Statistica Sinica **17**(2007), 1617-1642

# ASYMPTOTICS OF SAMPLE EIGENSTRUCTURE FOR A LARGE DIMENSIONAL SPIKED COVARIANCE MODEL

Debashis Paul

*University of California, Davis*

*Abstract:* This paper deals with a multivariate Gaussian observation model where the eigenvalues of the covariance matrix are all one, except for a finite number which are larger. Of interest is the asymptotic behavior of the eigenvalues of the sample covariance matrix when the sample size and the dimension of the observations both grow to infinity so that their ratio converges to a positive constant. When a population eigenvalue is above a certain threshold and of multiplicity one, the corresponding sample eigenvalue has a Gaussian limiting distribution. There is a "phase transition" of the sample eigenvectors in the same setting. Another contribution here is a study of the second order asymptotics of sample eigenvectors when corresponding eigenvalues are simple and sufficiently large.

*Key words and phrases:* Eigenvalue distribution, principal component analysis, random matrix theory.

## 1. Introduction

The study of eigenvalues and eigenvectors of sample covariance matrices has a long history. When the dimension $N$ is fixed, distributional aspects for both Gaussian and non-Gaussian observations have been dealt with at length by various authors. Anderson (1963), Muirhead (1982) and Tyler (1983) are among standard references. With dimension fixed, much of the study of the eigenstructure of sample covariance matrix is based on the fact that sample covariance approximates population covariance matrix well when sample size is large. However this is no longer the case when $N/n \to \gamma \in (0, \infty)$ as $n \to \infty$, where $n$ is the sample size. Under these circumstances it is known (see Bai (1999) for a review) that, if the true covariance is the identity matrix, then the Empirical Spectral Distribution (ESD) converges almost surely to the Marčenko-Pastur distribution, henceforth denoted by $F_\gamma$. When $\gamma \leq 1$, the support $F_\gamma$ is the set $[(1-\sqrt{\gamma})^2, (1+\sqrt{\gamma})^2]$, and when $\gamma > 1$ an isolated point zero is added to the support. It is known (Bai and Yin (1993)) that when the population covariance is the identity, the largest and the smallest eigenvalues, when $\gamma \leq 1$, converge *almost surely* to the respective boundaries of the support of $F_\gamma$. Johnstone (2001) derived the asymptotic distribution for the largest sample eigenvalue under the setting of an identity covariance under Gaussianity. Soshnikov (2002) proved the distributional limits under weaker assumptions, in addition to deriving distributional limits of the $k$th largest eigenvalue, for fixed but arbitrary $k$.

However, in recent years researchers in various fields have been using different versions of non-identity covariance matrices of growing dimension. Among these, a particularly interesting model has most of the eigenvalues one, and the few that are not are well-separated from the rest. This has been deemed the "spiked population model" by Johnstone (2001). It has also been observed that for certain types of data, e.g., in speech recognition (Buja, Hastie and Tibshirani (1995)), wireless communication (Telatar (1999)), statistical learning (Hoyle and Rattray (2003, 2004)), a few of the sample eigenvalues have limiting behavior that is different from the behavior when the covariance is the identity. The results of this paper lend understanding to these phenomena.

The literature on the asymptotics of sample eigenvalues when the covariance is not the identity is relatively recent. Silverstein and Choi (1995) derived the *almost sure* limit of the ESD under fairly general conditions. Bai and Silverstein (2004) derived the asymptotic distribution of certain linear spectral statistics. However, a systematic study of the individual eigenvalues has been conducted only recently by Péché (2003) and Baik, Ben Arous and Péché (2005). These authors deal with the situation where the observations are complex Gaussian and the covariance matrix is a finite rank perturbation of identity. Baik and Silverstein (2006) study the almost sure limits of sample eigenvalues when the observations are either real or complex, and under fairly weak distributional assumptions. They give almost sure limits of the $M$ largest and $M$ smallest (non-zero) sample eigenvalues, where $M$ is the number of non-unit population eigenvalues.

A crucial aspect of the work of the last three sets of authors is the discovery of a phase transition phenomenon. Simply put, if the non-unit eigenvalues are close to one, then their sample versions will behave in roughly the same way as if the true covariance were the identity. However, when the true eigenvalues are larger than $1+\sqrt{\gamma}$, the sample eigenvalues have a different asymptotic property. The results of Baik et al. (2005) show an $n^{2/3}$ scaling for the asymptotic distribution when a non-unit population eigenvalue lies below the threshold $1+\sqrt{\gamma}$, and an $n^{1/2}$ scaling for those above that threshold.

This paper is about the case of independently and identically distributed observations $X_1,\ldots,X_n$ from an $N$-variate real Gaussian distribution with mean zero and covariance $\Sigma = \text{diag}(\ell_1,\ell_2,\ldots,\ell_M,1,\ldots,1)$, where $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_M > 1$. Notice that, since observations are Gaussian, there is no loss of generality in assuming the covariance matrix to be diagonal: under an orthogonal transformation of the data, the sample eigenvalues are invariant and the sample eigenvectors are equivariant. The $N \times n$ matrix $\mathbf{X} = (X_1 : \ldots : X_n)$ is a double array, indexed by both $n$ and $N = N(n)$ on the same probability space, and with $N/n \to \gamma$, where $\gamma$ is a positive constant. Throughout it is assumed that $0 < \gamma < 1$, although much of the analysis can be extended to the case $\gamma \geq 1$ with a little extra work. The aim is to study the asymptotic behavior of the large eigenvalues of the sample covariance matrix $\mathbf{S} = (1/n)\mathbf{X}\mathbf{X}^T$ as $n \to \infty$.

In this context, the primary focus of study is the second order behavior of the $M$ largest eigenvalues of the sample covariance matrix. Distributional limits of the sample eigenvalues $\widehat{\ell_\nu}$ are derived when $\ell_\nu > 1 + \sqrt{\gamma}$, for the case when $\ell_\nu$ has multiplicity one. A comprehensive study of all possible scenarios is beyond the scope of this paper. The almost sure limits (see Theorem 1 and Theorem 2) of the sample eigenvalues, obtained by Baik and Silverstein (2006), are used in the proofs of some of the results. However, in the Gaussian case the same limits can be derived through the approach taken here in deriving the distributional limits of the eigenvalues and eigenvectors. For details refer to Paul (2004). This alternative approach gives a different perspective to the limits, in particular to their identification as certain linear functionals of the limiting Marčenko-Pastur law when the true eigenvalue is above $1+\sqrt{\gamma}$. Another aspect of the current approach is that it throws light on the behavior of the eigenvectors associated with the $M$ largest eigenvalues. The sample eigenvectors also undergo a phase transition. By performing a natural decomposition of the sample eigenvectors into "signal" and "noise" parts, it is shown that when $\ell_\nu > 1 + \sqrt{\gamma}$, the "signal" part of the eigenvectors is asymptotically normal. This paper also gives a reasonably thorough description of the "noise" part of the eigenvectors.

The results derived in this paper contain some important messages for inference on multivariate data. First, the phase transition phenomena described in this paper means that some commonly used tests for the hypothesis $\Sigma = I$, like the largest root test (Roy (1953)), may not reliably detect small departures from an identity covariance when the ratio $N/n$ is significantly larger than zero. At the same time, Theorem 3 can be used for making inference on the larger population eigenvalues. This is discussed further in Section 2.2. Second, important consequence of the results here are insights as to why it might not be such a good idea to use Principal Component Analysis (PCA) for dimension reduction in a high-dimensional setting, at least not in its standard form. This has been observed by Johnstone and Lu (2004), who show that when $N/n \to \gamma \in (0, \infty)$, the sample principal components are inconsistent estimates of the population principal components. Theorem 4 says exactly how bad this inconsistency is and its proof demonstrates clearly how this inconsistency originates. Theorem 5 and Theorem 6 are important to understanding the second order behavior of the sample eigenvectors, and have consequences for analyses of functional data. This is elaborated on in Section 2.4.

The rest of the paper is organized as follows. Section 2 has the main results. Section 3 has key quantities and expressions that are required to derive the results. Section 4 is devoted to deriving the asymptotic distribution of eigenvalues (Theorem 3). Section 5 concerns matrix perturbation analysis, which is a key ingredient in the proofs of Theorem 4-6. Some of the proofs are given in Appendix A and Appendix B.

## 2. Discussion of the Results

Throughout $\widehat{\ell}_\nu$ is used to denote the $\nu$th largest eigenvalue of $\mathbf{S}$, and $\Longrightarrow$ is used to denote convergence in distribution.

### 2.1. Almost sure limit of $M$ largest eigenvalues

The following results are due to Baik and Silverstein (2006), and are proved under finite fourth moment assumptions on the distribution of the random variables.

**Theorem 1.** *Suppose that $\ell_\nu \leq 1 + \sqrt{\gamma}$ and that $N/n \to \gamma \in (0,1)$, as $n \to \infty$. Then*

$$\widehat{\ell}_\nu \to (1+\sqrt{\gamma})^2, \qquad \text{almost surely as } \; n \to \infty. \tag{1}$$

**Theorem 2.** *Suppose that $\ell_\nu > 1 + \sqrt{\gamma}$ and that $N/n \to \gamma \in (0,1)$ as $n \to \infty$. Then*

$$\widehat{\ell}_\nu \to \ell_\nu \left(1 + \frac{\gamma}{\ell_\nu - 1}\right), \qquad \text{almost surely as } \; n \to \infty. \tag{2}$$

Denote the limit in (2) by $\rho_\nu := \ell_\nu \left(1 + \frac{\gamma}{\ell_\nu - 1}\right)$; $\rho_\nu$ appears (Lemma B.1) as a solution to the equation

$$\rho = \ell(1 + \gamma \int \frac{x}{\rho - x} dF_\gamma(x)) \tag{3}$$

with $\ell = \ell_\nu$. Since $F_\gamma$ is supported on $[(1-\sqrt{\gamma})^2, (1+\sqrt{\gamma})^2]$ for $\gamma \leq 1$ (with a single isolated point added to the support for $\gamma > 1$), the function on the RHS is monotonically decreasing in $\rho \in ((1+\sqrt{\gamma})^2, \infty)$, while the LHS is obviously increasing in $\rho$. So a solution to (3) exists only if $\ell_\nu \geq 1 + c_\gamma$, for some $c_\gamma > 0$. That $c_\gamma = \sqrt{\gamma}$ is a part of Lemma B.1. Note that when $\ell_\nu = 1 + \sqrt{\gamma}$, $\rho_\nu = (1+\sqrt{\gamma})^2$ is the almost sure limit of the $j$th largest eigenvalue (for $j$ fixed) in the identity covariance case.

### 2.2. Asymptotic normality of sample eigenvalues

When a non-unit eigenvalue of $\Sigma$ is of multiplicity one, and above the critical value $1+\sqrt{\gamma}$, it is shown that the corresponding sample eigenvalue is asymptotically normally distributed. Note that for the complex Gaussian case, a result in the analogous situation has been derived by (Baik et al. (2005, Thm 1.1(b))). They showed that when the largest eigenvalue is greater than $1+\sqrt{\gamma}$ and of multiplicity $k$, the largest sample eigenvalue, after similar centering and scaling, converges in distribution to the distribution of the largest eigenvalue of a $k \times k$ GUE (Gaussian Unitary Ensemble). They also derived the limiting distributions for the case when a (non-unit) population eigenvalue is smaller than $1+\sqrt{\gamma}$. Distributional aspect of a sample eigenvalue for the real case in the latter situation is beyond the scope of this paper. The method used in this paper differs substantially from the approach taken by Baik et al. (2005). They used the joint density of the eigenvalues of $\mathbf{S}$ to derive a determinantal representation of the distribution of the largest eigenvalue and then adopted the steepest descent method for the asymptotic analysis. However, this paper relies on a matrix analysis approach.

**Theorem 3.** *Suppose that $\ell_\nu > 1+\sqrt{\gamma}$ and that $\ell_\nu$ has multiplicity 1. Then as $n, N \to \infty$ so that $N/n - \gamma = o(n^{-1/2})$,*

$$\sqrt{n}(\widehat{\ell}_\nu - \rho_\nu) \Longrightarrow N(0, \sigma^2(\ell_\nu)) \tag{4}$$

*where, for $\ell > 1+\sqrt{\gamma}$ and $\rho(\ell) = \ell(1+\gamma/(\ell-1))$,*

$$\sigma^2(\ell) = \frac{2\ell\rho(\ell)}{1+\ell\gamma\int \frac{x}{(\rho(\ell)-x)^2} dF_\gamma(x)} = \frac{2\ell\rho(\ell)}{1+\frac{\ell\gamma}{(\ell-1)^2-\gamma}} = 2\ell^2\Big(1-\frac{\gamma}{(\ell-1)^2}\Big). \tag{5}$$

In the fixed $N$ case, when the $\nu$th eigenvalue has multiplicity 1, the $\nu$th sample eigenvalue is asymptotically $N(\ell_\nu, (1/n)2\ell_\nu^2)$ (Anderson (1963)). Thus the positivity of the dimension to sample size ratio creates a bias and reduces the variance. However, if $\gamma$ is much smaller compared to $\ell_\nu$, the variance $\sigma^2(\ell_\nu)$ is approximately $2\ell_\nu^2$ which is the asymptotic variance in the fixed $N$ case. This is what we expect intuitively, since the eigenvector associated with this sample eigenvalue, looking to maximize the quadratic form involving $\mathbf{S}$ (under orthogonality restrictions), will tend to put more mass on the $\nu$th coordinate. This is demonstrated even more clearly by Theorem 4.

Suppose that we test the hypothesis $\Sigma = I$ versus the alternative that $\Sigma = \mathrm{diag}(\ell_1, \ldots, \ell_M, 1, \ldots, 1)$ with $\ell_1 \geq \cdots \geq \ell_M > 1$, based on i.i.d. observations from $N(0, \Sigma)$. If $\ell_1 > 1+\sqrt{\gamma}$, it follows from Theorem 2 that the largest root test is asymptotically consistent. For the special case when $\ell_1$ is of multiplicity one, Theorem 3 gives an expression for the asymptotic power function, assuming that $N/n$ converges to $\gamma$ fast enough, as $n \to \infty$. One has to view this in context, since the result is derived under the assumption that $\ell_1, \ldots, \ell_M$ are all fixed, and we do not have a rate of convergence for the distribution of $\widehat{\ell}_1$ toward normality. However, Theorem 3 can be used to find confidence intervals for the larger eigenvalues under the non-null model.

### 2.3. Angle between true and estimated eigenvectors

It is well-known (see, for example, Muirhead (1982), or Anderson (1963)) that, when $\Sigma = I$ and the observations are Gaussian, the matrix of sample eigenvectors of $\mathbf{S}$ is Haar distributed. In the non-Gaussian situation, Silverstein (1990) showed weak convergence of random functions of this matrix. In the context of non-identity covariance, Hoyle and Rattray (2004) described a phase transition phenomenon in the asymptotic behavior of the angle between the true and estimated eigenvector associated with a non-unit eigenvalue $\ell_\nu$. They term this "the phenomenon of retarded learning". They derived this result at a level of rigor consistent with that in the physics literature. Their result can be rephrased in our context to mean that if $1 < \ell_\nu \leq 1 + \sqrt{\gamma}$ is a simple eigenvalue, then the cosine of the angle between the corresponding true and estimated eigenvectors converges almost surely to zero; yet, there is a strictly positive limit if $\ell_\nu > 1 + \sqrt{\gamma}$. Part (a) of Theorem 4, stated below and proved in Section 5, is a precise statement of the latter part of their result. This also readily proves a stronger version of the result regarding inconsistency of sample eigenvectors as is stated in Johnstone and Lu (2004).

**Theorem 4.** *Suppose that $N/n \to (0,1)$ as $n, N \to \infty$. Let $\widetilde{\mathbf{e}}_\nu$ denote the $N \times 1$ vector with $1$ in the $\nu$th coordinate and zeros elsewhere, and let $\mathbf{p}_\nu$ denote the eigenvector of $\mathbf{S}$ associated with the eigenvalue $\widehat{\ell}_\nu$.*

(a) *If $\ell_\nu > 1 + \sqrt{\gamma}$ and of multiplicity one,*

$$|\langle \mathbf{p}_\nu, \widetilde{\mathbf{e}}_\nu \rangle| \stackrel{a.s.}{\to} \sqrt{\left(1 - \frac{\gamma}{(\ell_\nu - 1)^2}\right) \Big/ \left(1 + \frac{\gamma}{\ell_\nu - 1}\right)} \qquad \text{as } n \to \infty. \tag{6}$$

(b) *If $\ell_\nu \leq 1 + \sqrt{\gamma}$,*

$$\langle \mathbf{p}_\nu, \widetilde{\mathbf{e}}_\nu \rangle \stackrel{a.s.}{\to} 0 \qquad \text{as } n \to \infty. \tag{7}$$

### 2.4. Distribution of sample eigenvectors

Express the eigenvector $\mathbf{p}_\nu$ corresponding to the $\nu$th sample eigenvalue as $\mathbf{p}_\nu = (p_{A,\nu}^T, p_{B,\nu}^T)^T$, where $p_{A,\nu}$ is the subvector corresponding to the first $M$ coordinates. Follow the convention that $\nu$th coordinate of $\mathbf{p}_\nu$ is nonnegative. Let $\mathbf{e}_k$ denote the $k$th canonical basis vector in $\mathbb{R}^M$. Then the following holds.

**Theorem 5.** *Suppose that $\ell_\nu > 1 + \sqrt{\gamma}$ and that $\ell_\nu$ has multiplicity* 1. *Then as $n, N \to \infty$ so that $N/n - \gamma = o(n^{-1/2})$,*

$$\sqrt{n}\Big(\frac{p_{A,\nu}}{\|p_{A,\nu}\|} - \mathbf{e}_\nu\Big) \Longrightarrow N_M(0, \Sigma_\nu(\ell_\nu)), \tag{8}$$

$$\Sigma_\nu(\ell_\nu) = \left(\frac{1}{1 - \frac{\gamma}{(\ell_\nu - 1)^2}}\right) \sum_{1 \le k \ne \nu \le M} \frac{\ell_k \ell_\nu}{(\ell_k - \ell_\nu)^2} \mathbf{e}_k \mathbf{e}_k^T. \tag{9}$$

The following is a non-asymptotic result about the behavior of the $\nu$th sample eigenvector. Here $\nu$ can be any number between 1 and $\min(n, N)$.

**Theorem 6.** *The vector $p_{B,\nu}/\|p_{B,\nu}\|$ is distributed uniformly on the unit sphere $\mathbb{S}^{N-M-1}$ and is independent of $\|p_{B,\nu}\|$.*

Theorem 6, taken in conjunction with Theorem 4 and Theorem 5, has interesting implications in the context of functional data analysis (FDA). One approach in FDA involves summarizing the data in terms of the first few principal components of the sample curves. A common technique here is to apply a smoothing to the curves before carrying out the PCA. Occassionally this is followed by smoothing of the estimated sample eigenvectors. Ramsay and Silverman (1997) detailed various methods of carrying out a functional principal component analysis (FPCA).

Think of a situation where each individual observation is a random function whose domain is an interval. Further, suppose that these functions are corrupted with additive and isotropic noise. If the true functions are smooth and belong to a finite-dimensional linear space, then it is possible to analyze these data by transforming the noisy curves in a suitable orthogonal basis, e.g., a wavelet basis or a Fourier basis. If the data are represented in terms of first $N$ basis functions (where $N$ is the resolution of the model, or the number of equally spaced points where the measurements are taken), then the matrix of coefficients in the basis representation can be written as $\mathbf{X}$, whose columns are $N$-dimensional vectors of wavelet or Fourier coefficients of individual curves. The corresponding model can then be described, under the assumption of Gaussianity, in terms of an $N(\mathbf{m}, \Sigma)$ model, where $\mathbf{m}$ is the mean vector and $\Sigma$ is the covariance matrix with eigenvalues $\ell_1 \ge \ell_2 \ge \cdots \ge \ell_M > \sigma^2 = \cdots = \sigma^2$. Here $\sigma^2$ is the variance of the isotropic noise associated with the sample curves. Our results show that when $\ell_\nu > \sigma^2(1 + \sqrt{\gamma})$ and $N/n$ converges to some $\gamma \in (0, 1)$, it is possible to give a fairly accurate approximation of the sample eigenvectors associated with the "signal" eigenvalues (up to a convention on sign). The similarity of the resulting expressions to the standard "signal plus noise" models prevalent in the nonparametric literature, and its implications in the context of estimating the eigenstructure of the sample curves, are being investigated by the author.

## 3. Representation of the Eigenvalues of S

Throughout assume that $n$ is large enough so that $N/n < 1$. Partition the matrix $\mathbf{S}$ as

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{AA} & \mathbf{S}_{AB} \\ \mathbf{S}_{BA} & \mathbf{S}_{BB} \end{bmatrix},$$

where the suffix $A$ corresponds to the set of coordinates $\{1, \ldots, M\}$, and $B$ corresponds to the set $\{M+1, \ldots, N\}$. As before, use $\widehat{\ell}_\nu$ and $\mathbf{p}_\nu$ to denote the $\nu$th largest sample eigenvalue and the corresponding sample eigenvector. To avoid any ambiguity, follow the convention that the $\nu$th element of $\mathbf{p}_\nu$ is nonnegative. Write $\mathbf{p}_\nu$ as $\mathbf{p}_\nu^T = (p_{A,\nu}^T, p_{B,\nu}^T)$ and denote the norm $\|p_{B,\nu}\|$ by $R_\nu$. Then almost surely $0 < R_\nu < 1$.

With this setting in place, express the first $M$ eigen-equations for $\mathbf{S}$ as

$$\mathbf{S}_{AA} p_{A,\nu} + \mathbf{S}_{AB} p_{B,\nu} = \widehat{\ell}_\nu p_{A,\nu}, \qquad \nu = 1, \ldots, M, \tag{10}$$

$$\mathbf{S}_{BA} p_{A,\nu} + \mathbf{S}_{BB} p_{B,\nu} = \widehat{\ell}_\nu p_{B,\nu}, \qquad \nu = 1, \ldots, M, \tag{11}$$

$$p_{A,\nu}^T p_{A,\nu'} + p_{B,\nu}^T p_{B,\nu'} = \delta_{\nu,\nu'}, \qquad 1 \leq \nu, \nu' \leq M, \tag{12}$$

where $\delta_{\nu\nu'}$ is the Kronecker symbol. Denote the vector $p_{A,\nu}/\|p_{A,\nu}\| = p_{A,\nu}/\sqrt{1-R_\nu^2}$ by $a_\nu$. Thus, $\|a_\nu\| = 1$. Similarly define $q_\nu := p_{B,\nu}/R_\nu$, and again $\|q_\nu\| = 1$.

Since almost surely $0 < R_\nu < 1$, and $\widehat{\ell}_\nu I - \mathbf{S}_{BB}$ is invertible, it follows from (11) that

$$q_\nu = \frac{\sqrt{1-R_\nu^2}}{R_\nu} (\widehat{\ell}_\nu I - \mathbf{S}_{BB})^{-1} \mathbf{S}_{BA} a_\nu. \tag{13}$$

Divide both sides of (10) by $\sqrt{1-R_\nu^2}$ and substitute the expression for $q_\nu$, to yield

$$(\mathbf{S}_{AA} + \mathbf{S}_{AB}(\widehat{\ell}_\nu I - \mathbf{S}_{BB})^{-1} \mathbf{S}_{BA}) a_\nu = \widehat{\ell}_\nu a_\nu, \qquad \nu = 1, \ldots, M. \tag{14}$$

This equation is important since it shows that $\widehat{\ell}_\nu$ is an eigenvalue of the matrix $K(\widehat{\ell}_\nu)$, where $K(x) := \mathbf{S}_{AA} + \mathbf{S}_{AB}(xI - \mathbf{S}_{BB})^{-1}\mathbf{S}_{BA}$, with corresponding eigenvector $a_\nu$. This observation is the building block for all the analyses that follow. However, it is more convenient to express the quantities in terms of the spectral elements of the data matrix $\mathbf{X}$.

Let $\Lambda$ denote the diagonal matrix $\mathrm{diag}(\ell_1, \ldots, \ell_M)$. Because of normality, the observation matrix $\mathbf{X}$ can be reexpressed as $\mathbf{X}^T = [\mathbf{Z}_A^T \Lambda^{1/2} : \mathbf{Z}_B^T]$, $\mathbf{Z}_A$ is $M \times n$, $\mathbf{Z}_B$ is $(N-M) \times n$. The entries of $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are i.i.d. $N(0,1)$, and $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are mutually independent. Also assume that $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are defined on the same probability space.

Write the singular value decomposition of $\mathbf{Z}_B/\sqrt{n}$ as

$$\frac{1}{\sqrt{n}}\mathbf{Z}_B = V\mathcal{M}^{\frac{1}{2}}H^T, \tag{15}$$

where $\mathcal{M}$ is the $(N-M)\times(N-M)$ diagonal matrix of the eigenvalues of $\mathbf{S}_{BB}$ in decreasing order; $V$ is the $(N-M)\times(N-M)$ matrix of eigenvectors of $\mathbf{S}_{BB}$; and $H$ is the $n\times(N-M)$ matrix of right singular vectors. Denote the diagonal elements of $\mathcal{M}$ by $\mu_1 > \cdots > \mu_{N-M}$, suppressing the dependence on $n$.

Note that the columns of $V$ form a complete orthonormal basis for $\mathbb{R}^{N-M}$, while the columns of $H$ form an orthonormal basis of an $(N-M)$ dimensional subspace (the rowspace of $\mathbf{Z}_B$) of $\mathbb{R}^n$.

Define $T := (1/\sqrt{n})H^T\mathbf{Z}_A^T$. $T$ is an $(N-M)\times M$ matrix with columns $t_1,\dots,t_M$. The most important property about $T$ is that the vectors $t_1,\dots,t_M$ are distributed as i.i.d. $N(0,\frac{1}{n}I_{N-M})$ and are independent of $\mathbf{Z}_B$. This is because the columns of $H$ form an orthonormal set of vectors, the rows of $\mathbf{Z}_A$ are i.i.d. $N_n(0,I)$ vectors, and $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are independent.

Thus, (14) can be expressed as,

$$(\mathbf{S}_{AA} + \Lambda^{\frac{1}{2}}T^T\mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-1}T\Lambda^{\frac{1}{2}})a_\nu = \widehat{\ell}_\nu a_\nu, \qquad \nu = 1,\dots,M. \tag{16}$$

Also, $K(x)$ can be expressed as

$$K(x) = \mathbf{S}_{AA} + \Lambda^{\frac{1}{2}}T^T\mathcal{M}(xI-\mathcal{M})^{-1}T\Lambda^{\frac{1}{2}}. \tag{17}$$

Rewrite (12) in terms of the vectors $\{a_\nu : \nu = 1,\dots,M\}$ as

$$\begin{aligned} &a_\nu^T[I + \Lambda^{\frac{1}{2}}T^T(\widehat{\ell}_\nu I - \mathcal{M})^{-1}\mathcal{M}(\widehat{\ell}_{\nu'} I - \mathcal{M})^{-1}T\Lambda^{\frac{1}{2}}]a_{\nu'} \\ &\quad = \frac{1}{1-R_\nu^2}\delta_{\nu\nu'}, \quad 1 \le \nu,\nu' \le M. \end{aligned} \tag{18}$$

Proofs of the theorems depend heavily on the asymptotic behavior of the largest eigenvalue, as well as the Empirical Spectral Distribution (ESD) of Wishart matrices in the null (i.e., identity covariance) case. Throughout, the ESD of $\mathbf{S}_{BB}$ is denoted by $\widehat{F}_{n,N-M}$. We know that (Bai (1999))

$$\widehat{F}_{n,N-M} \Longrightarrow F_\gamma, \quad \text{almost surely as } n\to\infty. \tag{19}$$

The following result, proved in Appendix A, is about the deviation of the largest eigenvalue of $\mathbf{S}_{BB}$ from its limiting value $\kappa_\gamma := (1+\sqrt{\gamma})^2$. The importance of this result is that, for proving the limit theorems, it is enough to do calculations by restricting attention to sets of the form $\mu_1 \le \kappa_\gamma + \delta$ for some suitably chosen $\delta > 0$.

**Proposition 1.** *For any* $0 < \delta < \kappa_\gamma/2$,

$$\mathbb{P}(\mu_1 - \kappa_\gamma > \delta) \leq \exp\left(-\frac{3n\delta^2}{64\kappa_\gamma}\right), \qquad \text{for } n \geq n_0(\gamma, \delta), \tag{20}$$

*where* $n_0(\gamma,\delta)$ *is an integer large enough that* $|(1+\sqrt{(N-m)/n})^2 - \kappa_\gamma| \leq \delta/4$ *for* $n \geq n_0(\gamma,\delta)$.

## 4. Proof of Theorem 3

The first step here is to utilize the eigen-equation (16) to get

$$\widehat{\ell}_\nu = a_\nu^T(\mathbf{S}_{AA} + \Lambda^{\frac{1}{2}} T^T \mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-1} T \Lambda^{\frac{1}{2}})a_\nu. \tag{21}$$

From this, after some manipulations (see Section 5.1 for details),

$$\begin{aligned} &\sqrt{n}(\widehat{\ell}_\nu - \rho_\nu)(1 + \ell_\nu t_\nu^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-2} t_\nu + d_\nu) \\ &\qquad = \sqrt{n}(s_{\nu\nu} + \ell_\nu t_\nu^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-1} t_\nu - \rho_\nu) + o_P(1), \end{aligned} \tag{22}$$

where $d_\nu = -\ell_\nu(\widehat{\ell}_\nu - \rho_\nu)(t_\nu^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-2}(\widehat{\ell}_\nu I - \mathcal{M})^{-1} t_\nu + O_P(1))$, and $s_{\nu\nu}$ is the $(\nu,\nu)$th element of $\mathbf{S}$. It follows readily that $d_\nu = o_P(1)$. It will be shown that the term on the RHS of (22) converges in distribution to a Gaussian random variable with zero mean and variance

$$2\ell_\nu\rho_\nu\left(1 + \ell_\nu\gamma\int\frac{x}{(\rho_\nu - x)^2}dF_\gamma(x)\right). \tag{23}$$

Next, from Proposition 2 stated below (and proved in Appendix B), it follows that

$$t_\nu^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-2} t_\nu \xrightarrow{a.s.} \gamma \int \frac{x}{(\rho_\nu - x)^2} dF_\gamma(x). \tag{24}$$

Hence (4), with $\sigma^2(\ell)$ given by the first expression in (5), follows from (23), (24), (22) and Slutsky's Theorem. Application of (58) gives the second equality in (5), and the third follows from simple algebra.

**Proposition 2.** *Suppose that* $N/n \to \gamma \in (0,1)$ *as* $n \to \infty$. *Let* $\delta, \epsilon > 0$ *satisfy* $\delta < [4(\kappa_\gamma + \epsilon/2)/(\rho - \kappa_\gamma - \epsilon/2)^2][(N-M)/n]$ *and* $\rho \geq \kappa_\gamma + \epsilon$. *Then there is*

$n_*(\rho,\delta,\epsilon,\gamma)$ *such that for all* $n \geq n_*(\rho,\delta,\epsilon,\gamma)$,

$$
\begin{aligned}
&\mathbb{P}\Big(|t_j^T\mathcal{M}(\rho I-\mathcal{M})^{-2}t_j-\gamma\int\frac{x}{(\rho-x)^2}dF_\gamma(x)|>\delta,\ \mu_1<\kappa_\gamma+\frac{\epsilon}{2}\Big)\\
&\quad\leq 2\exp\left(-\frac{n}{N-M}\frac{n(\frac{\delta}{4})^2(\rho-\kappa_\gamma-\frac{\epsilon}{2})^4}{6(\kappa_\gamma+\frac{\epsilon}{2})^2}\right)\\
&\qquad+2\exp\left(-\frac{n}{n+N-M}\frac{n^2(\frac{\delta}{4})^2}{2}\frac{(\rho-\kappa_\gamma-\frac{\epsilon}{2})^6}{16\rho^2(\kappa_\gamma+\frac{\epsilon}{2})}\right)\\
&\qquad+2\exp\left(-\frac{n}{n+N-M}\frac{n^2(\frac{\delta}{4})^2}{2}\frac{(\rho-\kappa_\gamma-\frac{\epsilon}{2})^4}{4(\kappa_\gamma+\frac{\epsilon}{2})}\right),\qquad 1\leq j\leq M.
\end{aligned}
$$

The main term on the RHS of (22) can be expressed as $W_n+W_n'$, where

$$
\begin{aligned}
W_n&=\sqrt{n}(s_{\nu\nu}-(1-\gamma)\ell_\nu+\ell_\nu t_\nu^T\mathcal{M}(\rho_\nu I-\mathcal{M})^{-1}t_\nu-\ell_\nu\rho_\nu\frac{1}{n}trace((\rho_\nu I-\mathcal{M})^{-1})),\\
W_n'&=\sqrt{n}\ell_\nu(\rho_\nu\frac{1}{n}trace((\rho_\nu I-\mathcal{M})^{-1})-\frac{\gamma\ell_\nu}{\ell_\nu-1}).
\end{aligned}
$$

Note that by (57), $\gamma\ell_\nu/(\ell_\nu-1)=\gamma[1+(1/(\ell_\nu-1))]=\gamma\int[\rho_\nu/(\rho_\nu-x)]dF_\gamma(x)$ . On the other hand

$$
\rho_\nu\frac{1}{n}trace((\rho_\nu I-\mathcal{M})^{-1})=\frac{N-M}{n}\int\frac{\rho_\nu}{\rho_\nu-x}\widehat{F}_{n,N-M}(x).
$$

Since the function $1/(\rho_\nu-z)$ is analytic in an open set containing the interval $[(1-\sqrt{\gamma})^2,(1+\sqrt{\gamma})^2]$, from (Bai and Silverstein (2004, Thm 1.1)) the sequence $W_n'=o_P(1)$ because $(N/n)-\gamma=o(n^{-1/2})$.

### 4.1. Asymptotic normality of $W_n$

Recall that by definition of $T$, $t_\nu=(1/\sqrt{n})H^TZ_{A,\nu}$ where $Z_{A,\nu}^T$ is the $\nu$th row of $\mathbf{Z}_A$. Since $N-M<n$, and the columns of $H$ are orthonormal, one can extend them to an orthonormal basis of $\mathbb{R}^n$ given by the matrix $\widetilde{H}=[H:H_c]$, where $H_c$ is $n\times(n-N+M)$. Thus, $\widetilde{H}\widetilde{H}^T=\widetilde{H}^T\widetilde{H}=I_n$. Write

$$
\begin{aligned}
s_{\nu\nu}&=\ell_\nu\frac{1}{n}Z_{A,\nu}^TZ_{A,\nu}=\ell_\nu\frac{1}{n}Z_{A,\nu}^T\widetilde{H}\widetilde{H}^TZ_{A,\nu}\\
&=\ell_\nu(\|\frac{1}{\sqrt{n}}H^TZ_{A,\nu}\|^2+\|\frac{1}{\sqrt{n}}H_c^TZ_{A,\nu}\|^2)\\
&=\ell_\nu(\|t_\nu\|^2+\|w_\nu\|^2),
\end{aligned}
$$

with $w_\nu:=(1/\sqrt{n})H_c^TZ_{A,\nu}$. Thus $w_\nu\sim N(0,I_{n-N+M}/n)$; $t_\nu\sim N(0,I_{N-M}/n)$; and these are mutually independent and independent of $\mathbf{Z}_B$. Therefore, one can represent $W_n$ as a sum of two independent random variables $W_{1,n}$ and $W_{2,n}$, where

$$W_{1,n} = \ell_\nu \sqrt{n}(\|w_\nu\|^2 - (1-\gamma)),$$
$$W_{2,n} = \ell_\nu \rho_\nu \sqrt{n}(t_\nu^T(\rho_\nu I - \mathcal{M})^{-1} t_\nu - \frac{1}{n} trace((\rho_\nu I - \mathcal{M})^{-1})).$$

Since $n\|w_\nu\|^2 \sim \chi^2_{n-N+M}$ and $N/n - \gamma = o(n^{-1/2})$, it follows that $W_{1,n} \Longrightarrow N(0, 2\ell_\nu^2(1-\gamma))$. Later, it is shown that

$$W_{2,n} \Longrightarrow N(0, 2\ell_\nu^2 \gamma \int \frac{\rho_\nu^2}{(\rho_\nu - x)^2} dF_\gamma(x)). \tag{25}$$

Therefore, the asymptotic normality of $W_n$ is established. $W_n' = o_P(1)$ then implies asymptotic normality of the RHS of (22). The expression (23) for asymptotic variance is then deduced from the identity

$$\int \frac{\rho_\nu^2}{(\rho_\nu - x)^2} dF_\gamma(x) = 1 + \frac{1}{\ell_\nu - 1} + \rho_\nu \int \frac{x}{(\rho_\nu - x)^2} dF_\gamma(x),$$

which follows from (57). Therefore, the asymptotic variance of $W_n$ is

$$2\ell_\nu^2(1-\gamma) + 2\ell_\nu^2\gamma \int \frac{\rho_\nu^2}{(\rho_\nu - x)^2} dF_\gamma(x)$$
$$= 2\ell_\nu^2(1 + \frac{\gamma}{\ell_\nu - 1}) + 2\ell_\nu^2 \rho_\nu \gamma \int \frac{x}{(\rho_\nu - x)^2} dF_\gamma(x),$$

from which (23) follows since $\ell_\nu(1 + \gamma/(\ell_\nu - 1)) = \rho_\nu$.

## 4.2. Proof of (25)

Let $t_\nu = (t_{\nu,1}, \ldots, t_{\nu,N-M})^T$, $t_{\nu,j} \overset{i.i.d.}{\sim} N(0, 1/n)$ and independent of $\mathcal{M}$. Hence, if $y_j = \sqrt{n} t_{\nu,j}$, then

$$W_{2,n} = \ell_\nu \rho_\nu \frac{1}{\sqrt{n}} \Big( \sum_{j=1}^{N-M} \frac{1}{\rho_\nu - \mu_j} y_j^2 - \sum_{j=1}^{N-M} \frac{1}{\rho_\nu - \mu_j} \Big),$$

where $\{y_j\}_{j=1}^{N-M} \overset{i.i.d.}{\sim} N(0,1)$, and $\{y_j\}_{j=1}^{N-M}$ is independent of $\mathcal{M}$. To establish (25) it suffices to show that, for all $t \in \mathbb{R}$,

$$\phi_{W_{2,n}}(t) := \mathbb{E}\exp(itW_{2,n}) \to \phi_{\widetilde{\sigma}^2(\ell_\nu)}(t) := \exp\Big( -\frac{t^2 \widetilde{\sigma}^2(\ell_\nu)}{2} \Big), \text{ as } n \to \infty,$$

where $\widetilde{\sigma}^2(\ell) = 2\ell^2\gamma \int \rho^2(\ell)/(\rho^2(\ell) - x)^2 dF_\gamma(x)$ for $\ell > 1 + \sqrt{\gamma}$. Define $J_\gamma(\delta) := \{\mu_1 \leq \kappa_\gamma + \delta\}$, where $\delta > 0$ is any number such that $\rho_\nu > \kappa_\gamma + 2\delta$. Note that

$J_\gamma(\delta)$ is a measurable set that depends on $n$, and $\mathbb{P}(J_\gamma(\delta)) \to 1$ as $n \to \infty$ by Proposition 1. Thus we need only establish that for all $t \in \mathbb{R}$,

$$\mathbb{E}\Big[\Big|\mathbb{E}(e^{itW_{2,n}} \mid \mathcal{M}) \exp\Big(\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\Big) - 1\Big|, \mu_1 \le \kappa_\gamma + \delta\Big] \to 0, \text{ as } n \to \infty, \tag{26}$$

where the outer expectation is with respect to the distribution of $\mathcal{M}$. Since the characteristic function of a $\chi_1^2$ random variable is $\psi(x) = 1/\sqrt{1-2ix}$, on the set $\{\mu_1 \le \kappa_\gamma + \delta\}$, the inner conditional expectation is

$$\begin{aligned} &\prod_{j=1}^{N-M} \psi\Big(\frac{t\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \mu_j)}\Big) \exp\Big(-\frac{it\ell_\nu\rho_\nu}{\sqrt{n}} \sum_{j=1}^{N-M} \frac{1}{\rho_\nu - \mu_j}\Big) \\ &\quad = \prod_{j=1}^{N-M} \left(1 - \frac{2it\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \mu_j)}\right)^{-\frac{1}{2}} \exp\Big(-\frac{it\ell_\nu\rho_\nu}{\sqrt{n}} \sum_{j=1}^{N-M} \frac{1}{\rho_\nu - \mu_j}\Big). \end{aligned} \tag{27}$$

Let $\log z$ $(z \in \mathbb{C})$ be the principal branch of the complex logarithm. Then

$$\left(1 - \frac{2it\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \mu_j)}\right)^{-\frac{1}{2}} = \exp\left(-\frac{1}{2}\log\left(1 - \frac{2it\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \mu_j)}\right)\right).$$

In view of the Taylor series expansion of $\log(1+z)$ (valid for $|z| < 1$), for $n \ge n_*(\nu, \gamma, \delta)$, large enough so that $(|t|\ell_\nu\rho_\nu)/(\sqrt{n}(\rho_\nu - \kappa_\gamma - \delta)) < 1/2$, the conditional expectation (27) is

$$\exp\left(\frac{1}{2}\sum_{j=1}^{N-M}\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{2it\ell_\nu\rho_\nu}{\sqrt{n}}\frac{1}{\rho_\nu - \mu_j}\Big)^k - \frac{it\ell_\nu\rho_\nu}{\sqrt{n}}\sum_{j=1}^{N-M}\frac{1}{\rho_\nu - \mu_j}\right).$$

The inner sum is dominated by a geometric series and hence is finite for $n \ge n_*(\nu, \gamma, \delta)$ on the set $J_\gamma(\delta)$. Interchanging the order of summations, on $J_\gamma(\delta)$ the term within the exponent becomes

$$-\frac{t^2}{2}\Big[2\ell_\nu^2\rho_\nu^2\frac{1}{n}\sum_{j=1}^{N-M}\frac{1}{(\rho_\nu - \mu_j)^2}\Big] + \frac{1}{2}\sum_{k=3}^{\infty}\frac{1}{k}\left(\frac{2it\ell_\nu\rho_\nu}{\sqrt{n}}\right)^k \sum_{j=1}^{N-M}\frac{1}{(\rho_\nu - \mu_j)^k}. \tag{28}$$

Denote the first term of (28) by $a_n(t)$ and the second term by $\widetilde{r}_n(t)$. For $n \ge n_*(\nu, \gamma, \delta)$, on $J_\gamma(\delta)$,

$$\begin{aligned} |\widetilde{r}_n(t)| &\le \frac{t^2}{3}\Big[2\ell_\nu^2\rho_\nu^2\frac{1}{n}\sum_{j=1}^{N-M}\frac{1}{(\rho_\nu - \mu_j)^2}\Big]\left(\frac{2|t|\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \kappa_\gamma - \delta)}\right) \\ &\quad \times \left(1 - \frac{2|t|\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \kappa_\gamma - \delta)}\right)^{-1}. \end{aligned} \tag{29}$$

Let $G_2(\cdot\ ;\rho,\gamma,\delta)$ to be the bounded function ($\rho > \kappa_\gamma + \delta$) defined through (55). Then on $J_\gamma(\delta)$, $(1/n)\sum_{j=1}^{N-M}[1/(\rho_\nu-\mu_j)^2] = ((N-M)/n)\int G_2(x;\rho_\nu,\gamma,\delta)\, d\widehat{F}_{n,N-M}(x)$ and the quantity on the RHS converges almost surely to $\gamma\int G_2(x;\rho_\nu,\gamma,\delta)dF_\gamma\ (x) = \gamma\int[1/(\rho_\nu-x)^2]dF_\gamma(x)$ because of the continuity of $G_2(\cdot;\rho_\nu,\gamma,\delta)$ and (19). Moreover, on $J_\gamma(\delta)$, $a_n(t)$ and $\widetilde{r}_n(t)$ are bounded for $n \geq n_*(\nu,\gamma,\delta)$. Therefore, from this observation, and (27) and (28), the sequence in (26) is bounded by

$$\mathbb{E}\left[\exp\Big(a_n(t)+\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\Big)(\exp(|\widetilde{r}_n(t)|)-1)\ \mathbb{I}_{J_\gamma(\delta)}\right] \\ +\mathbb{E}\left[\Big|\exp\Big(a_n(t)+\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\Big)-1\Big|\ \mathbb{I}_{J_\gamma(\delta)}\right],$$

which converges to zero by the Bounded Convergence Theorem.

## 5. Approximation to the Eigenvectors

This section deals with an asymptotic expansion of the eignvector $a_\nu$ of the matrix $K(\widehat{\ell}_\nu)$ associated with the eigenvalue $\widehat{\ell}_\nu$, when $\ell_\nu$ is greater than $1+\sqrt{\gamma}$ and has multiplicity 1. This expansion has already been used in the proof of Theorem 3. An important step, presented through Lemma 1, is to provide a suitable bound for the remainder in the expansion. The approach taken here follows the perturbation analysis approach in Kneip and Utikal (2001), (see also Kato (1980, Chap. 2)). For the benefit of the readers, the steps leading to the expansion are outlined below.

First observe that $\rho_\nu$ is *the* eigenvalue of $(\rho_\nu/\ell_\nu)\Lambda$ associated with *the* eigenvector $\mathbf{e}_\nu$. Define

$$\mathcal{R}_\nu = \sum_{k\neq\nu}^{M}\frac{\ell_\nu}{\rho_\nu(\ell_k-\ell_\nu)}\mathbf{e}_k\mathbf{e}_k^T. \tag{30}$$

Note that $\mathcal{R}_\nu$ is the resolvent of $(\rho_\nu/\ell_\nu)\Lambda$ "evaluated" at $\rho_\nu$. Then utilize the defining equation (16) to write

$$(\frac{\rho_\nu}{\ell_\nu}\Lambda-\rho_\nu I)a_\nu = -(K(\widehat{\ell}_\nu)-\frac{\rho_\nu}{\ell_\nu}\Lambda)a_\nu + (\widehat{\ell}_\nu-\rho_\nu)a_\nu.$$

Define $D_\nu = K(\widehat{\ell}_\nu)-(\rho_\nu/\ell_\nu)\Lambda_\nu$; premultiply both sides by $\mathcal{R}_\nu$; and observe that $\mathcal{R}_\nu((\rho_\nu/\ell_\nu)\Lambda-\rho_\nu I) = I_M - \mathbf{e}_\nu\mathbf{e}_\nu^T := P_\nu^\perp$. Then

$$P_\nu^\perp a_\nu = -\mathcal{R}_\nu D_\nu a_\nu + (\widehat{\ell}_\nu-\rho_\nu)\mathcal{R}_\nu a_\nu. \tag{31}$$

As a convention, suppose that $\langle\mathbf{e}_\nu,a_\nu\rangle \geq 0$. Then write $a_\nu = \langle\mathbf{e}_\nu,a_\nu\rangle\mathbf{e}_\nu + P_\nu^\perp a_\nu$ and observe that $\mathcal{R}_\nu\mathbf{e}_\nu = 0$. Hence

$$a_\nu - \mathbf{e}_\nu = -\mathcal{R}_\nu D_\nu\mathbf{e}_\nu + r_\nu, \tag{32}$$

where $r_\nu = -(1 - \langle \mathbf{e}_\nu, a_\nu \rangle)\mathbf{e}_\nu - \mathcal{R}_\nu D_\nu (a_\nu - \mathbf{e}_\nu) + (\widehat{\ell}_\nu - \rho_\nu)\mathcal{R}_\nu (a_\nu - \mathbf{e}_\nu)$. Now, define

$$\alpha_\nu = \|\mathcal{R}_\nu D_\nu\| + |\widehat{\ell}_\nu - \rho_\nu| \|\mathcal{R}_\nu\|, \text{ and } \beta_\nu = \|\mathcal{R}_\nu D_\nu \mathbf{e}_\nu\|. \tag{33}$$

The following lemma gives a bound on the residual $r_\nu$. The proof can be found in Paul (2004).

**Lemma 1.** $r_\nu$ *satisfies*

$$\|r_\nu\| \leq \begin{cases} \beta_\nu \left( \frac{\alpha_\nu(1+\alpha_\nu)}{1-\alpha_\nu(1+\alpha_\nu)} + \frac{\beta_\nu}{(1-\alpha_\nu(1+\alpha_\nu))^2} \right) & \textit{if } \alpha_\nu < \frac{\sqrt{5}-1}{2}, \\ \alpha_\nu^2 + 2\alpha_\nu & \textit{always.} \end{cases} \tag{34}$$

The next task is to establish that $\beta_\nu = o_P(1)$ and $\alpha_\nu = o_P(1)$. First, write

$$\begin{aligned} D_\nu = (\mathbf{S}_{AA} - \Lambda) &+ \Lambda^{\frac{1}{2}} \left( T^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-1} T - \frac{1}{n}\, trace(\mathcal{M}(\rho_\nu I - \mathcal{M})^{-1}) I \right) \Lambda^{\frac{1}{2}} \\ &+ \left( \frac{1}{n}\, trace(\mathcal{M}(\rho_\nu I - \mathcal{M})^{-1}) - \gamma \int \frac{x}{\rho_\nu - x} dF_\gamma(x) \right) \Lambda \\ &+ (\rho_\nu - \widehat{\ell}_\nu) \Lambda^{\frac{1}{2}} T^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-1} (\widehat{\ell}_\nu I - \mathcal{M})^{-1} T \Lambda^{\frac{1}{2}}. \end{aligned} \tag{35}$$

Since $\widehat{\ell}_\nu \overset{a.s.}{\to} \rho_\nu > \kappa_\gamma$ and $\mu_1 \overset{a.s.}{\to} \kappa_\gamma$, in view of the analysis carried out in Section 4, it is straightforward to see that $\|D_\nu\| \overset{a.s.}{\to} 0$. Therefore, $\alpha_\nu \overset{a.s.}{\to} 0$ and $\beta_\nu \overset{a.s.}{\to} 0$ from (33). However, it is possible to get a much better bound for $\beta_\nu$.

Define $V^{(i,\nu)} := T^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-i} T - (1/n) trace(\mathcal{M}(\rho_\nu I - \mathcal{M})^{-i}) I$ for $i = 1, 2$. Expand $D_\nu$ up to second order around $\rho_\nu$, and observe that $\mathcal{R}_\nu \Delta \mathbf{e}_\nu = 0$ for *any* diagonal matrix $\Delta$. From this, and (17), it follows that

$$\mathcal{R}_\nu D_\nu \mathbf{e}_\nu = \mathcal{R}_\nu (K(\rho_\nu) - \frac{\rho_\nu}{\ell_\nu}\Lambda)\mathbf{e}_\nu + (\rho_\nu - \widehat{\ell}_\nu)\mathcal{R}_\nu \overline{K}(\rho_\nu)\mathbf{e}_\nu + (\widehat{\ell}_\nu - \rho_\nu)^2 \overline{r}_\nu, \tag{36}$$

where $\overline{K}(\rho_\nu) = \Lambda^{1/2} V^{(2,\nu)} \Lambda^{1/2}$; and $\|\overline{r}_\nu\| = O_P(1)$. Note that (i) all terms, except those on the diagonal, of the matrix $\Lambda^{1/2} T^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-i} T \Lambda^{1/2} \mathbf{e}_\nu$ are $O_P(n^{-1/2})$ for $i = 1, 2$ (from an inequality similar to (41)), (ii) all except the diagonal of $\mathbf{S}_{AA}$ are $O_P(n^{-1/2})$, and (iii) $\mathcal{R}_\nu$ is diagonal with $(\nu, \nu)$th entry equal to 0. It now follows easily that

$$\beta_\nu = O_P(n^{-\frac{1}{2}}) + (\widehat{\ell}_\nu - \rho_\nu)^2 O_P(1). \tag{37}$$

### 5.1. Explanation for expansion (22)

The RHS of (21) can be written as

$$\mathbf{e}_\nu^T K(\widehat{\ell}_\nu)\mathbf{e}_\nu + 2\mathbf{e}_\nu^T K(\widehat{\ell}_\nu)(a_\nu - \mathbf{e}_\nu) + (a_\nu - \mathbf{e}_\nu)^T K(\widehat{\ell}_\nu)(a_\nu - \mathbf{e}_\nu). \tag{38}$$

The first term in (38) is the major component of (22), and it can be written as

$$\begin{aligned} &s_{\nu\nu} + \ell_\nu t_\nu^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-1} t_\nu + (\rho_\nu - \widehat{\ell}_\nu)\ell_\nu t_\nu^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-2} t_\nu \\ &\quad + (\rho_\nu - \widehat{\ell}_\nu)^2 \ell_\nu t_\nu^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-2}(\widehat{\ell}_\nu I - \mathcal{M})^{-1} t_\nu. \end{aligned}$$

Again, from (32), (33), (34) and (37),

$$\begin{aligned} (a_\nu - \mathbf{e}_\nu)^T K(\widehat{\ell}_\nu)(a_\nu - \mathbf{e}_\nu) &= \|a_\nu - \mathbf{e}_\nu\|^2 O_P(1) = \beta_\nu^2\, O_P(1) \\ &= O_P(n^{-1}) + (\widehat{\ell}_\nu - \rho_\nu)^2 O_P(n^{-\frac{1}{2}}) + (\widehat{\ell}_\nu - \rho_\nu)^4 O_P(1). \end{aligned}$$

To check the negligibility of the second term in (38), observe that by (32),

$$\begin{aligned} \mathbf{e}_\nu^T K(\widehat{\ell}_\nu)(a_\nu - \mathbf{e}_\nu) &= -\mathbf{e}_\nu^T D_\nu \mathcal{R}_\nu D_\nu \mathbf{e}_\nu + \mathbf{e}_\nu^T K(\widehat{\ell}_\nu) r_\nu \\ &= -\mathbf{e}_\nu^T D_\nu \mathcal{R}_\nu D_\nu \mathbf{e}_\nu + o_P(n^{-\frac{1}{2}}) + (\widehat{\ell}_\nu - \rho_\nu)^2 o_P(1), \end{aligned}$$

where the last equality is due to (34), (37) and $\alpha_\nu = o_P(1)$. Use (36), and the definition of $\mathcal{R}_\nu$, to get the expression

$$\begin{aligned} &\mathbf{e}_\nu^T D_\nu \mathcal{R}_\nu D_\nu \mathbf{e}_\nu \\ &= \sum_{j \neq \nu}^{M} \frac{\ell_\nu}{\rho_\nu} \left(\frac{\ell_j \ell_\nu}{\ell_j - \ell_\nu}\right) \left[\frac{s_{j\nu}}{\sqrt{\ell_j \ell_\nu}} + V_{j\nu}^{(1,\nu)} + (\rho_\nu - \widehat{\ell}_\nu) V_{j\nu}^{(2,\nu)} + (\rho_\nu - \widehat{\ell}_\nu)^2 \widetilde{V}_{j\nu}^{(3,\nu)}\right]^2, \end{aligned}$$

where $\widetilde{V}^{(3,\nu)} = T^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-2}(\widehat{\ell}_\nu I - \mathcal{M})^{-1} T$. Observe that for $j \neq \nu$, each of the terms $s_{j\nu}$, $V_{j\nu}^{(1,\nu)}$ and $V_{j\nu}^{(2,\nu)}$ is $O_P(n^{-1/2})$ and $\widetilde{V}_{j\nu}^{(3,\nu)} = O_P(1)$. It follows that

$$\mathbf{e}_\nu^T D_\nu \mathcal{R}_\nu D_\nu \mathbf{e}_\nu = O_P(n^{-1}) + (\widehat{\ell}_\nu - \rho_\nu)^2 O_P(n^{-1/2}) + (\widehat{\ell}_\nu - \rho_\nu)^4 O_P(1).$$

### 5.2. Proof of Theorem 4

**Part (a).** As a convention, choose $\langle \mathbf{p}_\nu, \widetilde{\mathbf{e}}_\nu \rangle \geq 0$. First, note that with $p_{A,\nu}$ as in (10), $\langle \mathbf{p}_\nu, \widetilde{\mathbf{e}}_\nu \rangle = \langle p_{A,\nu}, \mathbf{e}_\nu \rangle = \sqrt{1 - R_\nu^2}\, \langle a_\nu, \mathbf{e}_\nu \rangle$. Since $\beta_\nu \overset{a.s.}{\to} 0$ and $\alpha_\nu \overset{a.s.}{\to} 0$, from (32) and (34), it follows that $\langle a_\nu, \mathbf{e}_\nu \rangle \overset{a.s.}{\to} 1$. Therefore, from (18), (24), Theorem 2 and the above display,

$$\frac{1}{1 - R_\nu^2} \overset{a.s.}{\to} 1 + \ell_\nu \gamma \int \frac{x}{(\rho_\nu - x)^2} dF_\gamma(x),$$

from which (6) follows in view of Lemma B.2.

**Part (b).** From (18), it is clear that for (7) to hold, either $a_\nu^T \Lambda^{1/2} T^T \mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-2} T \Lambda^{1/2} a_\nu \overset{a.s.}{\to} \infty$, or $\langle a_\nu, \mathbf{e}_\nu \rangle \overset{a.s.}{\to} 0$. Hence, it suffices to show that the smallest eigenvalue of the matrix $E := T^T \mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-2} T$ diverges to infinity almost surely. The approach is to show that given $\epsilon > 0$, there is a $C_\epsilon > 0$ such that the probability $\mathbb{P}(\lambda_{min}(E) \le C_\epsilon)$ is summable over $n$, and that $C_\epsilon \to \infty$ as $\epsilon \to 0$.

Denote the rows of $T$ by $\mathbf{t}_j^T$, $j = 1, \ldots, N - M$ (treated as an $1 \times M$ vector); $\mathbf{t}_j$'s are to be distinguished from the vectors $t_1, \ldots, t_M$, the columns of $T$. In fact, $\mathbf{t}_j^T = (t_{j1}, \ldots, t_{jM})$. Then

$$E = \sum_{j=1}^{N-M} \frac{\mu_j}{(\widehat{\ell}_\nu - \mu_j)^2}\mathbf{t}_j\mathbf{t}_j^T \ge \sum_{j=\nu}^{N-M} \frac{\mu_j}{(\widehat{\ell}_\nu - \mu_j)^2}\mathbf{t}_j\mathbf{t}_j^T =: E_\nu, \quad \text{say,}$$

in the sense of inequalities between positive semi-definite matrices. Thus $\lambda_{min}(E) \ge \lambda_{min}(E_\nu)$. Then on the set $\overline{J}_{1,\nu} := \{\widehat{\ell}_\nu < \kappa_\gamma + \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2\}$,

$$E_\nu \ge \sum_{j=\nu}^{N-M} \frac{\mu_j}{(\kappa_\gamma + \epsilon - \mu_j)^2}\mathbf{t}_j\mathbf{t}_j^T =: \overline{E}_\nu$$

since $\widehat{\ell}_\nu \ge \mu_\nu$, by the interlacing inequality of eigenvalues of symmetric matrices (Rao (1973, Sec. 1f)). Thus, in view of Proposition 1, we need only provide a lower bound for the smallest eigenvalue of $\overline{E}_\nu$. However, it is more convenient to work with the matrix

$$\overline{E} = \sum_{j=1}^{N-M} \frac{\mu_j}{(\kappa_\gamma + \epsilon - \mu_j)^2}\mathbf{t}_j\mathbf{t}_j^T = T^T\mathcal{M}((\kappa_\gamma + \epsilon)I - \mathcal{M})^{-2}T. \tag{39}$$

Proving summability of $\mathbb{P}(\lambda_{min}(\overline{E}) \le C_\epsilon, \overline{J}_{1,\nu})$ (where $C_\epsilon \to \infty$ as $\epsilon \to 0$) suffices, because it is easy to check that $\mathbb{P}(\|\overline{E}_\nu - \overline{E}\| > \epsilon, \overline{J}_{1,\nu})$ is summable.

By Proposition 2, given $0 < \delta < [16(\kappa_\gamma + \epsilon/2)/\epsilon^2][(N - M)/n]$, there is an $n_*(\delta, \epsilon, \gamma)$ such that, for all $j = 1, \ldots, M$, for all $n \ge n_*(\delta, \epsilon, \gamma)$,

$$\mathbb{P}(|t_j^T\mathcal{M}((\kappa_\gamma+\epsilon)I-\mathcal{M})^{-2}t_j-\gamma\int\frac{x}{(\kappa_\gamma+\epsilon-x)^2}dF_\gamma(x)|>\delta, \overline{J}_{1,\nu})\le\varepsilon_1(n), \tag{40}$$

where $\varepsilon_1(n)$ is summable in $n$. On the other hand, since $\sqrt{n}t_j \sim N(0, I_{N-M})$ for $j = 1, \ldots, M$, and on $J_\gamma(\epsilon/2)$

$$\|\mathcal{M}((\kappa_\gamma + \epsilon)I - \mathcal{M})^{-2}\| = \frac{\mu_1}{(\kappa_\gamma + \epsilon - \mu_1)^2} \le \frac{\kappa_\gamma + \frac{\epsilon}{2}}{(\frac{\epsilon}{2})^2} \ ,$$

Lemma A.1 implies that, for $j \ne k$ and $0 < \delta < [2(\kappa_\gamma + \epsilon/2)/\epsilon^2][(N - M)/n]$,

$$\mathbb{P}(|t_j^T\mathcal{M}((\kappa_\gamma+\epsilon)I-\mathcal{M})^{-2}t_k|>\delta,\ \overline{J}_{1,\nu})\le 2\exp\left(-\frac{n}{N-M}\frac{n\delta^2(\frac{\epsilon}{2})^4}{3(\kappa_\gamma+\frac{\epsilon}{2})^2}\right). \tag{41}$$

Denote the RHS of (41) by $\varepsilon_2(n)$. Notice that $|\lambda_{min}(\overline{E}) - \lambda_{min}(D_{\overline{E}})| \leq \|\overline{E} - D_{\overline{E}}\|_{HS}$, where $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm and $D_{\overline{E}}$ is the diagonal of $\overline{E}$. Hence, from (40) and (41), with $\delta_M = \delta(1 + \sqrt{M(M-1)})$,

$$\mathbb{P}(\lambda_{min}(\overline{E}) \leq \gamma \int \frac{x}{(\kappa_\gamma + \epsilon - x)^2} dF_\gamma(x) - \delta_M, \overline{J}_{1,\nu}) \leq M\varepsilon_1(n) + \frac{M(M-1)}{2}\varepsilon_2(n) \tag{42}$$

for $0 < \delta < [2(\kappa_\gamma + \epsilon/2)/\epsilon^2][(N-M)/n]$, and for all $n \geq n_*(\delta, \epsilon, \gamma)$.

If $0 < \epsilon < 2\gamma$, then it can be checked that $\int x/(\kappa_\gamma + \epsilon - x)^2 dF_\gamma(x) > (1/16\sqrt{\gamma}\pi)(1/\sqrt{\epsilon})$. Therefore, set $\delta = \epsilon(1 + \sqrt{M(M-1)})^{-1}$ and choose $\epsilon$ small enough so that $(\sqrt{\gamma}/16\pi)(1/\sqrt{\epsilon}) - \epsilon > 0$. Call the last quantity $C_\epsilon$, and observe that $C_\epsilon \to \infty$ as $\epsilon \to 0$. By (42), the result follows.

### 5.3. Proof of Theorem 5

Use the fact (due to Theorem 3) that $\widehat{\ell}_\nu - \rho_\nu = O_P(n^{-1/2})$ and equations (32), (34) and (37) to get

$$\sqrt{n}(a_\nu - \mathbf{e}_\nu) = -\sqrt{n}\mathcal{R}_\nu D_\nu \mathbf{e}_\nu + o_P(1). \tag{43}$$

Since $\widehat{\ell}_\nu - \rho_\nu = O_P(n^{-1/2})$, and since the off-diagonal elements of the matrix $V^{(2,\nu)}$ are $O_P(n^{-1/2})$, from (36),

$$\sqrt{n}(a_\nu - \mathbf{e}_\nu) = -\sum_{k\neq\nu}^{M} \frac{\ell_\nu}{\rho_\nu} \frac{\sqrt{\ell_k\ell_\nu}}{(\ell_k - \ell_\nu)} \left[\frac{\sqrt{n}s_{k\nu}}{\sqrt{\ell_k\ell_\nu}} + \sqrt{n}t_k^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-1} t_\nu\right] \mathbf{e}_k + o_P(1). \tag{44}$$

Now, to complete the proof use the same trick as in the proof of Theorem 3. Using the same notation write $s_{k\nu} = \sqrt{\ell_k\ell_\nu}[t_k^T t_\nu + w_k^T w_\nu]$, where for $1 \leq k \leq M$, $w_k \overset{i.i.d.}{\sim} N(0, (1/n)I_{n-N+M})$, $t_k \overset{i.i.d.}{\sim} N(0, (1/n)I_{N-M})$, and $\{w_k\}$ and $\{t_k\}$ are mutually independent, and independent of $\mathbf{Z}_B$.

The asymptotic normality of $\sqrt{n}(a_\nu - \mathbf{e}_\nu)$ can then be established by proving the asymptotic normality of $W_{3,n} := \sum_{k\neq\nu}^{M}(\ell_\nu/\rho_\nu)[\sqrt{\ell_k\ell_\nu}/(\ell_k - \ell_\nu)]\sqrt{n}w_k^T w_\nu \mathbf{e}_k$ and $W_{4,n} := \sum_{k\neq\nu}^{M}(\ell_\nu/\rho_\nu)[\sqrt{\ell_k\ell_\nu}/(\ell_k - \ell_\nu)]\sqrt{n}t_k^T(I + \mathcal{M}(\rho_\nu I - \mathcal{M})^{-1})t_\nu \mathbf{e}_k$ separately. It can be shown that,

$$W_{3,n} \Longrightarrow N_M(0, (1-\gamma)\frac{\ell_\nu^2}{\rho_\nu^2}\sum_{k\neq\nu}^{M} \frac{\ell_k\ell_\nu}{(\ell_k - \ell_\nu)^2}\mathbf{e}_k\mathbf{e}_k^T).$$

To prove the asymptotic normality of $W_{4,n}$, fix an arbitrary vector $\mathbf{b} \in \mathbb{R}^M$, and consider the characteristic function of $\mathbf{b}^T W_{4,n}$, conditional on $\mathcal{M}$. As in the proof of Theorem 3, it is enough that, for some pre-specified $\delta > 0$,

$$\mathbb{E}\left[\left|\mathbb{E}(e^{i\mathbf{a}^T W_{4,n}}|\mathcal{M})\exp\left(\frac{1}{2}\mathbf{b}^T\widetilde{\Sigma}_\nu(\ell_\nu)\mathbf{b}\right)-1\right|, \mu_1 \le \kappa_\gamma+\delta\right] \to 0 \text{ as } n\to\infty, \quad (45)$$

where

$$\begin{aligned}\widetilde{\Sigma}_\nu(\ell_\nu) &= \Sigma_\nu(\ell_\nu) - (1-\gamma)\frac{\ell_\nu^2}{\rho_\nu^2}\sum_{1\le k\ne\nu\le M}\frac{\ell_k\ell_\nu}{(\ell_k-\ell_\nu)^2}\mathbf{e}_k\mathbf{e}_k^T \\ &= \gamma\ell_\nu^2\int\frac{1}{(\rho_\nu-x)^2}dF_\gamma(x)\sum_{1\le k\ne\nu\le M}\frac{\ell_k\ell_\nu}{(\ell_k-\ell_\nu)^2}\mathbf{e}_k\mathbf{e}_k^T.\end{aligned}$$

Define $J_\gamma(\delta) := \{\mu_1 \le \kappa_\gamma + \delta\}$ where $\delta$ is small enough so that $\rho_\nu > \kappa_\gamma + 2\delta$. Define

$$\mathbf{C}_\nu = \ell_\nu^2\sum_{1\le k\ne\nu\le M}\frac{\ell_k\ell_\nu}{(\ell_k-\ell_\nu)^2}\mathbf{e}_k\mathbf{e}_k^T.$$

Choose $n$ large enough so that $\mathbf{b}^T\mathbf{C}_\nu\mathbf{b} < n(\rho_\nu - \kappa_\gamma - \delta)^2$. Then observe that on $J_\gamma(\delta)$,

$$\begin{aligned}\mathbb{E}(e^{i\mathbf{b}^T W_{4,n}}|\mathcal{M}) &= \mathbb{E}[\mathbb{E}(e^{i\mathbf{b}^T W_{4,n}}|t_\nu,\mathcal{M})|\mathcal{M}] \\ &= \mathbb{E}\left[\exp\left(-\frac{1}{2}\mathbf{b}^T\mathbf{C}_\nu\mathbf{b}t_\nu^T(\rho_\nu I-\mathcal{M})^{-2}t_\nu\right)|\mathcal{M}\right] \\ &= \prod_{j=1}^{N-M}\left(1+\frac{\mathbf{b}^T\mathbf{C}_\nu\mathbf{b}}{n(\rho_\nu-\mu_j)^2}\right)^{-\frac{1}{2}}, \qquad (46)\end{aligned}$$

where the first and the last steps owe to the fact that the $t_k \sim N(0, I_{N-M}/n)$ and are independent for $k = 1, \ldots, M$. The rest of the proof imitates the arguments used in the proof of Theorem 3 and is omitted.

### 5.4. Proof of Theorem 6

An invariance approach is taken to prove the result. Use the notation of Section 3. Write $p_{B,\nu}$ as $p_{B,\nu} = p_{B,\nu}(\Lambda, \mathbf{Z}_A, \mathbf{Z}_B)$. That is, treat $p_{B,\nu}$ as a map from $\mathbb{R}^M \times \mathbb{R}^{M\times n} \times \mathbb{R}^{(N-M)\times n}$ into $\mathbb{R}^{N-M}$ that maps $(\Lambda, \mathbf{Z}_A, \mathbf{Z}_B)$ to the vector which is the $B$-subvector of the $\nu$th eigenvector of $\mathbf{S}$. Also, recall that $R_\nu = \|p_{B,\nu}\|$.

Denote the class of $(N-M)\times(N-M)$ orthogonal matrices by $\mathcal{O}_{N-M}$. From the decomposition of $\mathbf{S}$ in Section 3, it follows that, for $G \in \mathcal{O}_{N-M}$ arbitrary, $Gp_{B,\nu} = p_{B,\nu}(\Lambda, \mathbf{Z}_A, G\mathbf{Z}_B)$, i.e., for every fixed $\mathbf{Z}_A = z_A$, as a function of $\mathbf{Z}_B$,

$$Gp_{B,\nu}(\Lambda, z_A, \mathbf{Z}_B) = p_{B,\nu}(\Lambda, z_A, G\mathbf{Z}_B). \qquad (47)$$

Define, for $r > 0$, $\mathcal{A}_r := \mathcal{A}_r(z_A) = \{z_B \in \mathbb{R}^{(N-M)\times n} : \|p_{B,\nu}(\Lambda, z_A, z_B)\| = r\}$. Then (47) implies that

$$G\mathcal{A}_r = \mathcal{A}_r \quad \text{for all } G \in \mathcal{O}_{N-M}, \text{ for all } r > 0. \tag{48}$$

Thus, for any $0 < r_1 < r_2 < 1$, the set

$$\{\|p_{B,\nu}(\Lambda, z_A, z_B)\| \in [r_1, r_2)\} = \bigcup_{r_1 \le r < r_2} \mathcal{A}_r(z_A),$$

is invariant under rotation on the left. Note that $\bigcup_{r_1 \le r < r_2} \mathcal{A}_r(z_A)$ is the $z_A$-section of the set $\{R_\nu \in [r_1, r_2)\}$ and hence is measurable (w.r.t. the $\sigma$-algebra generated by $\mathbf{Z}_B$).

From this it follows that for every fixed $\mathbf{Z}_A = z_A$, and every Borel measurable subset $H$ of the unit sphere $\mathbb{S}^{N-M-1}$, for all $G \in \mathcal{O}_{N-M}$ it is meaningful to write

$$\begin{aligned}
&\mathbb{P}(R_\nu \in [r_1, r_2), G\frac{p_{B,\nu}(\Lambda, z_A, \mathbf{Z}_B)}{R_\nu} \in H | \mathbf{Z}_A = z_A) \\
&= \mathbb{P}(\mathbf{Z}_B \in \bigcup_{r_1 \le r < r_2} \mathcal{A}_r(z_A), G\frac{p_{B,\nu}(\Lambda, z_A, \mathbf{Z}_B)}{\|p_{B,\nu}(\Lambda, z_A, \mathbf{Z}_B)\|} \in H | \mathbf{Z}_A = z_A) \\
&= \mathbb{P}(G\mathbf{Z}_B \in \bigcup_{r_1 \le r < r_2} \mathcal{A}_r(z_A), \frac{p_{B,\nu}(\Lambda, z_A, G\mathbf{Z}_B)}{\|p_{B,\nu}(\Lambda, z_A, G\mathbf{Z}_B)\|} \in H | \mathbf{Z}_A = z_A) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(by (47) and (48))} \\
&= \mathbb{P}(\mathbf{Z}_B \in \bigcup_{r_1 \le r < r_2} \mathcal{A}_r(z_A), \frac{p_{B,\nu}(\Lambda, z_A, \mathbf{Z}_B)}{\|p_{B,\nu}(\Lambda, z_A, \mathbf{Z}_B)\|} \in H | \mathbf{Z}_A = z_A) \\
&\qquad\qquad\qquad \text{(since } G\mathbf{Z}_B \stackrel{\mathcal{L}}{=} \mathbf{Z}_B, \text{ independent of } \mathbf{Z}_A) \\
&= \mathbb{P}(R_\nu \in [r_1, r_2), \frac{p_{B,\nu}(\Lambda, z_A, \mathbf{Z}_B)}{R_\nu} \in H | \mathbf{Z}_A = z_A).
\end{aligned}$$

From the equality of the first and the last terms, by standard arguments we conclude that for all Borel subset $H$ of $\mathbb{S}^{N-M-1}$,

$$\mathbb{P}(G\frac{p_{B,\nu}}{R_\nu} \in H | R_\nu, \mathbf{Z}_A) = \mathbb{P}(\frac{p_{B,\nu}}{R_\nu} \in H | R_\nu, \mathbf{Z}_A)$$

for all $G \in \mathcal{O}_{N-M}$. This rotational invariance means that given $R_\nu$ and $\mathbf{Z}_A$, the conditional distribution of $q_\nu = p_{B,\nu}/R_\nu$ is uniform on $\mathbb{S}^{N-M-1}$. Moreover, since the conditional distribution of $q_\nu$ does not depend on $(R_\nu, \mathbf{Z}_A)$, $q_\nu$ and $(R_\nu, \mathbf{Z}_A)$ are independent. In particular, the marginal distribution of $q_\nu$ is uniform on $\mathbb{S}^{N-M-1}$.

## Acknowledgement

The author is supported in parts by the NIH grant #R01 EB1988-09. The author thanks Professor Iain Johnstone for some thoughtful discussions, and also for pointing to two important references; Professor Richard Olshen for some constructive suggestions regarding the presentation of the material; and Dr. Jinho Baik for some useful communications. He also thanks two anonymous referees for their suggestions that improved the presentation.

## Appendix A

### A.1. Weak concentration inequalities for random quadratic forms

The following two lemmas are referred to as weak concentration inequalities.

Suppose that $C : \mathcal{X} \to \mathbb{R}^{n\times n}$ is a measurable function. Let $Z$ be a random variable taking values in $\mathcal{X}$. Let $\|C\|$ denote the operator norm of $C$, i.e., the largest singular value of $K$.

**Lemma A.1.** *Suppose that $X$ and $Y$ are i.i.d. $N_n(0, I)$ independent of $Z$. Then for every $L > 0$ and $0 < \delta < 1$, for all $0 < t < \delta/(1-\delta)L$,*

$$\mathbb{P}\Big(\frac{1}{n}|X^T C(Z)Y| > t, \|C(Z)\| \leq L\Big) \leq 2\exp\left(-\frac{(1-\delta)nt^2}{2L^2}\right). \tag{49}$$

**Lemma A.2.** *Suppose that $X$ is distributed as $N_n(0, I)$ independent of $Z$. Also let $C(z) = C^T(z)$ for all $z \in \mathcal{X}$. Then for every $L > 0$ and $0 < \delta < 1$, for all $0 < t < 2\delta/(1-\delta)L$,*

$$\mathbb{P}(\frac{1}{n}|X^T C(Z)X - trace(C(Z))| > t, \|C(Z)\| \leq L) \leq 2\exp\left(-\frac{(1-\delta)nt^2}{4L^2}\right). \tag{50}$$

The proofs involve standard arguments and are omitted.

### A.2. Proof of Proposition 1

In order to prove this result the following result, a part of Theorem 2.13 of Davidson and Szarek (2001), is used.

**Lemma A.3.** *Let $Z$ be a $p \times q$ matrix of i.i.d. $N(0, 1/q)$ entries with $p \leq q$. Let $s_{\max}(Z)$ and $s_{\min}(Z)$ denote the largest and the smallest singular value of $Z$, respectively. Then*

$$\mathbb{P}(s_{\max}(Z) > 1 + \sqrt{\frac{p}{q}} + t) \leq e^{-\frac{qt^2}{2}}, \tag{51}$$

$$\mathbb{P}(s_{\min}(Z) < 1 - \sqrt{\frac{p}{q}} - t) \leq e^{-\frac{qt^2}{2}}. \tag{52}$$

Take $p = N - M$, $q = n$ and $Z = \mathbf{Z}_B/\sqrt{n}$. Note that $\sqrt{\mu_1} = s_{\max}\mathbf{Z}_B/\sqrt{n}$. Let $m_1 := (1 + \sqrt{(N-M)/n})^2$. For any $r > 0$, if $\sqrt{\mu_1} - \sqrt{m_1} \leq r$, then

$\mu_1 - m_1 \le (\sqrt{\mu_1} + \sqrt{m_1}) \max\{\sqrt{\mu_1} - \sqrt{m_1}, 0\} \le (2\sqrt{m_1} + r)r$ . This implies, by (51), that

$$e^{-\frac{nr^2}{2}} \ge \mathbb{P}(s_{\max}(\frac{1}{\sqrt{n}}\mathbf{Z}_B) > \sqrt{m_1} + r) \ge \mathbb{P}(\mu_1 - m_1 > r(2\sqrt{m_1} + r)).$$

Set $t = r(2\sqrt{m_1} + r) = (r + \sqrt{m_1})^2 - m_1$. Solving for $r$ one gets that, for $t > 0$, $r = \sqrt{t + m_1} - \sqrt{m_1}$. Substitute in the last display to get, for $t > 0$,

$$\mathbb{P}(\mu_1 - m_1 > t) \le \exp\left[-\frac{n}{2}(\sqrt{t + m_1} - \sqrt{m_1})^2\right]. \tag{53}$$

Now, to complete the proof, observe that

$$\sqrt{t + m_1} - \sqrt{m_1} = \frac{t}{\sqrt{t + m_1} + \sqrt{m_1}} > \frac{t}{2\sqrt{t + m_1}} \ge \frac{t}{2\sqrt{t + \kappa_\gamma + \delta/4}}$$

for $n \ge n_0(\gamma, \delta)$. Since $\delta < \kappa_\gamma/2$ implies that $\kappa_\gamma + \delta < 3\kappa_\gamma/2$, for $n \ge n_0(\gamma, \delta)$, one has $\sqrt{3\delta/4 + m_1} - \sqrt{m_1} > \sqrt{6}\delta/(8\sqrt{\kappa_\gamma})$. The result follows if $t = 3\delta/4$ is substituted in (53), since $|m_1 - \kappa_\gamma| \le \delta/4$.

### A.3. A concentration inequality for Lipschitz functionals

We restate Corollary 1.8(b) of Guionnet and Zeitouni (2000) in our context.

**Lemma A.4.** *Suppose that $\mathbf{Y}$ is an $m \times n$ matrix, $m \le n$, with independent (real or complex) entries $Y_{kl}$ with law $P_{kl}$, $1 \le k \le m, 1 \le l \le n$. Let $\mathbf{S}_\Delta = \mathbf{Y}\Delta\mathbf{Y}^*$ be a generalized Wishart matrix, where $\Delta$ is a diagonal matrix with real, nonnegative diagonal entries and spectral radius $\phi_\Delta > 0$. Suppose that the family $\{P_{kl} : 1 \le k \le m, 1 \le l \le n\}$ satisfies the logarithmic Sobolev inequality with uniformly bounded constant $c$. Then for any function $f$ such that $g(x) := f(x^2)$ is Lipschitz, for any $\delta > 0$,*

$$\mathbb{P}\left(\Big|\frac{1}{m}\mathit{trace}\ f(\frac{1}{m+n}\mathbf{S}_\Delta) - \mathbb{E}(\frac{1}{m}\mathit{trace}\ f(\frac{1}{m+n}\mathbf{S}_\Delta))\Big| > \delta\right) \le 2\exp\left(-\frac{m^2\delta^2}{2c\phi_\Delta|g|_{\mathcal{L}}^2}\right), \tag{54}$$

*where $|g|_{\mathcal{L}}$ is the Lipschitz norm of $g$.*

Define

$$G_k(x;\rho,\gamma,\epsilon) = \begin{cases} \frac{1}{(\rho-x)^k} & x \le \kappa_\gamma + \epsilon \\ \frac{1}{(\rho-\kappa_\gamma-\epsilon)^k} & x > \kappa_\gamma + \epsilon \end{cases}, \quad \text{where } \rho > \kappa_\gamma + \epsilon,\ k = 1, 2, \ldots. \tag{55}$$

Then define $f_k(x) = G_k(x;\rho,\gamma,\epsilon)$, $k = 1, 2$, $g_k(x) = f_k(x^2)$, and notice that $g_k(x)$ is Lipschitz with $|g_k|_{\mathcal{L}} = [2k(\kappa_\gamma + \epsilon)^{1/2}]/[(\rho - \kappa_\gamma - \epsilon)^{k+1}]$. Take $m = N - M$,

$\mathbf{Y} = \mathbf{Z}_B$ and $\Delta = [(m+n)/n]I_n$, and recall that $N(0,1)$ satisfies logarithmic Sobolev inequality with constant $c = 1$ (Bogachev (1998, Thm. 1.6.1)). Further, $\phi_\Delta = (m+n)/n$ and $\mathbf{S}_\Delta = (m+n)\mathbf{S}_{BB}$. Therefore Lemma A.4 implies the following.

**Proposition A.1.** *For $k = 1, 2$, and any $\delta > 0$,*

$$\mathbb{P}\left(\Big|\frac{1}{n} trace\ G_k(\mathbf{S}_{BB};\rho,\gamma,\epsilon) - \mathbb{E}(\frac{1}{n} trace\ G_k(\mathbf{S}_{BB};\rho,\gamma,\epsilon))\Big| > \delta\right) \leq 2\exp\left(-\frac{n}{n+N-M}\frac{n^2\delta^2}{2}\frac{(\rho-\kappa_\gamma-\epsilon)^{2(k+1)}}{4k^2(\kappa_\gamma+\epsilon)}\right). \tag{56}$$

## Appendix B

### B.1. Expression for $\rho_\nu$

**Lemma B.1.** *For $\ell_\nu \geq 1 + \sqrt{\gamma}$, $\rho_\nu = \ell_\nu\,(1 + [\gamma/(\ell_\nu - 1)])$ solves* (3).

**Proof.** First step is to show that for any $\ell > 1 + \sqrt{\gamma}$,

$$\int \frac{x}{\rho(\ell) - x} f_\gamma(x)dx = \frac{1}{\ell - 1}, \tag{57}$$

where $\rho(\ell) = \ell\,(1 + [\gamma/(\ell - 1)])$, and $f_\gamma(x)$ is the density of Marčenko-Pastur law with parameter $\gamma(\leq 1)$. This is obtained by direct computation. The case $\ell = 1 + \sqrt{\gamma}$ follows from this and the Monotone Convergence Theorem.

The following result is easily obtained from Lemma B.1.

**Lemma B.2.** *For $\ell > 1 + \sqrt{\gamma}$,*

$$\int \frac{x}{(\rho(\ell) - x)^2} dF_\gamma(x) = \frac{1}{(\ell - 1)^2 - \gamma}\ . \tag{58}$$

### B.2. Proof of Proposition 2

First, consider the expansion

$$\begin{aligned} &t_j^T \mathcal{M}(\rho I - \mathcal{M})^{-2} t_j - \gamma \int \frac{x}{(\rho - x)^2} dF_\gamma(x) \\ &= \left[t_j^T \mathcal{M}(\rho I - \mathcal{M})^{-2} t_j - \frac{1}{n}\ trace(\mathcal{M}(\rho I - \mathcal{M})^{-2})\right] \\ &\quad + \left[\frac{1}{n}\ trace(\mathcal{M}(\rho I - \mathcal{M})^{-2}) - \gamma \int \frac{x}{(\rho - x)^2} dF_\gamma(x)\right]. \end{aligned}$$

Split the second term further as

$$\rho\left[\frac{1}{n}\ trace((\rho I-\mathcal{M})^{-2})-\gamma\int\frac{1}{(\rho-x)^2}dF_\gamma(x)\right]$$
$$-\left[\frac{1}{n}\ trace((\rho I-\mathcal{M})^{-1})-\gamma\int\frac{1}{\rho-x}dF_\gamma(x)\right]$$
$$=(I)-(II),\quad \text{say.}$$

Define $J_\gamma(\epsilon/2):=\{\mu_1\le\kappa_\gamma+\epsilon/2\}$. In order to bound the first term, notice that $\sqrt{n}t_j\sim N(0,I_{N-M})$ for $j=1,\ldots,M$, and on $J_\gamma(\epsilon/2)$,

$$\|\mathcal{M}(\rho I-\mathcal{M})^{-2}\|=\frac{\mu_1}{(\rho-\mu_1)^2}\le\frac{\kappa_\gamma+\frac{\epsilon}{2}}{(\rho-\kappa_\gamma-\frac{\epsilon}{2})^2}.$$

Thus, apply Lemma A.2 to conclude that (setting $\delta=1/3$ in the lemma),

$$\mathbb{P}\Big(\Big|t_j^T\mathcal{M}(\rho I-\mathcal{M})^{-2}t_j-\frac{1}{n}trace\ (\mathcal{M}(\rho I-\mathcal{M})^{-2})\Big|\ge\frac{\delta}{4},J_\gamma(\frac{\epsilon}{2})\Big)$$
$$\le 2\exp\left(-\frac{n}{N-M}\frac{(\rho-\kappa_\gamma-\frac{\epsilon}{2})^4n(\frac{\delta}{4})^2}{6(\kappa_\gamma+\frac{\epsilon}{2})^2}\right)\text{ for }0<\delta<\frac{4(\kappa_\gamma+\frac{\epsilon}{2})}{(\rho-\kappa_\gamma-\frac{\epsilon}{2})^2}\left(\frac{N-M}{n}\right).\tag{59}$$

To provide bounds for $(I)$, observe that on $J_\gamma(\epsilon/2)$, $trace\ ((\rho I-\mathcal{M})^{-2})=trace\ G_2(\mathbf{S}_{BB};\rho,\gamma,\epsilon/2)$, where $G_2(\cdot;\cdot,\cdot,\cdot)$ is defined through (55). Therefore, Proposition A.1 implies that

$$\mathbb{P}\Big(\rho\Big|\frac{1}{n}trace\ ((\rho I-\mathcal{M})^{-2})-\mathbb{E}(\frac{1}{n}trace\ G_2(\mathbf{S}_{BB};\rho,\gamma,\epsilon/2))\Big|>\frac{\delta}{4},\ J_\gamma(\frac{\epsilon}{2})\Big)$$
$$\le 2\exp\left(-\frac{n}{n+N-M}\frac{n^2(\frac{\delta}{4})^2}{2}\frac{(\rho-\kappa_\gamma-\frac{\epsilon}{2})^6}{16\rho^2(\kappa_\gamma+\frac{\epsilon}{2})}\right).\tag{60}$$

Analogously, use $G_1(\cdot;\cdot,\cdot,\cdot)$ instead of $G_2(\cdot;\cdot,\cdot,\cdot)$ in the analysis of $(II)$, to obtain

$$\mathbb{P}\Big(\Big|\frac{1}{n}trace\ ((\rho I-\mathcal{M})^{-1})-\mathbb{E}(\frac{1}{n}trace\ G_1(\mathbf{S}_{BB};\rho,\gamma,\epsilon/2))\Big|>\frac{\delta}{4},\ J_\gamma(\frac{\epsilon}{2})\Big)$$
$$\le 2\exp\left(-\frac{n}{n+N-M}\frac{n^2(\frac{\delta}{4})^2}{2}\frac{(\rho-\kappa_\gamma-\frac{\epsilon}{2})^4}{4(\kappa_\gamma+\frac{\epsilon}{2})}\right).\tag{61}$$

Take $\overline{F}_{n,N-M}$ to be the expected ESD of $\mathbf{S}_{BB}$. Now, to tackle the remainders in $(I)$ and $(II)$ notice that, since $G_1$ and $G_2$ are continuous and bounded in their first argument, for $k=1,2$,

$$\mathbb{E}\Big(\frac{1}{n}trace\ G_k(\mathbf{S}_{BB};\rho,\gamma,\frac{\epsilon}{2})\Big)=\frac{N-M}{n}\int G_k(x;\rho,\gamma,\frac{\epsilon}{2})d\overline{F}_{n,N-M}(x).$$

Bai (1993) proved under fairly weak conditions (Bai (1993, Thm. 3.2), that if $\theta_1 \le p/n \le \theta_2$ where $0 < \theta_1 < 1 < \theta_2 < \infty$, then

$$\|\overline{F}_{n,p} - F_{\frac{p}{n}}\|_\infty \le C_1(\theta_1, \theta_2) n^{-\frac{5}{48}}, \tag{62}$$

where $F_{p/n}$ denotes the Marčenko-Pastur law with parameter $p/n$. Here $\|\cdot\|_\infty$ means the sup-norm and $C_1$ is a constant which depends on $\theta_1, \theta_2$.

Since $F_{(N-M)/n} \Longrightarrow F_\gamma$, as $n \to \infty$, and $G_1(\cdot; \rho, \gamma, \epsilon/2)$ is bounded and continuous, and $\int G_1(x; \rho, \gamma, \epsilon/2) dF_\gamma(x) = \int (\rho - x)^{-1} dF_\gamma(x)$ from (62), it follows that there is $n_1(\rho, \delta, \epsilon, \gamma) \ge 1$ such that, for $n \ge n_1(\rho, \delta, \epsilon, \gamma)$,

$$\Big|\mathbb{E}\Big(\frac{1}{n} trace\ G_1(\mathbf{S}_{BB}; \rho, \gamma, \frac{\epsilon}{2})\Big) - \gamma \int \frac{1}{\rho - x} dF_\gamma(x)\Big| \le \frac{\delta}{8}. \tag{63}$$

Similarly, there is $n_2(\rho, \delta, \epsilon, \gamma) \ge n_1(\rho, \delta, \epsilon, \gamma)$ such that for $n \ge n_2(\rho, \delta, \epsilon, \gamma)$,

$$\rho\Big|\mathbb{E}\Big(\frac{1}{n} trace\ G_2(\mathbf{S}_{BB}; \rho, \gamma, \frac{\epsilon}{2})\Big) - \gamma \int \frac{1}{(\rho - x)^2} dF_\gamma(x)\Big| \le \frac{\delta}{8}. \tag{64}$$

Combine (59), (60), (61), (64) and (63), and the result follows.

## References


Anderson, T. W. (1963). Asymptotic theory of principal component analysis. *Ann. Math. Statist.* **34**, 122-148.

Bai, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices. *Ann. Probab.* **21**, 649-672.

Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statist. Sinica* **9**, 611-677.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrix. *Ann. Probab.* **32**, 553-605.

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.* **21**, 1275-1294.

Baik, J., Ben Arous, G. and Péché (2005). Phase transition of the largest eigenvalue for non-null complex covariance matrices. *Ann. Probab.* **33**, 1643-1697.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97**, 1382-1408. (Also arXiv:math.ST/048165 v1).

Bogachev, V. I. (1998). *Gaussian Measures.* American Mathematical Society.

Buja, A., Hastie, T. and Tibshirani, R. (1995). Penalized discriminant analysis. *Ann. Statist.* **23**, 73-102.

Davidson, K. R. and Szarek, S. (2001). Local operator theory, random matrices and Banach spaces. In *"Handbook on the Geometry of Banach spaces", (***V. 1**, Johnson, W. B., Lendenstrauss, J. eds.), 317-366, Elsevier Science.

Guionnet, A. and Zeitouni, O. (2000). Concentration of the spectral measure for large matrices. *Electronic Communications in Probability* **5**, 119-136.

Hoyle, D. and Rattray, M. (2003). Limiting form of the sample covariance eigenspectrum in PCA and kernel PCA. *Advances in Neural Information Processing Systems* (*NIPS* 16).

Hoyle, D. and Rattray, M. (2004). Principal component analysis eigenvalue spectra from data with symmetry breaking structure. *Phys. Rev. E* **69**, 026124.

Johnstone, I. M. (2001). On the distribution of the largest principal component. *Ann. Statist.* **29**, 295-327.

Johnstone, I. M. and Lu, A. Y. (2004). Sparse principal component analysis. (To appear in *J. Amer. Statist. Assoc.*).

Kneip, A. and Utikal, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96**, 519-542.

Kato, T. (1980). *Perturbation Theory of Linear Operators*. Springer-Verlag, New York.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley, New York.

Paul, D. (2004). Asymptotics of the leading sample eigenvalues for a spiked covariance model. Technical report, Stanford University.
http://anson.ucdavis.edu/∼debashis/techrep/eigenlimit.pdf

Péché, S. (2003). Universality of local eigenvalue statistics for random sample covariance matrices. Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer-Verlag, New York.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley, New York.

Roy, S. N. (1953). On a heuristic method of test construction and its use in Multivariate analysis. *Ann. Math. Statist.* **24**, 220-238.

Silverstein, J. W. (1990). Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Ann. Probab.* **18**, 1174-1194.

Silverstein, J. W. and Choi, S. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* **54**, 295-309.

Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.* **108**, 1033-1056. (Also arXiv:math.PR/0104113 v2).

Telatar, E. (1999). Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications* **10**, 585-595.

Tyler, D. E. (1983). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika* **63**, 639-645.

Department of Statistics, University of California, Davis, One Shields Avenue, Davis, CA 95616.

E-mail: debashis@wald.ucdavis.edu

(Received June 2005; accepted February 2006)