SINGULAR VALUES OF LARGE NON-CENTRAL RANDOM MATRICES

WŁODEK BRYC AND JACK W. SILVERSTEIN

Abstract. We study largest singular values of large random matrices, each with mean of a fixed rank K. Our main result is a limit theorem as the number of rows and columns approach infinity, while their ratio approaches a positive constant. It provides a decomposition of the largest K singular values into the deterministic rate of growth, random centered fluctuations given as explicit linear combinations of the entries of the matrix, and a term negligible in probability. We use this representation to establish asymptotic normality of the largest singular values for random matrices with means that have block structure. We also deduce asymptotic normality for the largest eigenvalues of a random matrix arising in a model of population genetics.

1. Introduction

Finite rank perturbations of random matrices have been studied by numerous authors, starting with [5], [3]. This paper can be viewed as an extension of work done in [8] which describes the limiting behavior of the largest singular value, λ_1, of the $M \times N$ random matrix $\mathbf{D}^{(N)}$, consisting of i.i.d. random variables with common mean $\mu > 0$, and $M/N \to c$ as $N \to \infty$. Immediate results are obtained using known properties on the spectral behavior of centered matrices. Indeed, express $\mathbf{D}^{(N)}$ in the form

$$\mathbf{D}^{(N)} = \mathbf{C}^{(N)} + \mu \mathbf{1}_M \mathbf{1}_N^* \tag{1.1}$$

where $\mathbf{C}^{(N)}$ is an $M \times N$ matrix containing of i.i.d. mean 0 random variables having variance σ^2 and finite fourth moment, and $\mathbf{1}_k$ is the k dimensional vector consisting of 1's ($*$ denotes transpose). When the entries of $\mathbf{C}^{(N)}$ come from the first M rows and N columns of a doubly infinite array of random variables, then it is known ([11]) that the largest singular value of $\frac{1}{\sqrt{N}}\mathbf{C}$ converges a.s. to $\sigma(1+\sqrt{c})$ as $N \to \infty$. Noticing that the sole positive singular value of $\mu \mathbf{1}_M \mathbf{1}_N^*$ is $\mu\sqrt{MN}$, and using the fact that (see for example [9])

$$|\lambda_1 - \mu\sqrt{MN}| \le \|\mathbf{C}\|,$$

($\|\cdot\|$ denoting spectral norm on rectangular matrices) we have that almost surely, for all N

$$\lambda_1 = \mu\sqrt{MN} + O(\sqrt{N}).$$

From just considering the sizes of the largest singular values of $\mathbf{C}^{(N)}$ and $\mu \mathbf{1}_M \mathbf{1}_N^*$ one can take the view that $\mathbf{D}_N$ is a perturbation of a rank one matrix.

A result in [8] reveals that the difference between λ_1 and $\mu\sqrt{MN}$ is smaller than $O(\sqrt{N})$. It is shown that

$$\lambda_1 = \mu\sqrt{MN} + \frac{1}{2}\frac{\sigma^2}{\mu}\left(\sqrt{\frac{M}{N}} + \sqrt{\frac{N}{M}}\right) + \frac{1}{\sqrt{MN}}\mathbf{1}_M \mathbf{C}^{(N)} \mathbf{1}_N^* + \frac{1}{\sqrt{M}} Z_N, \tag{1.2}$$

where $\{Z_N\}$ is tight (i.e. stochastically bounded). Notice that the third term converges in distribution to an $N(0,\sigma^2)$ random variable.

This paper generalizes the result in [8] by both increasing the rank of the second term on the right of (1.1) while maintaining the same singular value dominance of this term over the random one, and relaxing the assumptions on the entries of $\mathbf{C}$. The goal is to cover the setting that is motivated by applications to population biology [7], [2], see also Section 3.3.

We use the following notation. We write $\mathbf{A} \in \mathcal{M}_{M\times N}$ to indicate the dimensions of a real matrix $\mathbf{A}$, and we denote by $\mathbf{A}^*$ its transpose; $[\mathbf{A}]_{r,s}$ denotes the (r,s) entry of matrix $\mathbf{A}$. $\mathbf{I}_d$ is the $d \times d$ identity matrix.

2010 *Mathematics Subject Classification.* Primary 60B10; 15A18; Secondary 92D10.

Key words and phrases. non-central random matrices; singular values; asymptotic normality.

We use the spectral norm $\|\mathbf{A}\| = \sup_{\{\mathbf{x}:\|\mathbf{x}\|=1\}} \|\mathbf{A}\mathbf{x}\|$ and occasionally the Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}\mathbf{A}\mathbf{A}^*}$. Recall that for an $M \times N$ matrix we have

$$\|\mathbf{A}\| \leq \|\mathbf{A}\|_F. \tag{1.3}$$

Vectors are denoted by lower case boldface letters like $\mathbf{x}$ and treated as column matrices so that the Euclidean length is $\|\mathbf{x}\|^2 = \mathbf{x}^*\mathbf{x}$.

Throughout the paper, we use the same letter C to denote various positive non-random constants that do not depend on N. The phrase "... holds for all N large enough" always means that there exists a non-random N_0 such that "..." holds for all $N > N_0$.

We fix $K \in \mathbb{N}$ and a sequence of integers $M = M^{(N)}$ such that

$$\lim_{N\to\infty} M/N = c > 0. \tag{1.4}$$

As a finite rank perturbation we take a sequence of deterministic rank K matrices $\mathbf{B} = \mathbf{B}^{(N)} \in \mathcal{M}_{M\times N}$ with K largest singular values $\rho_1^{(N)} \geq \rho_2^{(N)} \geq \cdots \geq \rho_K^{(N)}$.

Assumption 1.1. We assume that the limits

$$\gamma_r := \lim_{N\to\infty} \frac{\rho_r^{(N)}}{\sqrt{MN}} \tag{1.5}$$

exist and are distinct and strictly positive, $\gamma_1 > \gamma_2 > \cdots > \gamma_K > 0$.

Our second set of assumptions deals with randomness.

Assumption 1.2. Let $\mathbf{C} = \mathbf{C}^{(N)}$ be an $M \times N$ random matrix with real independent entries $[\mathbf{C}]_{i,j} = X_{i,j}$. (The distributions of the entries may differ, and may depend on N.) We assume that $\mathbb{E}(X_{i,j}) = 0$ and that there exists a constant C such that

$$\mathbb{E}(X_{i,j}^4) \leq C. \tag{1.6}$$

In particular, the variances

$$\sigma_{i,j}^2 = E(X_{i,j}^2)$$

are uniformly bounded.

We are interested in the asymptotic behavior of the K largest singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$ of the sequence of noncentered random $M \times N$ matrices

$$\mathbf{D} = \mathbf{D}^{(N)} = \mathbf{C} + \mathbf{B}. \tag{1.7}$$

Our main result represents each singular value λ_r as a sum of four terms, which represent the rate of growth, random centered fluctuation, deterministic shift, and a term negligible in probability. To state this result we need additional notation.

The rate of growth is determined by the singular values $\rho_1^{(N)} \geq \rho_2^{(N)} \geq \cdots \geq \rho_K^{(N)} > 0$ of the deterministic perturbation matrix $\mathbf{B}$. The singular value decomposition of $\mathbf{B}$ can be written as

$$\mathbf{B} = \mathbf{F}\mathrm{diag}(\rho_1, \ldots, \rho_k)\mathbf{V}^* = \mathbf{F}\mathbf{G}^*, \tag{1.8}$$

where $\mathbf{F} \in \mathcal{M}_{M\times K}$ has orthonormal columns and $\mathbf{G} = \mathbf{V}\mathrm{diag}(\rho_1, \ldots, \rho_k) \in \mathcal{M}_{N\times K}$ has orthogonal columns of lengths $\rho_1, \ldots, \rho_K$.

The random centered fluctuation term for the r-th singular value λ_r is

$$Z_r^{(N)} = \frac{[\mathbf{Z}_0]_{r,r}}{\gamma_r}, \tag{1.9}$$

where $\mathbf{Z}_0$ is a random centered $K \times K$ matrix given by

$$\mathbf{Z}_0 = \frac{1}{\sqrt{MN}} \mathbf{G}^*\mathbf{C}^*\mathbf{F}. \tag{1.10}$$

Note that $\mathbf{Z}_0$ implicitly depends on N.

The expression for the constant shift depends on the variances of the entries of $\mathbf{C}$. To write the expression we introduce a diagonal $N\times N$ matrix $\mathbf{\Delta}_R=\mathbb{E}(\mathbf{C}^*\mathbf{C})/M$ and a diagonal $M\times M$ matrix $\mathbf{\Delta}_S=\mathbb{E}(\mathbf{C}\mathbf{C}^*)/N$. The diagonal entries of these matrices are

$$[\mathbf{\Delta}_R]_{j,j}=\frac{1}{M}\sum_{i=1}^M\sigma_{i,j}^2,\quad [\mathbf{\Delta}_S]_{i,i}=\frac{1}{N}\sum_{j=1}^N\sigma_{i,j}^2.$$

Let $\mathbf{\Sigma}_r=\mathbf{\Sigma}_R^{(N)}$ and $\mathbf{\Sigma}_S=\mathbf{\Sigma}_S^{(N)}$ be deterministic $K\times K$ matrices given by

$$\mathbf{\Sigma}_R=\mathbf{G}^*\mathbf{\Delta}_R\mathbf{G} \tag{1.11}$$

and

$$\mathbf{\Sigma}_S=\mathbf{F}^*\mathbf{\Delta}_S\mathbf{F} \tag{1.12}$$

Define

$$m_r^{(N)}=\frac{1}{2}\left[\frac{\sqrt{c}}{\gamma_r^3MN}\mathbf{\Sigma}_R+\frac{1}{\sqrt{c}\gamma_r}\mathbf{\Sigma}_S\right]_{r,r}. \tag{1.13}$$

Theorem 1.1. *With the above notation, there exist $\varepsilon_1^{(N)}\to 0,\dots,\varepsilon_K^{(N)}\to 0$ in probability such that for $1\le r\le K$ we have*

$$\lambda_r=\rho_r^{(N)}+Z_r^{(N)}+m_r^{(N)}+\varepsilon_r^{(N)}. \tag{1.14}$$

Expression (1.14) is less precise than (1.2) (where the negligible term $Z_N/\sqrt{M}\to 0$ in probability at known rate as Z_n is stochastically bounded), but it is strong enough to establish asymptotic normality under appropriate additional conditions. Such applications require additional assumptions and are worked out in Section 3.

Remark 1.1. In our motivating example in Section 3.3, the natural setting is factorization of finite rank perturbation as

$$\mathbf{B}=\widetilde{\mathbf{F}}\widetilde{\mathbf{G}}^*=\sum_{s=1}^K\widetilde{\mathbf{f}}_s\widetilde{\mathbf{g}}_s^*, \tag{1.15}$$

where $\widetilde{\mathbf{F}}\in\mathcal{M}_{M\times K}$ has orthonormal columns $\widetilde{\mathbf{f}}_s$, but the columns, $\widetilde{\mathbf{g}}_s$, of $\widetilde{\mathbf{G}}\in\mathcal{M}_{N\times K}$ are not necessarily orthogonal. These matrices are a natural input for the problem so we would like to maintain their roles and recast Theorem 1.1 in terms of such matrices. We introduce matrices $\widetilde{\mathbf{R}}_0=\widetilde{\mathbf{R}}_0^{(N)}\in\mathcal{M}_{K\times K}$ by

$$\widetilde{\mathbf{R}}_0=\widetilde{\mathbf{G}}^*\widetilde{\mathbf{G}} \tag{1.16}$$

with eigenvalues $\rho_1^2\ge\dots\ge\rho_K^2$ and we denote the corresponding orthonormal eigenvectors by $\widetilde{\mathbf{u}}_1,\dots,\widetilde{\mathbf{u}}_K$. We claim that (1.9) is

$$Z_r^{(N)}=\frac{1}{\gamma_r}\widetilde{\mathbf{u}}_r^*\widetilde{\mathbf{Z}}_0\widetilde{\mathbf{u}}_r, \tag{1.17}$$

with

$$\widetilde{\mathbf{Z}}_0=\frac{1}{\sqrt{MN}}\widetilde{\mathbf{G}}^*\mathbf{C}^*\widetilde{\mathbf{F}}, \tag{1.18}$$

and similarly that (1.13) is

$$m_r^{(N)}=\frac{1}{2\sqrt{c}\gamma_r}\widetilde{\mathbf{u}}_r^*\left(\frac{c}{\gamma_r^2MN}\widetilde{\mathbf{\Sigma}}_R+\widetilde{\mathbf{\Sigma}}_S\right)\widetilde{\mathbf{u}}_r. \tag{1.19}$$

with $\widetilde{\mathbf{G}}$ and $\widetilde{\mathbf{F}}$ used in expressions (1.11) and (1.12) which define $\widetilde{\mathbf{\Sigma}}_R$ and $\widetilde{\mathbf{\Sigma}}_S$.

Indeed, let $\widetilde{\mathbf{G}}^*=\widetilde{\mathbf{U}}\operatorname{diag}(\rho_r)\widetilde{\mathbf{V}}^*$ be its singular value decomposition. If N is large enough so that the singular values are distinct, we may assume that $\widetilde{\mathbf{u}}_1,\dots,\widetilde{\mathbf{u}}_K$ are the columns of $\widetilde{\mathbf{U}}$. Then $\widetilde{\mathbf{V}}$ is determined uniquely and $\mathbf{B}=\widetilde{\mathbf{F}}\widetilde{\mathbf{U}}\operatorname{diag}(\rho_r)\widetilde{\mathbf{V}}^*$. Since by assumption our singular values are distinct, with proper alignment we may assume that $\mathbf{G}=\operatorname{diag}(\rho_r)\widetilde{\mathbf{V}}^*$ and then necessarily $\mathbf{F}=\widetilde{\mathbf{F}}\widetilde{\mathbf{U}}$ in (1.8). So for any $\mathbf{A}\in\mathcal{M}_{N\times M}$ we have

$$\widetilde{\mathbf{u}}_r^*\widetilde{\mathbf{G}}^*\mathbf{A}\widetilde{\mathbf{F}}\widetilde{\mathbf{u}}_r=[\mathbf{G}^*\mathbf{A}\mathbf{F}]_{r,r},$$

which we apply three times with $\mathbf{A} = \mathbf{C}^*$, $\mathbf{A} = \widetilde{\mathbf{\Sigma}}_R$ and $\mathbf{A} = \widetilde{\mathbf{\Sigma}}_S$.

2. Proof of Theorem 1.1

Throughout the proof, we assume that all our random variables are defined on a single probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In the initial parts of the proof we will be working on subsets $\Omega_N \subset \Omega$ such that $\mathbb{P}(\Omega_N) \to 1$.

2.1. **Singular value criterion.** In this section, we fix $r \in \{1, \dots, K\}$ and let $\lambda = \lambda_r$. For $\|\mathbf{C}\|^2 < \lambda^2$, matrices $\mathbf{I}_M - \frac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*$ and $\mathbf{I}_N - \frac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}$ are invertible, so we consider the following $K \times K$ matrices:

$$\mathbf{Z} = \frac{1}{\lambda}\mathbf{G}^*\left(\mathbf{I}_N - \tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{C}^*\mathbf{F}\,, \tag{2.1}$$

$$\mathbf{S} = \mathbf{F}^*\left(\mathbf{I}_M - \tfrac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}\mathbf{F}\,, \tag{2.2}$$

$$\mathbf{R} = \mathbf{G}^*\left(\mathbf{I}_N - \tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{G}\,. \tag{2.3}$$

(These auxiliary random matrices depend on N and λ, and are well defined only on a subset of the probability space Ω. We will see that these matrices are critical for our subsequent analysis.)

Lemma 2.1. *If $\|\mathbf{C}\|^2 < \lambda^2/2$ then*

$$\det\begin{bmatrix}\mathbf{Z} - \lambda\mathbf{I}_K & \mathbf{R}\\ \mathbf{S} & \mathbf{Z}^* - \lambda\mathbf{I}_K\end{bmatrix} = 0. \tag{2.4}$$

Proof. The starting point is the singular value decomposition $\mathbf{D} = \mathbf{U}\Lambda\mathbf{V}^*$. We choose the r-th columns $\mathbf{u} \in \mathbb{R}^M$ of $\mathbf{U}$ and $\mathbf{v} \in \mathbb{R}^N$ of $\mathbf{V}$ that correspond to the singular value $\lambda = \lambda_r$. Then (1.7) gives

$$\mathbf{C}\mathbf{v} + \sum_{s=1}^K (\mathbf{g}_s^*\mathbf{v})\mathbf{f}_s = \lambda\mathbf{u}\,, \tag{2.5}$$

$$\mathbf{C}^*\mathbf{u} + \sum_{s=1}^K (\mathbf{f}_s^*\mathbf{u})\mathbf{g}_s = \lambda\mathbf{v}\,. \tag{2.6}$$

where $\mathbf{f}_s \in \mathbb{R}^M$ is the s-th column of $\mathbf{F}$ and $\mathbf{g}_r \in \mathbb{R}^N$ is the r-th column of $\mathbf{G}$. Multiplying from the left by $\mathbf{C}^*$ or $\mathbf{C}$ we get

$$\begin{aligned}\mathbf{C}^*\mathbf{C}\mathbf{v} + \sum_{s=1}^K (\mathbf{g}_s^*\mathbf{v})\mathbf{C}^*\mathbf{f}_s &= \lambda\mathbf{C}^*\mathbf{u},\\ \mathbf{C}\mathbf{C}^*\mathbf{u} + \sum_{s=1}^K (\mathbf{f}_s^*\mathbf{u})\mathbf{C}\mathbf{g}_s &= \lambda\mathbf{C}\mathbf{v}.\end{aligned}$$

Using (2.5) and (2.6) to the expressions on the right, we get

$$\begin{aligned}\mathbf{C}^*\mathbf{C}\mathbf{v} + \sum_{s=1}^K (\mathbf{g}_s^*\mathbf{v})\mathbf{C}^*\mathbf{f}_s &= \lambda^2\mathbf{v} - \lambda\sum_{s=1}^K (\mathbf{f}_s^*\mathbf{u})\mathbf{g}_s\,,\\ \mathbf{C}\mathbf{C}^*\mathbf{u} + \sum_{s=1}^K (\mathbf{f}_s^*\mathbf{u})\mathbf{C}\mathbf{g}_s &= \lambda^2\mathbf{u} - \lambda\sum_{s=1}^K (\mathbf{g}_s^*\mathbf{v})\mathbf{f}_s\,.\end{aligned}$$

which we rewrite as

$$\begin{aligned}\lambda^2\mathbf{v} - \mathbf{C}^*\mathbf{C}\mathbf{v} &= \sum_{s=1}^K (\mathbf{g}_s^*\mathbf{v})\mathbf{C}^*\mathbf{f}_s + \lambda\sum_{s=1}^K (\mathbf{f}_s^*\mathbf{u})\mathbf{g}_s\,,\\ \lambda^2\mathbf{u} - \mathbf{C}\mathbf{C}^*\mathbf{u} &= \lambda\sum_{s=1}^K (\mathbf{g}_s^*\mathbf{v})\mathbf{f}_s + \sum_{s=1}^K (\mathbf{f}_s^*\mathbf{u})\mathbf{C}\mathbf{g}_s\,.\end{aligned}$$

This gives

$$\begin{aligned}\mathbf{v} &= \frac{1}{\lambda^2}\sum_{s=1}^K(\mathbf{g}_s^*\mathbf{v})\left(\mathbf{I}_N-\tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{C}^*\mathbf{f}_s+\frac{1}{\lambda}\sum_{s=1}^K(\mathbf{f}_s^*\mathbf{u})\left(\mathbf{I}_N-\tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{g}_s\,,\\ \mathbf{u} &= \frac{1}{\lambda}\sum_{s=1}^K(\mathbf{g}_s^*\mathbf{v})\left(\mathbf{I}_M-\tfrac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}\mathbf{f}_s+\frac{1}{\lambda^2}\sum_{s=1}^K(\mathbf{f}_s^*\mathbf{u})\left(\mathbf{I}_M-\tfrac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}\mathbf{C}\mathbf{g}_s\,.\end{aligned}$$

Notice that since $\mathbf{u},\mathbf{v}$ are unit vectors, some among the $2K$ the numbers $(\mathbf{f}_1^*\mathbf{u}),\dots,(\mathbf{f}_K^*\mathbf{u})$, $(\mathbf{g}_1^*\mathbf{v}),\dots,(\mathbf{g}_K^*\mathbf{v})$ must be non-zero. Since for $1\le t\le K$ we have

$$\begin{aligned}(\mathbf{g}_t^*\mathbf{v}) &= \frac{1}{\lambda^2}\sum_{s=1}^K(\mathbf{g}_s^*\mathbf{v})\mathbf{g}_t^*\left(\mathbf{I}_N-\tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{C}^*\mathbf{f}_s+\frac{1}{\lambda}\sum_{s=1}^K(\mathbf{f}_s^*\mathbf{u})\mathbf{g}_t^*\left(\mathbf{I}_N-\tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{g}_s\,,\\ (\mathbf{f}_t^*\mathbf{u}) &= \frac{1}{\lambda}\sum_{s=1}^K(\mathbf{g}_s^*\mathbf{v})\mathbf{f}_t^*\left(\mathbf{I}_M-\tfrac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}\mathbf{f}_s+\frac{1}{\lambda^2}\sum_{s=1}^K(\mathbf{f}_s^*\mathbf{u})\mathbf{f}_t^*\left(\mathbf{I}_M-\tfrac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}\mathbf{C}\mathbf{g}_s\,,\end{aligned}$$

noting that the entries of $\mathbf{Z}^*$ can be written as

$$[\mathbf{Z}^*]_{s,t}=\frac{1}{\lambda}\mathbf{f}_s^*(\mathbf{I}_M-\tfrac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*)^{-1}\mathbf{C}\mathbf{g}_t\,.$$

we see that the block matrix

$$\begin{bmatrix}\frac{1}{\lambda}\mathbf{Z} & \frac{1}{\lambda}\mathbf{R}\\ \frac{1}{\lambda}\mathbf{S} & \frac{1}{\lambda}\mathbf{Z}^*\end{bmatrix}$$

has eigenvalue 1. Thus $\det\begin{bmatrix}\frac{1}{\lambda}\mathbf{Z}-\mathbf{I}_K & \frac{1}{\lambda}\mathbf{R}\\ \frac{1}{\lambda}\mathbf{S} & \frac{1}{\lambda}\mathbf{Z}^*-\mathbf{I}_K\end{bmatrix}=0$ which for $\lambda>0$ is equivalent to (2.4).

□

Proposition 2.2 (Singular value criterion). *If* $\|\mathbf{C}\|^2<\lambda^2/4$ *then* $\mathbf{S}$ *is invertible and*

$$\det\left((\lambda\mathbf{I}_K-\mathbf{Z})\mathbf{S}^{-1}(\lambda\mathbf{I}_K-\mathbf{Z}^*)-\mathbf{R}\right)=0. \tag{2.7}$$

Similar equations that involve a $K\times K$ determinant appear in other papers on rank-K perturbations of random matrices, compare [10, Lemma 2.1 and Remark 2.2]. Note however that in our case λ enters the equation in a rather complicated way through $\mathbf{Z}=\mathbf{Z}(\lambda),\mathbf{R}=\mathbf{R}(\lambda),\mathbf{S}=\mathbf{S}(\lambda)$. The dependence of these matrices on N is also suppressed in our notation.

Proof. Note that if $\|\mathbf{C}\|^2\le\lambda^2/2$ then the norms of $\left(\mathbf{I}_M-\frac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}$ and $\left(\mathbf{I}_N-\frac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}$ are bounded by 2. Indeed, $\|\left(\mathbf{I}_M-\frac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}\|\le\sum_{k=0}^\infty\|(\mathbf{C}\mathbf{C}^*)^k\|/\lambda^{2k}\le\sum_{k=0}^\infty 1/2^k$. Since $\left(\mathbf{I}_M-\frac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}=\mathbf{I}_M+\frac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\left(\mathbf{I}_M-\frac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}$ and vectors $\mathbf{f}_1,\dots,\mathbf{f}_K$ are orthonormal, we get $\mathbf{F}^*\mathbf{F}=\mathbf{I}_K$ and

$$\mathbf{S}-\mathbf{I}_K=\frac{1}{\lambda^2}\mathbf{F}^*\mathbf{C}^*\mathbf{C}\left(\mathbf{I}_M-\tfrac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*\right)^{-1}\mathbf{F}\,.$$

Since $\|\mathbf{F}\|=1$ we have

$$\|\mathbf{S}-\mathbf{I}_K\|\le 2\|\mathbf{C}\|^2/\lambda^2. \tag{2.8}$$

We see that if $\|\mathbf{C}\|^2\le\lambda^2/4$ then $\|\mathbf{S}-\mathbf{I}_K\|\le 1/2$, so the inverse $\mathbf{S}^{-1}=\sum_{k=0}^\infty(\mathbf{I}-\mathbf{S})^k$ exists. For later reference we also note that

$$\|\mathbf{S}^{-1}\|\le 2. \tag{2.9}$$

Since

$$\begin{bmatrix}\mathbf{R} & \mathbf{Z}-\lambda\mathbf{I}_K\\ \mathbf{Z}^*-\lambda\mathbf{I}_K & \mathbf{S}\end{bmatrix}=\begin{bmatrix}\mathbf{Z}-\lambda\mathbf{I}_K & \mathbf{R}\\ \mathbf{S} & \lambda\mathbf{Z}^*-\lambda\mathbf{I}_K\end{bmatrix}\times\begin{bmatrix}0 & \mathbf{I}_K\\ \mathbf{I}_K & 0\end{bmatrix},$$

we see that (2.4) is equivalent to

$$\det\begin{bmatrix}\mathbf{R} & \mathbf{Z}-\lambda\mathbf{I}_K\\ \mathbf{Z}^*-\lambda\mathbf{I}_K & \mathbf{S}\end{bmatrix}=0.$$

Noting that $\lambda > 0$ by assumption, we see that (2.4) holds and gives (2.7), as by Schur's complement formula

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det D \det(A - BD^{-1}C).$$

□

2.2. **Equation for λ_r.** As was done previously we fix $r \in \{1, \dots, K\}$, and write $\lambda = \lambda_r$.

The main step in the proof of Theorem 1.1 is the following expression.

Proposition 2.3. *There exists a random sequence $\varepsilon_c^{(N)} \to 0$ in probability as $N \to \infty$ such that*

$$\lambda - \rho_r^{(N)} = \frac{N}{\lambda + \rho_r^{(N)}}[\mathbf{\Sigma}_S]_{r,r} + \frac{2\sqrt{MN}}{\lambda + \rho_r^{(N)}}[\mathbf{Z}_0]_{r,r} + \frac{M}{(\lambda + \rho_r^{(N)})\lambda^2}[\mathbf{\Sigma}_R]_{r,r} + \varepsilon_c^{(N)}. \tag{2.10}$$

The proof of Proposition 2.3 is technical and lengthy.

2.2.1. *Subset Ω_N.* With $\gamma_{K+1} := 0$, let

$$\delta := \min_{1 \le s \le K} (\gamma_s^2 - \gamma_{s+1}^2). \tag{2.11}$$

Assumption 1.1(iii) says that $\delta > 0$.

In the following the powers of N used are sufficient to conclude our arguments. Sharper bounds would not reduce moment assumption (1.6), which we use for [6] in the proof of Lemma 2.4.

Definition 2.1. Let $\Omega_N \subset \Omega$ be such that

$$\|\mathbf{C}\|^2 \le N^{5/4} \tag{2.12}$$

and

$$\max_{1 \le s \le K} |\lambda_s^2 - \gamma_s^2 MN| \le \tfrac{\delta}{4} MN. \tag{2.13}$$

We assume that N is large enough so that Ω_N is a non-empty set. In fact, $\mathbb{P}(\Omega_N) \to 1$, see Lemma 2.4.

We note that (2.13) implies that $c\sqrt{MN} \le \lambda_r \le C\sqrt{MN}$ with $c = \sqrt{\gamma_r^2 - \delta/4} \ge \sqrt{\gamma_K^2 - \delta/4} > \sqrt{\delta}/2$ and $C = \sqrt{\gamma_r^2 + \delta/4} < 2\gamma_1$. For later reference we state these bounds explicitly:

$$\frac{\sqrt{\delta MN}}{2} \le \lambda_r \le 2\,\gamma_1 \sqrt{MN}. \tag{2.14}$$

We also note that inequalities (2.12) and (2.14) imply that

$$\|\mathbf{C}\|^2 < \frac{\lambda_K^2}{4} \tag{2.15}$$

for all N large enough. (That is, for all $N > N_0$ with nonrandom constant N_0.) Thus matrices $\mathbf{Z}$, $\mathbf{R}$, $\mathbf{S}$ are well defined on Ω_N for large enough N and (2.7) holds.

Lemma 2.4. *For $1 \le r \le K$ we have $\lambda_r/\sqrt{MN} \to \gamma_r$ in probability. Furthermore, $\mathbb{P}(\Omega_N) \to 1$ as $N \to \infty$.*

Proof. From (1.7) we have $\mathbf{D} - \mathbf{B} = \mathbf{C}$, so by Weil-Mirsky theorem [9, page 204, Theorem 4.11], we have a bound $|\lambda_r - \rho_r^{(N)}| \le \|\mathbf{C}\|$ for the differences between the K largest singular values of $\mathbf{B}$ and $\mathbf{D}$. From [6, Theorem 2] we see that there is a constant C that does not depend on N such that $E\|\mathbf{C}\| \le C(\sqrt{M} + \sqrt{N})$. Thus

$$\frac{\lambda_r - \rho_r^{(N)}}{\sqrt{MN}} \to 0 \text{ in probability.}$$

Since $\rho_r^{(N)}/\sqrt{MN} \to \gamma_r > 0$ this proves the first part of the conclusion.

To prove the second part, we use the fact that continuous functions preserve convergence in probability, so $\lambda_r^2/(MN) \to \gamma_r^2$ in probability for $1 \le r \le K$. Thus

$$\mathbb{P}(\Omega_N') \le \mathbb{P}(\|\mathbf{C}\| > N^{5/8}) + \sum_{s=1}^K \mathbb{P}(|\lambda_s^2 - \gamma_s^2 MN| > \delta MN/4)$$
$$\le C\frac{\sqrt{M} + \sqrt{N}}{N^{5/8}} + \sum_{s=1}^K \mathbb{P}(|\lambda_s^2/(MN) - \gamma_s^2| > \delta/4) \to 0 \text{ as } N \to \infty.$$

□

2.2.2. *Proof of Proposition 2.3.* In view of (2.15), equation (2.7) holds on Ω_N if N is large enough. It implies that there is a (random) unit vector $\mathbf{x}_r^{(N)} = \mathbf{x} \in \mathbb{R}^K$ such that

$$(\lambda \mathbf{I}_K - \mathbf{Z})\mathbf{S}^{-1}(\lambda \mathbf{I}_K - \mathbf{Z}^*)\mathbf{x} = \mathbf{R}\mathbf{x}.$$

We further choose $\mathbf{x}$ with non-negative r-th component. Using diagonal matrix

$$\mathbf{R}_0 = \mathbf{G}^*\mathbf{G} = \mathrm{diag}(\rho_r^2)$$

we rewrite this as follows.

$$(\lambda^2 \mathbf{I}_K - \mathbf{R}_0)\mathbf{x} = \big(\lambda^2(\mathbf{I}_K - \mathbf{S}^{-1}) + \lambda(\mathbf{Z}\mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{Z}^*) - \mathbf{Z}\mathbf{S}^{-1}\mathbf{Z}^* + (\mathbf{R} - \mathbf{R}_0)\big)\,\mathbf{x}. \tag{2.16}$$

We now rewrite this equation using the (nonrandom) singular values $\rho_1^{(N)} \geq \rho_2^{(N)} \geq \cdots \geq \rho_K^{(N)} > 0$ of $\mathbf{B}$ and standard basis $\mathbf{e}_1, \dots, \mathbf{e}_K \in \mathbb{R}^K$.

Suppressing dependence on r and N in the notation, we insert

$$\mathbf{x} = (\alpha_1, \dots, \alpha_K)^* = \sum \alpha_s \mathbf{e}_s$$

into (2.16) and look at the s-th component. This shows that $\lambda = \lambda_r$ satisfies the following system of K equations

$$(\lambda^2 - \rho_s)\alpha_s = \lambda^2 \mathbf{e}_s^*(\mathbf{I}_K - \mathbf{S}^{-1})\mathbf{x} + \lambda \mathbf{e}_s^*(\mathbf{Z}\mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{Z}^*)\mathbf{x} + \mathbf{e}_s^*(\mathbf{R} - \mathbf{R}_0)\mathbf{x} - \mathbf{e}_s^*\mathbf{Z}\mathbf{S}^{-1}\mathbf{Z}^*\mathbf{x}, \tag{2.17}$$

where $1 \leq s \leq K$. (Recall that this is a system of highly nonlinear equations, as matrices $\mathbf{S}$, $\mathbf{Z}$ and $\mathbf{R}$, and the coefficients $\alpha_1, \dots \alpha_K$, depend implicitly on λ.)

It turns out that for our choice of $\lambda = \lambda_r$ random variable $\alpha_r = \alpha_r^{(N)}$ is close to its extreme value 1 while the other coefficients are asymptotically negligible. Since this only holds on Ω_N a more precise statement is as follows.

Lemma 2.5. *There exist deterministic constants C and N_0 such that for all $N > N_0$ and $\omega \in \Omega_N$ we have*

$$1 - CN^{-3/8} \leq \alpha_r \leq 1 \tag{2.18}$$

and

$$|\alpha_s| \leq (C/\sqrt{K-1})N^{-3/8} \text{ for } s \neq r. \tag{2.19}$$

Proof. Since $\sum \alpha_s^2 = 1$, inequality (2.18) is a consequence of (2.19). Indeed, $\alpha_r^2 = 1 - \sum_{s \neq r} \alpha_s^2 \geq 1 - C^2 N^{-3/4}$ and we use elementary inequality $\sqrt{1-x} \geq 1 - \sqrt{x}$ for $0 \leq x \leq 1$.

To prove (2.19), we use (2.17). By assumption, $\rho_j^2/(MN) \to \gamma_j^2$. Using (2.11), we choose N large enough so that $|\rho_s^2 - \gamma_s^2 MN| \leq \delta MN/4$. Then, with $s \neq r$ we get

$$\begin{aligned}|\lambda^2 - \rho_s^2| &= |(\lambda^2 - \gamma_r^2 MN) + (\gamma_r^2 - \gamma_s^2)MN + (\gamma_s^2 MN - \rho_s^2)| \\ &\geq |\gamma_r^2 - \gamma_s^2|MN - |\lambda^2 - \gamma_r^2 MN| - |\rho_s^2 - \gamma_s^2 MN| \geq (|\gamma_r^2 - \gamma_s^2| - \delta/2)MN \geq \frac{\delta}{2}MN.\end{aligned}$$

From (2.14) and (2.17) we get

$$\frac{\delta}{2}MN|\alpha_s| \leq 4\gamma_1^2 MN\|\mathbf{I}_K - \mathbf{S}^{-1}\| + 4\gamma_1\sqrt{MN}\|\mathbf{Z}\|\,\|\mathbf{S}^{-1}\| + \|\mathbf{R} - \mathbf{R}_0\| + \|\mathbf{Z}\|^2\|\mathbf{S}^{-1}\|.$$

Since $\mathbf{I}_K - \mathbf{S}^{-1} = \mathbf{S}^{-1}(\mathbf{S} - \mathbf{I}_K)$ using (2.9) we get

$$|\alpha_s| \leq \frac{16\gamma_1^2}{\delta}\|\mathbf{S} - \mathbf{I}_K\| + \frac{16\gamma_1}{\delta\sqrt{MN}}\|\mathbf{Z}\| + \frac{2}{\delta MN}\|\mathbf{R} - \mathbf{R}_0\| + \frac{4}{\delta MN}\|\mathbf{Z}\|^2. \tag{2.20}$$

We now estimate the norms of the $K \times K$ matrices on the right hand side. From (2.8) using (2.12) and (2.14) we get

$$\|\mathbf{S} - \mathbf{I}_K\| \leq \frac{2\|\mathbf{C}\|^2}{\lambda^2} \leq \frac{8N^{5/4}}{\delta MN} = \frac{8}{\delta}N^{1/4}M^{-1}. \tag{2.21}$$

Next, we bound $\|\mathbf{Z}\|$ using (2.1). Recall that $\| \left(\mathbf{I}_N - \frac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1} \| \leq 2$ for large enough N. From (1.5) we have$\|\mathbf{G}\| \leq 2\sqrt{MN}\gamma_1$, for large enough N. Using this, (2.12) and (2.14) we get

$$\|\mathbf{Z}\| \leq \|\mathbf{G}\|\frac{2}{\lambda}\|\mathbf{C}\| \leq \frac{4\sqrt{MN}\gamma_1}{\lambda}N^{5/8} \leq \frac{8\gamma_1}{\sqrt{\delta}}N^{5/8} \tag{2.22}$$

for large enough N.

Next we note that

$$\mathbf{R} - \mathbf{R}_0 = \frac{1}{\lambda^2}\mathbf{G}^*\mathbf{C}^*\mathbf{C}\left(\mathbf{I}_N - \tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{G}\,. \tag{2.23}$$

Thus (2.23) with bounds (2.14), (2.12) and the above bound on $\|\mathbf{G}\|$ give us for large enough N

$$\|\mathbf{R} - \mathbf{R}_0\| \leq \frac{8MN\gamma_1^2}{\lambda^2}\|\mathbf{C}\|^2 \leq \frac{32\gamma_1^2}{\delta}N^{5/4}. \tag{2.24}$$

Putting these bounds into (2.20) we get

$$|\alpha_s| \leq \frac{128\gamma_1^2}{\delta^2}N^{1/4}M^{-1} + \frac{128\gamma_1^2}{\delta\sqrt{\delta}}N^{1/8}M^{-1/2} + \frac{64\gamma_1^2}{\delta^2}N^{1/4}M^{-1} + \frac{256\gamma_1^2}{\delta^2}N^{1/4}M^{-1}.$$

In view of assumption (1.4), this proves (2.19).

□

The next step is to use Lemma 2.5 to rewrite the r-th equation in (2.17) to identify the "contributing terms" and the negligible "remainder" $\mathcal{R}$ which is of lower order than λ on Ω_N. We will accomplish this in several steps, so we will use the subscripts $a, b, c, \dots$ for bookkeeping purposes.

Define $\mathbf{x}_{(r)}$ by

$$\mathbf{x}_{(r)} = \sum_{s\neq r}\alpha_s\mathbf{e}_s.$$

We assume that N is large enough so that the conclusion of Lemma 2.5 holds and furthermore that $\alpha_r \geq 1/2$. Notice then that

$$\|\mathbf{x}_{(r)}\| \leq CN^{-3/8}. \tag{2.25}$$

Dividing (2.17) with $s = r$ by α_r we get

$$\lambda^2 - \rho_r^2 = \lambda^2\mathbf{e}_r^*(\mathbf{I}_K - \mathbf{S}^{-1})\mathbf{e}_r + 2\lambda\mathbf{e}_r^*\mathbf{Z}\mathbf{S}^{-1}\mathbf{e}_r + \mathbf{e}_r^*(\mathbf{R} - \mathbf{R}_0)\mathbf{e}_r - \mathbf{e}_r^*\mathbf{Z}\mathbf{S}^{-1}\mathbf{Z}^*\mathbf{e}_r + \sqrt{MN}\varepsilon_a^{(N)} + \mathcal{R}_a^{(N)},$$

where

$$\varepsilon_a^{(N)} = \frac{1}{\alpha_r}\mathbf{e}_r^*(\mathbf{Z}_0 + \mathbf{Z}_0^*)\mathbf{x}_{(r)}$$

and

$$\mathcal{R}_a^{(N)} = \lambda^2\frac{1}{\alpha_r}\mathbf{e}_r^*(\mathbf{I}_K - \mathbf{S}^{-1})\mathbf{x}_{(r)} + \lambda\frac{1}{\alpha_r}\mathbf{e}_r^*(\mathbf{Z}\mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{Z}^* - \frac{\sqrt{MN}}{\lambda}(\mathbf{Z}_0 + \mathbf{Z}_0^*))\mathbf{x}_{(r)} + \frac{1}{\alpha_r}\mathbf{e}_r^*(\mathbf{R} - \mathbf{R}_0)\mathbf{x}_{(r)} - \frac{1}{\alpha_r}\mathbf{e}_r^*\mathbf{Z}\mathbf{S}^{-1}\mathbf{Z}^*\mathbf{x}_{(r)}\,.$$

Here we slightly simplified the equation noting that since $\mathbf{S}$ is symmetric, $\mathbf{e}_r^*\mathbf{Z}\mathbf{S}^{-1}\mathbf{e}_r = \mathbf{e}_r^*\mathbf{S}^{-1}\mathbf{Z}^*\mathbf{e}_r$.

Our first task is to derive a deterministic bound for $\mathcal{R}_a^{(N)}$ on Ω_N.

Lemma 2.6. *There exist non-random constants C and N_0 such that on Ω_N for $N > N_0$ we have*

$$|\mathcal{R}_a^{(N)}| \leq CN^{7/8}.$$

Proof. The constant will be given by a complicated expression that will appear at the end of the proof. Within the proof, C denotes the constant from Lemma 2.5.

Notice that

$$\mathbf{Z} - \frac{\sqrt{MN}}{\lambda}\mathbf{Z}_0 = \frac{1}{\lambda^3}\mathbf{G}^*\mathbf{C}^*\mathbf{C}\left(\mathbf{I}_N - \tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{C}^*\mathbf{F}\,,$$

so for large enough N we get

$$\|\mathbf{Z} - \frac{\sqrt{MN}}{\lambda}\mathbf{Z}_0\| \leq \frac{4}{\lambda^3}\sqrt{MN}\gamma_1\|\mathbf{C}\|^3 \leq \frac{32\gamma_1}{\delta^{3/2}}N^{7/8}M^{-1}. \tag{2.26}$$

Recall (2.9) and recall that N is large enough so that $\alpha_r > 1/2$. Using (2.25) and writing $\mathbf{Z}\mathbf{S}^{-1} = \mathbf{Z}\mathbf{S}^{-1}(\mathbf{S}-\mathbf{I}_K) + \mathbf{Z}$ we get

$$|\mathcal{R}_a^{(N)}| \le 4C\lambda^2 N^{-3/8}\|\mathbf{S}-\mathbf{I}_K\| + 8C\lambda N^{-3/8}\|\mathbf{Z}\|\|\mathbf{S}-\mathbf{I}_K\| + 4C\lambda N^{-3/8}\|\mathbf{Z} - \frac{\sqrt{MN}}{\lambda}\mathbf{Z}_0\|$$
$$+ 2CN^{-3/8}\|\mathbf{R}-\mathbf{R}_0\| + 2CN^{-3/8}\|\mathbf{Z}\|^2$$
$$\le \frac{128C\gamma_1^2}{\delta}N^{7/8} + \frac{1024C\gamma_1^2}{\delta^{3/2}}M^{-1/2}N + \frac{256C\gamma_1^2}{\delta^{3/2}}M^{-1/2}N$$
$$+ \frac{64C\gamma_1^2}{\delta}N^{7/8} + \frac{128C\gamma_1^2}{\delta}N^{7/8}.$$

(Here we used (2.21), (2.22), then (2.26), (2.24) and (2.22) again.) This concludes the proof. □

Lemma 2.7. *For every $\eta > 0$, we have*

$$\lim_{N\to\infty} \mathbb{P}\left(\left\{\left|\varepsilon_a^{(N)}\right| > \eta\right\} \cap \Omega_N\right) = 0. \tag{2.27}$$

Proof. We first verify that each entry of the matrix $N^{-3/8}\mathbf{Z}_0$, which is well defined on Ω, converges in probability to 0. To do so, we bound the second moment of random variable $\xi = \mathbf{f}_r^*\mathbf{C}\mathbf{g}_s$. Since the entries of $\mathbf{C}$ are independent and centered random variables,

$$\mathbb{E}\xi^2 = \sum_{i=1}^M\sum_{j=1}^N [\mathbf{f}_r]_i^2\sigma_{i,j}^2[\mathbf{g}_s]_j^2 \le \sup_{i,j}\sigma_{i,j}^2\|\mathbf{g}_r\|^2 \le CMN. \tag{2.28}$$

Thus, see (1.10), each entry of matrix $\mathbf{Z}_0$ has bounded second moment, so

$$\zeta_N := N^{-3/8}\max_s|\mathbf{e}_r^*(\mathbf{Z}_0+\mathbf{Z}_0^*)\mathbf{e}_s| \to 0 \text{ in probability.}$$

To end the proof we note that by (2.19), for large enough N we have $\left|\varepsilon_a^{(N)}\right| \le 2CK\zeta_N$ on Ω_N, so

$$\mathbb{P}\left(\left\{\left|\varepsilon_a^{(N)}\right| > \eta\right\}\cap\Omega_N\right) \le \mathbb{P}\left(\{|\zeta_N| > \tfrac{\eta}{2CK}\}\cap\Omega_N\right) \le \mathbb{P}\left(|\zeta_N| > \tfrac{\eta}{CK}\right) \to 0.$$

□

Using the identity

$$\mathbf{I} - \mathbf{S}^{-1} = (\mathbf{S}-\mathbf{I}) - (\mathbf{S}-\mathbf{I})^2\mathbf{S}^{-1}$$

to the first term we rewrite (2.2.2) as

$$\lambda^2 - (\rho_r^{(N)})^2 = \lambda^2\mathbf{e}_r^*(\mathbf{S}-\mathbf{I}_K)\mathbf{e}_r + 2\lambda\mathbf{e}_r^*\mathbf{Z}\mathbf{e}_r + \mathbf{e}_r^*(\mathbf{R}-\mathbf{R}_0)\mathbf{e}_r - \mathbf{e}_r^*\mathbf{Z}\mathbf{Z}^*\mathbf{e}_r + \mathcal{R}_b^{(N)} + \sqrt{MN}\varepsilon_a^{(N)} + \mathcal{R}_a^{(N)}, \tag{2.29}$$

with

$$\mathcal{R}_b^{(N)} = -\lambda^2\mathbf{e}_r^*(\mathbf{S}-\mathbf{I}_K)^2\mathbf{S}^{-1}\mathbf{e}_r + 2\lambda\mathbf{e}_r^*\mathbf{Z}(\mathbf{I}_K-\mathbf{S})\mathbf{S}^{-1}\mathbf{e}_r - \mathbf{e}_r^*\mathbf{Z}(\mathbf{I}_K-\mathbf{S})\mathbf{S}^{-1}\mathbf{Z}^*\mathbf{e}_r\,.$$

Lemma 2.8. *There exist non-random constants C and N_0 such that on Ω_N for $N > N_0$ we have*

$$|\mathcal{R}_b^{(N)}| \le CN^{7/8}.$$

Proof. Using (2.14) and previous norm estimates (2.21) and (2.22), we get

$$|\mathcal{R}_b^{(N)}| \le 2\lambda^2\|\mathbf{S}-\mathbf{I}_K\|^2 + 4\lambda\|\mathbf{Z}\|\|\mathbf{S}-\mathbf{I}_K\| + 2\|\mathbf{Z}\|^2\|\mathbf{S}-\mathbf{I}_K\|$$
$$\le \frac{512\gamma_1^2}{\delta^2}N^{3/2}M^{-1} + \frac{512\gamma_1^2}{\delta^{3/2}}N^{11/8}M^{-1/2} + \frac{1024\gamma_1^2}{\delta^2}N^{3/2}M^{-1}.$$

This ends the proof. □

Define $K\times K$ random matrices

$$\mathbf{R}_1 = \mathbf{G}^*\mathbf{C}^*\mathbf{C}\mathbf{G} \tag{2.30}$$

and

$$\mathbf{S}_1 = \mathbf{F}^*\mathbf{C}\mathbf{C}^*\mathbf{F}\,. \tag{2.31}$$

Recall that $\mathbb{E}(\mathbf{R}_1) = M\mathbf{\Sigma}_R$ and $\mathbb{E}(\mathbf{S}_1) = N\mathbf{\Sigma}_S$, see (1.11) and (1.12).

Lemma 2.9. *There exist non-random constants C and N_0 such that on Ω_N for $N > N_0$ we have*

$$\|\mathbf{S} - \mathbf{I}_K - \frac{1}{\lambda^2}\mathbf{S}_1\| \leq CN^{-3/2}$$

and

$$\|\mathbf{R} - \mathbf{R}_0 - \tfrac{1}{\lambda^2}\mathbf{R}_1\| \leq CN^{1/2}.$$

Proof. Notice that

$$\mathbf{S} - \mathbf{I}_K - \frac{1}{\lambda^2}\mathbf{S}_1 = \frac{1}{\lambda^4}\mathbf{F}^*(\mathbf{C}\mathbf{C}^*)^2(\mathbf{I}_M - \tfrac{1}{\lambda^2}\mathbf{C}\mathbf{C}^*)^{-1}\mathbf{F}.$$

For large enough N (so that (2.12) and (2.14) hold), this gives

$$\|\mathbf{S} - \mathbf{I}_K - \frac{1}{\lambda^2}\mathbf{S}_1\| \leq \frac{2}{\lambda^4}\|\mathbf{C}\|^4 \leq \frac{32}{\delta^2}M^{-2}N^{1/2}.$$

Similarly, since

$$\mathbf{R} - \mathbf{R}_0 - \tfrac{1}{\lambda^2}\mathbf{R}_1 = \frac{1}{\lambda^4}\mathbf{G}^*(\mathbf{C}^*\mathbf{C})^2\left(\mathbf{I}_N - \tfrac{1}{\lambda^2}\mathbf{C}^*\mathbf{C}\right)^{-1}\mathbf{G}$$

for large enough N, we get

$$\|\mathbf{R} - \mathbf{R}_0 - \tfrac{1}{\lambda^2}\mathbf{R}_1\| \leq \max_{r,s}\frac{2}{\lambda^4}\|\mathbf{G}\|^2\|\mathbf{C}\|^4 \leq \frac{128\gamma_1^2}{\delta^2}M^{-1}N^{3/2}.$$

□

We now rewrite (2.29) as follows.

$$\lambda^2 - (\rho_r^{(N)})^2$$
$$= \mathbf{e}_r^*(\mathbf{S}_1)\mathbf{e}_r + 2\sqrt{MN}\mathbf{e}_r^*\mathbf{Z}_0\mathbf{e}_r + \frac{1}{\lambda^2}\mathbf{e}_r^*(\mathbf{R}_1)\mathbf{e}_r - \frac{MN}{\lambda^2}\mathbf{e}_r^*\mathbf{Z}_0\mathbf{Z}_0^*\mathbf{e}_r + \mathcal{R}_c^{(N)} + \mathcal{R}_b^{(N)} + \sqrt{MN}\varepsilon_a^{(N)} + \mathcal{R}_a^{(N)},$$

where

$$\mathcal{R}_c^{(N)} = \mathbf{e}_r^*\left(\lambda^2(\mathbf{S} - \mathbf{I}_K - \frac{1}{\lambda^2}\mathbf{S}_1) + 2(\lambda\mathbf{Z} - \sqrt{MN}\mathbf{Z}_0) + (\mathbf{R} - \mathbf{R}_0 - \frac{1}{\lambda^2}\mathbf{R}_1) + (\frac{MN}{\lambda^2}\mathbf{Z}_0\mathbf{Z}_0^* - \mathbf{Z}\mathbf{Z}^*)\right)\mathbf{e}_r\,.$$

Lemma 2.10. *There exist non-random constants C and N_0 such that on Ω_N for $N > N_0$ we have*

$$|\mathcal{R}_c^{(N)}| \leq CN^{7/8}.$$

Proof. As in the proof of Lemma 2.6, the final constant C can be read out from the bound at the end of the proof. In the proof, C is a constant from Lemma 2.9. By the triangle inequality, Lemma 2.9, (2.26) and (2.22), we have

$$\begin{aligned}|\mathcal{R}_c^{(N)}| &\leq \lambda^2\|\mathbf{S} - \mathbf{I}_K - \frac{1}{\lambda^2}\mathbf{S}_1\| + 2\|\lambda\mathbf{Z} - \sqrt{MN}\mathbf{Z}_0\| \\ &\quad + \|\mathbf{R} - \mathbf{R}_0 - \frac{1}{\lambda^2}\mathbf{R}_1\| + \|\mathbf{Z} - \frac{\sqrt{MN}}{\lambda}\mathbf{Z}_0\|(\|\mathbf{Z}\| + \frac{\sqrt{MN}}{\lambda}\|\mathbf{Z}_0\|) \\ &\leq 4C\gamma_1^2MN^{-1/2} + \frac{128\gamma_1^2}{\delta^{3/2}}M^{-1/2}N^{11/8} + CN^{1/2} + \frac{256\gamma_1^2}{\delta^2}M^{-1}N^{3/2} + \frac{128\gamma_1^2}{\delta^{3/2}}M^{-1}N^{3/2}.\end{aligned}$$

(Here we used the bound $\|\mathbf{Z}_0\| \leq 2\gamma_1 N^{5/8}$, which is derived similarly to (2.22).) □

The following holds on Ω. (Recall that expressions (2.31), (1.10) and (2.30) are well defined on Ω.)

Proposition 2.11. *There exists a random sequence $\varepsilon_b^{(N)} \to 0$ in probability as $N \to \infty$ such that*

$$\lambda - \rho_r^{(N)} = \frac{1}{\lambda + \rho_r^{(N)}}\mathbf{e}_r^*(\mathbf{S}_1)\mathbf{e}_r + \frac{2\sqrt{MN}}{\lambda + \rho_r^{(N)}}\mathbf{e}_r^*\mathbf{Z}_0\mathbf{e}_r + \frac{1}{(\lambda + \rho_r^{(N)})\lambda^2}\mathbf{e}_r^*(\mathbf{R}_1)\mathbf{e}_r + \varepsilon_b^{(N)}. \tag{2.32}$$

Proof. Let

$$\varepsilon_b^{(N)} = \begin{cases} \frac{1}{\lambda+\rho_r^{(N)}}\left(-\frac{MN}{\lambda^2}\mathbf{e}_r^*\mathbf{Z}_0\mathbf{Z}_0^*\mathbf{e}_r + \mathcal{R}_c^{(N)} + \mathcal{R}_b^{(N)} + \sqrt{MN}\varepsilon_a^{(N)} + \mathcal{R}_a^{(N)}\right) & \text{on } \Omega_N, \\ \lambda - \rho_r^{(N)} - \frac{1}{\lambda+\rho_r^{(N)}}\mathbf{e}_r^*(\mathbf{S}_1)\mathbf{e}_r - \frac{2\lambda}{\lambda+\rho_r^{(N)}}\mathbf{e}_r^*\mathbf{Z}_0\mathbf{e}_r - \frac{1}{(\lambda+\rho_r^{(N)})\lambda^2}\mathbf{e}_r^*(\mathbf{R}_1)\mathbf{e}_r & \text{otherwise.} \end{cases}$$

By Lemma 2.4 we have $\mathbb{P}(\Omega_N') \to 0$, so it is enough to show that given $\eta > 0$ we have $\mathbb{P}(\{|\varepsilon_b^{(N)}| > 5\eta\} \cap \Omega_N) \to 0$ as $N \to \infty$. Since the event $|\xi_1 + \cdots + \xi_5| > 5\eta$ is included in the union of events $|\xi_1| > \eta, \ldots, |\xi_5| > \eta$, in view of Lemmas 2.6, 2.8, 2.10, and Lemma 2.7 (recalling that expressions $\sqrt{MN}/(\lambda + \rho_r^{(N)})$ and MN/λ^2 are bounded by a non-random constant on Ω_N, see (2.14)) we only need to verify that

$$\mathbb{P}\left(\left\{\frac{\mathbf{e}_r^*\mathbf{Z}_0\mathbf{Z}_0^*\mathbf{e}_r}{\lambda + \rho_r^{(N)}} > \eta\right\} \cap \Omega_N\right) \le \mathbb{P}\left(\frac{\mathbf{e}_r^*\mathbf{Z}_0\mathbf{Z}_0^*\mathbf{e}_r}{\rho_r^{(N)}} > \eta\right) \to 0 \text{ as } N \to \infty.$$

Since for large enough N, we have $\rho_r^2 \ge \delta MN$, convergence follows from

$$\mathbb{E}\mathbf{e}_r^*\mathbf{Z}_0\mathbf{Z}_0^*\mathbf{e}_r \le \mathbb{E}\|\mathbf{Z}_0\|^2 \le \mathbb{E}\|\mathbf{Z}_0\|_F^2 \le K^2C, \tag{2.33}$$

where C is a constant from (2.28).

□

Proof of Proposition 2.3. Recall that $\mathbb{E}\mathbf{S}_1 = N\mathbf{\Sigma}_S$ and $\mathbb{E}\mathbf{R}_1 = M\mathbf{\Sigma}_R$. So expression (2.10) differs from (2.32) only by two terms:

$$\frac{1}{\lambda + \rho_r^{(N)}}\mathbf{e}_r^*(\mathbf{S}_1 - \mathbb{E}\mathbf{S}_1)\mathbf{e}_r$$

and

$$\frac{1}{(\lambda + \rho_r^{(N)})\lambda^2}\mathbf{e}_r^*(\mathbf{R}_1 - \mathbb{E}\mathbf{R}_1)\mathbf{e}_r.$$

Since $\rho_r^{(N)}/\sqrt{MN} \to \gamma_r > 0$ and by Lemma 2.4 we have $\lambda/\sqrt{MN} \to \gamma_r$ in probability, to end the proof we show that $\frac{1}{N}\|\mathbf{S}_1 - \mathbb{E}\mathbf{S}_1\|_F \to 0$ and $\frac{1}{N^3}\|\mathbf{R}_1 - \mathbb{E}\mathbf{R}_1\|_F \to 0$ in probability. To do so, we bound the second moments of the entries of the matrices. Recalling (2.31), we have

$$f_r^*(\mathbf{CC}^* - \mathbb{E}(\mathbf{CC}^*))f_s = \sum_{k=1}^N\sum_{i\ne j}[\mathbf{f}_r]_i X_{i,k}X_{j,k}[\mathbf{f}_s]_j + \sum_{k=1}^N\sum_{i=1}^M[\mathbf{f}_r]_i(X_{i,k}^2 - \sigma_{i,k}^2)[\mathbf{f}_s]_i = A_N + B_N \text{ (say)}.$$

By independence, we have

$$\mathbb{E}(A_N^2) = \sum_{k=1}^N\sum_{i\ne j}[\mathbf{f}_r]_i^2[\mathbf{f}_s]_j^2\sigma_{i,k}^2\sigma_{j,k}^2 \le C\sum_{k=1}^N\sum_{i=1}^N\sum_{j=1}^N[\mathbf{f}_r]_i^2[\mathbf{f}_s]_j^2 \le CN.$$

Next,

$$\mathbb{E}(B_N^2) = \sum_{k=1}^N\sum_{i=1}^M[\mathbf{f}_r]_i^2[\mathbf{f}_s]_i^2\mathbb{E}(X_{i,k}^2 - \sigma_{i,k}^2)^2 \le NC\sqrt{\sum_{i=1}^M[\mathbf{f}_r]_i^4}\sqrt{\sum_{i=1}^M[\mathbf{f}_s]_i^4} \le CN.$$

This shows that (with a different C) we have $\mathbb{E}\,|f_r^*(\mathbf{CC}^* - \mathbb{E}(\mathbf{CC}^*)f_s)|^2 \le CN$ and hence $\frac{1}{N}\|\mathbf{S}_1 - \mathbb{E}\mathbf{S}_1\|_F \to 0$ in mean square and in probability.

Similarly, recalling (2.30) we have

$$\mathbf{g}_r^*(\mathbf{C}^*\mathbf{C} - \mathbb{E}(\mathbf{C}^*\mathbf{C}))\mathbf{g}_s = \sum_{k=1}^M\sum_{i\ne j}[\mathbf{g}_r]_i[\mathbf{g}_s]_jX_{k,i}X_{k,j} + \sum_{k=1}^M\sum_{i=1}^N[\mathbf{g}_r]_i[\mathbf{g}_s]_i(X_{k,i}^2 - \sigma_{k,i}^2) = \widetilde{A}_N + \widetilde{B}_N \text{ (say)}.$$

Using independence of entries again, we get

$$\mathbb{E}(\widetilde{A}_N^2) = \sum_{k=1}^M\sum_{i\ne j}[\mathbf{g}_r]_i^2[\mathbf{g}_s]_j^2\sigma_{k,i}^2\sigma_{k,j}^2 \le CM\|\mathbf{g}_s\|^2\|\mathbf{g}_r\|^2 \le 2C\gamma_1^4M^3N^2$$

for large enough N. Similarly,

$$\mathbb{E}(\widetilde{B}_N^2) = \sum_{k=1}^{M}\sum_{i=1}^{N}[\mathbf{g}_r]_i^2[\mathbf{g}_s]_i^2\mathbb{E}(X_{k,i}^2 - \sigma_{k,i}^2)^2 \leq CM\|\mathbf{g}_r\|^2\|\mathbf{g}_s\|^2 \leq 2C\gamma_1^4 M^3 N^2.$$

This shows that (with a different C) we have $\mathbb{E}\left|\mathbf{g}_r^*(\mathbf{C}^*\mathbf{C} - \mathbb{E}(\mathbf{C}^*\mathbf{C}))\mathbf{g}_s\right|^2 \leq CN^5$ and hence $\frac{1}{N^3}\|\mathbf{R}_1 - \mathbb{E}\mathbf{R}_1\|_F \to 0$ in mean square and in probability. □

2.2.3. *Conclusion of proof of Theorem 1.1.* Theorem 1.1 is essentially a combination of (2.10), and convergence in probability from Lemma 2.4.

Proof of Theorem 1.1. We need to do a couple more approximations to the right hand side of (2.10). Indeed, we see that

$$\frac{N}{\lambda + \rho_r^{(N)}} = \sqrt{N/M}\frac{\sqrt{MN}}{\lambda + \rho_r^{(N)}} \to \frac{\sqrt{c}}{2\gamma_r}$$

$$\frac{2\sqrt{MN}}{\lambda + \rho_r^{(N)}} \to \frac{1}{\gamma_1}$$

$$\frac{M}{(\lambda + \rho_r^{(N)})\lambda^2} \sim \frac{\sqrt{M/N}}{2\gamma_r}\frac{1}{\gamma_r^2 MN} \sim \frac{\sqrt{c}}{2\gamma_r^3 MN}$$

To conclude the proof, we note that sequences $\{[\mathbf{\Sigma}_S^{(N)}]_{r,r}\}_N$, $\{[\mathbf{\Sigma}_R^{(N)}]_{r,r}/(MN)\}$ are bounded and $\{[\mathbf{Z}_0^{(N)}]_{r,r}\}$ is stochastically bounded by (2.28). □

Remark 2.1. Recall (1.15) and (1.16) from Remark 1.1. Examples in Section 3.3 have the additional property that

$$\frac{1}{MN}\widetilde{\mathbf{R}}_0^{(N)} \to \mathbf{Q}. \tag{2.34}$$

Under Assumption 1.1, the eigenvalues of $\mathbf{Q}$ are $\gamma_1^2 > \gamma_2^2 > \cdots > \gamma_K^2 > 0$. Denoting by $\mathbf{v}_r$ the corresponding orthonormal eigenvectors, we may assume that the first non-zero component of $\mathbf{v}_r$ is positive. After choosing the appropriate sign, without loss of generality we may assume that the same component of $\widetilde{\mathbf{u}}_r^{(N)}$ is non-negative for all N. Since by assumption eigenspaces of $\widetilde{\mathbf{R}}_0$ are one-dimensional for large enough N, we have

$$\widetilde{\mathbf{u}}_r^{(N)} \to \mathbf{v}_r \text{ as } N \to \infty.$$

We claim that in (1.17) and in (1.19) we can replace vectors $\widetilde{\mathbf{u}}_r$ by the corresponding eigenvectors $\mathbf{v}_r$ of $\mathbf{Q}$. Indeed, as in the proof of Theorem 1.1 the entries of the $K \times K$ matrices $\widetilde{\mathbf{\Sigma}}_R/(MN)$ and $\widetilde{\mathbf{\Sigma}}_S$ are bounded as $N \to \infty$. Also, we note that each entry $\xi = \widetilde{\mathbf{f}}_r^*\mathbf{C}\widetilde{\mathbf{g}}_s$ of matrix $\widetilde{\mathbf{Z}}_0$ is stochastically bounded due to uniform bound the second moment:

$$\mathbb{E}\xi^2 = \sum_{i=1}^{M}\sum_{j=1}^{N}[\mathbf{f}_r]_i^2\sigma_{i,j}^2[\mathbf{g}_s]_j^2 \leq \sup_{i,j}\sigma_{i,j}^2\|\mathbf{g}_r\|^2 \leq CMN.$$

(Compare (2.28).) This allows for the replacement of the $\widetilde{\mathbf{u}}_r$ with the $\mathbf{v}_r$.

3. Asymptotic normality of singular values

In this section we apply Theorem 1.1 to deduce asymptotic normality. To reduce technicalities involved, we begin with the simplest case of mean with rank 1. An example with mean of rank 2 is worked out in Section 3.2. A more involved application to population biology appears in Section 3.3. We use simulations to illustrate these results.

3.1. **Rank 1 perturbation.** The following is closely related to [8, Theorem 1.3] that was mentioned in the introduction.

Proposition 3.1. *Fix an infinite sequence $\{\mu_j\}$ such that the limit $\gamma^2 = \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^N \mu_j^2$ exists and is strictly positive. Consider the case $K = 1$, and assume that entries of random matrix $\mathbf{D} \in \mathcal{M}_{M\times N}$ are independent, with the same mean μ_j in the j-th column, the same variance σ^2, and uniformly bounded fourth moments. For the largest singular value λ of $\mathbf{D}$, we have $\lambda - \sqrt{M}\left(\sum_{j=1}^N \mu_j^2\right)^{1/2} \Rightarrow Z$ where Z is normal with mean $\frac{\sigma^2}{2\gamma}(\sqrt{c}+1/\sqrt{c})$ and variance σ^2. (Here $\Rightarrow$ denotes convergence in distribution.)*

Proof. In this setting $\mathbf{B} = \mathbf{f}\mathbf{g}^*$ with $\mathbf{f} = M^{-1/2}[1,\dots,1]^*$, $\mathbf{g} = \sqrt{M}[\mu_1,\dots,\mu_N]^*$. We get $\rho_1^2 = M\sum_{j=1}^N \mu_j^2$, $\gamma_1 = \gamma$, $\mathbf{\Sigma}_R = \sigma^2 M \sum_{j=1}^N \mu_j^2$, $\mathbf{\Sigma}_S = \sigma^2$, and

$$\mathbf{Z}_0 = \frac{1}{\sqrt{MN}}\sum_{i=1}^M\sum_{j=1}^N X_{i,j}\mu_j, \text{ so (1.9) gives } Z_1 = \frac{1}{\sqrt{MN}\gamma}\sum_{i=1}^M\sum_{j=1}^N X_{i,j}\mu_j.$$

Thus, the largest singular value of $\mathbf{D}$ can be written as

$$\lambda = \sqrt{M}\left(\sum_{j=1}^N \mu_j^2\right)^{1/2} + \frac{\sigma^2(\sqrt{c}+1/\sqrt{c})}{2\gamma} + \frac{1}{\sqrt{MN}\gamma}\sum_{i=1}^M\sum_{j=1}^N X_{i,j}\mu_j + \varepsilon^{(N)},$$

where $\varepsilon^{(N)} \to 0$ in probability. We have

$$\mathrm{Var}(Z_1) = \frac{\sigma^2}{\gamma^2}\frac{1}{N}\sum_{j=1}^N \mu_j^2 \to \sigma^2$$

and the sum of the fourth moments of the terms in Z_1 is

$$\frac{1}{M^2N^2\gamma^4}\sum_{i=1}^M\sum_{j=1}^N \mu_j^4 \mathbb{E}X_{i,j}^4 \le \frac{C}{M}\left(\frac{1}{N}\sum_{j=1}^N \mu_j^2\right)^2 \to 0.$$

So Z_1 is asymptotically normal by Lyapunov's theorem [1, Theorem 27.3]. □

3.2. **Block matrices.** Consider $(2M)\times(2N)$ block matrices

$$\mathbf{D} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix},$$

where $\mathbf{A}_1,\dots,\mathbf{A}_4$ are independent random $M\times N$ matrices. We assume that the entries of $\mathbf{A}_j$ are independent real identically distributed random variables with mean μ_j, variance σ_j^2 and with finite fourth moment. Then $\mathbf{B} = \mathbb{E}(\mathbf{D}) = \widetilde{\mathbf{f}}_1\widetilde{\mathbf{g}}_1^* + \widetilde{\mathbf{f}}_2\widetilde{\mathbf{g}}_2^*$ is of rank $K = 2$ with orthonormal

$$[\widetilde{\mathbf{f}}_1]_i = \begin{cases} 1/\sqrt{M} & \text{for } 1 \le i \le M \\ 0 & \text{for } M+1 \le i \le 2M \end{cases}$$

$$[\widetilde{\mathbf{f}}_2]_i = \begin{cases} 0 & \text{for } 1 \le i \le M \\ 1/\sqrt{M} & \text{for } M+1 \le i \le 2M \end{cases}$$

and with

$$[\widetilde{\mathbf{g}}_1]_j = \begin{cases} \sqrt{M}\mu_1 & \text{for } 1 \le j \le N \\ \sqrt{M}\mu_2 & \text{for } N+1 \le j \le 2N \end{cases}$$

$$[\widetilde{\mathbf{g}}_2]_j = \begin{cases} \sqrt{M}\mu_3 & \text{for } 1 \le j \le N \\ \sqrt{M}\mu_4 & \text{for } N+1 \le j \le 2N \end{cases}$$

So $\widetilde{\mathbf{G}} = \sqrt{M}\begin{bmatrix}\mu_1 & \mu_3\\ \mu_1 & \mu_3\\ \vdots & \\ \mu_1 & \mu_3\\ \mu_2 & \mu_4\\ \vdots & \\ \mu_2 & \mu_4\end{bmatrix}$ and $\widetilde{\mathbf{R}}_0 = \widetilde{\mathbf{G}}^*\widetilde{\mathbf{G}} = \begin{bmatrix}\|\widetilde{\mathbf{g}}_1\|^2 & \widetilde{\mathbf{g}}_1^*\widetilde{\mathbf{g}}_2\\ \widetilde{\mathbf{g}}_1^*\widetilde{\mathbf{g}}_2 & \|\widetilde{\mathbf{g}}_2\|^2\end{bmatrix} = MN\begin{bmatrix}\mu_1^2+\mu_2^2 & \mu_1\mu_3+\mu_2\mu_4\\ \mu_1\mu_3+\mu_2\mu_4 & \mu_3^2+\mu_4^2\end{bmatrix}$.

Denote by $\lambda_1 \geq \lambda_2$ the largest singular values of $\mathbf{D}$.

Proposition 3.2. *Suppose $\widetilde{\mathbf{g}}_1$ and $\widetilde{\mathbf{g}}_2$ are linearly independent and either $\widetilde{\mathbf{g}}_1^*\widetilde{\mathbf{g}}_2 \neq 0$, or if $\widetilde{\mathbf{g}}_1^*\widetilde{\mathbf{g}}_2 = 0$ then $\|\widetilde{\mathbf{g}}_1\| \neq \|\widetilde{\mathbf{g}}_2\|$. Then there exist constants $c_1 > c_2$ such that*

$$(\lambda_1 - c_1\sqrt{MN}, \lambda_2 - c_2\sqrt{MN}) \Rightarrow (Z_1, Z_2), \tag{3.1}$$

where (Z_1, Z_2) is a (noncentered) bivariate normal random variable.

Proof. To use Theorem 1.1, we first verify that Assumption 1.1 holds. We have

$$\widetilde{\mathbf{R}}_0 = 4MN\mathbf{Q}, \quad \text{where } \mathbf{Q} = \frac{1}{4}\begin{bmatrix}\mu_1^2+\mu_2^2 & \mu_1\mu_3+\mu_2\mu_4\\ \mu_1\mu_3+\mu_2\mu_4 & \mu_3^2+\mu_4^2\end{bmatrix}.$$

Noting that $\det(\mathbf{Q}) = (\mu_2\mu_3-\mu_1\mu_4)^2/16$, we see that $\gamma_1 \geq \gamma_2 > 0$ provided that $\det\begin{bmatrix}\mu_1 & \mu_2\\ \mu_3 & \mu_4\end{bmatrix} \neq 0$, i.e. provided that $\widetilde{\mathbf{g}}_1$ and $\widetilde{\mathbf{g}}_2$ are linearly independent. The eigenvalues of $\mathbf{Q}$ are

$$\gamma_1^2 = \frac{1}{8}\left(\mu_1^2+\mu_2^2+\mu_3^2+\mu_4^2+\sqrt{((\mu_2+\mu_3)^2+(\mu_1-\mu_4)^2)((\mu_2-\mu_3)^2+(\mu_1+\mu_4)^2)}\right),$$
$$\gamma_2^2 = \frac{1}{8}\left(\mu_1^2+\mu_2^2+\mu_3^2+\mu_4^2-\sqrt{((\mu_2+\mu_3)^2+(\mu_1-\mu_4)^2)((\mu_2-\mu_3)^2+(\mu_1+\mu_4)^2)}\right),$$

so condition $\gamma_1 > \gamma_2$ is satisfied except when $\mu_1 = \pm\mu_4$ and $\mu_2 = \mp\mu_3$, i.e. except when $\widetilde{\mathbf{g}}_1$ and $\widetilde{\mathbf{g}}_2$ are orthogonal and of the same length.

We see that $\rho_r^{(N)} = 2\gamma_r\sqrt{MN}$, which determines the constants $c_r = 2\gamma_r$ for (3.1). Next, we determine the remaining significant terms in (1.14). First, we check that the shifts $m_r^{(N)}$ in (1.14) do not depend on N. To do so we compute the matrices:

$$\widetilde{\boldsymbol{\Sigma}}_R = \frac{MN}{2}\begin{bmatrix}\mu_1^2(\sigma_1^2+\sigma_3^2)+\mu_2^2(\sigma_2^2+\sigma_4^2) & \mu_1\mu_3(\sigma_1^2+\sigma_3^2)+\mu_2\mu_4(\sigma_2^2+\sigma_4^2)\\ \mu_1\mu_3(\sigma_1^2+\sigma_3^2)+\mu_2\mu_4(\sigma_2^2+\sigma_4^2) & \mu_3^2(\sigma_1^2+\sigma_3^2)+\mu_4^2(\sigma_2^2+\sigma_4^2)\end{bmatrix}$$

and

$$\widetilde{\boldsymbol{\Sigma}}_S = \frac{1}{2}\begin{bmatrix}\sigma_1^2+\sigma_2^2 & 0\\ 0 & \sigma_3^2+\sigma_4^2\end{bmatrix}.$$

Indeed, we have

$$[\boldsymbol{\Delta}_R]_{j,j} = \begin{cases}(\sigma_1^2+\sigma_3^2)/2 & j \leq N\\ (\sigma_2^2+\sigma_4^2)/2 & N+1 \leq j \leq 2N\end{cases} \text{ and } [\boldsymbol{\Delta}_S]_{i,i} = \begin{cases}(\sigma_1^2+\sigma_2^2)/2 & i \leq M\\ (\sigma_3^2+\sigma_4^2)/2 & M+1 \leq i \leq 2M\end{cases}$$

To verify normality of the limit, we show that the matrix $\mathbf{Z}_0$ is asymptotically centered normal, so formula (1.9) gives a bivariate normal distribution in the limit. Denoting as previously by $X_{i,j}$ the entries of matrix $\mathbf{C} = \mathbf{D} - \mathbb{E}\mathbf{D}$, (1.18) gives

$$\widetilde{\mathbf{Z}}_0 = \frac{1}{2\sqrt{MN}}\begin{bmatrix}\mu_1\sum_{i=1}^{M}\sum_{j=1}^{N}X_{i,j}+\mu_2\sum_{i=1}^{M}\sum_{j=N+1}^{2N}X_{i,j} & \mu_1\sum_{i=M+1}^{2M}\sum_{j=1}^{N}X_{i,j}+\mu_2\sum_{i=M+1}^{2M}\sum_{j=N+1}^{2N}X_{i,j}\\ \mu_3\sum_{i=1}^{M}\sum_{j=1}^{N}X_{i,j}+\mu_4\sum_{i=1}^{M}\sum_{j=N+1}^{2N}X_{i,j} & \mu_3\sum_{i=M+1}^{2M}\sum_{j=1}^{N}X_{i,j}+\mu_4\sum_{i=M+1}^{2M}\sum_{j=N+1}^{2N}X_{i,j}\end{bmatrix}$$
$$\Rightarrow \tfrac{1}{2}\begin{bmatrix}\mu_1\sigma_1\zeta_1+\mu_2\sigma_2\zeta_2 & \mu_1\sigma_3\zeta_3+\mu_2\sigma_4\zeta_4\\ \mu_3\sigma_1\zeta_1+\mu_4\sigma_2\zeta_2 & \mu_3\sigma_3\zeta_3+\mu_4\sigma_4\zeta_4\end{bmatrix}$$

with independent $N(0,1)$ random variables $\zeta_1,\dots,\zeta_4$.

In particular, the limit $(Z_1,Z_2)=(m_1,m_2)+(Z_1^\circ,Z_2^\circ)$ is normal with mean given by (1.19), and centered bivariate normal random variable

$$Z_1^\circ=\tfrac{1}{2\gamma_1}\mathbf{v}_1^*\widetilde{\mathbf{Z}}_0\mathbf{v}_1,\quad Z_2^\circ=\tfrac{1}{2\gamma_2}\mathbf{v}_2^*\widetilde{\mathbf{Z}}_0\mathbf{v}_2.$$

□

3.2.1. *Numerical example and simulations.* For a numerical example, suppose $\sigma_j^2=\sigma^2$, $\mu_j=\mu$, except $\mu_1=0$. Then $\gamma_1^2=\frac{\mu^2}{8}\left(3+\sqrt{5}\right),\gamma_2^2=\frac{\mu^2}{8}\left(3-\sqrt{5}\right)$ and

$$\mathbf{v}_1=\frac{1}{\sqrt{10}}\begin{bmatrix}\sqrt{5-\sqrt{5}}\\ \sqrt{5+\sqrt{5}}\end{bmatrix}\approx\begin{bmatrix}0.525731\\0.850651\end{bmatrix}$$

$$\mathbf{v}_2=\frac{1}{\sqrt{10}}\begin{bmatrix}\sqrt{5+\sqrt{5}}\\ -\sqrt{5-\sqrt{5}}\end{bmatrix}\approx\begin{bmatrix}-0.850651\\0.525731\end{bmatrix}$$

$$\widetilde{\boldsymbol{\Sigma}}_S=\sigma^2\mathbf{I}_2,\quad \widetilde{\boldsymbol{\Sigma}}_R=MN\mu^2\sigma^2\begin{bmatrix}1&1\\1&2\end{bmatrix}$$

so with $c=M/N$, formula (1.19) gives

$$m_1=\frac{(\sqrt{5}-1)\sigma^2(M+N)}{2\mu\sqrt{MN}},\quad m_2=\frac{(\sqrt{5}+1)\sigma^2(M+N)}{2\mu\sqrt{MN}}.$$

We get

$$Z_1^\circ=\frac{\left(2\sqrt{5}\zeta_1+\left(5+\sqrt{5}\right)\zeta_2+\left(5+\sqrt{5}\right)\zeta_3+\left(5+3\sqrt{5}\right)\zeta_4\right)}{5\left(1+\sqrt{5}\right)}\sigma,$$

$$Z_2^\circ=\frac{\left(-\left(5+\sqrt{5}\right)\zeta_1+2\sqrt{5}\zeta_2+2\sqrt{5}\zeta_3+\left(\sqrt{5}-5\right)\zeta_4\right)}{10}\sigma.$$

Thus $\lambda_1-\mu\sqrt{\frac12 MN(3+\sqrt{5})},\lambda_2-\mu\sqrt{\frac12 MN(3-\sqrt{5})}$ is approximately normal with mean (m_1,m_2) and covariance matrix $\sigma^2\mathbf{I}_2$. In particular, if the entries of matrices are independent uninform $U(-1,1)$ for block $\mathbf{A}_1$ and $U(0,2)$ for blocks $\mathbf{A}_2,\mathbf{A}_3,\mathbf{A}_4$, then $\sigma^2=1/3,\mu=1$. So with $M=20,N=50$ we get

$$\lambda_1\approx 51.6228+\frac{1}{\sqrt{3}}\zeta_1,\ \lambda_2\approx 20.7378+\frac{1}{\sqrt{3}}\zeta_2$$

with (new) independent normal N(0,1) random variables ζ_1,ζ_2. Figure 1 show the result of simulations for two sets of choices of M,N.

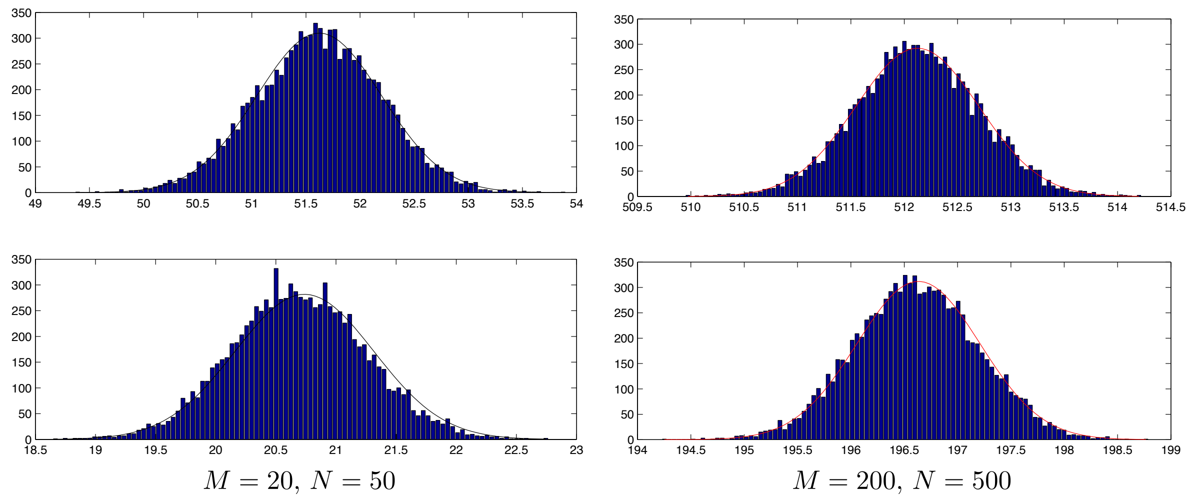

FIGURE 1. Histograms of simulations of 10,000 realizations, overlayed with normal density of variance 1/3. Top row: Largest singular value; Second row: second singular value. For small N, additional poorly controlled error arises from $\varepsilon^{(N)} \to 0$ in probability.

3.3. **Application to a model in population genetics.** Following [2] (see also [7]), we consider an $M \times N$ array $\mathbf{D}$ of genetic markers with rows labeled by individuals and columns labeled by polymorphic markers. The entries $[\mathbf{D}]_{i,j}$ are the number of alleles for marker j, individual i, are assumed independent, and take values $0, 1, 2$ with probabilities $(1-p)^2, 2p(1-p), p^2$ respectively, where p is the frequency of the j-th allele. We assume that we have data for M individuals from K subpopulation and that we have M_r individuals from the subpopulation labeled r. For our asymptotic analysis where $N \to \infty$ we assume (1.4) and that each subpopulation is sufficiently represented in the data so that

$$M_r/N \to c_r > 0,$$

where of course $c_1 + \cdots + c_K = c$. (Note that our notation for c_r is slightly different than the notation in [2].) We assume that allelic frequency for the j-th marker depends on the subpopulation of which the individual is a member but does not depend on the individual otherwise. Thus with the r-th subpopulation we associate the vector $\mathbf{p}_r \in (0,1)^N$ of allelic probabilities, where $p_r(j) := [\mathbf{p}_r]_j$ is the value of p for the j-th marker, $j = 1, 2, \ldots, N$.

We further assume that the allelic frequencies are fixed, but arise from some regular mechanism, which guarantees that for $d = 1, 2, 3, 4$ the following limits exist

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} p_{r_1}(j) p_{r_2}(j) \ldots p_{r_d}(j) = \pi_{r_1, r_2, \ldots, r_d},\ 1 \le r_1 \le \cdots \le r_d \le K.$$

This holds if the allelic probabilities $p_r(j)$ for the r-th population arise in ergodic fashion from joint allelic spectrum $\varphi(x_1, \ldots, x_K)$ [4] with

$$\pi_{r_1, r_2, \ldots, r_d} = \int_{[0,1]^K} x_{r_1} \ldots x_{r_d} \varphi(x_1, \ldots, x_K) dx_1 \ldots dx_K. \tag{3.2}$$

Under the above assumptions, the entries of $\mathbf{D}$ are independent Binomial random variables with the same number of trials 2, but with varying probabilities of success. Using the assumed distribution of the entries of $\mathbf{D}$ we have $\mathbf{B} = \mathbb{E}\mathbf{D} = 2\sum_{r=1}^{K} \widetilde{\mathbf{e}}_r \mathbf{p}_r^*$, where $\widetilde{\mathbf{e}}_r$ is the vector indicating the locations of the members of the r-th subpopulation, i.e. $[\widetilde{\mathbf{e}}_r]_i = 1$ when the i-th individual is a member of the r-th subpopulation. Assuming the entries of $\mathbf{D}$ are independent, we get $\mathbf{B} = \sum_{r=1}^{K} \widetilde{\mathbf{f}}_r \widetilde{\mathbf{g}}_r^* = \widetilde{\mathbf{F}}\widetilde{\mathbf{G}}^*$ with orthonormal vectors $\widetilde{\mathbf{f}}_r = \widetilde{\mathbf{e}}_r/\sqrt{M_r}$ and with $\widetilde{\mathbf{g}}_s = 2\sqrt{M_s}\mathbf{p}_s$, so we have (1.15). In this setting, Remark 2.1 applies. In (1.16), we have $[\widetilde{\mathbf{R}}_0]_{r,s} = 4\sqrt{M_r M_s}\mathbf{p}_r^*\mathbf{p}_s$ and $\widetilde{\mathbf{R}}_0/(MN) \to \mathbf{Q}$, where

$$[\mathbf{Q}]_{r,s} := 4\frac{\sqrt{c_r c_s}}{c}\pi_{r,s}, \tag{3.3}$$

so the eigenvalues of $\widetilde{\mathbf{R}}_0$ are $\rho_r^2 \sim \gamma_r^2 MN + o(N^2)$. As previously, we assume that the that $\mathbf{Q}$ has positive and distinct eigenvalues $\gamma_1^2 > \gamma_2^2 > \cdots > \gamma_K^2 > 0$ with corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^K$. (Due to change of notation, matrix (3.3) differs from [2, (2.6)] by a factor of 4.)

To state the result, for $1 \le t \le K$ we introduce matrices $\boldsymbol{\Sigma}_t \in \mathcal{M}_{K\times K}$ with entries

$$[\boldsymbol{\Sigma}_t]_{r,s} = \frac{\sqrt{c_r c_s}}{c}(\pi_{r,s,t} - \pi_{r,s,t,t}). \tag{3.4}$$

Proposition 3.3. *The K largest singular values of $\mathbf{D}$ are approximately normal,*

$$\begin{bmatrix} \lambda_1^{(N)} - \rho_1^{(N)} \\ \lambda_2^{(N)} - \rho_2^{(N)} \\ \vdots \\ \lambda_K^{(N)} - \rho_K^{(N)} \end{bmatrix} \Rightarrow \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_K \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \vdots \\ \zeta_K \end{bmatrix}$$

where

$$m_r = \frac{1}{\sqrt{c}\gamma_r}\sum_{t=1}^K [\mathbf{v}_r]_t^2(\pi_t - \pi_{t,t}) + \frac{4\sqrt{c}}{\gamma_r^3}\sum_{t=1}^K \frac{c_t}{c}\mathbf{v}_r^*\boldsymbol{\Sigma}_t\mathbf{v}_r \tag{3.5}$$

and $(\zeta_1, \dots, \zeta_K)$ is centered multivariate normal with the covariance

$$\mathbb{E}(\zeta_r\zeta_s) = \frac{8}{\gamma_r\gamma_s}\sum_{t=1}^K [\mathbf{v}_r]_t[\mathbf{v}_s]_t\mathbf{v}_r^*\boldsymbol{\Sigma}_t\mathbf{v}_s.$$

Proof. We apply Theorem 1.1 in the form stated as Remarks 1.1 and 2.1. The first step is to note that due to the form of vectors $\widetilde{\mathbf{f}}_k$, equation (1.18) gives a matrix $\widetilde{\mathbf{Z}}_0$ with independent columns. Our first task is to show that $\widetilde{\mathbf{Z}}_0$ is asymptotically normal by verifying that each of its independent columns is asymptotically normal.

Denote by $\mathcal{N}_k$ the index set for the k-th subpopulation (i.e., $[\widetilde{\mathbf{e}}_k]_i = 1$ if $i \in \mathcal{N}_k$). In this notation, the k-th column of $\widetilde{\mathbf{Z}}_0$ is

$$\frac{2}{\sqrt{MNM_k}}\begin{bmatrix} \sqrt{M_1}\sum_{i\in\mathcal{N}_k}\sum_{j=1}^N X_{i,j}p_1(j) \\ \sqrt{M_2}\sum_{i\in\mathcal{N}_k}\sum_{j=1}^N X_{i,j}p_2(j) \\ \vdots \\ \sqrt{M_K}\sum_{i\in\mathcal{N}_k}\sum_{j=1}^N X_{i,j}p_K(j) \end{bmatrix}.$$

To verify asymptotic normality and find the covariance, we fix $\mathbf{t} = [t_1, \dots, t_K]^*$. Then the dot product of $\mathbf{t}$ with the k-th column of $\widetilde{\mathbf{Z}}_0$ is

$$S_N = \sum_{i\in\mathcal{N}_k}\sum_{j=1}^N a_j(N)X_{i,j}$$

with

$$a_j(N) = \frac{2}{\sqrt{MNM_k}}\sum_{r=1}^K \sqrt{M_r}t_r p_r(j).$$

We first note that by independence

$$\begin{aligned}\operatorname{Var}(S_N) &= \sum_{i\in\mathcal{N}_k}\sum_{j=1}^N a_j^2(N)\mathbb{E}X_{i,j}^2 \\ &= 8\sum_{r_1,r_2=1}^K t_{r_1}t_{r_2}\frac{\sqrt{M_{r_1}M_{r_2}}}{M}\frac{1}{N}\sum_{j=1}^N p_{r_1}(j)p_{r_2}(j)p_k(j)(1-p_k(j)) \\ &\to 8\sum_{r_1,r_2=1}^K \frac{\sqrt{c_{r_1}c_{r_2}}}{c}(\pi_{r_1,r_2,k} - \pi_{r_1,r_2,k,k})t_{r_1}t_{r_2} = 8\mathbf{t}^*\boldsymbol{\Sigma}_k\mathbf{t}\end{aligned}$$

giving the covariance matrix for the k-column as 8 times (3.4).

Next we note that since $E(X_{i,j}^4) = 2p_k(j)(1-p_k(j))$, we have

$$\sum_{i\in\mathcal{N}_k}\sum_{j=1}^N a_j^4(N)\mathbb{E}X_{i,j}^4$$
$$= \frac{32}{NM_k}\sum_{r_1,r_2,r_3,r_4=1}^K \frac{\sqrt{M_{r_1}M_{r_2}M_{r_3}M_{r_4}}}{M^2} t_{r_1}t_{r_2}t_{r_3}t_{r_4}\frac{1}{N}\sum_{j=1}^N p_{r_1}(j)p_{r_2}(j)p_{r_3}(j)p_{r_4}(j)p_k(j)(1-p_k(j)) = O(1/N^2)\to 0.$$

By Lyapunov's theorem S_N is asymptotically normal. Thus the k-th column of $\widetilde{\mathbf{Z}}_0$ is asymptotically normal with covariance 8 times (3.4). Let $\mathbf{Z}_0^{(\infty)}$ denote the distributional limit of $\widetilde{\mathbf{Z}}_0$.

From (1.17) with $\widetilde{\mathbf{u}}_r$ replaced by $\mathbf{v}_r$ as in Remark 2.1, we see that $(Z_1^{(N)},\dots,Z_K^{(N)})$ converges in distribution to the multivariate normal r.v. $(\zeta_1,\dots,\zeta_K)$ with covariance

$$\mathbb{E}(\zeta_r\zeta_s) = \frac{1}{\gamma_r\gamma_s}\mathbb{E}\left(\mathbf{v}_r^*\mathbf{Z}_0^{(\infty)}\mathbf{v}_r\mathbf{v}_s^*\mathbf{Z}_0^{(\infty)}\mathbf{v}_s\right) = \frac{8}{\gamma_r\gamma_s}\sum_{t=1}^K[\mathbf{v}_r]_t[\mathbf{v}_s]_t\mathbf{v}_r^*\mathbf{\Sigma}_t\mathbf{v}_s.$$

Next, we use formula (1.19) to compute the shift. We first compute $\widetilde{\mathbf{\Sigma}}_S = \mathbb{E}(\widetilde{\mathbf{F}}^*\mathbf{C}\mathbf{C}^*\widetilde{\mathbf{F}})/N$. As already noted, $\mathbf{C}^*\widetilde{\mathbf{F}}\in\mathcal{M}_{N\times K}$ has K independent columns, with the k-th column

$$\frac{1}{\sqrt{M_k}}\begin{bmatrix}\sum_{i\in\mathcal{N}_k}X_{i,1}\\ \sum_{i\in\mathcal{N}_k}X_{i,2}\\ \vdots\\ \sum_{i\in\mathcal{N}_k}X_{i,N}\end{bmatrix}.$$

So $\widetilde{\mathbf{\Sigma}}_S$ is a diagonal matrix with

$$[\widetilde{\mathbf{\Sigma}}_S]_{rr} = \frac{2}{N}\sum_{j=1}^N p_r(j)(1-p_r(j)) \to 2(\pi_r-\pi_{rr}).$$

Next, we compute the limit of

$$\frac{c}{MN}\widetilde{\mathbf{\Sigma}}_R = \frac{c}{M^2N}\mathbb{E}(\widetilde{\mathbf{G}}^*\mathbf{C}^*\mathbf{C}\widetilde{\mathbf{G}}) \sim \frac{1}{MN^2}\mathbb{E}(\widetilde{\mathbf{G}}^*\mathbf{C}^*\mathbf{C}\widetilde{\mathbf{G}}).$$

Since

$$[\mathbb{E}(\widetilde{\mathbf{G}}^*\mathbf{C}^*\mathbf{C}\widetilde{\mathbf{G}})]_{rs} = 2\sum_{t=1}^K M_t\sum_{j=1}^N[\mathbf{g}_s]_j[\mathbf{g}_r]_jp_t(j)(1-p_t(j)) = 8\sqrt{M_rM_s}\sum_{t=1}^K M_t\sum_{j=1}^N p_s(j)p_r(j)p_t(j)(1-p_t(j))$$

we see that

$$\left[\frac{c}{MN}\widetilde{\mathbf{\Sigma}}_R\right]_{rs} \to \frac{8\sqrt{c_rc_s}}{c}\sum_{t=1}^K c_t(\pi_{r,s,t}-\pi_{r,s,t,t}).$$

This shows that

$$\frac{c}{MN}\widetilde{\mathbf{\Sigma}}_R \to 8\sum_{t=1}^K c_t\mathbf{\Sigma}_t.$$

From (1.19) we calculate

$$m_r = \sum_{s=1}^K\left(\frac{1}{\sqrt{c}\gamma_r}[\mathbf{v}_r]_s^2(\pi_s-\pi_{s,s}) + \frac{4c_s}{c^{3/2}\gamma_r^3}\sum_{t_1,t_2=1}^K[\mathbf{v}_r]_{t_1}[\mathbf{v}_r]_{t_2}\sqrt{c_{t_1}c_{t_2}}(\pi_{t_1,t_2,s}-\pi_{t_1,t_2,s,s})\right)$$

which is (3.5). □

Ref. [2] worked with the eigenvalues of the "sample covariance matrix" $(\mathbf{D}\mathbf{D}^*)/(\sqrt{M}+\sqrt{N})^2$, i.e., with the normalized squares of singular values $\Lambda_r = \lambda_r^2/(\sqrt{M}+\sqrt{N})^2$. Proposition 3.3 then gives the following normal approximation.

Proposition 3.4.

$$\begin{bmatrix} \Lambda_1 - \frac{(\rho_1^{(N)})^2}{(\sqrt{M}+\sqrt{N})^2} \\ \Lambda_2 - \frac{(\rho_2^{(N)})^2}{(\sqrt{M}+\sqrt{N})^2} \\ \vdots \\ \Lambda_K - \frac{(\rho_K^{(N)})^2}{(\sqrt{M}+\sqrt{N})^2} \end{bmatrix} \Rightarrow \begin{bmatrix} \tilde{m}_1 \\ \tilde{m}_2 \\ \vdots \\ \tilde{m}_K \end{bmatrix} + \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_K \end{bmatrix}$$

with recalculated shift

$$\tilde{m}_r = \frac{2}{(1+c^{1/2})^2} \sum_{t=1}^K [\mathbf{v}_r]_t^2 (\pi_t - \pi_{t,t}) + \frac{8c}{\gamma_r^2 (1+c^{1/2})^2} \sum_{t=1}^K \frac{c_t}{c} \mathbf{v}_r^* \mathbf{\Sigma}_t \mathbf{v}_r$$

and with recalculated centered multivariate normal random vector $(Z_1, \dots, Z_K)$ *with covariance*

$$\mathbb{E}(Z_r Z_s) = \frac{32c}{(1+\sqrt{c})^4} \sum_{t=1}^K [\mathbf{v}_r]_t [\mathbf{v}_s]_t \mathbf{v}_r^* \mathbf{\Sigma}_t \mathbf{v}_s.$$

Proof.

$$\Lambda_r - \frac{\rho_r^2}{(\sqrt{M}+\sqrt{N})^2} = (\lambda_r - \rho_r) \times \frac{\lambda_r + \rho_r}{\sqrt{MN}} \times \frac{\sqrt{MN}}{(\sqrt{M}+\sqrt{N})^2}.$$

Since

$$\frac{\lambda_r + \rho_r}{\sqrt{MN}} \to 2\gamma_r \text{ in probability, and } \frac{\sqrt{MN}}{(\sqrt{M}+\sqrt{N})^2} \to \frac{\sqrt{c}}{(1+\sqrt{c})^2},$$

see Lemma 2.4, the result follows. □

3.3.1. *Numerical illustration.* As an illustration of Theorem 3.3, we re-analyze the example from [2, Section 3.1]. In that example, the subpopulation sample sizes were drawn with proportions $c_1 = c/6$, $c_2 = c/3$, $c_3 = c/2$ where $c = M/N$ varied from case to case. The theoretical population proportions $p_r(j)$ at each location for each subpopulation were selected from the same uniform site frequency spectrum $\psi(x) = \sqrt{x}/2$. Following [2, Section 3.1], for our simulations we selected $p_1(j), p_2(j), p_3(j)$ independently at each location j, which corresponds to joint allelic spectrum $\psi(x,y,z) = \psi(x)\psi(y)\psi(z) = \sqrt{xyz}/8$ in (3.2).

In this setting, we can explicitly compute the theoretical matrix of moments (3.2) and matrix $\mathbf{Q}$ defined by (2.34):

$$[\pi_{r,s}] = \begin{bmatrix} 1/5 & 1/9 & 1/9 \\ 1/9 & 1/5 & 1/9 \\ 1/9 & 1/9 & 1/5 \end{bmatrix}, \ \mathbf{Q} = \frac{4}{c}[\sqrt{c_r c_s}\,\pi_{r,s}] = \begin{bmatrix} \frac{2}{15} & \frac{2\sqrt{2}}{27} & \frac{2}{9\sqrt{3}} \\ \frac{2\sqrt{2}}{27} & \frac{4}{15} & \frac{2\sqrt{2}}{9\sqrt{3}} \\ \frac{2}{9\sqrt{3}} & \frac{2\sqrt{2}}{9\sqrt{3}} & \frac{2}{5} \end{bmatrix} = \begin{bmatrix} 0.133333 & 0.104757 & 0.1283 \\ 0.104757 & 0.266667 & 0.181444 \\ 0.1283 & 0.181444 & 0.4 \end{bmatrix}$$

(Due to change of notation, this matrix and the eigenvalues are 4 times the corresponding values from [2, page 37].) The eigenvalues of the above matrix $\mathbf{Q}$ are

$$[\gamma_1^2, \gamma_2^2, \gamma_3^2] = [0.586836, 0.141985, 0.0711794]$$

and the corresponding eigenvectors are

$$\mathbf{v}_1 = \begin{bmatrix} 0.342425 \\ 0.545539 \\ 0.764939 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} -0.154523 \\ -0.770372 \\ 0.618586 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} -0.926751 \\ 0.33002 \\ 0.179496 \end{bmatrix}.$$

In order to apply the formulas, we use (3.2) to compute

$$\pi_{r,s,t} = \begin{cases} 1/27 & r \neq s \neq t \\ 1/15 & \text{if one pair of indexes is repeated} \\ 1/7 & r = s = t \end{cases}$$

and

$$\pi_{r,s,t,t} = \begin{cases} 1/45 & r \neq s \neq t \\ 1/25 & r = s \neq t \\ 1/21 & r \neq s = t \\ 1/9 & r = s = t \end{cases}$$

As an intermediate step towards (3.4), it is convenient to collect the above data into three auxiliary matrices

$$[\pi_{rs1} - \pi_{rs11}]_{r,s} = \begin{bmatrix} \frac{2}{63} & \frac{2}{105} & \frac{2}{105} \\ \frac{2}{105} & \frac{2}{75} & \frac{2}{135} \\ \frac{2}{105} & \frac{2}{135} & \frac{2}{75} \end{bmatrix}$$

$$[\pi_{rs2} - \pi_{rs22}]_{r,s} = \begin{bmatrix} \frac{2}{75} & \frac{2}{105} & \frac{2}{135} \\ \frac{2}{105} & \frac{2}{63} & \frac{2}{105} \\ \frac{2}{135} & \frac{2}{105} & \frac{2}{75} \end{bmatrix}$$

$$[\pi_{rs3} - \pi_{rs33}]_{r,s} = \begin{bmatrix} \frac{2}{75} & \frac{2}{135} & \frac{2}{105} \\ \frac{2}{135} & \frac{2}{75} & \frac{2}{105} \\ \frac{2}{105} & \frac{2}{105} & \frac{2}{63} \end{bmatrix}$$

From (3.4) we get

$$\mathbf{\Sigma}_1 = \begin{bmatrix} \frac{1}{189} & \frac{\sqrt{2}}{315} & \frac{1}{105\sqrt{3}} \\ \frac{\sqrt{2}}{315} & \frac{2}{225} & \frac{\sqrt{2}}{135\sqrt{3}} \\ \frac{1}{105\sqrt{3}} & \frac{\sqrt{2}}{135\sqrt{3}} & \frac{1}{75} \end{bmatrix} = \begin{bmatrix} 0.00529101 & 0.00448957 & 0.00549857 \\ 0.00448957 & 0.00888889 & 0.00604812 \\ 0.00549857 & 0.00604812 & 0.0133333 \end{bmatrix}$$

$$\mathbf{\Sigma}_2 = \begin{bmatrix} \frac{1}{189} & \frac{\sqrt{2}}{315} & \frac{1}{105\sqrt{3}} \\ \frac{\sqrt{2}}{315} & \frac{2}{225} & \frac{\sqrt{2}}{135\sqrt{3}} \\ \frac{1}{105\sqrt{3}} & \frac{\sqrt{2}}{135\sqrt{3}} & \frac{1}{75} \end{bmatrix} = \begin{bmatrix} 0.00444444 & 0.00448957 & 0.00427667 \\ 0.00448957 & 0.010582 & 0.00777616 \\ 0.00427667 & 0.00777616 & 0.0133333 \end{bmatrix}$$

$$\mathbf{\Sigma}_3 = \begin{bmatrix} \frac{1}{225} & \frac{\sqrt{2}}{405} & \frac{1}{105\sqrt{3}} \\ \frac{\sqrt{2}}{405} & \frac{2}{225} & \frac{\sqrt{2}}{105\sqrt{3}} \\ \frac{1}{105\sqrt{3}} & \frac{\sqrt{2}}{105\sqrt{3}} & \frac{1}{63} \end{bmatrix} = \begin{bmatrix} 0.00444444 & 0.00349189 & 0.00549857 \\ 0.00349189 & 0.00888889 & 0.00777616 \\ 0.00549857 & 0.00777616 & 0.015873 \end{bmatrix}$$

Using these expressions and (3.5) with $c = M/N = 120/2500$, we determine

$$[m_1, m_2, m_3] = [0.183948\sqrt{c} + \frac{0.174053}{\sqrt{c}}, 0.323598\sqrt{c} + \frac{0.353849}{\sqrt{c}}, 0.47449\sqrt{c} + \frac{0.49976}{\sqrt{c}}] = [0.834739, 1.68599, 2.38504].$$

We note that the shift is stronger for smaller singular values and is more pronounced for rectangular matrices with small (or large) c.

Finally, we compute the covariance matrix

$$[\mathbb{E}\zeta_r\zeta_s] = \begin{bmatrix} 0.306317 & 0.0293619 & 0.0225604 \\ 0.0293619 & 0.233577 & -0.00941692 \\ 0.0225604 & -0.00941692 & 0.235559 \end{bmatrix}.$$

The following figure illustrates that normal approximation works well.

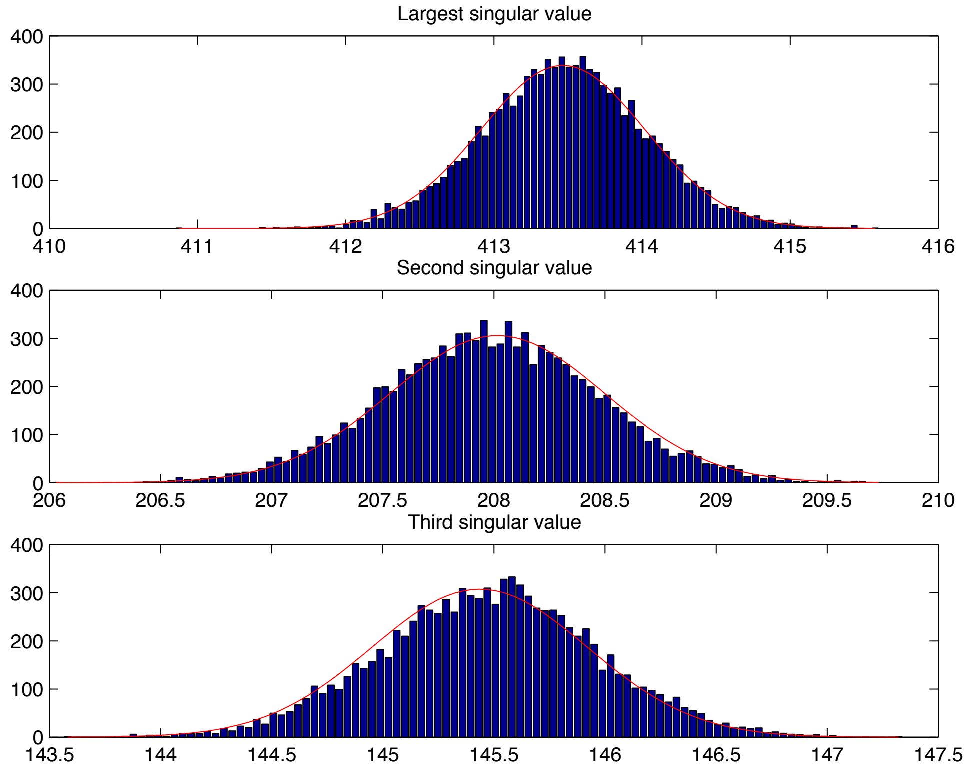

FIGURE 2. Histograms of three largest singular values with normal curves centered at $[\rho_1 + m_1, \rho_2 + m_2, \rho_3 + m_3] = [413.47, 208.02, 145.44]$ with variances $0.3063, 0.2336, 0.2356$. Based on 10,000 runs of simulations after a single randomization to choose vectors $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ with $M = 120, N = 2500$.

Based on simulations reported in Fig. 2, we conclude that normal approximation which uses the singular values of $\widetilde{\mathbf{R}}_0$ provides a reasonable fit.

Let us now turn to the question of asymptotic normality for the normalized squares of singular values $\Lambda_r = \lambda_r^2/(\sqrt{M} + \sqrt{N})^2$. With $M = 120, N = 2500$ (correcting the misreported values) Ref. [2] reported the observed values $(48.2, 11.5, 5.8)$ for $\Lambda_1, \Lambda_2, \Lambda_3$ versus the "theoretical estimates" $(47.4, 11.5, 5.7)$ calculated as $\gamma_r^2 MN/(\sqrt{M} + \sqrt{N})^2$ in our notation. While the numerical differences are small, Proposition 3.4 restricts the accuracy of such estimates due to their dependence on the eigenvalues of $\widetilde{\mathbf{R}}_0$ constructed from vectors of allelic probabilities $\mathbf{p}_r$. Figure 3 illustrates that different selections of such vectors from the same joint allelic spectrum (3.2) may yield quite different ranges for $\Lambda_1, \Lambda_2, \Lambda_3$. It is perhaps worth pointing out that different choices of allelic probabilities affect only the centering; the variance of normal approximation depends only on the allelic spectrum.

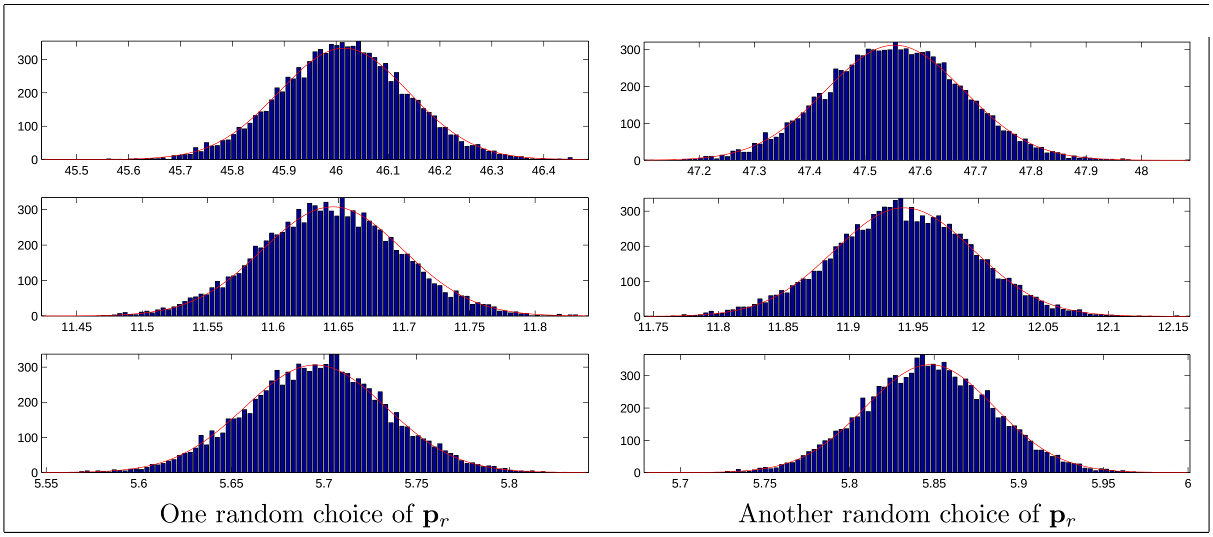

FIGURE 3. Two histograms of normalized squared singular values $(\Lambda_1, \Lambda_2, \Lambda_3)$, based on 10000 simulations, and the theoretical normal curves from Proposition 3.4 drawn in red. This is $M = 120$ individuals with $N = 2500$ markers. Although the numerical differences between $\Lambda_1, \Lambda_2, \Lambda_3$ on the left-hand-side and on the right-hand side are small, the histograms have practically disjoint supports.

Acknowledgement. We thank an anonymous referee for suggestions that lead to a stronger form of Theorem 1.1.

References

[1] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2012.

[2] K. Bryc, W. Bryc, and J. W. Silverstein. Separation of the largest eigenvalues in eigenanalysis of genotype data from discrete subpopulations. *Theoretical Population Biology*, 89:34–43, 2013. arXiv:1301.451.

[3] Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[4] M. Kimura. Diffusion models in population genetics. *Journal of Applied Probability*, 1(2):177–232, 1964.

[5] Donald W. Lang. Isolated eigenvalue of a random matrix. *Phys. Rev.*, 135:B1082–B1084, Aug 1964.

[6] Rafał Latała. Some estimates of norms of random matrices. *Proceedings of the American Mathematical Society*, 133(5):1273–1282, 2005.

[7] N. Patterson, A. L. Price, and D. Reich. Population structure and eigenanalysis. *PLoS Genet*, 2(12):e190, 2006.

[8] Jack W. Silverstein. The spectral radii and norms of large-dimensional non-central random matrices. *Comm. Statist. Stochastic Models*, 10(3):525–532, 1994.

[9] Gilbert W Stewart and Ji-Guang Sun. *Matrix Perturbation Theory (Computer Science and Scientific Computing)*. Academic Press Boston, 1990.

[10] Terence Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields*, 155(1-2):231–263, 2013.

[11] YQ Yin, ZD Bai, and PR Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, 78(4):509–521, 1988.

WŁODEK BRYC, DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF CINCINNATI, PO BOX 210025, CINCINNATI, OH 45221–0025, USA

Email address: wlodzimierz.bryc@uc.edu

JACK W. SILVERSTEIN, DEPARTMENT OF MATHEMATICS, BOX 8205, NORTH CAROLINA STATE UNIVERSITY, RALEIGH, NC 27695-8205, USA

Email address: jack@ncsu.edu