The Annals of Statistics
2021, Vol. 49, No. 1, 370–392
https://doi.org/10.1214/20-AOS1960

SINGULAR VECTOR AND SINGULAR SUBSPACE DISTRIBUTION FOR THE MATRIX DENOISING MODEL

BY ZHIGANG BAO[1,*], XIUCAI DING[2] AND AND KE WANG[1,†]

[1]*Department of Mathematics, Hong Kong University of Science and Technology*, [*]*mazgbao@ust.hk*; [†]*kewang@ust.hk*
[2]*Department of Statistics, University of California, Davis*, *xcading@ucdavis.edu*

In this paper, we study the matrix denoising model $Y = S + X$, where S is a low rank deterministic signal matrix and X is a random noise matrix, and both are $M \times n$. In the scenario that M and n are comparably large and the signals are supercritical, we study the fluctuation of the outlier singular vectors of Y, under fully general assumptions on the structure of S and the distribution of X. More specifically, we derive the limiting distribution of angles between the principal singular vectors of Y and their deterministic counterparts, the singular vectors of S. Further, we also derive the distribution of the distance between the subspace spanned by the principal singular vectors of Y and that spanned by the singular vectors of S. It turns out that the limiting distributions depend on the structure of the singular vectors of S and the distribution of X, and thus they are nonuniversal. Statistical applications of our results to singular vector and singular subspace inferences are also discussed.

1. Introduction.

Consider an $M \times n$ noisy matrix Y modeled as

$$Y = S + X, \tag{1.1}$$

where S is a low-rank deterministic matrix with fixed rank r and X is a real random noise matrix. We assume that S admits the singular value decomposition

$$S = UDV^* = \sum_{i=1}^{r} d_i \boldsymbol{u}_i \boldsymbol{v}_i^*, \tag{1.2}$$

where $D = \mathrm{diag}(d_1, \ldots, d_r)$ consists of the singular values of S and we assume $d_1 > \cdots > d_r > 0$; $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r) \in \mathbb{R}^{M\times r}$ and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r) \in \mathbb{R}^{n\times r}$ are the matrices consisting of the ℓ^2-normalized left and right singular vectors. For the noise matrix $X = (x_{ij})$ in (1.1), we assume that the entries x_{ij}'s are i.i.d. real random variables with

$$\mathbb{E}x_{ij} = 0, \qquad \mathbb{E}|x_{ij}|^2 = \frac{1}{n}. \tag{1.3}$$

For simplicity, we also assume the existence of all moments, that is, for every integer $q \geq 3$, there is some constant $C_q > 0$, such that

$$\mathbb{E}|\sqrt{n}x_{ij}|^q \leq C_q < \infty. \tag{1.4}$$

This condition can be weakened to the existence of some sufficiently high order moment. But we do not pursue this direction here. We remark that although we are primarily interested in the real case, our method also applies to the case when X is a complex noise matrix.

In practice, S is often called the *signal* matrix which contains the information of interest. In the high-dimensional setup, when M and n are comparably large, we are interested in the

Received December 2018; revised October 2019.
MSC2020 subject classifications. Primary 60B20, 62G10; secondary 62H10, 15B52, 62H25.
Key words and phrases. Random matrix, singular vector, singular subspace, matrix denoising model, signal-plus-noise model, nonuniversality.

inference of S or its left and right singular spaces, which are the subspaces spanned by $\boldsymbol{u}_i$'s or $\boldsymbol{v}_i$'s, respectively. Such a problem arises in many scientific applications such as matrix denoising [26, 28], multiple signal classification (MUSIC) [37, 62] and multidimensional scaling [31, 55]. We call the model in (1.1) the matrix denoising model, which is also known as the signal-plus-noise model in the literature. We refer to Section 1.2 for more introduction on the application aspects.

We denote the singular value decomposition of Y by

$$Y = \widehat{U}\Lambda\widehat{V}^* = \sum_{i=1}^{M\wedge n}\sqrt{\mu_i}\widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{v}}_i^*, \tag{1.5}$$

where $\mu_1 \geq \cdots \geq \mu_{M\wedge n}$ are the squares of the nontrivial singular values, and $\widehat{\boldsymbol{u}}_i$'s and $\widehat{\boldsymbol{v}}_i$'s are the ℓ^2-normalized sample singular vectors. Here, $\widehat{U} = (\widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_M)$ and $\widehat{V} = (\widehat{\boldsymbol{v}}_1, \dots, \widehat{\boldsymbol{v}}_n)$ and Λ is $M \times n$ with singular values on its main diagonal.

In this paper, we are interested in the distributions of the principal left and right singular vectors of Y and the subspaces spanned by them. On singular vectors, a natural quantity to look into is the projection of a sample principal singular vector onto its deterministic counterpart, that is, $|\langle\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i\rangle|$ and $|\langle\widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i\rangle|$, which characterizes the deviation of an original signal from the noisy one. On singular spaces, the natural estimators for U and V are their noisy counterparts

$$\widehat{U}_r = (\widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_r) \quad \text{and} \quad \widehat{V}_r = (\widehat{\boldsymbol{v}}_1, \dots, \widehat{\boldsymbol{v}}_r),$$

respectively, that is, the matrices consisting of the first r left and right singular vectors of Y, respectively. To measure the distance between $\widehat{U}_r$ and U, or $\widehat{V}_r$ and V, we consider the following matrix of the cosine principal angles between two subspaces (see [36], Section 6.4.3, for instance):

$$\cos\Theta(\widehat{V}_r, V) = \operatorname{diag}(\sigma_1^V, \dots, \sigma_r^V), \qquad \cos\Theta(\widehat{U}_r, U) = \operatorname{diag}(\sigma_1^U, \dots, \sigma_r^U),$$

where σ_i^V's and σ_i^U's are the singular values of the matrices $\widehat{V}_r^* V$ and $\widehat{U}_r^* U$, respectively. Therefore, an appropriate measure of the distance between the subspaces is $L := \|\cos\Theta(U, \widehat{U}_r)\|_F^2$ for the left singular subspace or $R := \|\cos\Theta(V, \widehat{V}_r)\|_F^2$ for the right singular subspace, where $\|\cdot\|_F^2$ stands for the Frobenius norm. Note that L and R can also be written as

$$L := \sum_{i,j=1}^{r}|\langle\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_j\rangle|^2 = \frac{1}{2}(2r - \|\widehat{U}_r\widehat{U}_r^* - UU^*\|_F^2), \tag{1.6}$$

$$R := \sum_{i,j=1}^{r}|\langle\widehat{\boldsymbol{v}}_i, \boldsymbol{v}_j\rangle|^2 = \frac{1}{2}(2r - \|\widehat{V}_r\widehat{V}_r^* - VV^*\|_F^2). \tag{1.7}$$

In this paper, we are interested in the following high-dimensional regime: for some small constant $\tau \in (0, 1)$, we have

$$M \equiv M(n), \qquad y \equiv y_n := \frac{M}{n} \to c \in [\tau, \tau^{-1}] \quad \text{as } n \to \infty. \tag{1.8}$$

Our main results are on the limiting distributions of individual $|\langle\widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i\rangle|^2$ (resp., $|\langle\widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i\rangle|^2$) and R (resp., L) when the signal strength, d_i's, are supercritical (cf. Assumption 2.1). They are detailed in Theorems 2.3, 2.9, after necessary notation are introduced. In the rest of this section, we review some related literature from both theoretical and applied perspectives.

1.1. *On finite-rank deformation of random matrices.* From the theoretical perspective, our model in (1.1) is in the category of the fixed-rank deformation of the random matrix models in the random matrix theory, which also includes the deformed Wigner matrix and the spiked sample covariance matrix as typical examples. There are a vast of work devoted to this topic and the primary interest is to investigate the limiting behavior of the extreme eigenvalues and the associated eigenvectors of the deformed models. Since the seminal work of Baik, Ben Arous and Péché [5], it is now well understood that the extreme eigenvalues undergo a so-called BBP transition along with the change of the strength of the deformation. Roughly speaking, there is a critical value such that the extreme eigenvalue of the deformed matrix will stick to the right end point of the limiting spectral distribution of the undeformed random matrix if the strength of the deformation is less than or equal to the critical value, and will otherwise jump out of the support of the limiting spectral distribution. In the latter case, we call the extreme eigenvalue as an outlier, and the associated eigenvector as an outlier eigenvector. Moreover, the fluctuation of the extreme eigenvalues in different regimes (subcritical, critical and supercritical) are also identified in [5] for the complex spiked covariance matrix. We also refer to [4, 6, 12, 13, 22, 26, 27, 41, 53] and the reference therein for the first-order limit of the extreme eigenvalue of various fixed-rank deformation models. The fluctuation of the extreme eigenvalues of various models have been considered in [3, 4, 9–11, 16, 17, 24, 25, 29, 35, 41, 42, 53, 54, 57]. Especially, the fluctuations of the outliers are shown to be nonuniversal for the deformed Wigner matrices, first in [24] under certain special assumptions on the structure of the deformation and the distribution of the matrix entries, and then in [41] in full generality.

The study on the behavior of the extreme eigenvectors has been mainly focused on the level of the first-order limit [12, 13, 21, 26, 34, 53]. In parallel to the results of the extreme eigenvalues, it is known that the eigenvectors are delocalized in the subcritical case and have a bias on the direction of the deformation in the supercritical case. It is recently observed in [15] that a deformation close to the critical regime will cause a bias even for the nonoutlier eigenvectors. On the level of the fluctuation, the limiting behavior of the extreme eigenvectors has not been fully studied yet. By establishing a general universality result of the eigenvectors of the sample covariance matrix in the null case, the authors of [15] are able to show that the law of the eigenvectors of the spiked covariance matrices are asymptotically Gaussian in the subcritical regime. More specifically, the generalized components of the eigenvectors (i.e., $\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{w} \rangle$ for any deterministic vector $\boldsymbol{w}$) are χ^2 distributed. For spiked Gaussian sample covariance matrices, in the supercritical regime, the fluctuation of a fixed-dimensional normalized subvector of the outlier eigenvector is proved to be Gaussian in [53], but this result cannot tell the distribution of $\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i \rangle$. Under some special assumptions on the structure of the deformation and the distribution of the random matrix entries, it is shown in [23] that the eigenvector distribution of a generalized deformed Wigner matrix model is nonuniversal in the supercritical regime. In the current work, we aim at establishing the nonuniversality for the outlier singular vectors for the matrix denoising model under fully general assumptions on the structure of the deformation S and the distribution of the random matrix X. This can be regarded as an eigenvector counterpart of the result on the outlying eigenvalue distribution in [41].

1.2. *On singular subspace inference.* From the applied perspective, our model (1.1) appears prominently in the study of signal processing [40, 51], machine learning [58, 61] and statistics [18, 19, 28, 33]. For instance, in the study of image denoising, S is treated as the true image [49] and in the problem of classification, S contains the underlying true mean vectors of samples [18]. In both situations, we need to understand the asymptotics of the singular vectors and subspace of S, given the observation Y. In addition, the statistics R and L

defined in (1.7) can be used for the inference of the structure of the singular subspace of S. We remark that these statistics have been used extensively to explore the properties of singular subspace. To name a few, in [39], the authors studied the problem of testing whether the sample singular subspace is equal to some given subspace; in [20], the authors studied the eigenvector inference problems for the correlated stochastic block model; in [38], the authors analyzed the impact of dimensionality reduction for subspace clustering algorithms; and in [18], the authors studied the high-dimensional clustering problem and the canonical correlation analysis. In the high-dimensional regime (1.8), to the best of our knowledge, the distributions of R and L have not been studied yet in the literature.

In the situation when M is fixed, the sample eigenvectors of XX^* are normally distributed [1]. When M diverges with n, many interesting results have been proposed under various assumptions. One line of the work is to derive the perturbation bounds for the perturbed singular vectors based on Davis–Kahan's theorem. For instance, in [52], the authors improve the perturbation bounds of Davis–Kahan theorem to be nearly optimal. In [18], the authors study similar problems and their related statistical applications. Most recently, in the papers [32, 33, 63], the authors derive the ℓ^∞ pertubation bounds assuming that the population vectors were delocalized (i.e., incoherent). The other line of the work is to study the asymptotic normality of the spectral projection under various regularity conditions. In such cases, the singular vectors of S can be estimated using those of Y and some Gaussian approximation technique can be employed. Considering the Gaussian data samples $\boldsymbol{x}_i \simeq \mathcal{N}(\mathbf{0}, \Sigma)$, $i = 1, 2, \ldots, n$ and $X = (\boldsymbol{x}_i)$, under the assumption that the order of $\frac{\operatorname{Tr}\Sigma}{\|\Sigma\|}$ is much smaller than n, in [44–46], the authors prove that the eigenvectors of XX^* are asymptotically normally distributed, whose variance depends the eigenvectors of Σ. Furthermore, in [59], assuming that m such random matrices X_i, $i = 1, 2, \ldots, m$ are available, the author shows that the singular vectors of S can be estimated via trace regression using matrix nuclear norm penalized least squares estimation (NNPLS). Under the assumption that $r^4 K \log^3 m = o(m)$, $K = \max\{M, n\}$, the author shows that the principal angles of the subspace estimated using NNPLS are asymptotically normal.

1.3. *Organization.* The rest of the paper is organized as follows. In Section 2, we state our main results and summarize our method for the proofs. In Section 3, we design Monte Carlo simulations to demonstrate the accuracy of our main results and briefly illustrate their applications through a hypothesis testing problem. In Section 4, we introduce some main technical results including the isotropic local law and also derive the Green function representation for our statistics. In Section 5, we prove Theorems 2.3, based on the recursive estimate in Proposition 5.2. We state more simulation results, further discussions of statistical applications, the proofs of Theorem 2.9 and some technical lemmas in the Supplementary Material [8].

2. Main results and methodology. In this section, we state our main results, and briefly summarize our proof strategy.

2.1. *Main results.* In this paper, the singular values of S are assumed to satisfy the following *supercritical* condition.

ASSUMPTION 2.1 (Supercritical condition). There exist a constant $C > 0$ and a (small) constant $\delta > 0$, such that

$$y^{1/4} + \delta \le d_r < \cdots < d_2 < d_1 \le C, \qquad \min_{1 \le j \ne i \le r} |d_i - d_j| \ge \delta.$$

REMARK 2.2. The first inequality above ensures that the first r singular values of Y are outliers, and the threshold $y^{1/4}$ is the analogous BBP transition point in [5]. The second inequality guarantees that the outliers of Y are well separated from each other. We also assume that $d_1, \dots, d_r$ are bounded by some constant C. All these conditions can be weakened. For instance, we do allow the existence of the subcritical and critical d_i's if we only focus on the outlier singular vectors. Also, the separation of d_i's by an order 1 distance δ is not necessary. In [15], a much weaker separation of order $n^{-1/2+\epsilon}$ is enough for the discussion of the eigenvalues. Moreover, we can also extend our results to the case when $d_1, \dots, d_r$ diverge with n. But we do not pursue these directions in the current paper.

In the sequel, we will only state the results for the right singular vectors and the right singular subspace. The results for the left ones can be obtained from the right ones by simply considering the transpose (with a rescaling) of our matrix model in (1.1). To state our results, we need more notation. First, we define

$$p(d) := \frac{(d^2+1)(d^2+y)}{d^2}. \tag{2.1}$$

For each $i \in [r]$, we will write $p_i \equiv p(d_i)$ for short. Recall (1.5). In [26], Theorem 3.4, it has been shown that p_i is the limit of μ_i. Further, we set

$$a(d) := \frac{d^4 - y}{d^2(d^2+1)}. \tag{2.2}$$

It has been proved in [26] that $a(d_i)$ are the limits of $|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i \rangle|^2$, respectively (see Lemma D.1 in [8]). We also denote by κ_l the lth cumulant of the random variables $\sqrt{n}x_{ij}$. For a vector $\boldsymbol{w} = (w(1), \dots, w(m))^T$ and $l \in \mathbb{N}$, we introduce the notation

$$\boldsymbol{s}_l(\boldsymbol{w}) := \sum_{i=1}^m w(i)^l.$$

Set

$$\theta(d) := \frac{d^4 + 2yd^2 + y}{d^3(d^2+1)^2}, \qquad \psi(d) := \frac{d^6 - 3yd^2 - 2y}{d^3(d^2+1)^2} \tag{2.3}$$

and

$$\begin{aligned} \mathcal{V}^E(d) := \frac{2}{d^4 - y} \biggl(& 2y(y+1)\theta(d)^2 - \frac{y(y-1)(5y+1)}{d(d^2+1)^2}\theta(d) \\ & + \frac{(d^4+y)(d^2+y)^2}{d^3(d^2+1)^2}\psi(d) + \frac{2y^2(y-1)^2}{d^2(d^2+1)^4} \biggr). \end{aligned} \tag{2.4}$$

For the right singular vectors, we have the following theorem.

THEOREM 2.3 (Right singular vectors). *Assume that* (1.3), (1.4), (1.8) *and Assumption* 2.1 *hold. For $i \in [r]$, define the random variable*

$$\Delta_i := -2\sqrt{n}\theta(d_i)\boldsymbol{u}_i^* X \boldsymbol{v}_i - \frac{2\psi(d_i)}{d_i^2}\biggl(\frac{\kappa_3}{n}\boldsymbol{s}_1(\boldsymbol{u}_i)\boldsymbol{s}_1(\boldsymbol{v}_i)\biggr), \tag{2.5}$$

and let $\mathcal{Z}_i \sim \mathcal{N}(0, \mathcal{V}_i)$ be a random variable independent of Δ_i, where

$$\begin{aligned} \mathcal{V}_i := {} & \mathcal{V}^E(d_i) - \frac{4}{d_i}\theta(d_i)\psi(d_i)\biggl(\frac{\kappa_3}{\sqrt{n}}\boldsymbol{s}_3(\boldsymbol{u}_i)\boldsymbol{s}_1(\boldsymbol{v}_i)\biggr) + \frac{4}{d_i}\theta(d_i)^2\biggl(\frac{\kappa_3}{\sqrt{n}}\boldsymbol{s}_1(\boldsymbol{u}_i)\boldsymbol{s}_3(\boldsymbol{v}_i)\biggr) \\ & + \frac{\psi(d_i)^2}{d_i^2}\kappa_4\boldsymbol{s}_4(\boldsymbol{u}_i) + \frac{y\theta(d_i)^2}{d_i^2}\kappa_4\boldsymbol{s}_4(\boldsymbol{v}_i). \end{aligned}$$

Then for any $i \in [r]$ and any bounded continuous function f, we have

$$\lim_{n\to\infty}\big(\mathbb{E}f\big(\sqrt{n}\big(|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2 - a(d_i)\big)\big) - \mathbb{E}f(\Delta_i + \mathcal{Z}_i)\big) = 0.$$

REMARK 2.4. In [41], the authors obtain the nonuniversality for the limiting distributions of the outliers (outlying eigenvalues) of the deformed Wigner matrices. The limiting distributions admit similar forms as the limiting distribution for the outlier singular vectors for our models. One might notice that the third or the fourth cumulants of the entries of the Wigner matrices are allowed to be different in [41]. An extension along this direction is also straightforward for our result.

We discuss a few special cases of interest. For simplicity, we assume that S has rank $r = 1$ and drop all the subindices.

REMARK 2.5. If the entries of $\sqrt{n}X$ are standard Gaussian random variables (i.e., $\kappa_3 = \kappa_4 = 0$), then $\Delta \simeq \mathcal{N}(0, 4\theta(d)^2)$ (see Definition 4.9 for the meaning of $\simeq$). Hence, we find $\Delta + \mathcal{Z}$ is asymptotically distributed as

$$\mathcal{N}\big(0, 4\theta(d)^2 + \mathcal{V}^E(d)\big).$$

REMARK 2.6. If both $\boldsymbol{u}$ and $\boldsymbol{v}$ are delocalized in the sense that $\|\boldsymbol{u}\|_\infty = o(1)$ and $\|\boldsymbol{v}\|_\infty = o(1)$. Then $\boldsymbol{s}_l(\boldsymbol{u}) = o(1)$ and $\boldsymbol{s}_l(\boldsymbol{v}) = o(1)$ for $l = 3, 4$. By (1.3), (1.4) and the fact $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1$, we find that $\mathbb{E}(\boldsymbol{u}^* X\boldsymbol{v}) = 0$ and $\mathbb{E}(\boldsymbol{u}^* X\boldsymbol{v})^2 = n^{-1}$. Then we conclude from Lyapunov's CLT for triangular array that

$$\Delta \simeq \mathcal{N}\bigg(-\frac{2\psi(d)}{d^2}\bigg(\frac{\kappa_3}{n}\boldsymbol{s}_1(\boldsymbol{u})\boldsymbol{s}_1(\boldsymbol{v})\bigg), 4\theta(d)^2\bigg), \tag{2.6}$$

and therefore $\Delta + \mathcal{Z}$ has asymptotically the same distribution as

$$\mathcal{N}\bigg(-\frac{2\psi(d)}{d^2}\bigg(\frac{\kappa_3}{n}\boldsymbol{s}_1(\boldsymbol{u})\boldsymbol{s}_1(\boldsymbol{v})\bigg), 4\theta(d)^2 + \mathcal{V}^E(d)\bigg).$$

The only difference from the Gaussian case is a shift caused by the nonvanishing third cumulant.

REMARK 2.7. If one of $\boldsymbol{u}$ and $\boldsymbol{v}$ is delocalized, say $\|\boldsymbol{u}\|_\infty = o(1)$, then Δ still has the limiting distribution in (2.6). Therefore, $\Delta + \mathcal{Z}$ has asymptotically the same distribution as a Gaussian random variable with mean

$$-\frac{2\psi(d)}{d^2}\bigg(\frac{\kappa_3}{n}\boldsymbol{s}_1(\boldsymbol{u})\boldsymbol{s}_1(\boldsymbol{v})\bigg)$$

and variance

$$4\theta(d)^2 + \mathcal{V}^E(d) + \frac{4}{d}\theta(d)^2\bigg(\frac{\kappa_3}{\sqrt{n}}\boldsymbol{s}_1(\boldsymbol{u})\boldsymbol{s}_3(\boldsymbol{v})\bigg) + y\frac{\theta(d)^2}{d^2}\kappa_4\boldsymbol{s}_4(\boldsymbol{v}).$$

REMARK 2.8. If neither $\boldsymbol{u}$ nor $\boldsymbol{v}$ is delocalized, then $\Delta + \mathcal{Z}$ is no longer Gaussian in general. For example, if $\boldsymbol{u} = \boldsymbol{e}_1$ and $\boldsymbol{v} = \boldsymbol{f}_1$ where $\boldsymbol{e}_1$ and $\boldsymbol{f}_1$ are the canonical basis vectors in $\mathbb{R}^M$ and $\mathbb{R}^n$ respectively, then $\Delta + \mathcal{Z}$ is asymptotically distributed as

$$-2\theta(d)\sqrt{n}X_{11} + \mathcal{N}\bigg(0, \mathcal{V}^E(d) + \kappa_4\frac{\psi(d)^2 + y\theta(d)^2}{d^2}\bigg),$$

which depends on the distribution of X_{11}, and thus is nonuniversal.

If the assumptions of Theorem 2.3 hold, we conclude from Remarks 2.6–2.8 that $|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2$ always has a Gaussian fluctuation if either the entries of X are Gaussian or one of $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ is delocalized in the sense $\|\boldsymbol{u}_i\|_\infty = o(1)$ or $\|\boldsymbol{v}_i\|_\infty = o(1)$. In the general setting when the noise matrix is non-Gaussian, the detailed distribution will rely on both the structure of the singular vectors and the noise matrix X.

Next, we study the distributions of the right singular space. For two vectors $\boldsymbol{w}_a = (w_a(1), \dots, w_a(m))^T$, $a = 1, 2$, we denote

$$\boldsymbol{s}_{k,l}(\boldsymbol{w}_1, \boldsymbol{w}_2) := \sum_{i=1}^{m} w_1(i)^k w_2(i)^l.$$

Recall R from (1.7). We have the following theorem.

THEOREM 2.9 (Right singular subspace). *Assume that* (1.3), (1.4), (1.8) *and Assumption* 2.1 *hold. Let* $\Delta = \sum_{i=1}^r \Delta_i$, *where* Δ_i *is defined in* (2.5). *Let* $\mathcal{Z}$ *be a random variable independent of* Δ *with law* $\mathcal{Z} \sim \mathcal{N}(0, \mathcal{V})$, *where*

$$\begin{aligned}\mathcal{V} := {} & \sum_{i=1}^{r} \mathcal{V}^E(d_i) + \kappa_4 \sum_{i,j=1}^{r} \left(\frac{\psi(d_i)\psi(d_j)}{d_i d_j} \boldsymbol{s}_{2,2}(\boldsymbol{u}_i, \boldsymbol{u}_j) + y \frac{\theta(d_i)\theta(d_j)}{d_i d_j} \boldsymbol{s}_{2,2}(\boldsymbol{v}_i, \boldsymbol{v}_j) \right) \\ & + \frac{\kappa_3}{\sqrt{n}} \sum_{i,j=1}^{r} \frac{4}{d_i} \theta(d_j) \big(\theta(d_i) \boldsymbol{s}_{2,1}(\boldsymbol{v}_i, \boldsymbol{v}_j) \boldsymbol{s}_1(\boldsymbol{u}_j) - \psi(d_i) \boldsymbol{s}_{2,1}(\boldsymbol{u}_i, \boldsymbol{u}_j) \boldsymbol{s}_1(\boldsymbol{v}_j) \big).\end{aligned}$$

Then for any bounded continuous function f, *we have that*

$$\lim_{n\to\infty} \left(\mathbb{E} f\left(\sqrt{n} \left(R - \sum_{i=1}^{r} a(d_i) \right) \right) - \mathbb{E} f(\Delta + \mathcal{Z}) \right) = 0.$$

2.2. *Proof strategy.* In this subsection, we briefly describe our proof strategy. We first review the method used in a related work [41], and then we highlight the novelty of our strategy.

As we mentioned, in [41], the authors derive the distribution of outliers (outlying eigenvalues) of the fixed-rank deformation of Wigner matrices. The main technical input is the isotropic local law for Wigner matrices, which provides a precise large deviation estimate for the quadratic form $\langle \boldsymbol{u}, (W - z)^{-1} \boldsymbol{v} \rangle$ for any deterministic vectors $\boldsymbol{u}, \boldsymbol{v}$. Here, W is a Wigner matrix. It turns out that an outlier of the deformed Wigner matrix can also be approximated by a quadratic form of the Green function, of the form $\langle \boldsymbol{u}, (W - z)^{-1} \boldsymbol{u} \rangle$. So one can turn to establish the law of the quadratic form of the Green function instead. In [41], the authors decompose the proof into three steps. First, the law is established for the GOE/GUE, the Gaussian Wigner matrix, for which orthogonal/unitary invariance of the matrix can be used to facilitate the proof. In the second step of going beyond Gaussian matrix, in order to capture the independence of the Gaussian part and the non-Gaussian part of the limiting distribution of the outliers, the authors construct an intermediate matrix in which most of the matrix entries are replaced by the Gaussian ones while those with coordinates corresponding to the large components of $\boldsymbol{u}$ are kept as generally distributed. The intermediate matrix allows one to use the nice properties of the Gaussian ensembles such as orthogonal/unitary invariance for the major part of the matrix, and meanwhile keeps the non-Gaussianity induced by the small amount of generally distributed entries. In the last step, the authors of [41] derive the law for the fully generally distributed Wigner matrix by further conducting a Green function comparison with the intermediate matrix.

For our problem, similarly, we will use the isotropic law of the sample covariance matrix in [14, 43] as a main technical input. It turns out that for the singular vectors, we can approximately represent $\sqrt{n}|\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_i\rangle|$ (after appropriate centering) in terms of a quantity of the form

$$\mathcal{Q}_i = \sqrt{n}\big(\operatorname{Tr}\big(G(p_i)\big) - \Pi_1(p_i)\big)A_i + \operatorname{Tr}\big(G'(p_i) - \Pi_1'(p_i)\big)B_i\big), \tag{2.7}$$

where G is the Green function of the linearization of the sample covariance matrix and Π_1 is the deterministic approximation of G; see (4.5) and (4.10) for the definitions. Here, both A_i and B_i are deterministic fixed-rank matrices. Hence, differently from the outlying eigenvalues or singular values, the Green function representation of the singular vectors also contains the derivative of the Green function. More importantly, instead of the three step strategy in [41], here we derive the law of the above $\mathcal{Q}_i$ directly for generally distributed matrix. Recall Δ_i defined in (2.5), whose random part is proportional to $\boldsymbol{u}_i^* X \boldsymbol{v}_i$, which is simply a linear combination of the entries of X. Inspired by [41], we decompose Δ_i into two parts, say $\widetilde{\Delta}_i$ and $\widehat{\Delta}_i$. The former contains the linear combination of $x_{k\ell}$'s for those indices k, ℓ corresponding to the large components u_{ik} and $v_{i\ell}$ in $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$. The latter contains the linear combinations of the rest of $x_{k\ell}$'s. Note that $\widehat{\Delta}_i$ is asymptotically normal by CLT since the coefficients of $x_{k\ell}$'s are small. However, $\widetilde{\Delta}_i$ may not be normal. The key idea of our strategy is to show the following recursive estimate: For any fixed $k \in \mathbb{N}$, we have

$$\mathbb{E}(\mathcal{Q}_i - \widetilde{\Delta}_i)^k \mathrm{e}^{\mathrm{i}t\widetilde{\Delta}_i} = (k-1)\widetilde{\mathcal{V}}_i \mathbb{E}(\mathcal{Q}_i - \widetilde{\Delta}_i)^{k-2} \mathrm{e}^{\mathrm{i}t\widetilde{\Delta}_i} + o(1), \tag{2.8}$$

for some positive number $\widetilde{\mathcal{V}}_i$. Choosing $t = 0$, we can derive the asymptotic normality of $\mathcal{Q}_i - \widetilde{\Delta}_i$ for (2.8) by the recursive moment estimate. Choosing t to be arbitrary, we can further deduce from (2.8) that

$$\mathbb{E}\mathrm{e}^{\mathrm{i}s(\mathcal{Q}_i - \widetilde{\Delta}_i) + \mathrm{i}t\widetilde{\Delta}_i} = \mathbb{E}\mathrm{e}^{\mathrm{i}s(\mathcal{Q}_i - \widetilde{\Delta}_i)} \mathbb{E}\mathrm{e}^{\mathrm{i}t\widetilde{\Delta}_i} + o(1).$$

Then asymptotic independence between $\mathcal{Q}_i - \widetilde{\Delta}_i$ and $\widetilde{\Delta}_i$ follows. Hence, we prove both the asymptotic normality and asymptotic independence from (2.8). The method of using the recursive estimate to get the large deviation bounds for Green function or some functional of the Green functions has been previously used in the context of the random matrix theory. For instance, we refer to [47]. However, as far as we know, it is the first time to use the recursive estimate to show the normality and the independence simultaneously for the functionals of the Green functions.

Moreover, we remark that the approach in this paper can also be applied to derive the distribution of the outlier eigenvectors of the spiked sample covariance matrix [7] and the deformed Wigner matrix.

Finally, we briefly compare the methods used in this paper and the related work [23]. In [23], the authors study the distribution of $|\langle \widehat{\boldsymbol{v}}, \boldsymbol{e}_1\rangle|^2$ of a deformed Wigner matrix whose deformation is a block diagonal deterministic Hermitian matrix containing one large spike $\theta \boldsymbol{e}_1 \boldsymbol{e}_1^*$ which creates one outlier of the deformed Wigner matrix. Here, $\widehat{\boldsymbol{v}}$ is the random outlier eigenvector. By the Helffer–Sjöstrand formula, they represent $|\langle \widehat{\boldsymbol{v}}, \boldsymbol{e}_1\rangle|^2$ in terms of an integral (over z) of $\boldsymbol{e}_1^*(W-z)^{-1}\boldsymbol{e}_1$. In contrast to our work, the major difference in [23] is that they establish the limiting distribution for the whole process $\boldsymbol{e}_1^*(W-z)^{-1}\boldsymbol{e}_1$ in z, and then use functional limit theorem to conclude the limit of the integral. In our work, relying on the isotropic law, we first integrate out the contour integral approximately. This results in the linear combination in (2.7), and then we only need to consider the joint distribution of the quadratic form of G and G' at a single point $p(d_i)$. Moreover, in [23], the authors decompose the quadratic form $\boldsymbol{e}_1^*(W-z)^{-1}\boldsymbol{e}_1$ into two parts using Schur's complement, where one of them can be proved to be Gaussian using an extension of the CLT for quadratic forms as in the previous work [24]. It is worth noticing that the independence between the Gaussian and non-Gaussian parts follows directly from the special structure of the model in [23]. However, in [41] and our work, since we do not have structural assumptions on S, we need to make more dedicated efforts for the independence (see [41], Proposition 7.12, and Proposition 5.1).

3. Simulations and statistical applications.

3.1. *Numerical simulations.* In this section, we present some numerical simulations for our results stated in Section 2.1. For the simulations, we consider two specific distributions for our noise matrix. We assume that $\sqrt{n}x_{ij}$'s are i.i.d. $\mathcal{N}(0,1)$ or i.i.d. with the distribution $\frac{1}{3}\delta_{\sqrt{2}} + \frac{2}{3}\delta_{-\frac{1}{\sqrt{2}}}$. We call these two types of noise as Gaussian noise and two-point noise, respectively. It is easy to check that the 3rd and 4th cumulants of the distribution $\frac{1}{3}\delta_{\sqrt{2}} + \frac{2}{3}\delta_{-\frac{1}{\sqrt{2}}}$ are $\kappa_3 = \frac{1}{\sqrt{2}}$ and $\kappa_4 = -\frac{3}{2}$. In the sequel, let $\{\boldsymbol{e}_i\}_{i=1}^M$ and $\{\boldsymbol{f}_j\}_{j=1}^n$ be the canonical basis of $\mathbb{R}^M$ and $\mathbb{R}^n$, respectively. Denote by $\mathbf{1}_m$ the all-one vector in $\mathbb{R}^m$.

Assume that S has rank $r = 1$ and admits the singular value decomposition $S = d\boldsymbol{u}^T\boldsymbol{v}$. Set the dimension ratio $y = M/n = 0.5$. We present the simulations corresponding to the special cases discussed in Remarks 2.5–2.8. Specifically, we consider following four cases: 1. Gaussian noise, $\boldsymbol{u} = \boldsymbol{e}_1$ and $\boldsymbol{v} = \boldsymbol{f}_1$; 2. Two-point noise, $\boldsymbol{u} = \mathbf{1}_M/\sqrt{M}$ and $\boldsymbol{v} = \mathbf{1}/\sqrt{n}$; 3. Two-point noise, $\boldsymbol{u} = \mathbf{1}_M/\sqrt{M}$ and $\boldsymbol{v} = \boldsymbol{f}_1$; 4. Two-point noise, $\boldsymbol{u} = \boldsymbol{e}_1$ and $\boldsymbol{v} = \boldsymbol{f}_1$. The normalization of $\sqrt{n}(|\langle\widehat{\boldsymbol{v}}, \boldsymbol{v}\rangle|^2 - a(d))$ listed in the above cases are chosen according to the calculations in Remarks 2.5–2.8. For case 4, we further subtract the non-Gaussian part $-2\theta(d)\sqrt{n}X_{11}$ from the statistic. Hence, in all four cases, we expect that the asymptotic distributions are normal. We denote the normalized statistics of the above four cases as $\mathcal{R}_g, \mathcal{R}_{dt}, \mathcal{R}_{pt}$ and $\mathcal{R}_{st}$, respectively, and we refer to the Supplementary Material [8], Section A, for more details on the definitions.

In Figure 1 of [8], we plot the ECDFs of of $\mathcal{R}_g, \mathcal{R}_{dt}, \mathcal{R}_{pt}, \mathcal{R}_{st}$ in subfigures (A), (B), (C), (D), respectively, for $n = 500$ and various values of $d = 2, 3, 5, 10$. The distributions of these quantities are fairly close to the standard normal distribution. In [8], Section A, we also record the probabilities for different quantiles of the empirical cumulative distributions (ECDFs) of the above statistics, they are fairly close to standard Gaussian even for a small sample size $n = 200$.

3.2. *Statistical applications.* In this section, we will briefly discuss the applications of our main results to the singular vector and singular subspace estimation and inference, and leave more details to the Supplementary Material [8].

We start with the estimation part and focus on the right singular vector and subspace. The estimation of singular vector and subspace is important in the recovery of low-rank matrix based on noisy observations (see, for instance, [18, 20, 28] and reference therein). It is clear that (see Lemma D.1 in [8]) the sample singular vector is concentrated on a cone with axis parallel to the true singular vector. The aperture of the cone is determined by the deterministic function $a(d)$ defined in (2.2). Further, when d increases, the sample singular vector will get closer to the true singular vector in ℓ^2 norm. It can be seen from the result in Theorem 2.3 that the variance of the fluctuation also decays when d increases. This phenomenon is recorded in Figure 2 in the Supplementary Material [8].

Empirically, it can be seen from Figure 2 in [8] that for a sequence of $y \in [\frac{1}{10}, 10]$, when $d > 5$, the variance part is already very small and hence the fluctuation can be ignored. Further, when $d > 7.5$, we can use the sample singular vector to estimate the true singular vector since their inner product is rather close to 1. Finally, note that the noise type will affect the variance of the fluctuation. Especially when the noise has negative κ_3 and κ_4, we can ignore the fluctuation for a smaller value of d. Once the singular vectors are estimated, the estimation of the singular subspace follows.

Next, we consider the inference of the singular vectors and subspace of S. Recall the decomposition in (1.2). For brevity, here we focus our discussion on the inference of V, assuming that U, D and the necessary parameters of X (e.g., cumulants of the entries of X) are known. In the Supplementary Material [8], we will also briefly discuss the possible extension of our results to adapt to the situation when D and the parameters of X are not known. Especially, using Theorem 2.3 we can test whether a singular vector $\boldsymbol{v}_i$ is equal to a given vector $\boldsymbol{v}_{i0}$, which can be formulated as

$$\text{(T0)} \qquad \boldsymbol{H}_0 : \boldsymbol{v}_i = \boldsymbol{v}_{i0}, \qquad \boldsymbol{H}_a : \boldsymbol{v}_i \neq \boldsymbol{v}_{i0},$$

and we can choose the testing statistic to be

$$\boldsymbol{S}_0 := \sqrt{n}\big(|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_{i0}\rangle|^2 - a(d_i)\big).$$

Further, using Theorem 2.9, one can test if the matrix V is equal to a given matrix, which can be formulated as

$$\text{(T1)} \qquad \boldsymbol{H}_0 : V = V_0, \qquad \boldsymbol{H}_a : V \neq V_0,$$

where $V_0 = (\boldsymbol{v}_{10}, \dots, \boldsymbol{v}_{i0})$ is a given matrix consisting of orthonormal columns. We can choose the testing statistic to be

$$\begin{aligned} \boldsymbol{S}_1 &= \sqrt{n}\left(\sum_{i,j=1}^{r} |\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_{j0}\rangle|^2 - \sum_{i=1}^{r} a(d_i)\right) \\ &= \sqrt{n}\left(\frac{1}{2}(2r - \|\widehat{V}_r \widehat{V}_r^* - V_0 V_0^*\|_F^2) - \sum_{i=1}^{r} a(d_i)\right). \end{aligned} \tag{3.1}$$

We remark here that in some cases like X is Gaussian, we can see from Theorem 2.9 that $\boldsymbol{S}_1$ is not a good statistic to distinguish V_0 from $V_0 O$ for some deterministic $r \times r$ orthogonal matrix O. Specifically, one cannot tell if $\widehat{V}_r$ is the matrix of the singular vectors of the model $X + UDV_0^*$ or $X + UD(V_0 O)^*$, since $V_0 V_0^* = (V_0 O)(V_0 O)^*$ in (3.1) and the limiting distribution of $\boldsymbol{S}_1$ does not depend on V when X is Gaussian. Hence, we do not expect the statistic $\boldsymbol{S}_1$ to be powerful for the test (T1) when the alternative is of the form $V_0 O$ in some cases like Gaussian noise. In other words, in this case, what one can test is if $VV^* = V_0 V_0^*$. Nevertheless, one can still do the test (T1) by using the testing statistic of the diagonal parts of $\boldsymbol{S}_1$ only, that is, $\boldsymbol{S}_{1d} = \sqrt{n}(\sum_i^r |\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_{i0}\rangle|^2 - \sum_{i=1}^r a(d_i))$. Under the null hypothesis, $\boldsymbol{S}_{1d}$ has the same distribution as $\boldsymbol{S}_1$ since it will be clear that $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_{j0}\rangle|^2$ is negligible if $i \neq j$, in the null case. But note that the limiting distribution of $\boldsymbol{S}_{1d}$ is no longer invariant under taking right orthogonal transformation for V_0. Hence, it can be used to test if $V = V_0$.

We mention that both (T0) and (T1) could be useful in many scientific disciplines, especially when the singular vectors of S are sparse and have practical meanings. For instance, an important goal of the study of gene expression data for cancer is to simultaneously identify related genes and subjects grouped together according to the cancer types [48], Section 2. For this purpose, the right singular vectors are used to visualize the gene grouping (see Figure 1 of [48]) and the left singular vectors are used to represent the subject grouping (see Figure 2 of [48]). Other examples include the study of the nutrition content data of different foods [48] and the mortality rate data after expanding on suitable basis functions [60], Section 3. In the literature, various algorithms have been proposed to estimate the sparse singular vectors; for instance, see [26, 48, 60, 61]. From the statistical perspective, with the above estimates, it is natural to do inference on the singular vectors. For instance, for the gene expression data of lung caner, researchers may be interested in testing whether a certain type of cancer is determined by a subset of genes and this is related to doing inference on the right singular vectors and right singular subspace.

Since we assume that U, D and the necessary parameters of X (e.g., cumulants of the entries of X) are known, we can carry out the z-score test to test $\boldsymbol{H}_0$ in both (T0) and (T1). Due the similarity of (T0) and (T1), we focus on (T1) and leave the detailed discussions and simulations to the Supplementary Material [8].

4. Techincal tools and Green function representations. This section is devoted to providing some basic notions and technical tools, which will be needed often in our proofs for the theorems. The basic notions are given in Section 4.1. A main technical input for our proof is the isotropic local law for the sample covariance matrix obtained in [14, 43]. It will be stated in Section 4.2. In Section 4.3, we represent (asymptotically) $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i\rangle|^2$'s and R (cf. (1.7)) in terms of the Green function. The discussion is based on [26], where the limits for $|\langle \widehat{\boldsymbol{u}}_i, \boldsymbol{u}_j\rangle|^2$ and $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_j\rangle|^2$ are studied. We then collect a few auxiliary definitions in Section 4.4.

4.1. *Basic notions*. For a positive integer n, we denote by $[n]$ the set $\{1, \dots, n\}$. Let $\mathbb{C}^+$ be the complex upper-half plane. Further, we define the following linearization for our model

$$\mathcal{Y}(z) := \mathcal{U}\mathcal{D}(z)\mathcal{U}^* + H(z), \quad z = E + \mathrm{i}\eta \in \mathbb{C}^+, \tag{4.1}$$

where

$$\mathcal{U} := \begin{pmatrix} U & \\ & V \end{pmatrix}, \qquad \mathcal{D}(z) := \sqrt{z}\begin{pmatrix} & D \\ D & \end{pmatrix}, \qquad H(z) := \sqrt{z}\begin{pmatrix} & X \\ X^* & \end{pmatrix}. \tag{4.2}$$

In the sequel, we will often omit z and simply write $\mathcal{Y} \equiv \mathcal{Y}(z), \mathcal{D} \equiv \mathcal{D}(z)$ and $H \equiv H(z)$ when there is no confusion.

We denote the empirical spectral distributions (ESD) of the matrices XX^* and X^*X by

$$F_1(x) := \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}_{\{\lambda_i(XX^*) \le x\}}, \qquad F_2(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\lambda_i(X^*X) \le x\}}.$$

$F_1(x)$ and $F_2(x)$ are known to satisfy the Marchenko–Pastur (MP) law [50]. More precisely, almost surely, $F_1(x)$ converges weakly to a nonrandom limit $F_{1y}(x)$ which has a density function given by

$$\rho_1(x) := \begin{cases} \frac{1}{2\pi x y}\sqrt{(\lambda_+ - x)(x - \lambda_-)} & \text{if } \lambda_- \le x \le \lambda_+, \\ 0 & \text{otherwise,} \end{cases}$$

and has a point mass $1 - 1/y$ at the origin if $y > 1$, where $\lambda_+ = (1 + \sqrt{y})^2$ and $\lambda_- = (1 - \sqrt{y})^2$. Further, the Stieltjes's transform of F_{1y} is given by

$$m_1(z) := \int \frac{1}{x - z}\,\mathrm{d}F_{1y}(x) = \frac{1 - y - z + \mathrm{i}\sqrt{(\lambda_+ - z)(z - \lambda_-)}}{2zy} \quad \text{for } z \in \mathbb{C}^+, \tag{4.3}$$

where the square root denotes the complex square root with a branch cut on the negative real axis. Similarly, almost surely, $F_2(x)$ converges weakly to a nonrandom limit $F_{2y}(x)$ which has a density function given by

$$\rho_2(x) := \begin{cases} \frac{1}{2\pi x}\sqrt{(\lambda_+ - x)(x - \lambda_-)} & \text{if } \lambda_- \le x \le \lambda_+, \\ 0 & \text{otherwise,} \end{cases}$$

and a point mass $1-y$ at the origin if $y<1$. The corresponding Stieltjes's transform is

$$m_2(z) := \int \frac{1}{x-z}\,\mathrm{d}F_{2y}(x) = \frac{y-1-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2z}. \tag{4.4}$$

Our estimation relies on the local MP law [56] and its isotropic version [14, 43], which provide sharp large deviation estimates for the Green functions

$$G(z) = (H-z)^{-1}, \qquad \mathcal{G}_1(z) = \big(XX^*-z\big)^{-1}, \qquad \mathcal{G}_2(z) = \big(X^*X-z\big)^{-1}.$$

Here, we recall the definition in (4.2). By Schur complement, one can derive

$$G(z) = \begin{pmatrix} \mathcal{G}_1(z) & z^{-1/2}\mathcal{G}_1(z)X \\ z^{-1/2}X^*\mathcal{G}_1(z) & \mathcal{G}_2(z) \end{pmatrix}. \tag{4.5}$$

The Stieltjes transforms for the ESD of XX^* and X^*X are defined by

$$\begin{aligned} m_{1n}(z) &= \frac{1}{M}\operatorname{Tr}\mathcal{G}_1(z) = \frac{1}{M}\sum_{i=1}^{M} G_{ii}(z), \\ m_{2n}(z) &= \frac{1}{n}\operatorname{Tr}\mathcal{G}_2(z) = \frac{1}{n}\sum_{\mu=M+1}^{M+n} G_{\mu\mu}(z). \end{aligned} \tag{4.6}$$

It is well known that $m_{1n}(z)$ and $m_{2n}(z)$ have nonrandom approximates $m_1(z)$ and $m_2(z)$, which are the Stieltjes transforms for the MP laws defined in (4.3) and (4.4). Specifically, for any fixed $z \in \mathbb{C}^+$, the following hold:

$$m_{1n}(z) - m_1(z) \xrightarrow{a.s.} 0, \qquad m_{2n}(z) - m_2(z) \xrightarrow{a.s.} 0.$$

Furthermore, one can easily check that $m_1(z)$ and $m_2(z)$ satisfy the following self-consistent equations (see [2] for instance)

$$m_1(z) + \frac{1}{z-(1-y)+zym_1(z)} = 0, \tag{4.7}$$

$$m_2(z) + \frac{1}{z+(1-y)+zm_2(z)} = 0. \tag{4.8}$$

We can also derive the following simple relation from the definitions:

$$m_1(z) = \frac{y^{-1}-1}{z} + y^{-1}m_2(z). \tag{4.9}$$

Next, we summarize some basic identities in the following lemma without proof. They can be checked from (4.3) and (4.4) via elementary calculations.

LEMMA 4.1. *Denote $p \equiv p(x)$ in (2.1). For any $x > y^{1/4}$, we have*

$$m_1(p) = \frac{-1}{x^2+y}, \qquad m_2(p) = \frac{-1}{x^2+1},$$

$$m_1'(p) = \frac{x^4}{(x^2+y)^2(x^4-y)}, \qquad m_2'(p) = \frac{x^4}{(x^2+1)^2(x^4-y)}.$$

Furthermore, denote by $\mathcal{T}(t) = tm_1(t)m_2(t)$. We have

$$\mathcal{T}(p) = x^{-2}, \qquad \mathcal{T}'(p) = \big(y-x^4\big)^{-1}.$$

In the sequel, we also need the following notion on high probability events.

DEFINITION 4.2 (High probability event). We say that an n-dependent event $E \equiv E(n)$ holds with high probability if, for any large $\varphi > 0$,

$$\mathbb{P}(E) \geq 1 - n^{-\varphi},$$

for sufficiently large $n \geq n_0(\varphi)$.

We also adopt the notion of *stochastic domination* introduced in [30].

DEFINITION 4.3 (Stochastic domination). Let

$$\mathsf{X} = \big(\mathsf{X}^{(n)}(u) : n \in \mathbb{N}, u \in \mathsf{U}^{(n)}\big), \qquad \mathsf{Y} = \big(\mathsf{Y}^{(n)}(u) : n \in \mathbb{N}, u \in \mathsf{U}^{(n)}\big),$$

be two families of nonnegative random variables, where $\mathsf{U}^{(n)}$ is a possibly n-dependent parameter set. We say that X is stochastically dominated by Y, uniformly in u, if for all small ϵ and large φ, we have

$$\sup_{u \in \mathsf{U}^{(n)}} \mathbb{P}\big(\mathsf{X}^{(n)}(u) > n^{\epsilon} \mathsf{Y}^{(n)}(u)\big) \leq n^{-\varphi},$$

for large enough $n \geq n_0(\epsilon, \varphi)$. In addition, we use the notation $\mathsf{X} = O_{\prec}(\mathsf{Y})$ if $|\mathsf{X}|$ is stochastically dominated by Y, uniformly in u. Throughout this paper, the stochastic domination will always be uniform in all parameters (mostly are matrix indices and the spectral parameter z) that are not explicitly fixed.

4.2. *Isotropic local laws*. The key ingredient in our estimation is a special case of the anisotropic local law derived in [43], which is essentially the isotropic local law previously derived in [14]. Let $\oplus$ be the direct sum of two matrices. Set

$$\Pi_1(z) := m_1(z) I_M \oplus m_2(z) I_n. \tag{4.10}$$

We will need the isotropic local law outside the spectrum of the MP law. For $\lambda_+ = (1 + y^{1/2})^2$, define the spectral domain

$$\boldsymbol{S}_o \equiv \boldsymbol{S}_o(\tau) := \{z = E + \mathrm{i}\eta \in \mathbb{C}^+ : \lambda_+ + \tau \leq E \leq \tau^{-1}, 0 \leq \eta \leq \tau^{-1}\}, \tag{4.11}$$

where $\tau > 0$ is a fixed small constant. Recall m_{1n} and m_{2n} defined in (4.6).

LEMMA 4.4 (Theorem 3.7 of [43], Theorem 3.12 of [14] and Theorem 3.1 of [56]). *Fix* $\tau > 0$, *for any unit deterministic* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{M+n}$, *we have*

$$\langle \boldsymbol{u}, \big(G(z) - \Pi_1(z)\big)\boldsymbol{v}\rangle = O_{\prec}\bigg(\sqrt{\frac{\operatorname{Im} m_2(z)}{n\eta}}\bigg), \tag{4.12}$$

$$\big|m_{1n}(z) - m_1(z)\big| = O_{\prec}\bigg(\frac{1}{n}\bigg), \qquad \big|m_{2n}(z) - m_2(z)\big| = O_{\prec}\bigg(\frac{1}{n}\bigg), \tag{4.13}$$

uniformly in $z \in \boldsymbol{S}_o$.

REMARK 4.5. The bounds in (4.13) cannot be directly read from any of Theorem 3.7 of [43], Theorem 3.12 of [14] or Theorem 3.1 of [56]. In all these theorems, a weaker bound $O_{\prec}(\frac{1}{n\eta})$ is stated for z both inside and outside of the support of the limiting spectral distribution. Here, since our parameter z can be real, we use the stronger bound $\frac{1}{n}$ instead of $\frac{1}{n\eta}$. For $z \in \boldsymbol{S}_o$, such a bound follows from the rigidity estimates of eigenvalues in [56] and the definition of the Stieltjes transform easily. Specifically, by (3.7) in [56], we know that for $a = 1, 2$, $\sup_{t \in \mathbb{R}} |F_a(t) - F_{ay}(t)| \prec \frac{1}{n}$, and further by (3.6) of [56] we know that $\sup_{t \in \mathbb{R}:|t| \geq 2 + n^{-\frac{2}{3}+\varepsilon}} |F_a(t) - F_{ay}(t)| = 0$ with high probability. Then using the integration by parts to $m_{an}(z) - m_a(z) = \int (t-z)^{-1}\, \mathrm{d}(F_a(t) - F_{ay}(t))$, one can easily conclude the bounds in (4.13).

Following from Lemma 4.4, by further using Cauchy's integral formula for derivatives, we have the following uniformly in $z \in \boldsymbol{S}_o$, for any given $l \in \mathbb{N}$,

$$\langle \boldsymbol{u}, (G^{(l)}(z) - \Pi_1^{(l)}(z))\boldsymbol{v}\rangle = O_{\prec}\left(\sqrt{\frac{\operatorname{Im} m_2(z)}{n\eta}}\right). \tag{4.14}$$

Denote by $\kappa = |E - \lambda_+|$. We summarize some basic estimates of $m_{1,2}(z)$ without proof. For any two numbers a_n and b_n (might be n-dependent), we write $a_n \sim b_n$ if there exist two positive constants C_1 and C_2 (independent of n) such that $C_1|b_n| \le |a_n| \le C_2|b_n|$.

LEMMA 4.6. *The following estimates hold uniformly in $z \in \boldsymbol{S}_o$*:

$$\left|m'_{1,2}(z)\right| \sim \left|m_{1,2}(z)\right| \sim 1, \tag{4.15}$$

$$\operatorname{Im} m_1(z) \sim \operatorname{Im} m_2(z) \sim \frac{\eta}{\sqrt{\kappa + \eta}}. \tag{4.16}$$

Given any deterministic bounded Hermitian matrix A with fixed rank, it is easy to see from Lemma 4.4 and Lemma 4.6, the spectral decomposition and (4.14) that the following estimates hold uniformly in $z \in \boldsymbol{S}_o$: For any fixed $k, \ell \in \mathbb{N}$,

$$\begin{gathered}
\max_{\mu,\nu}\left|\left(G^{(l)}(z)A\right)_{\mu\nu} - \left(\Pi_1^{(l)}(z)A\right)_{\mu\nu}\right| = O_{\prec}\left(\frac{1}{\sqrt{n}}\right), \\
\operatorname{Tr} G^{(l)}(z)A - \operatorname{Tr} \Pi_1^{(l)}(z)A = O_{\prec}\left(\frac{1}{\sqrt{n}}\right), \\
\max_{\mu,\nu}\left|\left(G^{(k)}(z)AG^{(l)}(z)\right)_{\mu\nu} - \left(\Pi_1^{(k)}A\Pi_1^{(l)}\right)_{\mu\nu}\right| = O_{\prec}\left(\frac{1}{\sqrt{n}}\right).
\end{gathered} \tag{4.17}$$

In our proof, we will rely on the estimates of powers of G, that is, $G^l, l = 2, 3, 4$. We have the following lemma whose proof is stated in [8].

LEMMA 4.7. *We have the following recursive relation*:

$$G^2 = 2G' + \frac{G}{z}, \qquad G^3 = (G^2)' + \frac{G^2}{z}, \qquad G^4 = \frac{2}{3}(G^3)' + \frac{G^3}{z}. \tag{4.18}$$

Recall Π_1 defined in (4.10) and further define

$$\Pi_2 := 2\Pi_1' + \frac{1}{z}\Pi_1, \qquad \Pi_3 := \Pi_2' + \frac{1}{z}\Pi_2, \qquad \Pi_4 := \frac{2}{3}\Pi_3' + \frac{1}{z}\Pi_3. \tag{4.19}$$

With Lemma 4.7, similar to (4.12) and (4.14), we can get the following estimates for $l = 1, 2, 3, 4$:

$$\langle \boldsymbol{u}, (G^l - \Pi_l)\boldsymbol{v}\rangle = O_{\prec}\left(\frac{1}{\sqrt{n}}\right), \tag{4.20}$$

uniformly in $z \in \boldsymbol{S}_o$. For brevity, in the sequel, we will use the notation

$$\Xi_l \equiv \Xi_l(z) := G^l(z) - \Pi_l(z), \quad l \in \mathbb{N}. \tag{4.21}$$

4.3. *Green function representation.* In this section, we represent (asymptotically) $|\langle \widehat{\boldsymbol{v}}_i, \boldsymbol{v}_i\rangle|^2$'s and R (cf. (1.7)) in terms of the Green function. The derivation relies on the results obtained in [26]. Recall $p(d)$ in (2.1) and $a(d)$ in (2.2). For $i \in [r]$, define

$$h_i(x) = \frac{x^4 p'(x)p(x)}{(x+d_i)^2}, \tag{4.22}$$

and we use the shorthand notation $\bar{i} = i + r$. To state results for the right singular vectors, we introduce a $2r \times 2r$ matrix function $W_i(x)$ for $x > 0$, which has only four nonzero entries given by

$$\begin{aligned} \big(W_i(x)\big)_{ii} &= m_2^2(x), \qquad \big(W_i(x)\big)_{\bar{i}\bar{i}} = \frac{1}{d_i^2 x}, \\ \big(W_i(x)\big)_{i\bar{i}} &= \big(W_i(x)\big)_{\bar{i}i} = -\frac{m_2(x)}{d_i\sqrt{x}}. \end{aligned} \tag{4.23}$$

We further denote the matrix function

$$M_i(x) = \mathcal{U} W_i(x) \mathcal{U}^*. \tag{4.24}$$

With the above notation, we further introduce two $(M+n)\times(M+n)$ matrices

$$\begin{aligned} A_i^R &= -d_i^2\big(h_i'(d_i)M_i(p_i) + h_i(d_i)p'(d_i)M_i'(p_i)\big), \\ B_i^R &= -d_i^2 h(d_i)p'(d_i)M_i(p_i). \end{aligned} \tag{4.25}$$

In light of the definition of $\mathcal{U}$ in (4.2), we have

$$A_i^R = \begin{pmatrix} \omega_{i1}\boldsymbol{u}_i\boldsymbol{u}_i^T & \omega_{i2}\boldsymbol{u}_i\boldsymbol{v}_i^T \\ \omega_{i3}\boldsymbol{v}_i\boldsymbol{u}_i^T & \omega_{i4}\boldsymbol{v}_i\boldsymbol{v}_i^T \end{pmatrix}, \qquad B_i^R = \begin{pmatrix} \varpi_{i1}\boldsymbol{u}_i\boldsymbol{u}_i^T & \varpi_{i2}\boldsymbol{u}_i\boldsymbol{v}_i^T \\ \varpi_{i3}\boldsymbol{v}_i\boldsymbol{u}_i^T & \varpi_{i4}\boldsymbol{v}_i\boldsymbol{v}_i^T \end{pmatrix}. \tag{4.26}$$

Here, we used the notation

$$\begin{aligned} \omega_{i1} &:= -d_i^2\big(h_i'(d_i)\big(W_i(p_i)\big)_{ii} + h_i(d_i)p'(d_i)\big(W_i'(p_i)\big)_{ii}\big), \\ \omega_{i4} &:= -d_i^2\big(h_i'(d_i)\big(W_i(p_i)\big)_{\bar{i}\bar{i}} + h_i(d_i)p'(d_i)\big(W_i'(p_i)\big)_{\bar{i}\bar{i}}\big), \\ \omega_{i2} = \omega_{i3} &:= -d_i^2\big(h_i'(d_i)\big(W_i(p_i)\big)_{i\bar{i}} + h_i(d_i)p'(d_i)\big(W_i'(p_i)\big)_{i\bar{i}}\big), \\ \varpi_{i1} &:= -d_i^2 h_i(d_i)p'(d_i)\big(W_i(p_i)\big)_{ii}, \\ \varpi_{i4} &:= -d_i^2 h_i(d_i)p'(d_i)\big(W_i(p_i)\big)_{\bar{i}\bar{i}}, \\ \varpi_{i2} = \varpi_{i3} &:= -d_i^2 h_i(d_i)p'(d_i)\big(W_i(p_i)\big)_{i\bar{i}}. \end{aligned}$$

Recall the notation introduced in (4.21). We have the following lemma whose proof is stated in [8].

LEMMA 4.8. *Under assumptions of* (1.3), (1.4), (1.8) *and Assumption* 2.1, *we have*

$$\big|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle\big|^2 = a(d_i) + \operatorname{Tr}\big(\Xi_1(p_i)A_i^R\big) + \operatorname{Tr}\big(\Xi_1'(p_i)B_i^R\big) + O_\prec\Big(\frac{1}{n}\Big).$$

Furthermore, we have

$$R = \sum_{i=1}^r a(d_i) + \sum_{i=1}^r \big(\operatorname{Tr}\big(\Xi_1(p_i)A_i^R\big) + \operatorname{Tr}\big(\Xi_1'(p_i)B_i^R\big)\big) + O_\prec\Big(\frac{1}{n}\Big). \tag{4.27}$$

4.4. *Auxiliary definitions.* It is convenient to introduce the following notion of convergence in distribution.

DEFINITION 4.9 ([41], Definition 7.3). Two sequences of random variables, $\{\mathsf{X}_n\}$ and $\{\mathsf{Y}_n\}$, are *asymptotically equal in distribution*, denoted as $\mathsf{X}_n \simeq \mathsf{Y}_n$, if they are tight and satisfy

$$\lim_{n\to\infty}\big(\mathbb{E}f(\mathsf{X}_n) - \mathbb{E}f(\mathsf{Y}_n)\big) = 0$$

for any bounded continuous function f.

We also collect some basic results on convergence and equivalence in distribution in the Supplementary Material [8], Lemma C.3.

The following notation from [41], Definition 7.11, will be convenient for us when we replace random variables with their i.i.d. copies.

DEFINITION 4.10. Let $\{\sigma_n\}$ be a sequence of bounded positive numbers. If X_n and Y_n are independent random variables with $\mathsf{Y}_n \simeq \mathcal{N}(0, \sigma_n^2)$, and if $\mathsf{S}_n \simeq \mathsf{X}_n + \mathsf{Y}_n$, we write $\mathsf{S}_n \simeq \mathsf{X}_n + \mathcal{N}(0, \sigma_n^2)$.

5. Proof of Theorems 2.3. For brevity, in this section, we omit the subindices of $d_i, \boldsymbol{u}_i, \boldsymbol{v}_i, \widehat{\boldsymbol{u}}_i, \widehat{\boldsymbol{v}}_i$ and write $d, \boldsymbol{u}, \boldsymbol{v}, \widehat{\boldsymbol{u}}, \widehat{\boldsymbol{v}}$ instead. Similarly, we write the matrices A_i^R and B_i^R (cf. (4.25)) as A and B, respectively. We also write $m_{1,2}(z)$ as $m_{1,2}$ for brevity.

By Lemma 4.8, we can reduce the problem to study

$$\mathcal{Q} \equiv \mathcal{Q}(z) := \sqrt{n}\big(\operatorname{Tr}\big(\Xi_1(z)A\big) + \operatorname{Tr}\big(\Xi_1'(z)B\big)\big), \tag{5.1}$$

at $z = p(d)$ (cf. (2.1)).

In the sequel, we will prove the limiting distribution of $\mathcal{Q}(z)$ at $z = p(d)$. The key task is to prove Proposition 5.1 below. In this section, we will show that Theorem 2.3 follows from Proposition 5.1. Let index $i \in [M]$ and $j \in [n]$. Denote the shorthand notation

$$j' = j + M. \tag{5.2}$$

For short, we also write $\sum_{i,j} = \sum_{i=1}^M \sum_{j=1}^n$.

In order to state Proposition 5.1, we first introduce some notation. For a fixed small constant $\nu > 0$, denote by

$$\mathcal{B}(\nu) := \big\{(i, j) \in [M] \times [n] : |\boldsymbol{u}(i)| > n^{-\nu}, |\boldsymbol{v}(j)| > n^{-\nu}\big\},$$

the set of the indices of those components with large magnitude. Since $\boldsymbol{u}$ and $\boldsymbol{v}$ are unit vectors, we have $|\mathcal{B}(\nu)| \le Cn^{4\nu}$ for some constant $C > 0$. Let $\mathcal{S}(\nu)$ be the complement of $\mathcal{B}(\nu)$, that is,

$$\mathcal{S}(\nu) = \big([M] \times [n]\big) \setminus \mathcal{B}(\nu). \tag{5.3}$$

For brevity, we introduce the notation

$$\mathcal{P}(\alpha_1, \dots, \alpha_m), \tag{5.4}$$

to represent the set of all the permutations of $(\alpha_1, \dots, \alpha_m)$, where α_i's can be alike. Recall (4.10) and (4.19). We set the deterministic quantity

$$\begin{aligned}\Delta_d &\equiv \Delta_d(z)\\ &:= -\frac{\kappa_3 z^{3/2}}{n}\sum_{i,j}\Big((\Pi_1)_{ii}(\Pi_1)_{j'j'}\big(2(\Pi_1 A \Pi_1)_{ij'} + \big(\Pi_1 B \Pi_1'\big)_{ij'} + \big(\Pi_1' B \Pi_1\big)_{ij'}\big)\\ &\quad + \frac{1}{2}\sum_{(a_1,a_2,a_3)\in\mathcal{P}(2,1,1)} (\Pi_{a_1})_{ii}(\Pi_{a_2})_{j'j'}\big((\Pi_1 B \Pi_{a_3})_{ij'} + (\Pi_{a_3} B \Pi_1)_{ij'}\big)\Big),\end{aligned} \tag{5.5}$$

and the random variable

$$\Delta_r \equiv \Delta_r(z) := \sqrt{nz} \sum_{(i,j)\in\mathcal{B}(\nu)} x_{ij} c_{ij}, \tag{5.6}$$

where

$$\begin{aligned} c_{ij} \equiv c_{ij}(z) := -\sum_{\substack{l_1,l_2\in\{i,j'\}\\ l_1\neq l_2}} \Big(&(\Pi_1 A \Pi_1)_{l_1 l_2} - \frac{1}{2z}(\Pi_1 B \Pi_1)_{l_1 l_2} \\ &+ \frac{1}{2}(\Pi_1 B \Pi_2)_{l_1 l_2} + \frac{1}{2}(\Pi_2 B \Pi_1)_{l_1 l_2}\Big). \end{aligned} \tag{5.7}$$

Define the $M \times n$ matrix function $S \equiv S(z) = (s_{ij})$ with

$$\begin{aligned} s_{ij} \equiv s_{ij}(z) := \sum_{\substack{l_1,\dots,l_4\in\{i,j'\}\\ l_1\neq l_4, l_2\neq l_3}} \Big(&(\Pi_1 A \Pi_1)_{l_1 l_2}(\Pi_1)_{l_3 l_4} - \frac{1}{2z}(\Pi_1 B \Pi_1)_{l_1 l_2}(\Pi_1)_{l_3 l_4} \\ &+ \frac{1}{2} \sum_{(a_1,a_2,a_3)\in\mathcal{P}(2,1,1)} (\Pi_{a_1} B \Pi_{a_2})_{l_1 l_2}(\Pi_{a_3})_{l_3 l_4}\Big). \end{aligned} \tag{5.8}$$

Further, we define the function

$$V \equiv V(z) := \mathcal{V}^E(z) + 2\frac{\kappa_3 z^{\frac{3}{2}}}{\sqrt{n}} \sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij} s_{ij} + \frac{\kappa_4 z^2}{n} \sum_{i,j} s_{ij}^2 + z \sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}^2, \tag{5.9}$$

where

$$\mathcal{V}^E \equiv \mathcal{V}^E(z) := -\sqrt{z} \sum_{\alpha=1,2} \Big(m_\alpha \mathfrak{a}_{1\alpha} + \frac{m_\alpha}{2} \tilde{\mathfrak{b}}_{1\alpha} + m'_\alpha \mathfrak{b}_{1\alpha} \Big). \tag{5.10}$$

Here, we refer to (E.9) in [8] for the definitions of $\mathfrak{a}_{1\alpha}$, $\mathfrak{b}_{1\alpha}$ and $\tilde{\mathfrak{b}}_{1\alpha}$ for $\alpha = 1, 2$.

With Δ_d and Δ_r defined in (5.5) and (5.6), we introduce the notation

$$\Delta \equiv \Delta(z) := \Delta_r(z) + \Delta_d(z) \tag{5.11}$$

and define

$$Q \equiv Q(z) := \mathcal{Q}(z) - \Delta(z). \tag{5.12}$$

PROPOSITION 5.1. *Under the assumptions of Theorem 2.3, we have that $Q(p_i)$ and $\Delta(p_i)$ are asymptotically independent. Furthermore,*

$$Q(p_i) \simeq \mathcal{N}\big(0, V(p_i)\big). \tag{5.13}$$

We first show how Proposition 5.1 implies Theorem 2.3.

PROOF OF THEOREM 2.3. By Lemma 4.8 and (5.1),

$$\sqrt{n}\big(|\langle \boldsymbol{v}_i, \widehat{\boldsymbol{v}}_i\rangle|^2 - a(d_i)\big) = \mathcal{Q}(p_i) + O_\prec\big(n^{-\frac{1}{2}}\big).$$

Here, $\mathcal{Q}(p_i)$ is defined in (5.1) with $(A, B) = (A_i^R, B_i^R)$ (cf. (4.25)). By Proposition 5.1, we have that at $z = p_i$,

$$\mathcal{Q} = \Delta_d + \Delta_r + Q \simeq \Delta_d + \sqrt{nz} \sum_{(i,j)\in\mathcal{B}(\nu)} x_{ij} c_{ij} + \mathcal{N}(0, V).$$

Next, by the central limit theorem and Lemma C.3 in [8], one has

$$\sqrt{nz}\sum_{i,j} x_{ij}c_{ij} \simeq \sqrt{nz}\sum_{(i,j)\in\mathcal{B}(\nu)} x_{ij}c_{ij} + \mathcal{N}\Big(0, z\sum_{(i,j)\in\mathcal{S}(\nu)} (c_{ij})^2\Big).$$

Furthermore, by the definition of $\mathcal{S}(\nu)$, we notice that

$$n^{-1/2}\sum_{(i,j)\in\mathcal{S}(\nu)} c_{ij}s_{ij} = n^{-1/2}\sum_{i,j} c_{ij}s_{ij} + O\big(n^{-\frac{1}{2}+4\nu}\big).$$

Let $C(z) = (c_{ij}(z))$ with $c_{ij}(z)$ defined in (5.7) and recall $S(z)$ from (5.8). Using Lemma C.3 in [8], we conclude that

$$\mathcal{Q}(p_i) \simeq \Delta_d(p_i) + \sqrt{np_i}\operatorname{Tr}\big(X^*C(p_i)\big) + \mathcal{N}\big(0, \mathcal{V}(p_i)\big),$$

where

$$\mathcal{V}(p_i) = \mathcal{V}^E(p_i) + 2\frac{\kappa_3 {p_i}^{3/2}}{\sqrt{n}}\operatorname{Tr}\big(C(p_i)^*S(p_i)\big) + \frac{\kappa_4 {p_i}^2}{n}\operatorname{Tr}\big(S(p_i)^*S(p_i)\big).$$

Denote

$$\Delta_i = \sqrt{np_i}\operatorname{Tr}\big(X^*C(p_i)\big) + \Delta_d(p_i)$$

and $\mathcal{Z}_i \sim \mathcal{N}(0, \mathcal{V}(p_i))$, which is independent of Δ_i. Next, plugging $z = p_i$ into (5.5), (5.7), (5.8), using Lemma 4.1 and taking into account the definitions of A_i^R, B_i^R in (4.25), we find that

$$\Delta_i = -\sqrt{n}\frac{2(d_i^4 + 2yd_i^2 + y)}{d_i^3(d_i^2+1)^2}\boldsymbol{u}_i^* X\boldsymbol{v}_i - \frac{2(d_i^6 - 3yd_i^2 - 2y)}{d_i^5(d_i^2+1)^2}\Big(\frac{\kappa_3}{n}\sum_{k,l}\boldsymbol{u}_i(k)\boldsymbol{v}_i(l)\Big).$$

The variance $\mathcal{V}(p_i)$ is the sum of

$$\begin{aligned}
&2\frac{\kappa_3}{\sqrt{n}}p_i^{3/2}\operatorname{Tr}\big(C(p_i)^*S(p_i)\big) + \frac{\kappa_4}{n}p_i^2\operatorname{Tr}\big(S(p_i)^*S(p_i)\big) \\
&\quad = -\frac{4(d_i^4+2yd_i^2+y)(d_i^6-3yd_i^2-2y)}{d_i^7(d_i^2+1)^4}\Big(\frac{\kappa_3}{\sqrt{n}}\sum_{k,l}\boldsymbol{u}_i(k)^3\boldsymbol{v}_i(l)\Big) \\
&\qquad + \frac{4(d_i^4+2yd_i^2+y)^2}{d_i^7(d_i^2+1)^4}\Big(\frac{\kappa_3}{\sqrt{n}}\sum_{k,l}\boldsymbol{u}_i(k)\boldsymbol{v}_i(l)^3\Big) \\
&\qquad + \frac{(d_i^6-3yd_i^2-2y)^2}{d_i^8(d_i^2+1)^4}\Big(\kappa_4\sum_k \boldsymbol{u}_i(k)^4\Big) + \frac{(d_i^4+2yd_i^2+y)^2}{d_i^8(d_i^2+1)^4}\Big(\kappa_4 y_n\sum_l \boldsymbol{v}_i(l)^4\Big)
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{V}^E(p_i) = \frac{2}{d_i^4 - y}\bigg(&2y(y+1)\Big(\frac{d^4+2yd^2+y}{d^3(d^2+1)^2}\Big)^2 - \frac{y(y-1)(5y+1)}{d_i(d_i^2+1)^2}\Big(\frac{d^4+2yd^2+y}{d^3(d^2+1)^2}\Big) \\
&+ \frac{(d_i^4+y)(d_i^2+y)^2}{d_i^3(d_i^2+1)^2}\Big(\frac{d^6-3yd^2-2y}{d^3(d^2+1)^2}\Big) + \frac{2y^2(y-1)^2}{d_i^2(d_i^2+1)^4}\bigg).
\end{aligned}$$

The last expression is obtained by using the definitions of $\mathfrak{a}_{1\alpha}$, $\mathfrak{b}_{1\alpha}$ and $\tilde{\mathfrak{b}}_{1\alpha}$ for $\alpha = 1, 2$ in (E.9) of [8] and performing tedious yet elementary calculations. Recall (2.3). The conclusion of Theorem 2.3 follows immediately by rewriting Δ_i and $\mathcal{V}(p_i)$ in terms of $\theta(d_i)$ and $\psi(d_i)$. $\square$

The rest of this section is devoted to the proof of Proposition 5.1. Our proof relies on the cumulant expansion in Lemma C.1 of [8], where we need to control the expectation. Throughout the proof, we will frequently use the estimates in (4.17). These estimates hold with high probability, which do not yield bounds for the expectations directly. In order to translate the high probability bounds into those for the expectations, one needs a crude deterministic bound for the Green function on the bad event with tiny probability. To this end, we will work with a slight modification of the real $z = p(d)$ for Green function. Specifically, in the proof of the following Proposition 5.2, we will also use the parameter

$$z = p(d) + \mathrm{i}n^{-C}, \tag{5.14}$$

for a large constant C. On the bad event, we will use the naive bound of the Green function $\|G\| \le N^C$, which will be compensated by the tiny probability of the bad event. At the end, by the continuity of $G(\tilde{z})$ at $\tilde{z}$ away from the support of the MP law, it is (asymptotically) equivalent to work with (5.14), for the proof of Proposition 5.1. We first claim that it suffices to establish the following recursive estimate.

PROPOSITION 5.2. *Suppose that the assumptions of Theorem 2.3 hold. Let $z_0 = p(d)$ and z be defined in (5.14). We have*

$$\mathbb{E}Q(z)e^{\mathrm{i}t\Delta(z_0)} = O_{\prec}\big(n^{-\frac{1}{2}+4\nu}\big), \tag{5.15}$$

and for any fixed integer $k \ge 2$,

$$\mathbb{E}Q^k(z)e^{\mathrm{i}t\Delta(z_0)} = (k-1)V\mathbb{E}Q^{k-2}(z)e^{\mathrm{i}t\Delta(z_0)} + O_{\prec}\big(n^{-\frac{1}{2}+4\nu}\big). \tag{5.16}$$

The proof of Proposition 5.2 is our main technical task, which will be stated in Section E of [8]. Now we first show the proof of Proposition 5.1 based on Proposition 5.2.

PROOF OF PROPOSITION 5.1. Recall the following elementary bound, for any $x \in \mathbb{R}$ and sufficiently large $N \in \mathbb{N}$, we have

$$\left| e^{\mathrm{i}x} - \sum_{k=0}^{N} \frac{(\mathrm{i}x)^k}{k!} \right| \le \min\left\{ \frac{|x|^{N+1}}{(N+1)!}, \frac{2|x|^N}{N!} \right\}. \tag{5.17}$$

First, we write $Q(z) = Q_R(z) + \mathrm{i}Q_I(z)$, where $Q_R(z)$ and $Q_I(z)$ stand for the real and imaginary parts of $Q(z)$, respectively. According to the choice of z in (5.14), we have the deterministic bound $|Q_I(z)| \le N^C$ for some large positive constant C. Moreover, by continuity of the Green function and the Stieltjes transform, one can easily check that $|Q_I(z)| \le N^{-C'}$, for some large positive constant C' with high probability. Using the small bound $N^{-C'}$ on the high probability event and the large deterministic bound N^C on the tiny probability event, one can easily derive from (5.15) and (5.16) that

$$\mathbb{E}Q_R(z)e^{\mathrm{i}t\Delta(z_0)} = O_{\prec}\big(n^{-\frac{1}{2}+4\nu}\big), \tag{5.18}$$

$$\mathbb{E}Q_R^k(z)e^{\mathrm{i}t\Delta(z_0)} = (k-1)V\mathbb{E}Q_R^{k-2}(z)e^{\mathrm{i}t\Delta(z_0)} + O_{\prec}\big(n^{-\frac{1}{2}+4\nu}\big). \tag{5.19}$$

For any $s, t \in \mathbb{R}$, by (5.17), we have

$$\mathbb{E}e^{\mathrm{i}sQ_R(z)+\mathrm{i}t\Delta(z_0)} = \sum_{k=0}^{2N-1} \frac{(\mathrm{i}s)^k}{k!}\mathbb{E}Q_R^k(z)e^{\mathrm{i}t\Delta(z_0)} + O\left(\frac{s^{2N}}{(2N)!}\mathbb{E}Q_R^{2N}(z)\right). \tag{5.20}$$

For the error term on the right-hand side of (5.20), using (5.19) recursively for $t = 0$, we first find

$$\mathbb{E}Q_R^{2N}(z) = (2N-1)!!V^N + O_{\prec}\big(n^{-\frac{1}{2}+4\nu}\big).$$

Thus, for arbitrarily small $\epsilon > 0$, by taking N sufficiently large, we have $\frac{(2N-1)!!V^N}{(2N)!} < \epsilon$ and it follows that

$$\left|\mathbb{E}e^{\mathrm{i}sQ_R(z)+\mathrm{i}t\Delta(z_0)} - \sum_{k=0}^{2N-1} \frac{(\mathrm{i}s)^k}{k!}\mathbb{E}Q_R^k(z)e^{\mathrm{i}t\Delta(z_0)}\right| < \epsilon + O_\prec\big(n^{-\frac{1}{2}+4\nu}\big). \tag{5.21}$$

Using (5.19), we get the following estimate:

$$\sum_{k=0}^{2N-1} \frac{(\mathrm{i}s)^k}{k!}\mathbb{E}Q_R^k(z)e^{\mathrm{i}t\Delta(z_0)} = \sum_{k=0}^{N-1} \frac{(\mathrm{i}s)^{2k}}{(2k)!!}V^k\mathbb{E}e^{\mathrm{i}t\Delta(z_0)} + O_\prec\big(n^{-\frac{1}{2}+4\nu}\big). \tag{5.22}$$

Next, combing (5.22) with the fact

$$\exp\left(\frac{x^2}{2}\right) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!!},$$

together with (5.21), we conclude that

$$\big|\mathbb{E}e^{\mathrm{i}sQ_R(z)+\mathrm{i}t\Delta(z_0)} - e^{-\frac{1}{2}Vs^2}\mathbb{E}e^{\mathrm{i}t\Delta(z_0)}\big| < 2\epsilon + O_\prec\big(n^{-\frac{1}{2}+4\nu}\big). \tag{5.23}$$

The asymptotic independence of $Q_R(z)$ and $\Delta(z_0)$ is a consequence of (5.23) and the fact ϵ is arbitrarily small. (5.13) can be proved by setting $s = 0$. Although Proposition 5.2 is proved under the choice (5.14), by continuity of G outside of the support of MP law, we know $Q(z_0) = Q_R(z) + O(N^{-C'})$ with high probability for some positive constant C'. This concludes the proof of Proposition 5.1. □

Acknowledgments. X. C. Ding would like to thank Wei Q. Deng for many helpful discussions on the applications in statistical genetics. The authors also would like to thank two anonymous referees, an Associate Editor and Editor for their suggestions and comments, which have significantly improved the paper. The authors also want to thank Jiang Hu for many helpful discussions.

Z. G. Bao was partially supported by Hong Kong RGC grant ECS 26301517, GRF 16300618, GRF 16301519 and NSFC 11871425. X. C. Ding was partially supported by NSERC of Canada (Jeremy Quastel) and was previously at University of Toronto and Duke University. K. Wang was partially supported by HKUST Initiation Grant IGN16SC05, Hong Kong RGC grants GRF 16301618 and GRF 16308219.

SUPPLEMENTARY MATERIAL

Supplement to "Singular vector and singular subspace distribution for the matrix denoising model" (DOI: 10.1214/20-AOS1960SUPP; .pdf). This file contains detailed simulation results, further discussions on statistical applications, auxiliary lemmas, the proofs of the theorems and lemmas of the paper.

REFERENCES

[1] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620 https://doi.org/10.1214/aoms/1177704248

[2] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

[3] 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

[4] BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009

[5] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

[6] 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

[7] BAO, Z., DING, X., WANG, J. and WANG, K. (2019). Principal components of spiked covariance matrices in the supercritical regime. Available at arXiv:1907.12251.

[8] BAO, Z., DING, X. and WANG, K. (2021). Supplement to "Singular vector and singular subspace distribution for the matrix denoising model." https://doi.org/10.1214/20-AOS1960SUPP

[9] BAO, Z., HU, J., PAN, G. and ZHOU, W. (2019). Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Ann. Statist.* **47** 612–640. MR3909944 https://doi.org/10.1214/18-AOS1704

[10] BAO, Z., PAN, G. and ZHOU, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864 https://doi.org/10.1214/14-AOS1281

[11] BENAYCH-GEORGES, F., GUIONNET, A. and MAIDA, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** 1621–1662. MR2835249 https://doi.org/10.1214/EJP.v16-929

[12] 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

[13] 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

[14] BLOEMENDAL, A., ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19** no. 33, 53. MR3183577 https://doi.org/10.1214/ejp.v19-3054

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

[16] BLOEMENDAL, A. and VIRÁG, B. (2013). Limits of spiked random matrices I. *Probab. Theory Related Fields* **156** 795–825. MR3078286 https://doi.org/10.1007/s00440-012-0443-2

[17] BLOEMENDAL, A. and VIRÁG, B. (2016). Limits of spiked random matrices II. *Ann. Probab.* **44** 2726–2769. MR3531679 https://doi.org/10.1214/15-AOP1033

[18] 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

[19] 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

[20] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. MR3988761 https://doi.org/10.1214/18-AOS1752

[21] CAPITAINE, M. (2018). Limiting eigenvectors of outliers for spiked information-plus-noise type matrices. In *Séminaire de Probabilités XLIX*. *Lecture Notes in Math.* **2215** 119–164. Springer, Cham. MR3837103

[22] CAPITAINE, M. and DONATI-MARTIN, C. (2017). Spectrum of deformed random matrices and free probability. In *Advanced Topics in Random Matrices*. *Panor. Synthèses* **53** 151–190. Soc. Math. France, Paris. MR3792626

[23] CAPITAINE, M. and DONATI-MARTIN, C. (2018). Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of Wigner matrices. Available at arXiv:1807.07773.

[24] 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

[25] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2012). Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 107–133. MR2919200 https://doi.org/10.1214/10-AIHP410

[26] DING, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli* **26** 387–417. MR4036038 https://doi.org/10.3150/19-BEJ1129

[27] DING, X. and YANG, F. (2019). Spiked separable covariance matrices and principal components. Available at arXiv:1905.13060.

[28] DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257

[29] EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592 https://doi.org/10.1214/009117906000000917

[30] ERDŐS, L., KNOWLES, A. and YAU, H.-T. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. MR3119922 https://doi.org/10.1007/s00023-013-0235-y

[31] FAN, J., SUN, Q., ZHOU, W.-X. and ZHU, Z. (2018). Principal component analysis for big data. Available at arXiv:1801.01602.

[32] 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

[33] FAN, J. and ZHONG, Y. (2018). Optimal subspace estimation using overidentifying vectors via generalized method of moments. Available at arXiv:1805.02826.

[34] FAN, Z., JOHNSTONE, I. and SUN, Y. (2018). Spiked covariances and principal components analysis in high-dimensional random effects models. Available at arXiv:1806.09529.

[35] 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

[36] GOLUB, G. H. and VAN LOAN, C. F. (2012). *Matrix Computations*, 3rd ed. *Johns Hopkins Studies in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD.

[37] HACHEM, W., LOUBATON, P., MESTRE, X., NAJIM, J. and VALLET, P. (2013). A subspace estimator for fixed rank perturbations of large random matrices. *J. Multivariate Anal.* **114** 427–447. MR2993897 https://doi.org/10.1016/j.jmva.2012.08.006

[38] HECKEL, R., TSCHANNEN, M. and BÖLCSKEI, H. (2017). Dimensionality-reduced subspace clustering. *Inf. Inference* **6** 246–283. MR3764525 https://doi.org/10.1093/imaiai/iaw021

[39] HUANG, J., QIU, Q. and CALDERBANK, R. (2016). The role of principal angles in subspace classification. *IEEE Trans. Signal Process.* **64** 1933–1945. MR3470590 https://doi.org/10.1109/TSP.2015.2500889

[40] KAY, S. M. (1998). *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice-Hall, New York.

[41] 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

[42] KNOWLES, A. and YIN, J. (2014). The outliers of a deformed Wigner matrix. *Ann. Probab.* **42** 1980–2031. MR3262497 https://doi.org/10.1214/13-AOP855

[43] KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770 https://doi.org/10.1007/s00440-016-0730-4

[44] KOLTCHINSKII, V., LÖFFLER, M. and NICKL, R. (2020). Efficient estimation of linear functionals of principal components. *Ann. Statist.* **48** 464–490. MR4065170 https://doi.org/10.1214/19-AOS1816

[45] KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1976–2013. MR3573302 https://doi.org/10.1214/15-AIHP705

[46] KOLTCHINSKII, V. and LOUNICI, K. (2017). Normal approximation and concentration of spectral projectors of sample covariance. *Ann. Statist.* **45** 121–157. MR3611488 https://doi.org/10.1214/16-AOS1437

[47] LEE, J. O. and SCHNELLI, K. (2018). Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields* **171** 543–616. MR3800840 https://doi.org/10.1007/s00440-017-0787-8

[48] LEE, M., SHEN, H., HUANG, J. Z. and MARRON, J. S. (2010). Biclustering via sparse singular value decomposition. *Biometrics* **66** 1087–1095. MR2758496 https://doi.org/10.1111/j.1541-0420.2010.01392.x

[49] LEVIN, A. and NADLER, B. (2011). Natural image denoising: Optimality and inherent bounds. In 2011 *IEEE Conference On Computer Vision And Pattern Recognition* (*CVPR*) **40**.

[50] MARČENKO, V. A. and PASTUR, L. A. (1967). The spectrum of random matrices. *Teor. Funkciĭ Funkcional. Anal. i Priložen. Vyp.* **4** 122–145. MR0221550

[51] NADLER, B. and JOHNSTONE, I. M. (2011). On the distribution of Roy's Largest Root Test in MANOVA and in signal detection in noise. Technical report, Department of Statistics, Stanford University.

[52] 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

[53] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[54] 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

[55] PETERFREUND, E. and GAVISH, M. (2018). Multidimensional scaling of noisy high dimensional data. Available at arXiv:1801.10229.

[56] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. MR3199978 https://doi.org/10.1214/13-AAP939

[57] PIZZO, A., RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 64–94. MR3060148 https://doi.org/10.1214/11-AIHP459

[58] WANG, W., MIGUEL, A. C.-P. and LU, Z. (2011). A denoising view of matrix completion. *Adv. Neural Inf. Process. Syst.* **24** 334–342.

[59] 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

[60] YANG, D., MA, Z. and BUJA, A. (2014). A sparse singular value decomposition method for high-dimensional data. *J. Comput. Graph. Statist.* **23** 923–942. MR3270704 https://doi.org/10.1080/10618600.2013.858632

[61] YANG, D., MA, Z. and BUJA, A. (2016). Rate optimal denoising of simultaneously sparse and low rank matrices. *J. Mach. Learn. Res.* **17** Paper No. 92, 27. MR3543498

[62] ZAHERNIA, A., DEHGHANI, M. J. and JAVIDAN, R. (2011). MUSIC algorithm for DOA estimation using MIMO arrays. In 2011 6*th International Conference on Telecommunication Systems, Services, and Applications* (*TSSA*) 149–153.

[63] ZHONG, Y. and BOUMAL, N. (2018). Near-optimal bounds for phase synchronization. *SIAM J. Optim.* **28** 989–1016. MR3782406 https://doi.org/10.1137/17M1122025