymbol{M}\|$, $\|\boldsymbol{M}\|_{\mathrm{F}}$ and $\|\boldsymbol{M}\|_\infty$ the spectral norm, the Frobenius norm and the entrywise ℓ_∞ norm (the largest magnitude of all entries) of $\boldsymbol{M}$, respectively. Let $[n] := \{1, \ldots, n\}$. In addition, the notation $f(n) = O(g(n))$ or $f(n) \lesssim g(n)$ means that there is a constant $c > 0$ such that $|f(n)| \le c|g(n)|$, $f(n) \gtrsim g(n)$ means that there is a constant $c > 0$ such that $|f(n)| \ge c|g(n)|$, and $f(n) \asymp g(n)$ means that there exist constants $c_1, c_2 > 0$ such that $c_1|g(n)| \le |f(n)| \le c_2|g(n)|$.

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 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$, where $\boldsymbol{V} \in \mathbb{C}^{n\times n}$ is a nonsingular eigenvector matrix and $\boldsymbol{\Lambda}$ is diagonal. Let $\tilde{\lambda}$ be an eigenvalue of $\boldsymbol{A} + \boldsymbol{H}$. Then there exists an eigenvalue λ of $\boldsymbol{A}$ such that*

$$|\lambda - \tilde{\lambda}| \le \|\boldsymbol{V}\|\|\boldsymbol{V}^{-1}\|\|\boldsymbol{H}\|. \tag{7}$$

In addition, if $\boldsymbol{A}$ is symmetric, then there exists an eigenvalue λ of $\boldsymbol{A}$ such that

$$|\lambda - \tilde{\lambda}| \le \|\boldsymbol{H}\|. \tag{8}$$

However, caution needs to be exercised as the Bauer–Fike theorem does not specify which eigenvalue of $\boldsymbol{A}$ is close to an eigenvalue of $\boldsymbol{A} + \boldsymbol{H}$. 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 $\boldsymbol{M}^\star$ is a rank-r symmetric matrix whose top-r eigenvalues obey $|\lambda_1^\star| \ge \cdots \ge |\lambda_r^\star| > 0$. If $\|\boldsymbol{H}\| < |\lambda_r^\star|/2$, then the top-$r$ eigenvalues $\lambda_1, \ldots, \lambda_r$ of $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$, sorted by modulus, obey that: for any $1 \le l \le r$,*

$$|\lambda_l - \lambda_j^\star| \le \|\boldsymbol{H}\| \quad \text{for some } 1 \le j \le r. \tag{9}$$

In addition, if $r = 1$, then both the leading eigenvalue and the leading eigenvector of $\boldsymbol{M}$ are real-valued.

This result, which we establish in Appendix 1.1 deferred to Supplementary Material (Chen, Cheng and Fan (2020)), 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 $\boldsymbol{H}$ contains independent random entries (see, e.g., Corollary 1).

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

THEOREM 2 (Neumann trick). *Consider the matrices $\boldsymbol{M}^\star$ and $\boldsymbol{M}$ (see (5) and (2)). Suppose $\|\boldsymbol{H}\| < |\lambda_l|$ for some $1 \le l \le n$. Then*

$$\boldsymbol{u}_l = \sum_{j=1}^{r} \frac{\lambda_j^\star}{\lambda_l} (\boldsymbol{u}_j^{\star\top} \boldsymbol{u}_l) \left\{ \sum_{s=0}^{\infty} \frac{1}{\lambda_l^s} \boldsymbol{H}^s \boldsymbol{u}_j^\star \right\}. \tag{10}$$

PROOF. We supply the proof in Appendix 1.2 for self-containedness. □

REMARK 3. In particular, if $\boldsymbol{M}^\star$ is a rank-1 matrix and $\|\boldsymbol{H}\| < |\lambda_1|$, then

$$\boldsymbol{u}_1 = \frac{\lambda_1^\star}{\lambda_1} (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1) \left\{ \sum_{s=0}^{\infty} \frac{1}{\lambda_1^s} \boldsymbol{H}^s \boldsymbol{u}_1^\star \right\}. \tag{11}$$

An immediate consequence of the Neumann trick is the following lemma, which asserts that each of the top-r eigenvectors of $\boldsymbol{M}$ resides almost within the top-r eigensubspace of $\boldsymbol{M}^\star$, provided that $\|\boldsymbol{H}\|$ is sufficiently small. The proof is deferred to Appendix 1.3.

LEMMA 3. *Suppose $\boldsymbol{M}^\star$ is a rank-r symmetric matrix with r nonzero eigenvalues obeying $1 = \lambda_{\max}^\star = |\lambda_1^\star| \ge \cdots \ge |\lambda_r^\star| = \lambda_{\min}^\star > 0$ and associated eigenvectors $\boldsymbol{u}_1^\star, \ldots, \boldsymbol{u}_r^\star$. Define $\kappa \triangleq \lambda_{\max}^\star / \lambda_{\min}^\star$. If $\|\boldsymbol{H}\| \le 1/(4\kappa)$, then the top-$r$ eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ of $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$ obey*

$$\sum_{j=1}^{r} |\boldsymbol{u}_j^{\star\top} \boldsymbol{u}_l|^2 \ge 1 - \frac{64\kappa^4}{9} \|\boldsymbol{H}\|^2, \quad 1 \le l \le r. \tag{12}$$

In addition, if $r = 1$, then one further has

$$\min\{\|\boldsymbol{u}_1 - \boldsymbol{u}_1^\star\|_2, \|\boldsymbol{u}_1 + \boldsymbol{u}_1^\star\|_2\} \le \frac{8\sqrt{2}}{3} \|\boldsymbol{H}\|. \tag{13}$$

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 $\boldsymbol{M}^\star$ is a symmetric rank-1 matrix. We shall start by presenting a master bound which, as we will see, immediately leads to our main findings.

4.1.1. *A master bound.* Our master bound is concerned with the perturbation of linear forms of eigenvectors, as stated below.

THEOREM 3 (Perturbation of linear forms of eigenvectors (rank-1)). *Consider a rank-1 symmetric matrix $\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} \in \mathbb{R}^{n\times n}$ with incoherence parameter μ (cf. Definition 1). Suppose the noise matrix $\boldsymbol{H}$ obeys Assumption 1, and assume the existence of some sufficiently small constant $c_1 > 0$ such that*

$$\max\{\sigma\sqrt{n \log n}, B \log n\} \le c_1 |\lambda^\star|. \tag{14}$$

Then for any fixed vector $\boldsymbol{a} \in \mathbb{R}^n$ with $\|\boldsymbol{a}\|_2 = 1$, with probability at least $1 - O(n^{-10})$ one has

$$\left| \boldsymbol{a}^\top \left(\boldsymbol{u} - \frac{\boldsymbol{u}^{\star\top} \boldsymbol{u}}{\lambda/\lambda^\star} \boldsymbol{u}^\star \right) \right| \lesssim \frac{\max\{\sigma\sqrt{n \log n}, B \log n\}}{|\lambda^\star|} \sqrt{\frac{\mu}{n}}. \tag{15}$$

REMARK 4 (The noise size). We would like to remark on the range of the noise size covered by our theory. If the incoherence parameter of the truth $\boldsymbol{M}^\star$ obeys $\mu \asymp 1$, then even the magnitude of the largest entry of $\boldsymbol{M}^\star$ cannot exceed the order of $|\lambda^\star|/n$. One can thus interpret the condition (14) in this case as

$$\sigma \lesssim \sqrt{\frac{n}{\log n}}\|\boldsymbol{M}^\star\|_\infty \quad \text{and} \quad B \lesssim \frac{n}{\log n}\|\boldsymbol{M}^\star\|_\infty.$$

In other words, the standard deviation σ of each noise component is allowed to be substantially larger (i.e., $\sqrt{n/\log n}$ times larger) than the magnitude of any of the true entries. In fact, this condition (14) matches, up to some log factor, the one required for spectral methods to perform noticeably better than random guessing.

In words, Theorem 3 tells us that: the quantity $\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda/\lambda^\star}\boldsymbol{a}^\top\boldsymbol{u}^\star$ serves as a remarkably accurate approximation of the linear form $\boldsymbol{a}^\top\boldsymbol{u}$. In particular, the approximation error is at most $O(1/\sqrt{n})$ under the condition (14) for incoherent matrices. Encouragingly, this approximation accuracy holds true for an arbitrary deterministic direction (reflected by $\boldsymbol{a}$). As a consequence, one can roughly interpret Theorem 3 as

$$\boldsymbol{u} \approx \frac{\lambda^\star}{\lambda}\boldsymbol{u}^\star\boldsymbol{u}^{\star\top}\boldsymbol{u} = \frac{1}{\lambda}\boldsymbol{M}^\star\boldsymbol{u}, \tag{16}$$

where such an approximation is fairly accurate along any fixed direction. Compared with the identity $\boldsymbol{u} = \frac{1}{\lambda}\boldsymbol{M}\boldsymbol{u} = \frac{1}{\lambda}(\boldsymbol{M}^\star + \boldsymbol{H})\boldsymbol{u}$, our results imply that $\boldsymbol{H}\boldsymbol{u}$ is exceedingly small along any fixed direction, even though $\boldsymbol{H}$ and $\boldsymbol{u}$ are highly dependent. As we shall explain in Section 4.3, this observation usually cannot happen when $\boldsymbol{H}$ is a symmetric random matrix or when one uses the leading singular vector instead, due to the significant bias resulting from symmetry.

This master theorem has several interesting implications, as we shall elucidate momentarily.

4.1.2. *Eigenvalue perturbation*. To begin with, Theorem 3 immediately yields a much sharper nonasymptotic perturbation bound regarding the leading eigenvalue λ of $\boldsymbol{M}$.

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

$$|\lambda - \lambda^\star| \lesssim \max\{\sigma\sqrt{n\log n}, B\log n\}\sqrt{\frac{\mu}{n}}. \tag{17}$$

PROOF. Without loss of generality, assume that $\lambda^\star = 1$. Taking $\boldsymbol{a} = \boldsymbol{u}^\star$ in Theorem 3, we get

$$|\boldsymbol{u}^{\star\top}\boldsymbol{u}|\frac{|\lambda - 1|}{\lambda} = \left|\boldsymbol{u}^{\star\top}\boldsymbol{u} - \boldsymbol{u}^{\star\top}\boldsymbol{u}^\star\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda}\right| \lesssim \max\{\sigma\sqrt{n\log n}, B\log n\}\sqrt{\frac{\mu}{n}}. \tag{18}$$

From Lemma 1 and the condition (14), we know $\|\boldsymbol{H}\| < 1/4$, which combines with Lemma 2 and Lemma 3 yields $\lambda \asymp |\boldsymbol{u}^{\star\top}\boldsymbol{u}| \asymp 1$. Substitution into (18) yields

$$\begin{aligned}|\lambda - 1| &\lesssim \left|\frac{\lambda}{\boldsymbol{u}^{\star\top}\boldsymbol{u}}\right| \max\{\sigma\sqrt{n\log n}, B\log n\}\sqrt{\frac{\mu}{n}} \\ &\lesssim \max\{\sigma\sqrt{n\log n}, B\log n\}\sqrt{\frac{\mu}{n}}.\end{aligned}$$

□

For the vast majority of applications we encounter, the maximum possible noise magnitude B (cf. Assumption 1) obeys $B \lesssim \sigma\sqrt{n/\log n}$, in which case the bound in Corollary 1 simplifies to

$$|\lambda - \lambda^\star| \lesssim \sigma\sqrt{\mu \log n}. \tag{19}$$

This means that the eigenvalue estimation error is not much larger than the variability of each noise component. In addition, we remind the reader that for a fairly broad class of noise (see Remark 4), the leading eigenvalue λ of $\boldsymbol{M}$ is guaranteed to be real-valued, an observation that has been made in Lemma 2. In practice, however, one might still encounter some scenarios where λ is complex-valued. As a result, we recommend the practitioner to use the real part of λ as the eigenvalue estimate, which clearly enjoys the same statistical guarantee as in Corollary 1.

Comparison to the vanilla SVD-based approach. In order to facilitate comparison, we denote by λ_{svd} the largest singular value of $\boldsymbol{M}$, and look at $|\lambda_{\mathsf{svd}} - \lambda^\star|$. Combining Weyl's inequality, Lemma 1 and the condition (14), we arrive at

$$|\lambda_{\mathsf{svd}} - \lambda^\star| \le \|\boldsymbol{H}\| \lesssim \max\{\sigma\sqrt{n \log n}, B \log n\}. \tag{20}$$

When $\mu \asymp 1$, this error bound w.r.t. this (unadjusted) singular value could be $\sqrt{n}$ times larger than the perturbation bound (17) derived for the leading eigenvalue. This corroborates our motivating experiments in Figure 1.

Comparison to vanilla eigendecomposition after symmetrization. The reader might naturally wonder what would happen if we symmetrize the data matrix before performing eigendecomposition. Consider, for example, the i.i.d. Gaussian noise case where $H_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, and assume $\lambda^\star > 0$ for simplicity. The leading eigenvalue λ_{sym} of the symmetrized matrix $(\boldsymbol{M} + \boldsymbol{M}^\top)/2$ has been extensively studied in the literature (Benaych-Georges and Nadakuditi (2011), Féral and Péché (2007), Füredi and Komlós (1981), Knowles and Yin (2013), Péché (2006), Renfrew and Soshnikov (2013), Yin, Bai and Krishnaiah (1988)). In particular, it has been shown (e.g., Capitaine, Donati-Martin and Féral (2009)) that, with probability approaching one,

$$\lambda_{\mathsf{sym}} = \lambda^\star + \frac{n\sigma^2}{2\lambda^\star} + O(\sigma\sqrt{\log n}). \tag{21}$$

If $\sigma = |\lambda^\star|\sqrt{1/(n \log n)}$ (which is the setting in our numerical experiment), then this can be translated into

$$\frac{\lambda_{\mathsf{sym}} - \lambda^\star}{\lambda^\star} = \frac{1}{2\log n} + O\Big(\frac{1}{\sqrt{n}}\Big).$$

This implies that λ_{sym} suffers from a substantially larger bias than the leading eigenvalue λ obtained without symmetrization, since in this case we have (cf. Corollary 1)

$$\left|\frac{\lambda - \lambda^\star}{\lambda^\star}\right| \lesssim \frac{1}{\sqrt{n}}. \tag{22}$$

Comparison to properly adjusted eigendecomposition and SVD-based methods. Armed with the approximation (21) in the i.i.d. Gaussian noise case, the careful reader might naturally suggest a properly corrected estimate $\lambda_{\mathsf{sym,c}}$ as follows (again assuming $\lambda > 0$)

$$\lambda_{\mathsf{sym,c}} = \frac{1}{2}\Big(\lambda_{\mathsf{sym}} + \sqrt{\lambda_{\mathsf{sym}}^2 - 2n\sigma^2}\Big), \tag{23}$$

which is a shrinkage-type estimate chosen to satisfy $\lambda_{\mathsf{sym}} = \lambda_{\mathsf{sym,c}} + \frac{n\sigma^2}{2\lambda_{\mathsf{sym,c}}}$. A little algebra reveals that: if $\sigma = 1/\sqrt{n \log n}$, then

$$\left| \frac{\lambda_{\mathsf{sym,c}} - \lambda^\star}{\lambda^\star} \right| \lesssim \frac{1}{\sqrt{n}},$$

thus matching the estimation accuracy of λ (cf. (22)). In addition, some sort of universality results has been established as well in the literature (Capitaine, Donati-Martin and Féral (2009)), implying that the same approximation and correction are applicable to a broad family of zero-mean noise with identical variance. As we shall illustrate numerically in Section 4.2, this approach (i.e., $\lambda_{\mathsf{sym,c}}$) performs almost identically to the one using vanilla eigendecomposition without symmetrization. In addition, very similar observations have been made for the SVD-based approach (Benaych-Georges and Nadakuditi (2012), Bryc and Silverstein (2018), Féral and Péché (2007), Péché (2006), Silverstein (1994), Yin, Bai and Krishnaiah (1988)); for the sake of brevity, we do not repeat the arguments here.

We would nevertheless like to single out a few statistical advantages of the eigendecomposition approach without symmetrization. To begin with, λ is obtained via vanilla eigendecomposition, and computing it does not rely on any kind of noise statistics. This is in stark contrast to the bias correction (23) in the presence of symmetric data, which requires prior knowledge about (or a very precise estimate of) the noise variance σ^2. Leaving out this prior knowledge matter, a more important issue is that the approximation formula (21) assumes identical variance of noise components across all entries (i.e., homoscedasticity). While an approximation of this kind has been found for more general cases beyond homoscedastic noise (e.g., Bryc and Silverstein (2018)), the approximation formula (e.g., Bryc and Silverstein (2018), Theorem 1.1) becomes fairly complicated, requires prior knowledge about all variance parameters and is thus difficult to implement in practice. In comparison, the vanilla eigendecomposition approach analyzed in Corollary 1 imposes no restriction on the noise statistics and is fully adaptive to heteroscedastic noise.

Lower bounds. To complete the picture, we provide a simple information-theoretic lower bound for the i.i.d. Gaussian noise case, which will be established in Appendix 2.

LEMMA 4. *Fix any small constant $\varepsilon > 0$. Suppose that $H_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. Consider three matrices*

$$\boldsymbol{M} = \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{H}, \qquad \widetilde{\boldsymbol{M}} = (\lambda^\star + \Delta)\boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{H}, \qquad \widehat{\boldsymbol{M}} = (\lambda^\star - \Delta)\boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{H}$$

with $\|\boldsymbol{u}^\star\|_2 = 1$. If $\Delta \le \sigma\sqrt{(\log_2 1.5 - \varepsilon)\log 2}$, then no algorithm can distinguish $\boldsymbol{M}$, $\widetilde{\boldsymbol{M}}$ and $\widehat{\boldsymbol{M}}$ with $p_{\mathrm{e}} \le \varepsilon$, where p_{e} is the minimax probability of error for testing three hypotheses (namely, the ones claiming that the true eigenvalues are $\lambda^\star$, $\lambda^\star + \Delta$, and $\lambda^\star - \Delta$, resp.).

In short, Lemma 4 asserts that one cannot possibly locate an eigenvalue to within a precision of Δ much better than σ, which reveals a fundamental limit that cannot be broken by any algorithm. In comparison, the vanilla eigendecomposition method based on asymmetric data achieves an accuracy of $|\lambda - \lambda^\star| \lesssim \sigma\sqrt{\log n}$ (cf. Corollary 1 and (19)) for the incoherent case, thus matching the information-theoretic lower bound up to some log factor. In fact, the extra $\sqrt{\log n}$ factor arises simply because we are aiming for a high-probability guarantee.

4.1.3. *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*

$$\min\{|\boldsymbol{a}^\top(\boldsymbol{u}-\boldsymbol{u}^\star)|, |\boldsymbol{a}^\top(\boldsymbol{u}+\boldsymbol{u}^\star)|\} \lesssim \left(|\boldsymbol{a}^\top\boldsymbol{u}^\star| + \sqrt{\frac{\mu}{n}}\right)\frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{|\lambda^\star|}. \tag{24}$$

PROOF. Without loss of generality, assume that $\boldsymbol{u}^{\star\top}\boldsymbol{u} \geq 0$ and that $\lambda^\star = 1$. Then one has

$$\begin{aligned}|\boldsymbol{a}^\top(\boldsymbol{u}-\boldsymbol{u}^\star)| &\leq \left|\boldsymbol{a}^\top\boldsymbol{u} - \boldsymbol{a}^\top\boldsymbol{u}^\star\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda}\right| + |\boldsymbol{a}^\top\boldsymbol{u}^\star|\left|\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda} - 1\right| \\ &\leq \max\{\sigma\sqrt{n\log n}, B\log n\}\sqrt{\frac{\mu}{n}} + |\boldsymbol{a}^\top\boldsymbol{u}^\star|\left|\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda} - 1\right|,\end{aligned}$$

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

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

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 emphasis is eigendecomposition—does not rely on the Gaussianality 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. (2020) for phase retrieval and blind deconvolution, although the vector $\boldsymbol{a}$ therein is specific to the problems (i.e., the design vectors) and cannot be made general.

REMARK 5. The perturbation theory for linear forms of eigenvectors has been substantially extended in our follow-up work; the interested reader is referred to Cheng, Wei and Chen (2020) for details.

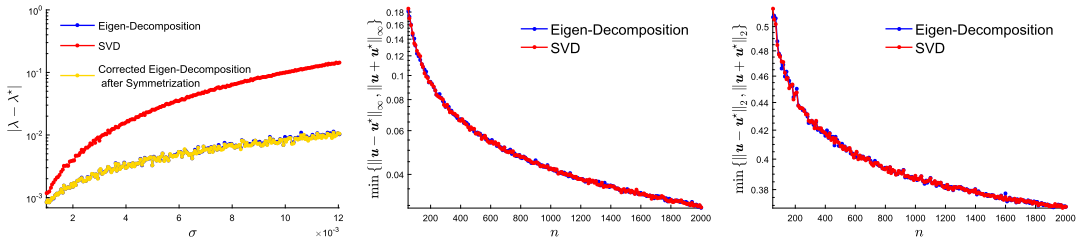
4.1.4. *Entrywise eigenvector perturbation.* A straightforward consequence of Corollary 2 that is worth emphasizing is sharp entrywise control of the leading eigenvector as follows.

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

$$\min\{\|\boldsymbol{u}-\boldsymbol{u}^\star\|_\infty, \|\boldsymbol{u}+\boldsymbol{u}^\star\|_\infty\} \lesssim \frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{|\lambda^\star|}\sqrt{\frac{\mu}{n}}. \tag{25}$$

PROOF. Recognizing that $\|\boldsymbol{u}-\boldsymbol{u}^\star\|_\infty = \max_i |\boldsymbol{e}_i^\top\boldsymbol{u} - \boldsymbol{e}_i^\top\boldsymbol{u}^\star|$ and recalling our assumption $|\boldsymbol{e}_i^\top\boldsymbol{u}| \leq \sqrt{\mu/n}$, we can invoke Corollary 2 and the union bound to establish this entrywise bound. □

We note that: while the ℓ_2 perturbation (or sin $\boldsymbol{\Theta}$ distance) of eigenvectors or singular vectors has been extensively studied (Cai and Zhang (2018), Davis and Kahan (1970), O'Rourke, Vu and Wang (2018), Vu (2011), Wang (2015), Wedin (1972)), the entrywise eigenvector



(a) eigenvalue perturbation (b) ℓ_∞ eigenvector perturbation (c) ℓ_2 eigenvector perturbation

FIG. 2. *Numerical simulation for rank-*1 *matrix estimation under i.i.d. Gaussian noise* $\mathcal{N}(0, \sigma^2)$*, where the rank-*1 *truth* $\boldsymbol{M}^\star$ *is generated randomly with leading eigenvalue* 1. (*a*): $|\lambda - \lambda^\star|$ *versus* σ *when* $n = 1000$; (*b*) *and* (*c*): ℓ_∞ *and* ℓ_2 *eigenvector estimation errors versus* n *with* $\sigma = 1/\sqrt{n \log n}$*, respectively. The blue* (*resp.*, *red*) *lines represent the average errors over* 100 *independent trials using the vanilla eigendecomposition* (*resp.*, *SVD*) *approach applied to* $\boldsymbol{M}$*. The orange line in* (*a*) *represents the average errors over* 100 *independent trials using the corrected leading eigenvalue* $\lambda_{\mathsf{sym,c}}$ *of the symmetrized matrix* $(\boldsymbol{M} + \boldsymbol{M}^\top)/2$ (*cf.* (23)).

behavior was much less explored. The prior literature contains only a few entrywise eigenvector perturbation analysis results for settings very different from ours, for example, the i.i.d. random matrix case (O'Rourke, Vu and Wang (2016), Vu and Wang (2015)), the symmetric low-rank case (Abbe et al. (2020), Eldridge, Belkin and Wang (2018), Fan, Wang and Zhong (2017)) and the case with transition matrices for reversible Markov chains (Chen et al. (2019b)). Our results add another instance to this body of works in providing entrywise eigenvector perturbation bounds.

4.2. *Applications*. We apply our main results to two concrete matrix estimation problems and examine the effectiveness of these bounds. As before, $\boldsymbol{M}^\star$ is a rank-1 matrix with incoherence parameter μ and leading eigenvalue $\lambda^\star$.

Low-rank matrix estimation from Gaussian noise. Suppose that $\boldsymbol{H}$ is composed of i.i.d. Gaussian random variables $\mathcal{N}(0, \sigma^2)$.[2] If $\sigma \lesssim \frac{1}{\sqrt{n \log n}}$, applying Corollaries 1–3 reveals that with high probability,

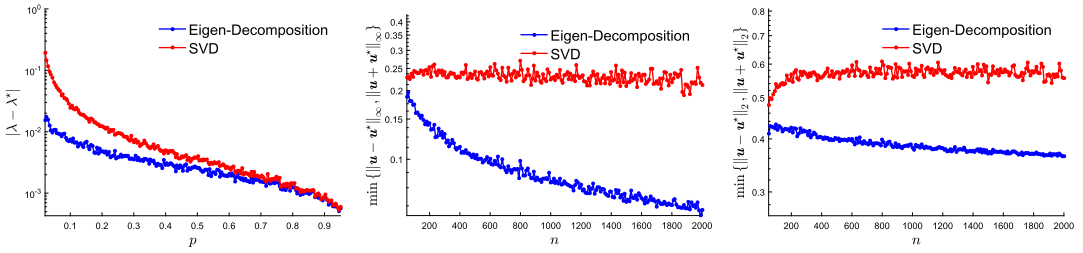
$$|\lambda - \lambda^\star| \lesssim \sigma \sqrt{\mu \log n}, \tag{26a}$$

$$\min\{\|\boldsymbol{u} - \boldsymbol{u}^\star\|_\infty, \|\boldsymbol{u} + \boldsymbol{u}^\star\|_\infty\} \lesssim \frac{\sigma \sqrt{\mu \log n}}{|\lambda^\star|}, \tag{26b}$$

$$\min\{|\boldsymbol{a}^\top(\boldsymbol{u} - \boldsymbol{u}^\star)|, |\boldsymbol{a}^\top(\boldsymbol{u} + \boldsymbol{u}^\star)|\} \lesssim \left(|\boldsymbol{a}^\top \boldsymbol{u}^\star| + \sqrt{\frac{\mu}{n}}\right) \frac{\sigma \sqrt{n \log n}}{|\lambda^\star|} \tag{26c}$$

for any fixed unit vector $\boldsymbol{a} \in \mathbb{R}^n$. We have conducted additional numerical experiments in Figure 2, which confirm our findings. It is also worth noting that empirically, eigendecomposition and SVD applied to $\boldsymbol{M}$ achieve nearly identical ℓ_2 and ℓ_∞ errors when estimating the leading eigenvector of $\boldsymbol{M}^\star$. In addition, we also include the numerical estimation error of the corrected eigenvalue $\lambda_{\mathsf{sym,c}}$ (cf. (23)) of the symmetrized matrix $(\boldsymbol{M} + \boldsymbol{M}^\top)/2$. As can be seen from Figure 2, vanilla eigendecomposition without symmetrization performs nearly identically to the one with symmetrization and proper correction.

[2] In this case, one can take $B \asymp \sigma \sqrt{\log n}$, which clearly satisfies $B \log n \ll \sqrt{n\sigma^2 \log n}$.



(a) eigenvalue perturbation (b) ℓ_∞ eigenvector perturbation (c) ℓ_2 eigenvector perturbation

FIG. 3. *Numerical simulation for rank-*1 *matrix completion, where the rank-*1 *truth* $\boldsymbol{M}^\star$ *is randomly generated with leading eigenvalue* 1 *and sampling rate is* $p = 3\log n/n$. (*a*) $|\lambda - \lambda^\star|$ *versus* p *when* $n = 1000$; (*b*) *and* (*c*): ℓ_∞ *and* ℓ_2 *eigenvector estimation errors versus* n, *respectively. The blue* (*resp.*, *red*) *lines represent the average errors over* 100 *independent trials using the eigendecomposition* (*resp.*, *SVD*) *approach.*

Low-rank matrix completion. Suppose that $\boldsymbol{M}$ is generated using random partial entries of $\boldsymbol{M}^\star$ as follows:

$$M_{ij} = \begin{cases} \dfrac{1}{p} M^\star_{ij} & \text{with probability } p, \\ 0 & \text{else,} \end{cases} \tag{27}$$

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

$$\frac{|\lambda - \lambda^\star|}{|\lambda^\star|} \lesssim \frac{1}{\sqrt{n}} \sqrt{\frac{\mu^3 \log n}{pn}}, \tag{28a}$$

$$\min\{\|\boldsymbol{u} - \boldsymbol{u}^\star\|_\infty, \|\boldsymbol{u} + \boldsymbol{u}^\star\|_\infty\} \lesssim \frac{1}{\sqrt{n}} \sqrt{\frac{\mu^3 \log n}{pn}}, \tag{28b}$$

$$\min\{|\boldsymbol{a}^\top(\boldsymbol{u} - \boldsymbol{u}^\star)|, |\boldsymbol{a}^\top(\boldsymbol{u} + \boldsymbol{u}^\star)|\} \lesssim \left(|\boldsymbol{a}^\top \boldsymbol{u}^\star| + \sqrt{\frac{\mu}{n}}\right) \sqrt{\frac{\mu^2 \log n}{pn}} \tag{28c}$$

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

Finally, we remark that all the above applications assume the availability of an asymmetric data matrix $\boldsymbol{M}$. One might naturally wonder whether there is anything useful we can say if only a symmetric matrix $\boldsymbol{M}$ is available. While this is in general difficult, our theory does have direct implications for both matrix completion and the case with i.i.d. Gaussian noise in the presence of symmetric data matrices; that is, it is possible to first asymmetrize the data matrix followed by eigendecomposition. The interested reader is referred to Appendix 10 for details.

4.3. *Why asymmetry helps?* We take a moment to develop some intuition underlying Theorem 3, focusing on the case with $\lambda^\star = 1$ for simplicity. The key ingredient is the Neumann trick stated in Theorem 2. Specifically, in the rank-1 case we can expand

$$\boldsymbol{u} = \frac{1}{\lambda}(\boldsymbol{u}^{\star\top} \boldsymbol{u}) \sum_{s=0}^{\infty} \frac{1}{\lambda^s} \boldsymbol{H}^s \boldsymbol{u}^\star.$$

A little algebra yields

$$\left|\boldsymbol{a}^\top\left(\boldsymbol{u}-\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda}\boldsymbol{u}^\star\right)\right| = \left|\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda}\sum_{s=1}^{\infty}\frac{\boldsymbol{a}^\top\boldsymbol{H}^s\boldsymbol{u}^\star}{\lambda^s}\right| \lesssim \sum_{s=1}^{\infty}\left|\frac{\boldsymbol{a}^\top\boldsymbol{H}^s\boldsymbol{u}^\star}{\lambda^s}\right|, \tag{29}$$

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

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

- *Asymmetric case*: when $\boldsymbol{H}$ is composed of independent zero-mean entries each with variance σ_n^2, one has
$$\mathbb{E}[\boldsymbol{a}^\top\boldsymbol{H}^2\boldsymbol{u}^\star] = \boldsymbol{a}^\top\mathbb{E}[\boldsymbol{H}^2]\boldsymbol{u}^\star = \boldsymbol{a}^\top(\sigma^2\boldsymbol{I})\boldsymbol{u}^\star = \sigma^2\boldsymbol{a}^\top\boldsymbol{u}^\star.$$
- *Symmetric case*: when $\boldsymbol{H}$ is symmetric and its upper triangular part consists of independent zero-mean entries with variance σ_n^2, it holds that
$$\mathbb{E}[\boldsymbol{a}^\top\boldsymbol{H}^2\boldsymbol{u}^\star] = \boldsymbol{a}^\top\mathbb{E}[\boldsymbol{H}^2]\boldsymbol{u}^\star = \boldsymbol{a}^\top(n\sigma^2\boldsymbol{I})\boldsymbol{u}^\star = n\sigma^2\boldsymbol{a}^\top\boldsymbol{u}^\star.$$

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

The vanilla 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 $\boldsymbol{M}\boldsymbol{M}^\top$. Given that $\boldsymbol{M}\boldsymbol{M}^\top - \boldsymbol{M}^\star\boldsymbol{M}^{\star\top}$ is also symmetric, the aforementioned bias issue arises as well. This explains why vanilla eigendecomposition might have an advantage over vanilla SVD when dealing with asymmetric matrices.

Finally, we remark that the aforementioned bias issue becomes less severe as $\|\boldsymbol{H}\|$ decreases. For example, when $\|\boldsymbol{H}\|$ is exceedingly small, the only dominant term on the right-hand side of (29) is $\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}^\star$, with all higher-order terms being vanishingly small. In this case, $\mathbb{E}[\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}^\star] = 0$ for both symmetric and asymmetric zero-mean noise matrices. As a consequence, the advantage of eigendecomposition becomes negligible when dealing with nearly-zero noise. This observation is also confirmed in the numerical experiments reported in Figure 2(a) and Figure 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 $\|\boldsymbol{H}\|$ has been studied in the literature (Eldridge, Belkin and Wang (2018), O'Rourke, Vu and Wang (2018), Vu (2011)). For example, it was shown in O'Rourke, Vu and Wang (2018) that when $\|\boldsymbol{H}\| \lesssim \frac{1}{\sqrt{n}}\|\boldsymbol{M}^\star\|$, 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 $\boldsymbol{H}$ is symmetric (cf. Eldridge, Belkin and Wang (2018), Theorem 6). By contrast, our eigenvalue perturbation results achieve this gain even when $\|\boldsymbol{H}\|$ is nearly as large as $\|\boldsymbol{M}^\star\|$ (up to some logarithmic factor).

4.4. *Proof outline of Theorem 3.* This subsection outlines the main steps for establishing Theorem 3. To simplify presentation, we shall assume without loss of generality that

$$\lambda^\star = 1. \tag{30}$$

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

$$\boldsymbol{M} = \boldsymbol{M}^\star + \tilde{\boldsymbol{H}} \tag{31}$$

with probability exceeding $1 - O(n^{-10})$, where $\tilde{H}_{ij} \triangleq H_{ij}\mathbb{1}_{\{|H_{ij}|\le B\}}$ 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 $\mathsf{Var}(\tilde{H}_{ij}) \le \sigma^2$, which coincides with the case obeying Conditions 1–3, 4(a) in Assumption 1.

As already mentioned in Section 4.3, everything boils down to controlling $|\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}^\star|$ for $s \ge 1$. This is accomplished via the following lemma.

LEMMA 5 (Bounding higher-order terms). *Consider any fixed unit vector $\boldsymbol{a} \in \mathbb{R}^n$ and any positive integers s, k satisfying $Bsk \le 2$ and $n\sigma^2 sk \le 2$. Under the assumptions of Theorem 3,*

$$\left|\mathbb{E}\left[\left(\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}^\star\right)^k\right]\right| \le \frac{sk}{2} \max\left\{(Bsk)^{sk}, (2n\sigma^2 sk)^{sk/2}\right\}\left(\sqrt{\frac{\mu}{n}}\right)^k. \tag{32}$$

PROOF. The proof of Lemma 5 is combinatorial in nature, which we defer to Appendix 3. □

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 5 into a high probability bound as follows.

COROLLARY 4. *Under the assumptions of Lemma 5, there exists some universal constant $c_2 > 0$ such that*

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

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

PROOF. See Appendix 6. □

In addition, in view of Lemma 1 and the condition (14), one has

$$\|\boldsymbol{H}\| \lesssim \max\left\{B\log n, \sqrt{n\sigma^2 \log n}\right\} < 1/10 \tag{33}$$

with probability $1 - O(n^{-10})$, which together with Lemma 2 implies $\lambda \ge 3\|\boldsymbol{H}\|$. This further leads to

$$\sum_{s:s\ge 20\log n}\left(\frac{\|\boldsymbol{H}\|}{\lambda}\right)^s \le \frac{\|\boldsymbol{H}\|}{\lambda}\sum_{s:s\ge 20\log n-1}\left(\frac{\|\boldsymbol{H}\|}{\lambda}\right)^s \le \frac{\|\boldsymbol{H}\|}{\lambda}\sum_{s:s\ge 20\log n-1}\frac{1}{3^s}$$

$$\lesssim \max\left\{B\log n, \sqrt{n\sigma^2\log n}\right\}\cdot n^{-10}.$$

Putting the above bounds together and using the fact that λ is real-valued and $\lambda \geq 1/2$ (cf. Lemma 2), we have

$$\begin{aligned}
&\left|\boldsymbol{a}^\top\left(\boldsymbol{u}-\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda}\boldsymbol{u}^\star\right)\right| \\
&\quad=\left|\frac{\boldsymbol{u}^{\star\top}\boldsymbol{u}}{\lambda}\sum_{s=1}^{+\infty}\frac{\boldsymbol{a}^\top\boldsymbol{H}^s\boldsymbol{u}^\star}{\lambda^s}\right| \\
&\quad\lesssim \sum_{s=1}^{20\log n}\frac{1}{\lambda^s}|\boldsymbol{a}^\top\boldsymbol{H}^s\boldsymbol{u}^\star|+\sum_{s=20\log n}^{+\infty}\left(\frac{\|\boldsymbol{H}\|}{\lambda}\right)^s \\
&\quad\leq\sqrt{\frac{\mu}{n}}\sum_{s=1}^{20\log n}\left(2c_2\max\{B\log n,\sqrt{n\sigma^2\log n}\}\right)^s+\frac{\max\{B\log n,\sqrt{n\sigma^2\log n}\}}{n^{10}} \\
&\quad\lesssim\max\{B\log n,\sqrt{n\sigma^2\log n}\}\sqrt{\frac{\mu}{n}},
\end{aligned}$$

as long as $\max\{B\log n,\sqrt{n\sigma^2\log n}\}$ is sufficiently small. Here, the last line also uses the fact that $\mu\geq 1$ (and hence $\sqrt{\mu/n}\gg n^{-10}$). This concludes 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 $\boldsymbol{M}^\star$ is symmetric and rank-r, as detailed in this section. As before, assume that the r nonzero eigenvalues of $\boldsymbol{M}^\star$ obey $\lambda^\star_{\max}=|\lambda^\star_1|\geq\cdots\geq|\lambda^\star_r|=\lambda^\star_{\min}$. Once again, we start with a master bound.

THEOREM 4 (Perturbation of linear forms of eigenvectors (rank-r)). *Consider a rank-r symmetric matrix $\boldsymbol{M}^\star\in\mathbb{R}^{n\times n}$ with incoherence parameter μ. Define $\kappa\triangleq\lambda^\star_{\max}/\lambda^\star_{\min}$. Suppose that*

$$\frac{\max\{\sigma\sqrt{n\log n},B\log n\}}{\lambda^\star_{\max}}\leq\frac{c_1}{\kappa} \tag{34}$$

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

$$\left|\boldsymbol{a}^\top\left(\boldsymbol{u}_l-\sum_{j=1}^{r}\frac{\lambda^\star_j\boldsymbol{u}^{\star\top}_j\boldsymbol{u}_l}{\lambda_l}\boldsymbol{u}^\star_j\right)\right|\lesssim\max\{\sigma\sqrt{n\log n},B\log n\}\frac{\kappa}{|\lambda_l|}\sqrt{\frac{\mu r}{n}} \tag{35}$$

$$\lesssim\frac{\max\{\sigma\sqrt{n\log n},B\log n\}}{\lambda^\star_{\max}}\kappa^2\sqrt{\frac{\mu r}{n}}. \tag{36}$$

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 κ 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 lth $(1\leq l\leq r)$ eigenvalue λ_l of $\boldsymbol{M}$. Under the assumptions of Theorem 4, with probability at least $1-O(n^{-10})$, there exists $1\leq j\leq r$ such that*

$$|\lambda_l-\lambda^\star_j|\lesssim\max\{\sigma\sqrt{n\log n},B\log n\}\kappa r\sqrt{\frac{\mu}{n}}, \tag{37}$$

provided that

$$\frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{\lambda_{\max}^{\star}} \le c_1/\kappa^2 \tag{38}$$

for some sufficiently small constant $c_1 > 0$.

PROOF. See Appendix 7. □

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

$$|\lambda_l - \lambda_j^{\star}| \le \|\boldsymbol{H}\| \lesssim \max\{\sigma\sqrt{n\log n}, B\log n\} \quad \text{for some } 1 \le j \le r. \tag{39}$$

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 eigensubspace.

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

$$\|\boldsymbol{a}^{\top}\boldsymbol{U}\|_2 \lesssim \kappa\sqrt{r}\|\boldsymbol{a}^{\top}\boldsymbol{U}^{\star}\|_2 + \frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{\lambda_{\max}^{\star}}\kappa^2 r\sqrt{\frac{\mu}{n}}. \tag{40}$$

PROOF. See Appendix 8. □

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

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

$$\|\boldsymbol{U}\|_{2,\infty} \lesssim \kappa r\sqrt{\frac{\mu}{n}}. \tag{41}$$

PROOF. Given that $\|\boldsymbol{U}\|_{2,\infty} = \max_{1\le i\le n}\|\boldsymbol{e}_i^{\top}\boldsymbol{U}\|_2$ and recalling our assumption implies $\|\boldsymbol{U}^{\star}\|_{2,\infty} \le \sqrt{\mu r/n}$, 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, Montanari and Oh (2010), and the above kind of incoherence property is crucial in guaranteeing fast convergence of the subsequent nonconvex iterative refinement procedures Ma et al. (2020).

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 $\boldsymbol{U}$ and $\boldsymbol{U}^{\star}$. 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.

REMARK 8. The theory for the rank-r case has recently been significantly improved; see our follow-up work (Cheng, Wei and Chen (2020)) for details.

5.2. *Application*: *Spectral estimation when* $\boldsymbol{M}^\star$ *is asymmetric and rank*-1. In some scenarios, the above general rank results allow us to improve spectral estimation when $\boldsymbol{M}^\star$ is asymmetric. Consider the case where $\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{v}^{\star\top} \in \mathbb{R}^{n_1 \times n_2}$ is an asymmetric rank-1 matrix with leading singular value $\lambda^\star$. Suppose that we observe two independent noisy copies of $\boldsymbol{M}^\star$, namely,

$$\boldsymbol{M}_1 = \boldsymbol{M}^\star + \boldsymbol{H}_1, \qquad \boldsymbol{M}_2 = \boldsymbol{M}^\star + \boldsymbol{H}_2, \tag{42}$$

where $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are independent noise matrices. The goal is to estimate the singular value and singular vectors of $\boldsymbol{M}^\star$ from $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$.

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

$$\boldsymbol{M}_{\mathsf{d}}^\star \triangleq \begin{bmatrix} \boldsymbol{0} & \boldsymbol{M}^\star \\ \boldsymbol{M}^{\star\top} & \boldsymbol{0} \end{bmatrix}, \qquad \boldsymbol{M}_{\mathsf{d}} \triangleq \begin{bmatrix} \boldsymbol{0} & \boldsymbol{M}_1 \\ \boldsymbol{M}_2^\top & \boldsymbol{0} \end{bmatrix}. \tag{43}$$

Here, we place $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ in two different subblocks, in order to "asymmetrize" the dilation matrix. The rationale is that $\boldsymbol{M}_{\mathsf{d}}^\star$ is a rank-2 symmetric matrix with exactly two nonzero eigenvalues

$$\lambda_1(\boldsymbol{M}_{\mathsf{d}}^\star) = \lambda^\star \quad \text{and} \quad \lambda_2(\boldsymbol{M}_{\mathsf{d}}^\star) = -\lambda^\star,$$

whose corresponding eigenvectors are given by

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

respectively. This motivates us to perform eigendecomposition of $\boldsymbol{M}_{\mathsf{d}}$, and use the top-2 eigenvalues and eigenvectors to estimate $\lambda^\star$, $\boldsymbol{u}^\star$ and $\boldsymbol{v}^\star$, respectively.

Eigenvalue perturbation analysis. As an immediate consequence of Corollary 5, the two leading eigenvalues of $\boldsymbol{M}_{\mathsf{d}}$ provide fairly accurate estimates of the leading singular value $\lambda^\star$ of $\boldsymbol{M}^\star$, as stated below.

COROLLARY 8. *Assume* $\boldsymbol{M}^\star \in \mathbb{R}^{n_1 \times n_2}$ *is a rank*-1 *matrix with leading singular value* $\lambda^\star$ *and incoherence parameter* μ. *Define* $n \triangleq n_1 + n_2$. *Suppose that* $\lambda_1^{\mathsf{d}} \geq \lambda_2^{\mathsf{d}}$ *are the two leading eigenvalues of* $\boldsymbol{M}_{\mathsf{d}}$ (*cf.* (43)), *and that* $\boldsymbol{H}_1$ *and* $\boldsymbol{H}_2$ *are independent and satisfy Assumption* 1. *Then with probability at least* $1 - O(n^{-10})$,

$$\max\{|\lambda_1^{\mathsf{d}} - \lambda^\star|, |\lambda_2^{\mathsf{d}} + \lambda^\star|\} \lesssim \max\{\sigma\sqrt{n\log n}, B\log n\}\sqrt{\frac{\mu}{n}}, \tag{44}$$

provided that

$$\frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{\lambda^\star} \leq c_1 \tag{45}$$

for some sufficiently small constant $c_1 > 0$.

PROOF. To begin with, it follows from Corollary 5 that both λ_1^{d} and λ_2^{d} are close to either $\lambda^\star$ or $-\lambda^\star$. 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, λ_1^{d} (resp., λ_2^{d}) is close to $\lambda^\star$ (resp., $-\lambda^\star$). □

Eigenvector perturbation analysis. We then move on to studying the eigenvector perturbation bounds. Specifically, denote by $\boldsymbol{u}_1^{\mathsf{d}}$ and $\boldsymbol{u}_2^{\mathsf{d}}$ the eigenvectors of $\boldsymbol{M}_{\mathsf{d}}$ associated with its

two leading eigenvalues λ_1^{d} and λ_2^{d}, respectively. Without loss of generality, we assume that $\lambda_1^{\mathsf{d}} \geq \lambda_2^{\mathsf{d}}$. If we write

$$\boldsymbol{u}_1^{\mathsf{dilation}} = \begin{pmatrix} \boldsymbol{u}_{1,1}^{\mathsf{d}} \\ \boldsymbol{u}_{1,2}^{\mathsf{d}} \end{pmatrix} \quad \text{with } \boldsymbol{u}_{1,1}^{\mathsf{d}} \in \mathbb{R}^{n_1}, \boldsymbol{u}_{1,2}^{\mathsf{d}} \in \mathbb{R}^{n_2},$$

then we can employ $\boldsymbol{u}_{1,1}^{\mathsf{d}}$ and $\boldsymbol{u}_{1,2}^{\mathsf{d}}$ to estimate $\boldsymbol{u}^\star$ and $\boldsymbol{v}^\star$ after proper normalization, namely,

$$\boldsymbol{u} \triangleq \frac{\boldsymbol{u}_{1,1}^{\mathsf{d}}}{\|\boldsymbol{u}_{1,1}^{\mathsf{d}}\|_2}, \qquad \boldsymbol{v} \triangleq \frac{\boldsymbol{u}_{1,2}^{\mathsf{d}}}{\|\boldsymbol{u}_{1,2}^{\mathsf{d}}\|_2}. \tag{46}$$

The following theorem develops error bounds for both $\boldsymbol{u}$ and $\boldsymbol{v}$, which we establish in Appendix 9. Here, we denote $\min \|\boldsymbol{x} \pm \boldsymbol{y}\|_2 = \min\{\|\boldsymbol{x} - \boldsymbol{y}\|_2, \|\boldsymbol{x} + \boldsymbol{y}\|_2\}$ and $\min \|\boldsymbol{x} \pm \boldsymbol{y}\|_\infty = \min\{\|\boldsymbol{x} - \boldsymbol{y}\|_\infty, \|\boldsymbol{x} + \boldsymbol{y}\|_\infty\}$.

THEOREM 5. *Suppose $\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{v}^{\star\top} \in \mathbb{R}^{n_1 \times n_2}$ is a rank-1 matrix with leading singular value $\lambda^\star$ and incoherence parameter μ, where $\|\boldsymbol{u}^\star\|_2 = \|\boldsymbol{v}^\star\|_2 = 1$. Define $n \triangleq n_1 + n_2$, and fix any unit vectors $\boldsymbol{a} \in \mathbb{R}^{n_1}$ and $\boldsymbol{b} \in \mathbb{R}^{n_2}$. Then with probability at least $1 - O(n^{-10})$, the estimates $\boldsymbol{u}$ and $\boldsymbol{v}$ (cf. (46)) obey*

$$\max\{\min\|\boldsymbol{u} \pm \boldsymbol{u}^\star\|_2, \min\|\boldsymbol{v} \pm \boldsymbol{v}^\star\|_2\} \lesssim \frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{\lambda^\star}, \tag{47a}$$

$$\max\{\min\|\boldsymbol{u} \pm \boldsymbol{u}^\star\|_\infty, \min\|\boldsymbol{v} \pm \boldsymbol{v}^\star\|_\infty\} \lesssim \frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{\lambda^\star}\sqrt{\frac{\mu}{n}}, \tag{47b}$$

$$\min|\boldsymbol{a}^\top(\boldsymbol{u} \pm \boldsymbol{u}^\star)| \lesssim \left(|\boldsymbol{a}^\top\boldsymbol{u}^\star| + \sqrt{\frac{\mu}{n}}\right)\frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{\lambda^\star}, \tag{47c}$$

$$\min|\boldsymbol{b}^\top(\boldsymbol{v} \pm \boldsymbol{v}^\star)| \lesssim \left(|\boldsymbol{b}^\top\boldsymbol{v}^\star| + \sqrt{\frac{\mu}{n}}\right)\frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{\lambda^\star}, \tag{47d}$$

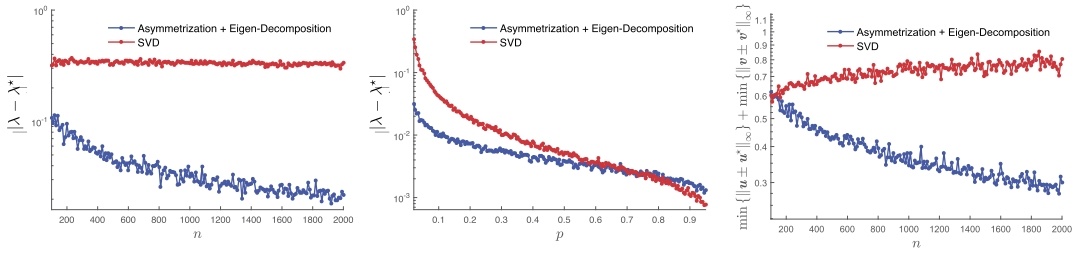
provided that there exists some some sufficiently small constant $c_1 > 0$ such that

$$\frac{\max\{\sigma\sqrt{n\log n}, B\log n\}}{\lambda^\star} \leq c_1. \tag{48}$$

Similar to the symmetric rank-1 case, the estimation errors of the estimates $\boldsymbol{u}$ and $\boldsymbol{v}$ 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 $\boldsymbol{M}^\star$ is a real-valued and rank-1 matrix.

Further, we conduct numerical experiments for matrix completion when $\boldsymbol{M}^\star$ is a rank-1 and asymmetric matrix in Figure 4. Here, we suppose that at most 1 sample is observed for each entry, and we estimate the singular value and singular vectors of $\boldsymbol{M}^\star$ via the above-mentioned dilation trick, coupled with the asymmetrization procedure discussed in Section 10. The numerical performance confirms that the proposed technique outperforms vanilla SVD in spectral estimation.

Finally, we remark that the asymptotic behavior of the eigenvalues of asymmetric random matrices has been extensively explored in the physics literature (e.g., (Brézin and Zee (1998), Feinberg and Zee (1997), Khoruzhenko (1996), Lytova and Tikhomirov (2018), Mehlig and Chalker (1998), Sommers et al. (1988))). Their focus, however, has largely been to pin down the asymptotic density of the eigenvalues, similar to the semicircle law in the symmetric case. Nevertheless, a sharp perturbation bound for the leading eigenvalue—particularly for the low-rank case—is beyond their reach. A few recent papers began to explore the locations of



(a) eigenvalue perturbation vs. n (b) eigenvalue perturbation vs. p (c) ℓ_∞ eigenvector perturbation

FIG. 4. *Numerical experiments for rank-1 matrix completion, where* $\boldsymbol{M}^\star = \boldsymbol{u}^\star \boldsymbol{v}^{\star\top} \in \mathbb{R}^{n_1 \times n_2}$ *is randomly generated with leading singular value* $\lambda^\star = 1$. *Let* $n = n_1 = 2n_2$. *Each entry is observed independently with probability* p. *(a)* $|\lambda - \lambda^\star|$ *versus* n *with* $p = 3\log n/n$; *(b)* $|\lambda - \lambda^\star|$ *versus* p *with* $n = 1000$; *(c)* ℓ_∞ *eigenvector estimation error versus* n *with* $p = 3\log n/n$. *The blue (resp., red) lines represent the average errors over* 100 *independent trials using using the eigendecomposition (resp., SVD) approach applied to* $\boldsymbol{M}_{\mathsf{dilation}}$ *(resp.,* $\boldsymbol{M}$*).*

eigenvalue outliers that fall outside the bulk predicted by the circular law (Benaych-Georges and Rochet (2016), Bordenave and Capitaine (2016), Rajagopalan (2015), Tao (2013)). The results reported therein either do not focus on obtaining the right convergence rate (e.g., providing only a bound like $|\lambda - \lambda^\star| = o(|\lambda^\star|)$) or are restricted to a special family of ground truth (e.g., the one with a diagonal block equal to identity) or i.i.d. noise. As a result, these prior results are insufficient to demonstrate the power and benefits of the eigen-decomposition method in the presence of data asymmetry.

5.3. *Proof of Theorem 4.* Without loss of generality, we shall assume $\lambda^\star_{\max} = \lambda^\star_1 = 1$ throughout the proof. To begin with, Lemma 2 implies that for all $1 \le l \le r$,

$$|\lambda_l| \ge \big|\lambda^\star_{\min}\big| - \|\boldsymbol{H}\| > 1/(2\kappa) > \|\boldsymbol{H}\| \tag{49}$$

as long as $\|\boldsymbol{H}\| < 1/(2\kappa)$. In view of the Neumann trick (Theorem 2), we can derive

$$\begin{aligned}
&\left|\boldsymbol{a}^\top \boldsymbol{u}_l - \sum_{j=1}^r \frac{\lambda^\star_j \boldsymbol{u}^{\star\top}_j \boldsymbol{u}_l}{\lambda_l} \boldsymbol{a}^\top \boldsymbol{u}^\star_j\right| \qquad (50)\\
&\quad = \left|\sum_{j=1}^r \frac{\lambda^\star_j}{\lambda_l}(\boldsymbol{u}^{\star\top}_j \boldsymbol{u}_l)\left\{\sum_{s=1}^\infty \frac{1}{\lambda^s_l} \boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}^\star_j\right\}\right| \\
&\quad \le \left(\sum_{j=1}^r \frac{|\lambda^\star_j|}{|\lambda_l|}|\boldsymbol{u}^{\star\top}_j \boldsymbol{u}_l|\right)\left\{\max_{1\le j\le r}\sum_{s=1}^\infty \frac{1}{|\lambda_l|^s}|\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}^\star_j|\right\} \\
&\quad \le \sqrt{r\sum_{j=1}^r |\boldsymbol{u}^{\star\top}_j \boldsymbol{u}_l|^2}\left\{\max_{1\le j\le r}\frac{|\lambda^\star_j|}{|\lambda_l|}\right\}\left\{\max_{1\le j\le r}\sum_{s=1}^\infty \frac{1}{|\lambda_l|^s}|\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}^\star_j|\right\} \\
&\quad \le \sqrt{r}\cdot\frac{1}{|\lambda_l|}\cdot\left\{\max_{1\le j\le r}\sum_{s=1}^\infty \frac{1}{|\lambda_l|^s}|\boldsymbol{a}^\top \boldsymbol{H}^s \boldsymbol{u}^\star_j|\right\}, \qquad (51)
\end{aligned}$$

where the third line follows since $\sum_{j=1}^r |\boldsymbol{u}^{\star\top}_j \boldsymbol{u}_l|^2 \le \|\boldsymbol{u}_l\|_2^2 = 1$, and the last inequality makes use of (49). Apply Corollary 4 to reach

$$(51) \le \frac{\sqrt{r}}{|\lambda_l|}\sum_{s=1}^\infty \big(2c_2\kappa \max\{B\log n, \sqrt{n\sigma^2\log n}\}\big)^s \sqrt{\frac{\mu}{n}}$$

$$\lesssim \frac{\kappa}{|\lambda_l|}\max\{B\log n, \sqrt{n\sigma^2\log n}\}\sqrt{\frac{\mu r}{n}}$$
$$\lesssim \kappa^2\max\{B\log n, \sqrt{n\sigma^2\log n}\}\sqrt{\frac{\mu r}{n}},$$

with the proviso that $|\lambda_l| > 1/(2\kappa)$ and $\max\{B\log n, \sqrt{n\sigma^2\log n}\} \le c_1/\kappa$ 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 (34).

6. Discussions. In this paper, we demonstrate the remarkable advantage of eigendecomposition over SVD in the presence of asymmetric noise matrices. This is in stark contrast to conventional wisdom, which is generally not in favor of eigendecomposition for asymmetric matrices. Our results only reflect the tip of an iceberg, and there are many outstanding issues left answered. 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, for example, 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 $\boldsymbol{M}$. 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.

A challenging signal-to-noise ratio regime. Take the rank-1 case for example: the present work focuses on the regime where $\|\boldsymbol{H}\| \lesssim \|\boldsymbol{M}^\star\|/\sqrt{\log n}$, and it is known that spectral methods fail to yield reliable estimation if $\|\boldsymbol{H}\| \gg \|\boldsymbol{M}^\star\|$. There is, however, a "gray" region (which includes, e.g., the case with $\|\boldsymbol{H}\| \approx \|\boldsymbol{M}^\star\|$) that has not been addressed. Developing nonasymptotic yet informative perturbation bounds for this regime is likely very challenging and requires new analysis techniques, which we leave for future investigation.

Correlated noise. The current theoretical development relies heavily on the assumption that the noise matrix $\boldsymbol{H}$ 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. (2019), 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 (Bai and Yao (2008), Bao, Ding and Wang (2018), Cai, Han and Pan (2017), Cape, Tang and Priebe (2019), Chen et al. (2019c), Johnstone (2001), Xia (2019)). Can we use the current framework to characterize the distributions of the leading eigenvalues as well as certain linear forms of the eigenvectors of $\boldsymbol{M}$ when the noise matrix is nonsymmetric?

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 (Cai et al. (2019), Zhang and Xia (2018)). Here, we conclude the paper with an example in *covariance estimation* (Baik and Silverstein (2006), Fan, Wang and Zhong (2017)). Imagine that we observe a collection of n independent Gaussian vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$, which have mean zero and covariance matrix

$$\boldsymbol{\Sigma}^\star = \boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{I}_d \tag{52}$$

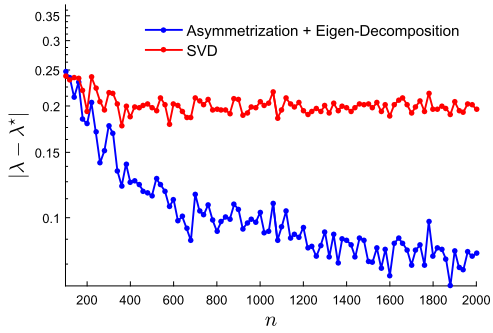


FIG. 5. *Numerical experiments for the spiked covariance model, where the sample vectors are zero-mean Gaussian vectors with covariance matrix* $\mathbf{\Sigma}^\star$ (*cf.* (52)). *We plot* $|\lambda - \lambda^\star|$ *versus* n *with* $d = n/10$. *The blue* (*resp.*, *red*) *lines represent the average errors over* 100 *independent trials when* λ *is the leading eigenvalue of* $\widehat{\mathbf{\Sigma}}_{\mathsf{asym}}$ (*resp.*, $\hat{\mathbf{\Sigma}}$).

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

$$\widehat{\mathbf{\Sigma}}_{\mathsf{asym}} = \frac{2}{n}\left(\sum_{i=1}^{n/2} \mathsf{Upper}\big(\boldsymbol{X}_i \boldsymbol{X}_i^\top\big) + \sum_{i=n/2+1}^{n} \mathsf{Lower}\big(\boldsymbol{X}_i \boldsymbol{X}_i^\top\big)\right), \tag{53}$$

where $\mathsf{Upper}(\cdot)$ (resp., $\mathsf{Lower}(\cdot)$) extracts out the upper (resp., lower) triangular part of the matrix, including (resp., excluding) the diagonal entries. As can be seen from Figure 5, the largest eigenvalue of the asymmetrized $\widehat{\mathbf{\Sigma}}_{\mathsf{asym}}$ is much closer to the true spectral norm of $\mathbf{\Sigma}^\star$, compared to the largest singular value of the sample covariance matrix $\hat{\mathbf{\Sigma}}$. We leave the theoretical understanding of such findings to future investigation.

Acknowledgments Y. Chen is supported in part by the AFOSR YIP award FA9550-19-1-0030, by the ARO grants W911NF-20-1-0097 and W911NF18-1-0303, by the ONR grant N00014-19-1-2120, by the NSF grants CCF-190766, IIS-1900140 and DMS-2014279.

J. Fan is supported in part by the NSF grants DMS-1662139 and DMS-1712591, by the ONR grant N00014-19-1-2120, and by the NIH grant 2R01-GM072611-13.

C. Cheng is supported in part by the Elite Undergraduate Training Program of School of Mathematical Sciences in Peking University, and by the William R. Hewlett Stanford graduate fellowship. We thank Cong Ma for helpful discussions, and Zhou Fan for showing us an example of asymmetrizing the Gaussian matrix.

Y. Chen is the corresponding author.

SUPPLEMENTARY MATERIAL

Additional proofs (DOI: 10.1214/20-AOS1963SUPP; .pdf). Additional proofs of the results in the paper can be found in the Supplementary Material.

REFERENCES

ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053 https://doi.org/10.1214/07-AIHP118

BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

BAO, Z., DING, X. and WANG, K. (2018). Singular vector and singular subspace distribution for the matrix denoising model. Preprint. Available at arXiv:1809.10476.

BAUER, F. L. and FIKE, C. T. (1960). Norms and exclusion theorems. *Numer. Math.* **2** 137–141. MR0118729 https://doi.org/10.1007/BF01386217

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** 120–135. MR2944410 https://doi.org/10.1016/j.jmva.2012.04.019

BENAYCH-GEORGES, F. and ROCHET, J. (2016). Outliers in the single ring theorem. *Probab. Theory Related Fields* **165** 313–363. MR3500273 https://doi.org/10.1007/s00440-015-0632-x

BORDENAVE, C. and CAPITAINE, M. (2016). Outlier eigenvalues for deformed i.i.d. random matrices. *Comm. Pure Appl. Math.* **69** 2131–2194. MR3552011 https://doi.org/10.1002/cpa.21629

BRÉZIN, E. and ZEE, A. (1998). Non-Hermitean delocalization: Multiple scattering and bounds. *Nuclear Phys. B* **509** 599–614. MR1487025 https://doi.org/10.1016/S0550-3213(97)00652-4

BRYC, W. and SILVERSTEIN, J. W. (2018). Singular values of large non-central random matrices. Preprint. Available at arXiv:1802.02960.

CAI, T., HAN, X. and PAN, G. (2017). Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices. Preprint. Available at arXiv:1711.00217.

CAI, T. T. and ZHANG, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. MR3766946 https://doi.org/10.1214/17-AOS1541

CAI, C., LI, G., CHI, Y., POOR, H. V. and CHEN, Y. (2019). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. Preprint. Available at arXiv:1910.04267.

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

CAPE, J., TANG, M. and PRIEBE, C. E. (2019). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. MR3912394 https://doi.org/10.1093/biomet/asy070

CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 https://doi.org/10.1214/08-AOP394

CHEN, Y., CHENG, C. and FAN, J. (2020). Supplement to "Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices." https://doi.org/10.1214/20-AOS1963SUPP

CHEN, Y., CHI, Y., FAN, J., MA, C. and YAN, Y. (2019a). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. Available at arXiv:1902.07698.

CHEN, Y., FAN, J., MA, C. and WANG, K. (2019b). Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.* **47** 2204–2235. MR3953449 https://doi.org/10.1214/18-AOS1745

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019c). Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116

CHENG, C., WEI, Y. and CHEN, Y. (2020). Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. Preprint. Available at arXiv:2001.04620.

CHI, Y., LU, Y. M. and CHEN, Y. (2019). Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Trans. Signal Process.* **67** 5239–5269. MR4016283 https://doi.org/10.1109/TSP.2019.2937282

DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450 https://doi.org/10.1137/0707001

ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: Spectral analysis beyond Davis–Kahan. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 38. Proceedings of Machine Learning Research PMLR. MR3857310

ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.* **41** 2279–2375. MR3098073 https://doi.org/10.1214/11-AOP734

FAN, J., WANG, W. and ZHONG, Y. (2017). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** Paper No. 207, 42. MR3827095

FEINBERG, J. and ZEE, A. (1997). Non-Hermitian random matrix theory: Method of Hermitian reduction. *Nuclear Phys. B* **504** 579–608. MR1488584 https://doi.org/10.1016/S0550-3213(97)00502-6

FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807 https://doi.org/10.1007/s00220-007-0209-3

FÜREDI, Z. and KOMLÓS, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241. MR0637828 https://doi.org/10.1007/BF02579329

JAIN, P. and NETRAPALLI, P. (2015). Fast exact matrix completion with finite samples. In *Conference on Learning Theory* 1007–1034.

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **56** 2980–2998. MR2683452 https://doi.org/10.1109/TIT.2010.2046205

KHORUZHENKO, B. (1996). Large-N eigenvalue distribution of randomly perturbed asymmetric matrices. *J. Phys. A* **29** L165–L169. MR1395506 https://doi.org/10.1088/0305-4470/29/7/003

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

KOLTCHINSKII, V. and XIA, D. (2016). Perturbation of linear forms of singular vectors under Gaussian noise. In *High Dimensional Probability VII. Progress in Probability* **71** 397–423. Springer, Cham. MR3565274 https://doi.org/10.1007/978-3-319-40519-3_18

LI, X., LING, S., STROHMER, T. and WEI, K. (2019). Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Appl. Comput. Harmon. Anal.* **47** 893–934. MR3994997 https://doi.org/10.1016/j.acha.2018.01.001

LYTOVA, A. and TIKHOMIROV, K. (2018). On delocalization of eigenvectors of random non-Hermitian matrices. Preprint. Available at arXiv:1810.01590.

MA, C., WANG, K., CHI, Y. and CHEN, Y. (2020). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. *Found. Comput. Math.* **20** 451–632.

MEHLIG, B. and CHALKER, J. T. (1998). Eigenvector correlations in non-Hermitian random matrix ensembles. *Ann. Phys.* **7** 427–436. MR1666250 https://doi.org/10.1002/(SICI)1521-3889(199811)7:5/6<427::AID-ANDP427>3.0.CO;2-1

O'ROURKE, S., VU, V. and WANG, K. (2016). Eigenvectors of random matrices: A survey. *J. Combin. Theory Ser. A* **144** 361–442. MR3534074 https://doi.org/10.1016/j.jcta.2016.06.008

O'ROURKE, S., VU, V. and WANG, K. (2018). Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra Appl.* **540** 26–59. MR3739989 https://doi.org/10.1016/j.laa.2017.11.014

PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787 https://doi.org/10.1007/s00440-005-0466-z

RAJAGOPALAN, A. B. (2015). Outlier eigenvalue fluctuations of perturbed iid matrices. Preprint. Available at arXiv:1507.01441.

RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices II: Delocalized perturbations. *Random Matrices Theory Appl.* **2** 1250015, 36. MR3039820 https://doi.org/10.1142/S2010326312500153

SILVERSTEIN, J. W. (1994). The spectral radii and norms of large-dimensional non-central random matrices. *Comm. Statist. Stochastic Models* **10** 525–532. MR1284550 https://doi.org/10.1080/15326349408807308

SOMMERS, H.-J., CRISANTI, A., SOMPOLINSKY, H. and STEIN, Y. (1988). Spectrum of large random asymmetric matrices. *Phys. Rev. Lett.* **60** 1895–1898. MR0948613 https://doi.org/10.1103/PhysRevLett.60.1895

TAO, T. (2012). *Topics in Random Matrix Theory. Graduate Studies in Mathematics* **132**. Amer. Math. Soc., Providence, RI. MR2906465 https://doi.org/10.1090/gsm/132

TAO, T. (2013). Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields* **155** 231–263. MR3010398 https://doi.org/10.1007/s00440-011-0397-9

TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.

VU, V. (2011). Singular vectors under random perturbation. *Random Structures Algorithms* **39** 526–538. MR2846302 https://doi.org/10.1002/rsa.20367

VU, V. and WANG, K. (2015). Random weighted projections, random quadratic forms and random eigenvectors. *Random Structures Algorithms* **47** 792–821. MR3418916 https://doi.org/10.1002/rsa.20561

WANG, R. (2015). Singular vector perturbation under Gaussian noise. *SIAM J. Matrix Anal. Appl.* **36** 158–177. MR3310977 https://doi.org/10.1137/130938177

WEDIN, P. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111. MR0309968 https://doi.org/10.1007/bf01932678

XIA, D. (2016). Statistical inference for large matrices. Ph.D. thesis, Georgia Institute of Technology.

XIA, D. (2019). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Trans. Inf. Theory* **65** 7437–7459. MR4030894 https://doi.org/10.1109/TIT.2019.2924900

YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344 https://doi.org/10.1007/BF00353874

ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377