Asymptotics of the leading sample eigenvalues for a spiked covariance model

Debashis Paul

Department of Statistics, Stanford University

390 Serra Mall, Stanford, CA 94305

debashis@stat.stanford.edu

Abstract : We consider a multivariate Gaussian observation model where the covariance matrix is diagonal and the diagonal entries are all equal to one except for a finite number which are bigger. We address the question of asymptotic behaviour of the eigenvalues of the sample covariance matrix when the sample size and the dimension of the observations both grow to infinity in such a way that their ratio converges to a positive constant. We establish almost sure limits of the largest few sample eigenvalues. We also show that when a population eigenvalue is above a certain threshold and of multiplicity one, the corresponding sample eigenvalue has a Gaussian limiting distribution. We also demonstrate a phase transition phenomenon of the sample eigenvectors in the same setting.

Keywords : Principal component analysis, Eigenvalue distribution, Random matrix theory.

1 Introduction

Study of eigenvalues of sample covariance matrices has a long history. When the dimension N is fixed, the 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. In fixed dimension scenario much of the study of the eigenstructure of sample covariance matrix utilizes the fact that it is a good approximation of the population covariance matrix when sample size is large. However this is no longer the case when $\frac{N}{n} \to \gamma \in (0,\infty)$ as $n \to \infty$, where n is the sample size. Under these circustances 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 Marchenko-Pastur distribution, henceforth denoted by F_γ. When $\gamma \leq 1$, the support F_γ 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 identity, the largest and the smallest eigenvalues, when $\gamma \leq 1$, converge *almost surely* to the respective boundaries of the support of F_γ. Johnstone (2001) and Soshnikov (2002) have derived asymptotic distribution for largest, second largest etc sample eigenvalues under the same setting.

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 is when all except a few of the eigenvalues equal one and the few that are not are well-separated from the rest. This has been referred to as a "spiked population model" by Johnstone (2001). It has also been observed that for certain types of data e.g. in speech recognition (Buja *et al.*, 1995), wireless communication (Telatar, 1999), statistical learning (Hoyle and Rattray, 2003, 2004), a few of the sample eigenvalues have limiting behaviour that is different from the behaviour under identity covariance scenario. This paper attempts to contribute towards understanding these phenomena.

The literature on the asymptotics of sample eigenvalues for the non-identity covariance scenario is relatively recent. Silverstein and Choi (1995) derived *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), Baik, Ben Arous and Péché (2004) (henceforth Baik *et al.*, 2004). These authors deal with the situation when the observations are complex Gaussian and the covariance matrix is a finite rank perturbation of identity. When this paper was being written the author came to know about the work by Baik and Silverstein (2004), which studies 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 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 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.* (2004) show a $n^{2/3}$ scaling for the asymptotic distribution when a non-unit population eigenvalue lies below the threshold $1 + \sqrt{\gamma}$, and a $n^{1/2}$ scaling for those above that threshold.

In this paper we focus our attention on the case where we have independently and identically distributed real-valued observations $X_1, \ldots, X_n$ from an N-variate normal distribution with mean zero and covariance matrix $\Sigma = diag(\ell_1, \ell_2, \ldots, \ell_M, 1, \ldots, 1)$ where $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_M > 1$. We treat the $N \times n$ matrix $\mathbf{X} = (X_1 : \ldots : X_n)$ as a double array indexed by both n and $N = N(n)$ on the same probability space, such that $N/n \to \gamma$, where γ is a positive constant. Throughout we shall assume that $0 < \gamma < 1$ although much of the analysis can be extended to the case $\gamma \geq 1$ with a little extra work. Our aim is to study asymptotic behaviour of the large eigenvalues of the sample covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$ as $n \to \infty$. In this context we get the same almost sure limits for the M largest eigenvalues as those obtained by Baik and Silverstein (2004). However, while they derive these limits by studying the Stieltjes transform of the distribution which serves as the *almost sure* limit of the ESD, we rely on a matrix analysis approach and use properties of Gaussian distribution, including various concentration inequalities, as well as several known results about the limiting behaviour of the ESD for the null (or identity covariance) model. The advantage of this approach is that it gives a different perspective to the limits, in particular their identification as certain linear functionals of the limiting Marchenko-Pastur law when the true eigenvalue is above $1+\sqrt{\gamma}$. This analysis also allows us to derive distributional limits of the sample eigenvalues $\widehat{\ell}_\nu$ when $\ell_\nu > 1+\sqrt{\gamma}$. We do this only for the case when ℓ_ν has multiplicity one. A comprehensive study of all possible scenarios is beyond the scope of this paper. Another aspect of our approach is that it throws light on the behaviour of the eigenvectors associated with the M largest eigenvalues. We show that the sample eigenvectors also undergo a phase transition. We would like to emphasize that, even though our method is not suitable for analyzing the distributional limits for the case $\ell_\nu \leq 1+\sqrt{\gamma}$, it does afford a more probabilistic interpretation of the results in the other scenario, and may be applied to study similar problems in other contexts.

The results derived in this paper contain two important messages about the inferential aspect of dealing with large dimensional multivariate data. First, and most notably, the phase transition phenomenon described in this paper means that some commonly used tests for the hypothesis $\Sigma = I$, like the largest root test (Roy, 1953), may not be able to detect comparatively small departures from idenitity covariance when the ratio N/n is significantly larger than zero. At the same time, our distributional convergence result (*Theorem 3*) can be used to approximate the power of the largest root test against alternatives where the departure from the null model of identity covariance is through perturbations by positive semidefinite matrices of finite rank. We discuss this further in *Section 2.2*. At a more practical level, these results show that exploratory data analytic techniques like "scree plot" to determine number of significant eigenvalues may be of rather limited use when dealing with certain types of near-isotropic high dimensional data. In such circumstances, even the somewhat more sophisticated technique of comparing the sample eigenvalues with the quantiles of the limiting Marchenko-Pastur law, as advocated by Wachter (1976), may not be particularly successful because of the phase transition. The second important consequence of our results is that it gives some insight 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 already 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. Moreover, our method of proof clearly demonstrates how this inconsistency originates.

The rest of the paper is organized as follows. In *Section 2* we describe the main results and point to their salient features. In *Section 3* we define the key quantities and expressions that will help us derive the results. *Section 4* is devoted to proving the almost sure limits of eigenvalues. In *Section 5* we derive the asymptotic distribution result (*Theorem 3*). *Section 6* describes the matrix perturbation analysis approach which is a key ingredient in the proof of *Theorem 3* and *Theorem 4*. Proofs of some of the auxilliary results are given in the two appendices (*Appendix A* and *Appendix B*).

2 Main results

In this section we describe the four main results of this paper. The first two pertain to the *almost sure* limits of sample eigenvalues, the third describes their asymptotic distribution under certain restrictions, while the fourth describes a result about the asymptotic behaviour of sample eigenvectors. We use $\widehat{\ell}_\nu$ to denote the ν-th largest eigenvalue of $\mathbf{S}$.

2.1 Almost sure limit of M largest eigenvalues

We have the following results about the *almost sure* limits of M largest sample eigenvalues. These were independently derived by Baik and Silverstein (2004) for non-Gaussian observations.

Theorem 1 : *Suppose* $\ell_\nu \leq 1+\sqrt{\gamma}$*, then with* $\frac{N}{n} \to \gamma \in (0,1)$ *as* $n \to \infty$ *we have*

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

Theorem 2 : *Suppose* $\ell_\nu > 1+\sqrt{\gamma}$*, then with* $\frac{N}{n} \to \gamma \in (0,1)$ *as* $n \to \infty$ *we have*

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

Let us discuss a little about the limits appearing in (1) and (2). We shall denote the limit in (2) by $\rho_\nu := \ell_\nu \left(1 + \frac{\gamma}{\ell_\nu - 1}\right)$. It turns out, via *Lemma B.1*, that ρ_ν appears as a solution to the following equation

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

with $\ell = \ell_\nu$. Since F_γ 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 $\rho \in ((1+\sqrt{\gamma})^2, \infty)$ and the LHS is obviously increasing in ρ. 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 $\ell_\nu = 1+\sqrt{\gamma}$, $\rho_\nu = (1+\sqrt{\gamma})^2$, 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 Σ is simple, i.e. of multiplicity one, and above the critical value $1+\sqrt{\gamma}$, we show that the corresponding sample eigenvalue is asymptotically normally distributed. While a generalization of this result for the multiplicity greater than one case seems interesting, we do not pursue it here. We note that for the complex Gaussian case a result in the analogous situation has been derived by Baik *et al.* (2004, Theorem 1.1(b)), where 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.

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

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

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

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

In the fixed N case, when the ν-th eigenvalue has multiplicity 1, the ν-th sample eigenvalue is asymptoctically $N(\ell_\nu, \frac{1}{n}2\ell_\nu^2)$. This is a special case of a more general result by Anderson (1963). Thus the fact that the dimension to sample size ratio is positive, contributes towards the bias and a reduction in variance. However, if γ is much smaller compared to ℓ_ν, 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 from invovling $\mathbf{S}$ (under orthogonality restrictions), will tend to put more mass on the ν-th coordinate. This is demonstrated even more clearly by *Theorem 4* that we state later. But before that, we give a brief account of the importance of *Theorem 1-3* from a statistical perspective.

As we already noted in *Section 1*, one possible application of *Theorem 3* is in the calculation of asymptotic power for the largest root test. The latter refers to the testing problem where the null hypothesis says that the covariance matrix is identity. And the test rejects the null hypothesis at level $\alpha \in (0,1)$ if the largest eigenvalue of $\mathbf{S}$ is above a critical level $c_{n,N,\alpha}$, say. Johnstone (2001) proposed a conservative test of this type for large (n, N) data based on the quantiles of Tracy-Widom distribution. His proposal means that the cutoff value, for large n, can be approximated as

$$c_{n,N,\alpha} \approx (1 + \sqrt{\frac{N}{n}})^2 + N^{-1/6} n^{-1/2} (1 + \sqrt{\frac{N}{n}})^{4/3} \tau_\alpha, \qquad \text{for } \ \alpha \in (0,1)$$

where τ_α is the $(1-\alpha)$ quantile of Tracy-Widom law of order 1.

Now suppose we consider the alternative hypothesis that the population covariance matrix is $\Sigma = diag(\ell_1, \ldots, \ell_M, 1, \ldots, 1)$ with $\ell_1 \geq \ldots \geq \ell_M > 1$. If $\ell_1 > 1 + \sqrt{\gamma}$, *Theorem 2* shows that the largest root test is asymptotically consistent. For the special case when ℓ_1 is of multiplicity one, *Theorem 3* immediately gives an expression for the asymptotic power function, assuming that $\frac{N}{n}$ converges to γ fast enough, as $n \to \infty$. But one has to view this in proper context, since our 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 $\widehat{\ell}_1$ towards normality. A detailed analysis of power properties against local alternatives is beyond the scope of this paper. However, *Theorem 1* indicates that the largest root test may fail to detect a departure from the null model of identity covariance if ℓ_1 is less than $1 + \sqrt{\gamma}$.

It is important to point to a potential advantage of such a test as compared to some other well-known tests for the same hypothesis. Ledoit and Wolf (2002) give a nice overview of different tests of sphericity used in high-dimensional setting. They consider tests based on statistics U, and W given below.

$$U = \frac{1}{N} trace \left[\left(\frac{\mathbf{S}}{\frac{1}{N} trace(\mathbf{S})} - I \right)^2 \right], \quad W = \frac{1}{N} trace[(\mathbf{S} - I)^2] - \frac{N}{n} \left[\frac{1}{N} trace(\mathbf{S}) \right]^2 + \frac{N}{n}$$

The statistic U is used to test sphericity, i.e. $\Sigma = \mathbf{c} I$ for some $\mathbf{c} > 0$ unknown. Their results (Ledoit and Wolf, 2002, Proposition 1-7) show that if $\beta_N = \frac{1}{N} trace(\Sigma)$ and $\theta_N^2 = \frac{1}{N} trace(\Sigma - I)^2$ are fixed, at values $\beta > 0$ and θ, say, as $\frac{N}{n} \to \gamma \in (0, \infty)$, then the test based on W is consistent for testing $H_0 : (\beta - 1)^2 + \theta^2 = 0$ against $H_A : (\beta - 1)^2 + \theta^2 > 0$. Whereas the test based on U is consistent for $H_0 : \theta^2/\beta^2 = 0$ against $H_A : \theta^2/\beta^2 > 0$. Their results can be easily extended to the case where $\beta_N \to \beta$ and $\theta_N \to \theta$ rather than being fixed quantities. Notice that even when Σ is a finite rank perturbation of identity, $\beta = 1$ and $\theta = 0$. Under this setting these tests cannot distinguish between H_0 and H_A. We expect similar sort of asymptotic behvaiour from any test that relies upon traces of powers of $\mathbf{S}$ and statistics derived from them. In contrast the test described in the previous paragraph can separate the null from the alternative in the same scenario under the rather mild requirement that $\lambda_1(\Sigma) > 1 + \sqrt{\gamma}$. We treat this comparison as a way of emphasizing the following point: our results show that for signal detection problems in high dimension, when the signal is rather feeble, leaning on tests based on the extreme eigenvalues may be more meaningful than depending on tests which are based on the bulk of the eigenvalue specturm.

2.3 Angle between true and estimated eigenvectors

Hoyle and Rattray (2004) mention about a phase transition phenomenon in the asymptotic behaviour of the angle between the true and estimated eigenvector associated with a non-unit eigenvalue ℓ_ν. They term this "the phenomenon of retarded learning". They derived this result at a physical level of rigour. 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 *almost surely* converges to zero, whereas one gets strictly positive limit if $\ell_\nu > 1 + \sqrt{\gamma}$. Part (a) of *Theorem 4*, stated below and proved in *Section 6*, 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 stated in Johnstone and Lu (2004).

Theorem 4 : *Let $\widetilde{\mathbf{e}}_\nu$ denote the $N \times 1$ vector with 1 in the ν-th coordinate and zeros elsewhere, and $\mathbf{p}_\nu$ denote the eigenvector of $\mathbf{S}$ associated with the eigenvalue $\widehat{\ell}_\nu$.*

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

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

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

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

In order to prove this result we use a specific decmposition of the eigenvectors as explained in *Section 3*. Proceeding along this line it is possible to study the behaviour of the sample eigenvectors in more detail. But we shall give it a full treatment elsewhere and hence do not deal with this issue in the current paper.

3 Representation of the eigenvalues of S

Throughout we assume that n is large enough so that $\frac{N}{n} < 1$. In order to proceed further we introduce some notations that will help us in later stages. First we 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 we use $\widehat{\ell}_\nu$ and $\mathbf{p}_\nu$ to denote the ν-th largest sample eigenvalue and the correpsonding sample eigenvector. We shall follow the convention that the ν-th element of $\mathbf{p}_\nu$ is nonnegative to avoid any ambiguity. We shall write $\mathbf{p}_\nu$ as $\mathbf{p}_\nu^T = (p_{A,\nu}^T, p_{B,\nu}^T)$ and denote the norm $\parallel p_{B,\nu} \parallel$ by R_ν. Then *almost surely* $0 < R_\nu < 1$.

With this setting in place, now we can express the first M eigenequations for $\mathbf{S}$ as

$$\mathbf{S}_{AA} p_{A,\nu} + \mathbf{S}_{AB} p_{B,\nu} = \widehat{\ell}_\nu p_{A,\nu}, \qquad \nu = 1, \ldots, M, \tag{8}$$

$$\mathbf{S}_{BA} p_{A,\nu} + \mathbf{S}_{BB} p_{B,\nu} = \widehat{\ell}_\nu p_{B,\nu}, \qquad \nu = 1, \ldots, M, \tag{9}$$

$$p_{A,\nu}^T p_{A,\nu'} + p_{B,\nu}^T p_{B,\nu'} = \delta_{\nu,\nu'}, \qquad 1 \leq \nu, \nu' \leq M. \tag{10}$$

Here $\delta_{\nu\nu'}$ is the Kronecker symbol. Now denote the vector $p_{A,\nu} / \parallel p_{A,\nu} \parallel = p_{A,\nu}/\sqrt{1 - R_\nu^2}$ by b_ν. Thus $\parallel b_\nu \parallel = 1$. Similarly define $q_\nu := p_{B,\nu}/R_\nu$ and again $\parallel q_\nu \parallel = 1$.

With all the relevant quantities about the problem now defined we can express the eigenequations in a more suitable form that will allow us to make useful observations about the relationship among the empirical eigenvalues $\widehat{\ell}_1, \ldots, \widehat{\ell}_M$. First, changing sides in (9) to collect terms involving q_ν, and noticing that *almost surely* $0 < R_\nu < 1$, and $\widehat{\ell}_\nu I - \mathbf{S}_{BB}$ is invertible,

$$q_\nu = \frac{\sqrt{1 - R_\nu^2}}{R_\nu} (\widehat{\ell}_\nu I - \mathbf{S}_{BB})^{-1} \mathbf{S}_{BA} b_\nu \tag{11}$$

Now, dividing both sides of (8) by $\sqrt{1 - R_\nu^2}$ and substituting the expression for q_ν, we get

$$(\mathbf{S}_{AA} + \mathbf{S}_{AB}(\widehat{\ell}_\nu I - \mathbf{S}_{BB})^{-1} \mathbf{S}_{BA}) b_\nu = \widehat{\ell}_\nu b_\nu, \qquad \nu = 1, \ldots, M. \tag{12}$$

This equation is quite remarkable 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 b_ν. This particular observation will be the building block for all our analysis. However, we shall find it more convenient to express the quantities in terms of the spectral elements of the data matrix $\mathbf{X}$.

Let Λ denote the diagonal matrix $diag(\ell_1, \ldots, \ell_M)$. Because of normality assumption, the observation matrix $\mathbf{X}$ can be reexpressed as

$$\mathbf{X}^T = [\mathbf{Z}_A^T \Lambda^{1/2} : \mathbf{Z}_B^T], \qquad \mathbf{Z}_A \text{ is } M \times n, \qquad \mathbf{Z}_B \text{ is } (N - M) \times n,$$

and 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. We can also assume that $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are defined on the same probability space.

Let the singular value decomposition of $\frac{1}{\sqrt{n}} \mathbf{Z}_B$ be given as

$$\frac{1}{\sqrt{n}} \mathbf{Z}_B = V \mathcal{M}^{1/2} H^T \tag{13}$$

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. We shall denote the diagonal elements of $\mathcal{M}$ by $\mu_1 > \ldots > \mu_{N-M}$, suppressing the dependence on n.

Let the columns of V be denoted by $v_1, \ldots, v_{N-M}$ and the columns of H be denoted by $h_1, \ldots, h_{N-M}$. Note that $\{v_1, \ldots, v_{N-M}\}$ is a complete orthonormal basis for $\mathbb{R}^{N-M}$, whereas $h_1, \ldots, h_{N-M}$ form an orthonormal basis of an $(N - M)$ dimensional subspace (viz. the rowspace of $\mathbf{Z}_B$) of $\mathbb{R}^n$.

Observe that $q_\nu = V \zeta_\nu = \sum_{j=1}^{N-M} \zeta_{\nu j} v_j$ for some unit vector ζ_ν. Also, define by T the matrix $\frac{1}{\sqrt{n}} H^T \mathbf{Z}_A^T$. T is an $(N - M) \times M$ matrix and let its columns be denoted by $t_1, \ldots, t_M$. The most important property about T that we shall use repeatedly is that the vectors $t_1, \ldots, t_M$ 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 and the rows of $\mathbf{Z}_A$ are i.i.d. $N_n(0, I)$ vectors, and moreover, $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are independently distributed.

Thus, we obtain the following equations by simple linear transformations of (11) and (12), respectively.

$$\zeta_\nu = \frac{\sqrt{1-R_\nu^2}}{R_\nu}(\widehat{\ell}_\nu I - \mathcal{M})^{-1}\mathcal{M}^{1/2}T\Lambda^{1/2}b_\nu, \qquad \nu = 1,\ldots,M. \tag{14}$$

$$(\mathbf{S}_{AA} + \Lambda^{1/2}T^T\mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-1}T\Lambda^{1/2})b_\nu = \widehat{\ell}_\nu b_\nu, \qquad \nu = 1,\ldots,M. \tag{15}$$

Also note that $K(x)$ can be expressed as

$$K(x) = \mathbf{S}_{AA} + \Lambda^{1/2}T^T\mathcal{M}(xI - \mathcal{M})^{-1}T\Lambda^{1/2} \tag{16}$$

We conclude this section by rewriting equation (10) in terms of the vectors $\{b_\nu : \nu = 1,\ldots,M\}$ as

$$b_\nu^T[I + \Lambda^{1/2}T^T(\widehat{\ell}_\nu I - \mathcal{M})^{-1}\mathcal{M}(\widehat{\ell}_{\nu'} I - \mathcal{M})^{-1}T\Lambda^{1/2}]b_{\nu'} = \frac{1}{1-R_\nu^2}\delta_{\nu\nu'}, \qquad 1 \le \nu,\nu' \le M, \tag{17}$$

which is same as

$$b_\nu^T[I + \mathbf{S}_{AB}(\widehat{\ell}_\nu I - \mathbf{S}_{BB})^{-1}(\widehat{\ell}_{\nu'} I - \mathbf{S}_{BB})^{-1}\mathbf{S}_{BA}]b_{\nu'} = \frac{1}{1-R_\nu^2}\delta_{\nu\nu'}, \qquad 1 \le \nu,\nu' \le M. \tag{18}$$

4 Almost sure limits

In this section we prove *Theorem 1* and *Theorem 2*. Proofs of these two theorems depend heavily on the asymptotic behaviour of the largest eigenvalue of a Wishart matrix in the null (i.e. identity covariance) case, as well as on the limiting behaviour of the *Empirical Spectral Distribution* of Wishart matrices. Throughout, the ESD of $\mathbf{S}_{BB}$ is denoted by $\widehat{F}_{n,N-M}$. Then we know that (cf. Bai, 1999)

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

where $\Longrightarrow$ denotes distributional convergence.

Our proof relies upon essentially showing the following fact

$$\begin{aligned} t_j^T\mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-1}t_k &\to 0, \quad \textit{almost surely} \quad 1 \le j \ne k \le M, \\ t_j^T\mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-1}t_j &\to \gamma\int\frac{x}{\rho_\nu - x}dF_\gamma(x), \quad \textit{almost surely.} \end{aligned}$$

The rest of the section is organized as follows. We shall establish first *Theorem 2* and then *Theorem 1*. The proofs of these two results use the same technique in that they use the *interlacing inequality* for eigenvalues of symmetric matrices to derive upper and lower bounds for $\widehat{\ell}_\nu$ which may fail to hold with negligible probability.

However, it turns out that the derivation of the lower bound for $\widehat{\ell}_\nu$ in the proof of *Theorem 2* becomes much easier if one has a suitable preliminary lower bound on $\widehat{\ell}_\nu$. To be more specific, one needs to ensure that the set $C_\nu = \{\widehat{\ell_\nu} > \mu_1\}$ has very high probability when $\ell_\nu > 1 + \sqrt{\gamma}$. This is comparatively a lot easier in the case when ℓ_ν is in fact greater than $(1+\sqrt{\gamma})^2$. This is established via *Proposition 1* and *Proposition 2*. Incidentally, *Proposition 2* gives a general purpose bound for the j-th eigenvalue μ_j of $\mathbf{S}_{BB}$ for every fixed j. However, in the general case (i.e. when simply $\ell_\nu > 1 + \sqrt{\gamma}$), we explicitly construct a lower bound for $\widehat{\ell_\nu}$ using equations (14) and (15). This requires a lot more work and to keep the exposition simpler we have deferred this result (*Proposition B.2*) till *Appendix B*.

It is comparatively easier to derive sharp upper bounds for $\widehat{\ell_\nu}$ and the same technique can be used to derive bounds for the cases when $\ell_\nu > 1 + \sqrt{\gamma}$ and when $1 < \ell_\nu \le 1 + \sqrt{\gamma}$. For the time being we assume that either $\ell_\nu \le 1 + \sqrt{\gamma}$ or $\ell_\nu > (1+\sqrt{\gamma})^2$.

4.1 Bounds on the eigenvalues $\widehat{\ell_\nu}$

We make use of the *interlacing inequality* for eigenvalues of symmetric matrices, (see e.g. Section 1f of Rao, 1973). We introduce some notations for later use.

4.1.1 Notations

We denote the quantity $(1+\sqrt{\gamma})^2$ by κ_γ. Throughout, unless otherwise specified, $\lambda_j(C)$ for any symmetric matrix C will denote the j-th largest eigenvalue of C. For any $m \times m$ matrix C , if $G \subset \{1, \ldots, m\}$, then by C_G we shall denote the submatrix of C deleting the rows and columns that are in G. Also, we shall use $\| \;\; \|$ to denote both l^2 norm of vectors, as well as the 2-norm, or the largest singular value, of matrices. $\| \;\; \|_{HS}$ will mean the Hilbert-Schmidt norm for matrices. For $\rho > \kappa_\gamma$, we define

$$\Lambda_G(\rho) = (1 + \gamma \int \frac{x}{\rho - x} dF_\gamma(x))\Lambda_G \tag{19}$$

4.1.2 Interlacing inequalities

By the *interlacing inequality* we have

$$\lambda_1(\mathbf{S}_G) \ge \lambda_{|G|+1}(\mathbf{S}) \qquad \text{and} \qquad \lambda_k(\mathbf{S}) \ge \lambda_k(\mathbf{S}_G), \qquad \text{for all } G \subset \{1, \ldots, N\}, \;\; \text{all } k \tag{20}$$

Define $\Gamma_\nu = \{1, \ldots, \nu\}$ and $\overline{\Gamma}_\nu = \{\nu+1, \ldots, M\}$ for $1 \le \nu \le M$ and $\Gamma_0 = \phi = \overline{\Gamma}_M$. Then (20) implies

$$\lambda_1(\mathbf{S}_{\Gamma_{\nu-1}}) \ge \widehat{\ell_\nu} \ge \lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu}) \tag{21}$$

4.1.3 Eigenvalues of submatrices

Observe that $\lambda_1(\mathbf{S}_{\Gamma_{\nu-1}})$ and $\lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu})$ are *some* eigenvalues of $K_{\Gamma_{\nu-1}}(\lambda_1(\mathbf{S}_{\Gamma_{\nu-1}}))$ and $K_{\overline{\Gamma}_\nu}(\lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu}))$, respectively, where

$$K_G(x) = S_{AA,G} + \Lambda_G^{1/2} T_G^T \mathcal{M}(xI - \mathcal{M})^{-1} T_G \Lambda_G^{1/2}, \qquad \text{for } G \subset \{1, \ldots, M\}. \tag{22}$$

Here by T_G we denote the submatrix of T with all columns in set G deleted. This follows by noting that for $G \subset \{1, \ldots, M\}$,

$$\mathbf{S}_G = \begin{bmatrix} \mathbf{S}_{AA,G} & \mathbf{S}_{AB,G} \\ \mathbf{S}_{BA,G} & \mathbf{S}_{BB} \end{bmatrix}, \quad \text{with } \mathbf{S}_{AB,G} = \frac{1}{n}\Lambda_G^{1/2}\mathbf{Z}_{A,G}\mathbf{Z}_B^T, \;\; \mathbf{S}_{AA,G} = \frac{1}{n}\Lambda_G^{1/2}\mathbf{Z}_{A,G}\mathbf{Z}_{A,G}^T\Lambda_G^{1/2},$$

and $\mathbf{S}_{BA,G} = \mathbf{S}_{AB,G}^T$, where $\mathbf{Z}_{A,G}$ denotes the submatrix of $\mathbf{Z}_A$ with rows in the set G deleted.

4.1.4 Preliminary bounds

To begin with let us assume that $\ell_\nu > (1+\sqrt{\gamma})^2$. This situation is simpler to deal with. In the following $\overline{\mathbf{S}}_{AA,\nu}$ denotes the submatrix of $\mathbf{S}_{AA}$ consisting of only the first ν rows and ν columns. We show that the eigenvalues of $\mathbf{S}_{AA,\nu}$ concentrate around their population counterparts, and by the *interlacing inequalities* we directly have

$$\lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu}) \geq \lambda_\nu(\overline{\mathbf{S}}_{AA,\nu}), \tag{23}$$

Therefore the required lower bound for $\widehat{\ell}_\nu$ is easily obtained if we apply (21). We formally state the following:

Proposition 1 : *Let $\kappa_\gamma = (1+\sqrt{\gamma})^2$ and $0 < \epsilon < \frac{\ell_1 \nu}{2}$ be any number such that $\ell_\nu > \kappa_\gamma + 2\epsilon$. Then,*

$$\mathbb{P}(\lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu}) \leq \kappa_\gamma + \epsilon) \leq 2\nu \exp\left(-\frac{n\epsilon^2}{6\ell_1^2\nu^2}\right) + \nu(\nu-1)\exp\left(-\frac{n\epsilon^2}{3\ell_1^2\nu^2}\right) \tag{24}$$

Proof : In view of (23) we only need to establish the inequality for $\lambda_\nu(\overline{\mathbf{S}}_{AA,\nu})$. Let $\overline{\Lambda}_\nu = diag(\ell_1, \ldots, \ell_\nu)$. Then for $\nu' = 1, \ldots, \nu$,

$$|\lambda_{\nu'}(\overline{\mathbf{S}}_{AA,\nu}) - \ell_{\nu'}| \leq \| \overline{\mathbf{S}}_{AA,\nu} - \overline{\Lambda}_\nu \| \tag{25}$$

In certain circumstances this upper bound can be improved with a more careful analysis, but we do not need that here. The bound appearing on the RHS can be majorized by the Hilbert-Schmidt norm:

$$\| \overline{\mathbf{S}}_{AA,\nu} - \overline{\Lambda}_\nu \|_{HS} = \sqrt{\sum_{k=1}^{\nu} |s_{kk} - \ell_k|^2 + \sum_{j \neq k}^{\nu} |s_{jk}|^2}$$

where s_{jk} is the (j,k)-th element of $\mathbf{S}_{AA}$, $1 \leq j, k \leq M$. In order to bound the terms appearing inside square roots we use large deviation inequalities for quadratic forms of Gaussian random variables. Observe that $s_{jk} = \sqrt{\ell_j \ell_k} \frac{1}{n} Z_{A,j}^T Z_{A,k}$ where $Z_{A,j}^T$ is the j-th row of $\mathbf{Z}_A$. Taking $X = Z_{A,j}$, $Y = Z_{A,k}$, $C(Z) = I$ and $L = 1$ in *Lemma A.1* we get

$$\mathbb{P}(|s_{kj}| > \sqrt{\ell_j \ell_k} t) \leq 2\exp(-\frac{(1-\delta)nt^2}{2}), \qquad 0 < t < \frac{\delta}{1-\delta} \tag{26}$$

Similarly, applying *Lemma A.2* with $X = Z_{A,k}$, $C(Z) = I$ and $L = 1$ we get

$$\mathbb{P}(|s_{kk} - \ell_k| > \ell_k t) \leq 2\exp(-\frac{(1-\delta)nt^2}{4}), \qquad 0 < t < \frac{2\delta}{1-\delta} \tag{27}$$

Thus, taking $\delta = \frac{1}{3}$ in (26) and (27) and setting $t = \frac{\epsilon}{\ell_1 \nu}$, since $0 < \epsilon < \frac{\ell_1 \nu}{2}$ we get, after using (25) and the expression for $\| \overline{\mathbf{S}}_{AA,\nu} - \overline{\Lambda}_\nu \|_{HS}$

$$\mathbb{P}(|\lambda_\nu(\overline{\mathbf{S}}_{AA,\nu}) - \ell_\nu| > \epsilon) \leq 2\nu \exp\left(-\frac{n\epsilon^2}{6\ell_1^2\nu^2}\right) + \nu(\nu-1)\exp\left(-\frac{n\epsilon^2}{3\ell_1^2\nu^2}\right) \tag{28}$$

From this (24) follows.

Next we state a result about the concentration of largest few eigenvalues of $\mathbf{S}_{BB}$ around κ_γ. This is proved in *Appendix A*.

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

$$\mathbb{P}(|\mu_1 - \kappa_\gamma| \geq \delta) \leq 2\exp\left(-\frac{n\delta^2}{32\kappa_\gamma}\right), \qquad \textit{for } n \geq n_0(\gamma, \delta) \tag{29}$$

where $n_0(\gamma, \delta)$ *is an integer large enough such that* $|Median(\mu_1) - \kappa_\gamma| \leq \frac{\delta}{4}$ *for* $n \geq n_0(\gamma, \delta)$.

Remark : The proof relies on the asymptotic distribution of the largest eigenvalue of a sample covariance matrix in the identity covariance case (Johnstone, 2001) and a concentration inequality for singular values of Gaussian random matrices. Soshnikov (2002) proved that when centered and scaled by the same numbers, a similar type of limiting law holds for any leading eigenvalue (i.e. any μ_j with j fixed). The details of these distributions can be found in Tracy and Widom (1994), (1996). So the same proposition applies to any μ_j for j fixed.

4.2 Upper bound for $\widehat{\ell}_\nu$

First we derive a tight upper bound for $\widehat{\ell}_\nu$. Our strategy is to utilize the upper bound in (21). For this we do not need (24). However, we need the bound (29). For simplicity of notations, we shall use $\widehat{\lambda}_{1,\nu}$ to mean $\lambda_1(\mathbf{S}_{\Gamma_{\nu-1}})$. Our aim is to prove the following:

Proposition 3 : *Let* $\ell_\nu > 1 + \sqrt{\gamma}$ *for some* $1 \leq \nu \leq M$. *Then given* $\epsilon > 0$ *there exists* $n_1(\nu, \epsilon, \Lambda, \gamma)$ *large enough such that, for* $n \geq n_1(\nu, \epsilon, \Lambda, \gamma)$,

$$\mathbb{P}(\widehat{\ell}_\nu > \rho_\nu + \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2) \leq (M - \nu + 1)\varepsilon_1(n, \epsilon, \Lambda, \gamma) + (M-\nu)(M-\nu+1)\varepsilon_2(n, \epsilon, \Lambda, \gamma), \tag{30}$$

where ε_1 and ε_2 are the terms appearing on the RHS of equations (47) and (39), respectively.

Proof : First, from (21) we have

$$\mathbb{P}(\widehat{\ell}_\nu > \rho_\nu + \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2) \le \mathbb{P}(\widehat{\lambda}_{1,\nu} > \rho_\nu + \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2)$$

Since $\widehat{\lambda}_{1,\nu}$ is an eigenvalue of $K_{\Gamma_{\nu-1}}(\widehat{\lambda}_{1,\nu})$ where $K_G(\cdot)$ is defined through (22),

$$\widehat{\lambda}_{1,\nu} \le \lambda_1(K_{\Gamma_{\nu-1}}(\rho_\nu + \epsilon)) \qquad \text{on the set } \; J_{1,\nu} := \{\widehat{\lambda}_{1,\nu} > \rho_\nu + \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2\} \tag{31}$$

To verify (31) observe that on $J_{1,\nu}$ the inequalities $\frac{\mu_j}{\widehat{\lambda}_{1,\nu} - \mu_j} \le \frac{\mu_j}{\rho_\nu + \epsilon - \mu_j}$ hold for all $j = 1, \ldots, N - M$. This implies the inequality for positive semidefinite matrices: $K_{\Gamma_{\nu-1}}(\widehat{\lambda}_{1,\nu}) \le K_{\Gamma_{\nu-1}}(\rho_\nu + \epsilon)$ (since, for any $a \in \mathbb{R}^{M-|G|}$, $a^T K_G(x) a = a^T \mathbf{S}_{AA,G} a + \sum_{j=1}^{N-M} c_j^2 \mu_j (x - \mu_j)^{-1}$ where $c = T_G \Lambda_G^{1/2} a$).

Let $\Lambda_G(\rho)$ be defined through (19). Then by (31) and a simple inequality for eigenvalues of symmetric matrices, on the set $J_{1,\nu}$,

$$\widehat{\lambda}_{1,\nu} \le \ell_\nu (1 + \gamma \int \frac{x}{\rho_\nu + \epsilon - x} dF_\gamma(x)) + \parallel K_{\Gamma_{\nu-1}}(\rho_\nu + \epsilon) - \Lambda_{\Gamma_{\nu-1}}(\rho_\nu + \epsilon) \parallel_{HS} \tag{32}$$

since $\ell_\nu(1 + \gamma \int \frac{x}{\rho_\nu + \epsilon - x} dF_\gamma(x))$ is the largest eigenvalue of $\Lambda_{\Gamma_{\nu-1}}(\rho_\nu + \epsilon)$.

Now, let $\delta := \delta(\epsilon) > 0$ be such that

$$\ell_\nu (1 + \gamma \int \frac{x}{\rho_\nu + \epsilon - x} dF_\gamma(x)) + \delta = \rho_\nu + \epsilon \tag{33}$$

Indeed we can find such a δ because if we define $\ell_{\nu,\epsilon}$ to be $\ell_{\nu,\epsilon} = (\rho_\nu + \epsilon)(1 + \gamma \int \frac{x}{\rho_\nu + \epsilon - x} dF_\gamma(x))^{-1}$, then we have $\ell_{\nu,\epsilon} > \ell_\nu$ (since, by definition, ρ_ν is the solution to equation (3)), and hence

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

Denote the matrix appearing inside $\parallel \quad \parallel_{HS}$ on the RHS of (32) by $\widetilde{D}_\nu = ((\widetilde{D}_{\nu,jk}))_{j,k=1}^{M-\nu+1}$. From our construction, if

$$|\widetilde{D}_{\nu,jk}| < \delta_{jk}, \;\; 1 \le j,k \le M - \nu + 1, \qquad \text{where } \; \delta_{jk} > 0, \; \forall \; j,k, \; \text{and} \; \sum_{j=1}^{M-\nu+1} \sum_{k=1}^{M-\nu+1} \delta_{jk}^2 \le \delta^2,$$

then from (32) we get, on $J_{1,\nu}$, $\widehat{\lambda}_{1,\nu} < \rho_\nu + \epsilon$ which is an impossibility. Hence after taking the union bound,

$$\mathbb{P}(J_{1,\nu}) \le \sum_{j=1}^{M-\nu+1} \mathbb{P}(|\widetilde{D}_{\nu,jj}| \ge \delta_{jj}, \mu_1 < \kappa_\gamma + \epsilon/2) + \sum_{j<k}^{M-\nu+1} \mathbb{P}(|\widetilde{D}_{\nu,jk}| \ge \delta_{jk}, \mu_1 < \kappa_\gamma + \epsilon/2) \tag{34}$$

We set $\delta_{jj} = \ell_{j+\nu-1} \widetilde{\delta}_1$, $j = 1, \ldots, M - \nu + 1$ and $\delta_{jk} = \sqrt{\ell_{j+\nu-1}\ell_{k+\nu-1}}\; \widetilde{\delta}_2$, $1 \le j < k \le M - \nu + 1$ where $\widetilde{\delta}_1, \widetilde{\delta}_2 > 0$ are such that $\delta^2 = \widetilde{\delta}_1^2 \sum_{j=\nu}^M \ell_j^2 + \widetilde{\delta}_2^2 (\sum_{j=\nu}^M \ell_j)^2$. To be specific, we take $\widetilde{\delta}_2 = \frac{\delta}{\sqrt{2}} (\sum_{j=\nu}^M \ell_j)^{-1}$ and $\widetilde{\delta}_1 = \frac{\delta}{\sqrt{2}} (\sum_{j=\nu}^M \ell_j^2)^{-1/2}$.

Remark : Notice that the bound we are using here is rather crude. In specific situations, e.g. when $\ell_1, \ldots, \ell_\nu$ are distinct, one may be able to get better bounds.

Observe that for $j = 1, \ldots, M - \nu + 1$,

$$\begin{aligned}\widetilde{D}_{\nu,jj} &= \ell_{j+\nu-1}(\frac{1}{n} Z_{A,j+\nu-1}^T Z_{A,j+\nu-1} - 1) \\ &\quad + \ell_{j+\nu-1}(t_{j+\nu-1}^T \mathcal{M}((\rho_\nu + \epsilon)I - \mathcal{M})^{-1} t_{j+\nu-1} - \frac{1}{n} trace\ (\mathcal{M}((\rho_\nu + \epsilon)I - \mathcal{M})^{-1})) \\ &\quad + \ell_{j+\nu-1}(\frac{1}{n} trace\ (\mathcal{M}((\rho_\nu + \epsilon)I - \mathcal{M})^{-1}) - \gamma \int \frac{x}{\rho_\nu + \epsilon - x} dF_\gamma(x)) \qquad (35)\end{aligned}$$

whereas, for $1 \leq j \neq k \leq M - \nu + 1$,

$$\widetilde{D}_{\nu,jk} = \sqrt{\ell_{j+\nu-1}\ell_{k+\nu-1}}[\frac{1}{n} Z_{A,j+\nu-1}^T Z_{A,k+\nu-1} + t_{j+\nu-1}^T \mathcal{M}((\rho_\nu + \epsilon)I - \mathcal{M})^{-1} t_{k+\nu-1}] \qquad (36)$$

Define $J_\gamma(\epsilon) := \{\mu_1 \leq \kappa_\gamma + \epsilon\}$. Then to bound $\mathbb{P}(|\widetilde{D}_{\nu,jk}| \geq \delta_{jk}, J_\gamma(\epsilon/2))$, for $j \neq k$, observe that from (26) we have

$$\mathbb{P}(|\frac{1}{n} Z_{A,j+\nu-1}^T Z_{A,k+\nu-1}| \geq \widetilde{\delta}_2/2) \leq 2 \exp\left(-\frac{n\widetilde{\delta}_2^2}{12}\right), \qquad \text{for } \ 0 < \widetilde{\delta}_2 < 1 \qquad (37)$$

Since $\sqrt{n} t_j \sim N(0, I_{N-M})$ for $j = 1, \ldots, M$, and

$$\| \mathcal{M}((\rho_\nu + \epsilon)I - \mathcal{M})^{-1} \| = \frac{\mu_1}{\rho_\nu + \epsilon - \mu_1} \leq \frac{\kappa_\gamma + \epsilon/2}{\rho_\nu + \epsilon/2 - \kappa_\gamma}$$

on $J_\gamma(\epsilon/2)$, we can apply *Lemma A.1* to conclude that (taking $\delta = \frac{1}{3}$ in the lemma), for $j \neq k$,

$$\begin{aligned}&\mathbb{P}(|t_{j+\nu-1}^T \mathcal{M}((\rho_\nu + \epsilon)I - \mathcal{M})^{-1} t_{k+\nu-1}| \geq \widetilde{\delta}_2/2, J_\gamma(\epsilon/2)) \\ &= 2 \exp\left(-\frac{n}{N-M} \frac{(\rho_\nu + \epsilon/2 - \kappa_\gamma)^2 n\widetilde{\delta}_2^2}{12(\kappa_\gamma + \epsilon/2)^2}\right) \\ &= 2 \exp\left(-\frac{1}{\gamma} \frac{(\rho_\nu + \epsilon/2 - \kappa_\gamma)^2 n\widetilde{\delta}_2^2}{12(\kappa_\gamma + \epsilon/2)^2}(1 + o(1))\right), \qquad \text{for } 0 < \widetilde{\delta}_2 < \frac{\kappa_\gamma + \epsilon/2}{\rho_\nu + \epsilon/2 - \kappa_\gamma} \qquad (38)\end{aligned}$$

Combining (36), (37) and (38), for $0 < \widetilde{\delta}_2 < \min\{1, \frac{\kappa_\gamma + \epsilon/2}{\rho_\nu + \epsilon/2 - \kappa_\gamma}\}$,

$$\begin{aligned}\mathbb{P}(|\widetilde{D}_{\nu,jk}| \geq \delta_{jk}, \mu_1 < \kappa_\gamma + \epsilon/2) &\leq 2 \exp\left(-\frac{n\widetilde{\delta}_2^2}{12}\right) + 2 \exp\left(-\frac{n}{N-M} \frac{(\rho_\nu + \epsilon/2 - \kappa_\gamma)^2 n\widetilde{\delta}_2^2}{12(\kappa_\gamma + \epsilon/2)^2}\right) \qquad (39) \\ &\quad \text{for all } \quad 1 \leq j < k \leq M - \nu + 1\end{aligned}$$

In order to obtain a similar bound for $\widetilde{D}_{\nu,jj}$, first observe that from (27),

$$\mathbb{P}(|\frac{1}{n} Z_{A,j+\nu-1}^T Z_{A,j+\nu-1} - 1| \geq \widetilde{\delta}_1/4) \leq 2 \exp\left(-\frac{n\widetilde{\delta}_1^2}{96}\right), \qquad \text{for } \ 0 < \widetilde{\delta}_1 < 4 \qquad (40)$$

Again, argument silimlar to that leading to (38) implies (this time using *Lemma A.2*),

$$\begin{aligned}
&\mathbb{P}(|t_{j+\nu-1}^T \mathcal{M}((\rho_\nu+\epsilon)I-\mathcal{M})^{-1}t_{j+\nu-1} - \frac{1}{n} trace\ (\mathcal{M}((\rho_\nu+\epsilon)I-\mathcal{M})^{-1})| \geq \widetilde{\delta}_1/4, J_\gamma(\epsilon/2)) \\
\leq\ & 2\exp\left(-\frac{n}{N-M}\frac{(\rho_\nu+\epsilon/2-\kappa_\gamma)^2 n\widetilde{\delta}_1^2}{96(\kappa_\gamma+\epsilon/2)^2}\right) \\
=\ & 2\exp\left(-\frac{1}{\gamma}\frac{(\rho_\nu+\epsilon/2-\kappa_\gamma)^2 n\widetilde{\delta}_1^2}{96(\kappa_\gamma+\epsilon/2)^2}(1+o(1))\right), \qquad \text{for } 0<\widetilde{\delta}_1<\frac{4(\kappa_\gamma+\epsilon/2)}{\rho_\nu+\epsilon/2-\kappa_\gamma} \qquad (41)
\end{aligned}$$

To provide a bound for the remaining terms, we observe that on $J_\gamma(\epsilon/2)$,

$$\begin{aligned}
trace\ (\mathcal{M}((\rho_\nu+\epsilon)I-\mathcal{M})^{-1}) + (N-M) &= (\rho_\nu+\epsilon)\ trace\ (((\rho_\nu+\epsilon)I-\mathcal{M})^{-1}) \\
&= (\rho_\nu+\epsilon)\ trace\ G_1(\mathbf{S}_{BB};\rho_\nu+\epsilon,\gamma,\epsilon/2),
\end{aligned}$$

where the function $G_1(\cdot;\cdot,\cdot,\cdot)$ is defined through (90) in *Appendix A*. Therefore we can apply *Proposition A.1* (in the *Appendix*) to get

$$\begin{aligned}
&\mathbb{P}(|\frac{1}{n} trace\ (\mathcal{M}((\rho_\nu+\epsilon)I-\mathcal{M})^{-1}) - \\
&\qquad \mathbb{E}(\frac{1}{n}(\rho_\nu+\epsilon)\ trace\ G_1(\mathbf{S}_{BB};\rho_\nu+\epsilon,\gamma,\epsilon/2)) + \frac{N-M}{n}| > \widetilde{\delta}_1/4,\ J_\gamma(\epsilon/2)) \\
=\ & \mathbb{P}(|\frac{1}{n} trace\ G_1(\mathbf{S}_{BB};\rho_\nu+\epsilon,\gamma,\epsilon/2) - \mathbb{E}(\frac{1}{n} trace\ G_1(\mathbf{S}_{BB};\rho_\nu+\epsilon,\gamma,\epsilon/2))| \\
&\qquad > (\rho_\nu+\epsilon)^{-1}\widetilde{\delta}_1/4,\ J_\gamma(\epsilon/2)) \\
\leq\ & 2\exp\left(-\frac{n}{n+N-M}\frac{n^2\widetilde{\delta}_1^2}{2}\frac{(\rho_\nu+\epsilon/2-\kappa_\gamma)^4}{64(\rho_\nu+\epsilon)^2(\kappa_\gamma+\epsilon/2)}\right) \\
=\ & 2\exp\left(-\frac{n^2\widetilde{\delta}_1^2}{2(1+\gamma)}\frac{(\rho_\nu+\epsilon/2-\kappa_\gamma)^4}{64(\rho_\nu+\epsilon)^2(\kappa_\gamma+\epsilon/2)}(1+o(1))\right) \qquad (42)
\end{aligned}$$

Now to tackle the remainder we notice that

$$\mathbb{E}(\frac{1}{n} trace\ G_1(\mathbf{S}_{BB};\rho_\nu+\epsilon,\gamma,\epsilon/2)) = \frac{N-M}{n}\mathbb{E}\int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)d\widehat{F}_{n,N-M}(x)$$

where $\widehat{F}_{n,N-M}$ denotes the *ESD* of the matrix $\mathbf{S}_{BB}$. Note that G_1 is bounded above and monotone in its first argument. Further, defining $\overline{F}_{n,N-M}$ to be the expected *ESD*, by linearity of expectation, $\mathbb{E}\int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)d\widehat{F}_{n,N-M}(x) = \int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)d\overline{F}_{n,N-M}(x)$. It is well-known that $\overline{F}_{n,N-M} \Longrightarrow F_\gamma$ as $n\to\infty$, where F_γ is the Marchenko-Pastur law with parameter γ. Bai (1993) proved under fairly weak conditions that (Bai, 1993, Theorem 3.2) if $\theta_1 \leq \frac{p}{n} \leq \theta_2$ where $0<\theta_1<1<\theta_2<\infty$, then

$$\parallel \overline{F}_{n,p} - F_{p/n} \parallel_\infty \leq C_1(\theta_1,\theta_2)n^{-5/48}. \qquad (43)$$

When $0<\theta_1<\theta_2<1$, he also showed (Bai, 1993, Theorem 3.1) that

$$\parallel \overline{F}_{n,p} - F_{p/n} \parallel_\infty \leq C_2(\theta_1,\theta_2)n^{-1/4}. \qquad (44)$$

Here $\| \cdot \|_\infty$ means the sup-norm and C_1, C_2 are constants with values depending on θ_1, θ_2.

Then utilizing the fact that F_γ has bounded support, $G_1(\cdot; \rho_\nu + \epsilon, \gamma, \epsilon/2)$ is bounded, nondecreasing and differentiable everywhere except at $x = \kappa_\gamma + \epsilon/2$ with $G_1'(x) \equiv 0$ on $(\kappa_\gamma + \epsilon/2, \infty)$, and integrating by parts, we get from (43), for any $0 < \theta_1 < 1$,

$$\begin{aligned}
& |\int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)d\overline{F}_{n,N-M}(x) - \int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)dF_\gamma(x)| \\
\leq\ & |\int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)d\overline{F}_{n,N-M}(x) - \int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)dF_{(N-M)/n}(x)| \\
& \quad + |\int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)dF_{(N-M)/n}(x) - \int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)dF_\gamma(x)| \\
\leq\ & C_3(\theta_1)(n^{-5/48} + \| F_{(N-M)/n} - F_\gamma \|_\infty) \int_0^{\kappa_\gamma+\epsilon/2} |G_1'(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)|dx \\
\leq\ & C_3'(\theta_1)(n^{-5/48} + C_4(\frac{N-M}{n} - \gamma)) \frac{(\kappa_\gamma+\epsilon/2)}{(\rho_\nu+\epsilon/2-\kappa_\gamma)^2}
\end{aligned} \tag{45}$$

uniformly in $\theta_1 \leq \frac{N-M}{n} \leq 1$, where C_3, C_3' are constants depending on θ_1 and $C_4(\cdot)$ is a nonnegative function converging to 0 at 0. Observe that

$$\int G_1(x;\rho_\nu+\epsilon,\gamma,\epsilon/2)dF_\gamma(x) = \int \frac{1}{\rho_\nu+\epsilon-x} dF_\gamma(x)$$

Therefore $\exists\ n_1(\nu, \epsilon, \Lambda, \gamma) \geq 1$ such that for $n \geq n_1(\nu, \epsilon, \Lambda, \gamma)$,

$$|(\rho_\nu+\epsilon)\mathbb{E}(\frac{1}{n} trace\ G_1(\mathbf{S}_{BB};\rho_\nu+\epsilon,\gamma,\epsilon/2)) - \frac{N-M}{n} - \gamma \int \frac{x}{\rho_\nu+\epsilon-x} dF_\gamma(x)| \leq \widetilde{\delta}_1/4 \tag{46}$$

Combining (40), (41), (42) and (46), for $\widetilde{\delta}_1 < 4\min\{1, \frac{\kappa_\gamma+\epsilon/2}{\rho_\nu+\epsilon/2-\kappa_\gamma}\}$,

$$\begin{aligned}
& \mathbb{P}(|\widetilde{D}_{\nu,jj}| \geq \delta_{jj}, \mu_1 < \kappa_\gamma + \epsilon/2) \\
\leq\ & 2\exp\left(-\frac{n\widetilde{\delta}_1^2}{96}\right) + 2\exp\left(-\frac{n}{N-M}\frac{(\rho_\nu+\epsilon/2-\kappa_\gamma)^2 n\widetilde{\delta}_1^2}{96(\kappa_\gamma+\epsilon/2)^2}\right) \\
& +2\exp\left(-\frac{n}{n+N-M}\frac{n^2\widetilde{\delta}_1^2}{2}\frac{(\rho_\nu+\epsilon/2-\kappa_\gamma)^4}{64(\rho_\nu+\epsilon)^2(\kappa_\gamma+\epsilon/2)}\right) \quad \text{for } 1 \leq j \leq M-\nu+1
\end{aligned} \tag{47}$$

for all $n \geq n_1(\nu, \epsilon, \Lambda, \gamma)$.

Remark : Since the upper bounds in (39) and (47) involve quantities $\widetilde{\delta}_2$ and $\widetilde{\delta}_1$, respectively, it is important to clarify their behaviour vis-a-vis ϵ when $\epsilon \to 0$. Since $\widetilde{\delta}_1, \widetilde{\delta}_2$ are proportional to $\delta(\epsilon)$, defined by (33), we study the latter. From (33), we get

$$\frac{d\delta(\epsilon)}{d\epsilon} = 1 + \ell_\nu\gamma \int \frac{x}{(\rho_\nu+\epsilon-x)^2} dF_\gamma(x) \to 1 + \frac{\ell_\nu\gamma}{(\ell_\nu-1)^2-\gamma} \qquad \text{as } \epsilon \to 0,$$

by *Lemma B.2.* This shows that $\exists\ 0 < c_1 < c_2 < \infty$ such that $c_1\epsilon \leq \delta(\epsilon) \leq c_2\epsilon$ for ϵ small enough.

4.3 Lower bound for $\widehat{\ell}_\nu$

Now, we derive a sharp lower bound for $\widehat{\ell}_\nu$ under the restriction that $\{\widehat{\ell}_\nu > \kappa_\gamma + \epsilon/2 > \mu_1\}$ for $\epsilon > 0$ small enough so that $\kappa_\gamma + 2\epsilon < \rho_\nu$. Then utilizing the lower bound in (21) in a way very similiar to the proof of *Proposition 3*, we obtain :

Proposition 4 : *Let $\ell_\nu > 1 + \sqrt{\gamma}$ for some $1 \leq \nu \leq M$. Let $\epsilon > 0$ be such that $\rho_\nu > \kappa_\gamma + 2\epsilon$. Then there exists $n_2(\nu, \epsilon, \Lambda, \gamma)$ large enough such that, for $n \geq n_2(\nu, \epsilon, \Lambda, \gamma)$,*

$$\mathbb{P}(\widehat{\ell}_\nu < \rho_\nu - \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2, \lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu}) > \kappa_\gamma + \epsilon/2) \leq \nu\widetilde{\varepsilon}_1(n, \epsilon, \Lambda, \gamma) + \nu(\nu-1)\widetilde{\varepsilon}_2(n, \epsilon, \Lambda, \gamma) \quad (48)$$

where $\widetilde{\varepsilon}_1$ and $\widetilde{\varepsilon}_2$ are given by the RHS of (53) and (52), respectively.

Proof : Define $\widehat{\lambda}_{\nu,\nu} = \lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu})$. From (21) we have

$$\mathbb{P}(\widehat{\ell}_\nu < \rho_\nu - \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2, \widehat{\lambda}_{\nu,\nu} > \kappa_\gamma + \epsilon/2) \leq \mathbb{P}(\widehat{\lambda}_{\nu,\nu} < \rho_\nu - \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2, \widehat{\lambda}_{\nu,\nu} > \kappa_\gamma + \epsilon/2)$$

Since $\widehat{\lambda}_{\nu,\nu}$ is an eigenvalue of $K_{\overline{\Gamma}_\nu}(\widehat{\lambda}_{\nu,\nu})$,

$$\widehat{\lambda}_{\nu,\nu} \geq \lambda_\nu(K_{\overline{\Gamma}_\nu}(\rho_\nu - \epsilon)) \qquad \text{on the set} \quad J_{\nu,\nu} := \{\kappa_\gamma + \epsilon/2 < \widehat{\lambda}_{\nu,\nu} < \rho_\nu - \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2\} \quad (49)$$

Then, with $\Lambda_G(\rho)$ defined through (19), on the set $J_{\nu,\nu}$,

$$\widehat{\lambda}_{\nu,\nu} \geq \ell_\nu(1 + \gamma \int \frac{x}{\rho_\nu - \epsilon - x} dF_\gamma(x)) - \parallel K_{\overline{\Gamma}_\nu}(\rho_\nu - \epsilon) - \Lambda_{\overline{\Gamma}_\nu}(\rho_\nu - \epsilon) \parallel_{HS} \quad (50)$$

since $\ell_\nu(1 + \gamma \int \frac{x}{\rho_\nu - \epsilon - x} dF_\gamma(x))$ is the smallest eigenvalue of $\Lambda_{\overline{\Gamma}_\nu}(\rho_\nu - \epsilon)$.

Now, let $\overline{\delta} := \overline{\delta}(\epsilon) > 0$ be such that $\ell_\nu(1 + \gamma \int \frac{x}{\rho_\nu - \epsilon - x} dF_\gamma(x)) - \overline{\delta} = \rho_\nu - \epsilon$. We can find such a $\overline{\delta}$ because if we define $\overline{\ell}_{\nu,\epsilon}$ to be $\overline{\ell}_{\nu,\epsilon} = (\rho_\nu - \epsilon)(1 + \gamma \int \frac{x}{\rho_\nu - \epsilon - x} dF_\gamma(x))^{-1}$, then we have $\overline{\ell}_{\nu,\epsilon} < \ell_\nu$ (since, by definition, ρ_ν is the solution to equation (3)), and hence

$$\ell_\nu(1 + \gamma \int \frac{x}{\rho_\nu - \epsilon - x} dF_\gamma(x)) = \frac{\ell_\nu}{\overline{\ell}_{\nu,\epsilon}}(\rho_\nu - \epsilon) > \rho_\nu - \epsilon$$

Note also that $\overline{\delta} \downarrow 0$ as $\epsilon \downarrow 0$. Denote the matrix appearing inside $\parallel \quad \parallel_{HS}$ on the RHS of (50) by $\overline{D}_\nu = ((\overline{D}_{\nu,jk}))_{j,k=1}^\nu$. From our construction, if

$$|\overline{D}_{\nu,jk}| \leq \overline{\delta}_{jk}, \quad 1 \leq j, k \leq \nu, \qquad \text{where } \overline{\delta}_{jk} > 0, \ \forall\, j, k, \text{ and } \sum_{j=1}^{\nu} \sum_{k=1}^{\nu} \overline{\delta}_{jk}^2 \leq \overline{\delta}^2,$$

then from (50) we get, on $J_{\nu,\nu}$, $\widehat{\lambda}_{\nu,\nu} > \rho_\nu - \epsilon$ which is impossible. Hence after taking the union bound,

$$\mathbb{P}(J_{\nu,\nu}) \leq \sum_{j=1}^{\nu} \mathbb{P}(|\overline{D}_{\nu,jj}| \geq \overline{\delta}_{jj}, \mu_1 < \kappa_\gamma + \epsilon/2) + \sum_{j<k}^{\nu} \mathbb{P}(|\overline{D}_{\nu,jk}| \geq \overline{\delta}_{jk}, \mu_1 < \kappa_\gamma + \epsilon/2) \quad (51)$$

We set $\overline{\delta}_{jj} = \ell_j \widetilde{\delta}_3,\ j = 1, \ldots, \nu$ and $\overline{\delta}_{jk} = \sqrt{\ell_j \ell_k}\ \widetilde{\delta}_4,\ 1 \le j < k \le \nu$ where $\widetilde{\delta}_3, \widetilde{\delta}_4 > 0$ are such that $\overline{\delta}^2 = \widetilde{\delta}_3^2 \sum_{j=1}^{\nu} \ell_j^2 + \widetilde{\delta}_4^2 (\sum_{j=1}^{\nu} \ell_j)^2$. To be specific, we take $\widetilde{\delta}_4 = \frac{\overline{\delta}}{\sqrt{2}} (\sum_{j=1}^{\nu} \ell_j)^{-1}$ and $\widetilde{\delta}_3 = \frac{\overline{\delta}}{\sqrt{2}} (\sum_{j=1}^{\nu} \ell_j^2)^{-1/2}$.

Therefore by derivations similar to (39), we have for $0 < \widetilde{\delta}_4 < \min\{1, \frac{\kappa_\gamma + \epsilon/2}{\rho_\nu - 3\epsilon/2 - \kappa_\gamma}\}$,

$$\mathbb{P}(|\overline{D}_{\nu,jk}| \ge \overline{\delta}_{jk}, \mu_1 < \kappa_\gamma + \epsilon/2) \quad \le \quad 2 \exp\left(-\frac{n\widetilde{\delta}_4^2}{12}\right) + 2\exp\left(-\frac{n}{N-M} \frac{(\rho_\nu - 3\epsilon/2 - \kappa_\gamma)^2 n \widetilde{\delta}_4^2}{12(\kappa_\gamma + \epsilon/2)^2}\right) \quad \text{for all} \quad 1 \le j < k \le \nu \tag{52}$$

Similarly, $\exists\ n_2(\nu, \epsilon, \Lambda, \gamma)$ such that for $\widetilde{\delta}_3 < 4 \min\{1, \frac{\kappa_\gamma + \epsilon/2)}{\rho_\nu - 3\epsilon/2 - \kappa_\gamma}\}$,

$$\begin{aligned} & \mathbb{P}(|\overline{D}_{\nu,jj}| \ge \overline{\delta}_{jj}, \mu_1 < \kappa_\gamma + \epsilon/2) \\ \le \quad & 2\exp\left(-\frac{n\widetilde{\delta}_3^2}{96}\right) + 2\exp\left(-\frac{n}{N-M} \frac{(\rho_\nu - 3\epsilon/2 - \kappa_\gamma)^2 n\widetilde{\delta}_3^2}{96(\kappa_\gamma + \epsilon/2)^2}\right) \\ & +2\exp\left(-\frac{n}{n+N-M} \frac{n^2 \widetilde{\delta}_3^2}{2} \frac{(\rho_\nu - 3\epsilon/2 - \kappa_\gamma)^4}{64(\rho_\nu - \epsilon)^2 (\kappa_\gamma + \epsilon/2)}\right) \qquad \text{for } 1 \le j \le \nu \end{aligned} \tag{53}$$

for all $n \ge n_2(\nu, \epsilon, \Lambda, \gamma)$.

Proof of Theorem 2 : The remark following *Proposition 3* remains valid for *Proposition 4* as well (possibly with different constants). And so, the proof in the case when $\ell_\nu > (1+\sqrt{\gamma})^2$ now follows easily by combining *Proposition 1*, *Proposition 2*, *Proposition 3* and *Proposition 4* and applying first Borel-Cantelli lemma.

Proof for the general case is deduced by combining *Proposition 2*, *Proposition 3* and *Proposition B.2* and then applying first Borel-Cantelli lemma.

4.4 Proof of Theorem 1

By *interlacing inequality*, it follows that $\widehat{\ell}_\nu \ge \mu_\nu$. *Proposition 2* and the remark following that ensure that μ_ν concentrates around κ_γ. In view of this we only need to show that for every $\epsilon > 0$ the probability $\mathbb{P}(\widehat{\ell}_\nu > \kappa_\gamma + \epsilon)$ is summable over n, so that an Application of Borel-Cantelli lemma will complete the proof.

We take essentially the same approach as in proving *Proposition 3*. As before, we denote $\lambda_1(\mathbf{S}_{\Gamma_{\nu-1}})$ by $\widehat{\lambda}_{1,\nu}$ and use (21). So we only need to ensure that $\mathbb{P}(\widehat{\lambda}_{1,\nu} > \kappa_\gamma + \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2)$ is summable. As before, consider the set $\widetilde{J}_{1,\nu} := \{\widehat{\lambda}_{1,\nu} > \kappa_\gamma + \epsilon, \mu_1 < \kappa_\gamma + \epsilon/2\}$. Then, since $\widehat{\lambda}_{1,\nu}$ is an eigenvalue of $K_{\Gamma_{\nu-1}}(\widehat{\lambda}_{1,\nu})$, with $K_G(\cdot)$ defined by (22), we have

$$\widehat{\lambda}_{1,\nu} \le \lambda_1(K_{\Gamma_{\nu-1}}(\kappa_\gamma + \epsilon)) \qquad \text{on the set} \ \ \widetilde{J}_{1,\nu}. \tag{54}$$

Then on the set $\widetilde{J}_{1,\nu}$,

$$\widehat{\lambda}_{1,\nu} \le \ell_\nu(1 + \gamma \int \frac{x}{\kappa_\gamma + \epsilon - x} dF_\gamma(x)) + \parallel K_{\Gamma_{\nu-1}}(\kappa_\gamma + \epsilon) - \Lambda_{\Gamma_{\nu-1}}(\kappa_\gamma + \epsilon) \parallel_{HS} \tag{55}$$

since $\ell_\nu(1+\gamma\int\frac{x}{\kappa_\gamma+\epsilon-x}dF_\gamma(x))$ is the largest eigenvalue of $\Lambda_{\Gamma_{\nu-1}}(\kappa_\gamma+\epsilon)$, where the last quantity is defined through (19).

$\exists\ \delta := \delta(\epsilon) > 0$ such that

$$\ell_\nu(1+\gamma\int\frac{x}{\kappa_\gamma+\epsilon-x}dF_\gamma(x))+\delta=\kappa_\gamma+\epsilon \tag{56}$$

This is because, if we define $\kappa_{\gamma,\epsilon}=(\kappa_\gamma+\epsilon)(1+\gamma\int\frac{x}{\kappa_\gamma+\epsilon-x}dF_\gamma(x))^{-1}$, then $\kappa_{\gamma,\epsilon}>1+\sqrt{\gamma}\geq\ell_\nu$ since whenever $\rho>\kappa_\gamma$, $\exists$ a unique $\ell>1+\sqrt{\gamma}$ which solves the equation (3), so that

$$\ell_\nu(1+\gamma\int\frac{x}{\kappa_\gamma+\epsilon-x}dF_\gamma(x))=\frac{\ell_\nu}{\kappa_{\gamma,\epsilon}}(\kappa_\gamma+\epsilon)<\kappa_\gamma+\epsilon$$

As should be clear by now, the proof of *Proposition 3* can now be followed verbatim just by replacing the quantity $\rho_\nu+\epsilon$ by $\kappa_\nu+\epsilon$, and replacing $J_{1,\nu}$ by $\widetilde{J}_{1,\nu}$. Thus, skipping all the details we simply present the final result :

Proposition 5 : *Let $\ell_\nu\leq 1+\sqrt{\gamma}$. With $\widetilde{\delta}_1$ and $\widetilde{\delta}_2$ having the same definition as in the proof of Proposition 3, and $\delta=\delta(\epsilon)$ defined through (56) sufficiently small, $\exists\ n_3(\nu,\epsilon,\Lambda,\gamma)$ large enough so that for $n\geq n_3(\nu,\epsilon,\Lambda,\gamma)$*

$$\mathbb{P}(\widehat{\ell}_\nu>\kappa_\gamma+\epsilon,\mu_1<\kappa_\gamma+\epsilon/2)\leq(M-\nu+1)\overline{\varepsilon}_1(n,\epsilon,\Lambda,\gamma)+(M-\nu)(M-\nu+1)\overline{\varepsilon}_2(n,\epsilon,\Lambda,\gamma,) \tag{57}$$

where

$$\begin{aligned}\overline{\varepsilon}_1 &= 2\exp\left(-\frac{n\widetilde{\delta}_1^2}{96}\right)+2\exp\left(-\frac{n}{N-M}\frac{n\widetilde{\delta}_1^2\epsilon^2}{384(\kappa_\gamma+\epsilon/2)^2}\right)\\ &\quad+2\exp\left(-\frac{n}{n+N-M}\frac{n^2\widetilde{\delta}_1^2}{2}\frac{\epsilon^4}{1024(\kappa_\gamma+\epsilon)^2(\kappa_\gamma+\epsilon/2)}\right)\\ \overline{\varepsilon}_2 &= 2\exp\left(-\frac{n\widetilde{\delta}_2^2}{12}\right)+2\exp\left(-\frac{n}{N-M}\frac{n\widetilde{\delta}_2^2\epsilon^2}{48(\kappa_\gamma+\epsilon/2)^2}\right)\end{aligned}$$

Remark : It is important to take note of the behaviour of $\delta(\epsilon)$. When $\ell_\nu<1+\sqrt{\gamma}$, (56) implies that as $\epsilon\downarrow 0$, $\delta(\epsilon)\to\kappa_\gamma-\ell_\nu(1+\gamma\int\frac{x}{\kappa_\gamma-x}dF_\gamma(x))$, by Monotone Convergence Theorem. By *Lemma B.1*, the limit equals $\kappa_\gamma-\ell_\nu(1+\sqrt{\gamma})>0$. This shows that $\delta(\epsilon)$ is bounded below, and so the same is true for $\widetilde{\delta}_1$ and $\widetilde{\delta}_2$. However, if $\ell_\nu=1+\sqrt{\gamma}$ then $\delta(\epsilon)\to 0$ as $\epsilon\downarrow 0$ and therefore we need to do a more careful analysis to ensure that the bound in (57) is meaningful. Again using *Lemma B.1*, we can write, for $0 < \epsilon < 2\gamma$,

$$\begin{aligned}
\delta(\epsilon) &= \epsilon + (1+\sqrt{\gamma})\gamma\left[\int \frac{x}{\kappa_\gamma - x}dF_\gamma(x) - \int \frac{x}{\kappa_\gamma + \epsilon - x}dF_\gamma(x)\right] \\
&= \epsilon + (1+\sqrt{\gamma})\gamma\int_{(1-\sqrt{\gamma})^2}^{\kappa_\gamma} \frac{x\epsilon}{(\kappa_\gamma+\epsilon-x)(\kappa_\gamma - x)}\frac{1}{2\pi\gamma x}\sqrt{(\kappa_\gamma - x)(x-(1-\sqrt{\gamma})^2)}dx \\
&> \epsilon + \frac{\epsilon(1+\sqrt{\gamma})}{2\pi}\int_{\kappa_\gamma-\epsilon}^{\kappa_\gamma}\frac{\sqrt{\kappa_\gamma - \epsilon - (1-\sqrt{\gamma})^2}}{(\kappa_\gamma+\epsilon-x)\sqrt{\kappa_\gamma - x}}dx \\
&> \epsilon + \frac{\epsilon(1+\sqrt{\gamma})}{2\pi}\cdot\epsilon\cdot\frac{\sqrt{4\gamma-\epsilon}}{2\epsilon\sqrt{\epsilon}} \ > \ \sqrt{\epsilon}\left[\sqrt{\epsilon} + \frac{(1+\sqrt{\gamma})\sqrt{\gamma}}{2\sqrt{2}\pi}\right]
\end{aligned}$$

Which means that for $0 < \epsilon < \epsilon_0$, say, $\delta(\epsilon) > c\sqrt{\epsilon}$ for some constant $c > 0$ and so the bound (57) is strong enough.

Thus the proof of summability of $\mathbb{P}(\widehat{\ell}_\nu > \kappa_\gamma + \epsilon)$ is completed by combining *Proposition 5* with *Proposition 2*.

5 Proof of Theorem 3

The proof involves several parts. The first step is to utilize the eigen-equation (15) to get

$$\widehat{\ell}_\nu = b_\nu^T(\mathbf{S}_{AA} + \Lambda^{1/2}T^T\mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-1}T\Lambda^{1/2})b_\nu \tag{58}$$

Next step is to show that

$$b_\nu - \mathbf{e}_\nu = -[\mathcal{R}_\nu(K(\rho_\nu) - \frac{\rho_\nu}{\ell_\nu}\Lambda)\mathbf{e}_\nu + (\rho_\nu - \widehat{\ell}_\nu)\mathcal{R}_\nu\overline{K}_\nu(\rho_\nu)\mathbf{e}_\nu] + (\widehat{\ell}_\nu - \rho_\nu)^2 O_P(1) + o_P(n^{-1/2}) \tag{59}$$

where $K(x)$ is defined by (16), $\mathcal{R}_\nu$ is a deterministic diagonal matrix, and $\overline{K}(\rho_\nu)$ is a stochastic matrix with norm $O_P(n^{-1/2})$. This is done in *Section 6*. Then we can write (see *Section 6.1*), after expanding $K(\widehat{\ell}_\nu)$ appearing in (58) around ρ_ν, using (59), changing sides, and finally multiplying by $\sqrt{n}$,

$$\sqrt{n}(\widehat{\ell}_\nu - \rho_\nu)(1 + \ell_\nu t_\nu^T\mathcal{M}(\rho_\nu I - \mathcal{M})^{-2}t_\nu + d_\nu) = \sqrt{n}(s_{\nu\nu} + \ell_\nu t_\nu^T\mathcal{M}(\rho_\nu I - \mathcal{M})^{-1}t_\nu - \rho_\nu) + o_P(1) \tag{60}$$

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 (ν,ν)-th element of $\mathbf{S}$. It readily follows that $d_\nu = o_P(1)$. We first show that the term on the RHS of (60) converges in distribution to a Gaussian random variable with zero mean and variance given by

$$2\ell_\nu\rho_\nu\left(1 + \ell_\nu\gamma\int\frac{x}{(\rho_\nu - x)^2}dF_\gamma(x)\right) \tag{61}$$

Next, from *Proposition 6* stated below, it follows that

$$t_\nu^T\mathcal{M}(\rho_\nu I - \mathcal{M})^{-2}t_\nu \xrightarrow{a.s.} \gamma\int\frac{x}{(\rho_\nu - x)^2}dF_\gamma(x) \tag{62}$$

Hence (4), with $\sigma^2(\ell)$ given by the first expression in (5), follows from (61), (62) and (60) once we apply Slutsky's theorem. Application of (99) gives the second equality in (5) and the third follows from simple algebra.

Proposition 6 : *Suppose* $\frac{N}{n} \to \gamma \in (0,1)$ *as* $n \to \infty$. *Let* $\delta, \epsilon > 0$ *be such that* $\delta < \frac{16(\kappa_\gamma+\epsilon/2)}{\epsilon^2}$, *and* $\rho \geq \kappa_\gamma + \epsilon$. *Then* $\exists\ n_4(\rho,\delta,\epsilon,\gamma)$ *such that for all* $n \geq n_4(\rho,\delta,\epsilon,\gamma)$,

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

The proof of this proposition is given in *Appendix B*.

The main term on the RHS of (60) can be expressed as $W_n + W_n'$, where

$$W_n = \sqrt{n}(s_{\nu\nu} - (1-\gamma)\ell_\nu + \ell_\nu t_\nu^T \mathcal{M}(\rho_\nu I - \mathcal{M})^{-1} t_\nu - \ell_\nu \rho_\nu \frac{1}{n} trace((\rho_\nu I - \mathcal{M})^{-1}))$$

and

$$W_n' = \sqrt{n}\ell_\nu(\rho_\nu \frac{1}{n} trace((\rho_\nu I - \mathcal{M})^{-1}) - \frac{\gamma \ell_\nu}{\ell_\nu - 1})$$

Note that by (97),

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

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

Remark : The result of Bai and Silverstein (2004) is actually much stronger than what we need. They also show the result under fairly weak conditions. From their result one can deduce asymptotic normality of the sequence $\sqrt{n}W_n'$ if we replace γ by $\frac{N-M}{n}$. However, for our purpose we only need that $W_n' = o_P(1)$.

5.1 Asymptotic normality of W_n

First we recall that by definition of T, $t_\nu = \frac{1}{\sqrt{n}} H^T Z_{A,\nu}$ where $Z_{A,\nu}^T$ is the ν-th row of $\mathbf{Z}_A$. Since $N - M < n$, and columns of H are orthonormal, we can extend them to form 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, $\widetilde{H}\widetilde{H}^T = \widetilde{H}^T\widetilde{H} = I_n$. Then writing

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

with $w_\nu := \frac{1}{\sqrt{n}} H_c^T Z_{A,\nu}$, we have $w_\nu \sim N(0, \frac{1}{n} I_{n-N+M})$, $t_\nu \sim N(0, \frac{1}{n} I_{N-M})$, and these are mutually independent and independent of $\mathbf{Z}_B$ (since $\widetilde{H}$ is an orthonormal basis and $Z_{A,\nu} \sim N(0, I_n)$). Therefore we can decompose W_n as a sum of two independent random variables $W_{1,n}$ and $W_{2,n}$ where

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

Since $n \| w_\nu \|^2 \sim \chi^2_{n-N+M}$ and $\frac{N}{n} - \gamma = o(n^{-1/2})$, we get $W_{1,n} \Longrightarrow N(0, 2\ell_\nu^2(1-\gamma))$. In *Section 5.2* we prove that

$$W_{2,n} \Longrightarrow N(0, 2\ell_\nu^2 \gamma \int \frac{\rho_\nu^2}{(\rho_\nu - x)^2} dF_\gamma(x)). \tag{63}$$

The asymptotic normality of W_n is therefore established. Since $W_n' = o_P(1)$ this implies asymptotic normality of the RHS of (60). The expression (61) for asymptotic variance is then deduced as follows:

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

where in the last step we used (97). 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 (61) follows since $\ell_\nu(1 + \frac{\gamma}{\ell_\nu - 1}) = \rho_\nu$.

5.2 Proof of (63)

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

$$W_{2,n} = \ell_\nu \rho_\nu \frac{1}{\sqrt{n}} \left(\sum_{j=1}^{N-M} \frac{1}{\rho_\nu - \mu_j} y_j^2 - \sum_{j=1}^{N-M} \frac{1}{\rho_\nu - \mu_j} \right), \qquad \text{where } \{y_j\}_{j=1}^{N-M} \overset{i.i.d.}{\sim} N(0,1),$$

and $\{y_j\}_{j=1}^{N-M}$ are independent of $\mathcal{M}$. Thus, given $\mathcal{M}$, $W_{2,n}$ is a weighted sum of i.i.d. mean 0 random variables. To establish (63) we need to show that

$$\phi_{W_{2,n}}(t) := \mathbb{E}\exp(itW_{2,n}) \to \phi_{\widetilde{\sigma}^2(\ell_\nu)}(t) := \exp\left(-\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\right), \qquad \text{for all } t \in \mathbb{R}, \quad \text{as } n \to \infty$$

where $\widetilde{\sigma}^2(\ell) = 2\ell^2\gamma \int \frac{\rho^2(\ell)}{(\rho^2(\ell)-x)^2} dF_\gamma(x)$ for $\ell > 1+\sqrt{\gamma}$. It is enough to show that

$$\mathbb{E}\left|\mathbb{E}(e^{itW_{2,n}} \mid \mathcal{M}) \exp\left(\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\right) - 1\right| \to 0, \qquad \text{for all } t \in \mathbb{R}, \quad \text{as } n \to \infty$$

where the outer expectation is with respect to the distribution of $\mathcal{M}$. We break this expectation into two parts, one over the set $J_\gamma(\delta) := \{\mu_1 \leq \kappa_\gamma + \delta\}$ where $\delta > 0$ is any number such that $\rho_\nu > \kappa_\gamma + 2\delta$, and the complementary part over the set $J_\gamma^c(\delta) = \{\mu_1 > \kappa_\gamma + \delta\}$. Note that $J_\gamma(\delta)$ is a measurable set that depends on n and $\mathbb{P}(J_\gamma(\delta)) \to 1$ as $n \to \infty$. Since the inner expectation is a bounded r.v., the second term converges to zero. Thus we only need to establish that

$$\mathbb{E}\left[\left|\mathbb{E}(e^{itW_{2,n}} \mid \mathcal{M}) \exp\left(\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\right) - 1\right|, \mu_1 \leq \kappa_\gamma + \delta\right] \to 0, \qquad \text{for all } t \in \mathbb{R}, \quad \text{as } n \to \infty \tag{64}$$

Since characteristic function of a χ_1^2 random variable at any point t is given by $\psi(t) := \frac{1}{\sqrt{1-2it}}$, on the set $\{\mu_1 \leq \kappa_\gamma + \delta\}$ the inner conditional expectation is

$$\begin{aligned} &\prod_{j=1}^{N-M} \psi\left(\frac{t\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \mu_j)}\right) \exp\left(-\frac{it\ell_\nu\rho_\nu}{\sqrt{n}} \sum_{j=1}^{N-M} \frac{1}{\rho_\nu - \mu_j}\right) \\ = &\prod_{j=1}^{N-M} \left(1 - \frac{2it\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \mu_j)}\right)^{-1/2} \exp\left(-\frac{it\ell_\nu\rho_\nu}{\sqrt{n}} \sum_{j=1}^{N-M} \frac{1}{\rho_\nu - \mu_j}\right) \end{aligned} \tag{65}$$

Denoting by $\log z$ ($z \in \mathbb{C}$) the principal branch of the complex logarithm we have

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

Recalling the Taylor series expansion of $\log(1+z)$ (valid for $|z| < 1$), we can write, for $n \geq n_*(\nu,\gamma,\delta)$, large enough so that $\frac{|t|\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu - \kappa_\gamma - \delta)} < \frac{1}{2}$, the conditional expectation (65) as

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

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

$$\begin{aligned} &\frac{1}{2}\sum_{k=2}^{\infty}\frac{1}{k}\left(\frac{2it\ell_\nu\rho_\nu}{\sqrt{n}}\right)^k \sum_{j=1}^{N-M}\frac{1}{(\rho_\nu - \mu_j)^k} \\ = &-\frac{t^2}{2}\left[2\ell_\nu^2\rho_\nu^2\frac{1}{n}\sum_{j=1}^{N-M}\frac{1}{(\rho_\nu - \mu_j)^2}\right] + \frac{1}{2}\sum_{k=3}^{\infty}\frac{1}{k}\left(\frac{2it\ell_\nu\rho_\nu}{\sqrt{n}}\right)^k \sum_{j=1}^{N-M}\frac{1}{(\rho_\nu - \mu_j)^k} \end{aligned} \tag{66}$$

Denoting the first term of (66) by $a_n(t)$ and the second term by $\widetilde{r}_n(t)$, for $n \geq n_*(\nu, \gamma, \delta)$, on $J_\gamma(\delta)$,

$$\begin{aligned} |\widetilde{r}_n(t)| &\leq \frac{t^2}{3}\left[2\ell_\nu^2\rho_\nu^2\frac{1}{n}\sum_{j=1}^{N-M}\frac{1}{(\rho_\nu-\mu_j)^2}\right]\sum_{k=1}^{\infty}\left(\frac{2|t|\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu-\kappa_\gamma-\delta)}\right)^k \\ &= \frac{t^2}{3}\left[2\ell_\nu^2\rho_\nu^2\frac{1}{n}\sum_{j=1}^{N-M}\frac{1}{(\rho_\nu-\mu_j)^2}\right]\left(\frac{2|t|\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu-\kappa_\gamma-\delta)}\right)\left(1-\frac{2|t|\ell_\nu\rho_\nu}{\sqrt{n}(\rho_\nu-\kappa_\gamma-\delta)}\right)^{-1} \end{aligned} \quad (67)$$

Let $G_2(\cdot\,;\rho,\gamma,\delta)$ to be the bounded function (defined for $\rho > \kappa_\gamma + \delta$) defined through (90) in the appendix. Then on $J_\gamma(\delta)$, $\frac{1}{n}\sum_{j=1}^{N-M}\frac{1}{(\rho_\nu-\mu_j)^2} = \frac{N-M}{n}\int G_2(x;\rho_\nu,\gamma,\delta)d\widehat{F}_{n,N-M}(x)$ and the quantity on the RHS converges *almost surely* to $\gamma\int G_2(x;\rho_\nu,\gamma,\delta)dF_\gamma(x) = \gamma\int\frac{1}{(\rho_\nu-x)^2}dF_\gamma(x)$. 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 (65) and (66),

$$\begin{aligned} &\mathbb{E}\left[\left|\mathbb{E}(e^{itW_{2,n}} \mid \mathcal{M})\exp\left(\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\right)-1\right|, J_\gamma(\delta)\right] = \mathbb{E}\left[|\exp\left(a_n(t)+\widetilde{r}_n(t)+\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\right)-1|\mathbb{I}_{J_\gamma(\delta)}\right] \\ &\leq \mathbb{E}\left[\exp\left(a_n(t)+\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\right)(\exp(|\widetilde{r}_n(t)|)-1)\,\mathbb{I}_{J_\gamma(\delta)}\right]+\mathbb{E}\left[|\exp\left(a_n(t)+\frac{t^2\widetilde{\sigma}^2(\ell_\nu)}{2}\right)-1|\,\mathbb{I}_{J_\gamma(\delta)}\right] \\ &\to\ 0+0, \qquad \text{as } n\to\infty \end{aligned}$$

by bounded convergence theorem. Since $t \in \mathbb{R}$ is arbitrary, (64) follows.

6 Approximation to the eigenvectors

In this section we derive a first order asymptotic expansion of the vector b_ν associated with the eigenvalue $\widehat{\ell}_\nu$, when ℓ_ν is greater than $1+\sqrt{\gamma}$ and has multiplicity 1. This expansion has already been used in the proof of *Theorem 3*. We proceed with the standard perturbation analysis approach. Our construction follows Kneip and Utikal (2001), (see also Kato, 1980, Chapter 2). First observe that ρ_ν is *the* eigenvalue of $\frac{\rho_\nu}{\ell_\nu}\Lambda$ associated with *the* eigenvector $\mathbf{e}_\nu$. Define

$$\mathcal{R}_\nu = \sum_{k\neq\nu}^{M}\frac{\ell_\nu}{\rho_\nu(\ell_k-\ell_\nu)}\mathbf{e}_k\mathbf{e}_k^T \quad (68)$$

Note that $\mathcal{R}_\nu$ is the resolvent of $\frac{\rho_\nu}{\ell_\nu}\Lambda$ "evaluated" at ρ_ν. Then utilizing the defining equation (15) we can express

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

Defining $D_\nu = K(\widehat{\ell}_\nu)-\frac{\rho_\nu}{\ell_\nu}\Lambda_\nu$, premultiplying both sides by $\mathcal{R}_\nu$ and observing that $\mathcal{R}_\nu(\frac{\rho_\nu}{\ell_\nu}\Lambda-\rho_\nu I) = I_M-\mathbf{e}_\nu\mathbf{e}_\nu^T := P_\nu^\perp$ we get,

$$P_\nu^\perp b_\nu = -\mathcal{R}_\nu D_\nu b_\nu+(\widehat{\ell}_\nu-\rho_\nu)\mathcal{R}_\nu b_\nu \quad (69)$$

As a convention let us suppose $\langle \mathbf{e}_\nu, b_\nu \rangle \geq 0$. Then expressing $b_\nu = \langle \mathbf{e}_\nu, b_\nu \rangle \mathbf{e}_\nu + P_\nu^\perp b_\nu$ and observing that $\mathcal{R}_\nu \mathbf{e}_\nu = 0$, we get

$$b_\nu - \mathbf{e}_\nu = -\mathcal{R}_\nu D_\nu \mathbf{e}_\nu + r_\nu \tag{70}$$

where

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

Now, define

$$\alpha_\nu = \| \mathcal{R}_\nu D_\nu \| + |\widehat{\ell}_\nu - \rho_\nu| \, \| \mathcal{R}_\nu \| \qquad \text{and} \qquad \beta_\nu = \| \mathcal{R}_\nu D_\nu \mathbf{e}_\nu \| \tag{71}$$

Lemma 1 : r_ν *satisfies*

$$\| r_\nu \| \;\leq\; \begin{cases} \beta_\nu \left(\frac{\alpha_\nu(1+\alpha_\nu)}{1-\alpha_\nu(1+\alpha_\nu)} + \frac{\beta_\nu}{(1-\alpha_\nu(1+\alpha_\nu))^2} \right) & \text{if } \alpha_\nu < \frac{\sqrt{5}-1}{2} \\ \alpha_\nu^2 + 2\alpha_\nu & \text{always} \end{cases} \tag{72}$$

Proof : Rewriting (69) we get

$$P_\nu^\perp b_\nu = -\mathcal{R}_\nu D_\nu \mathbf{e}_\nu - \mathcal{R}_\nu D_\nu (b_\nu - \mathbf{e}_\nu) + (\widehat{\ell}_\nu - \rho_\nu)\mathcal{R}_\nu (b_\nu - \mathbf{e}_\nu) \tag{73}$$

From this

$$y_\nu := \| P_\nu^\perp b_\nu \| \;\leq\; \| \mathcal{R}_\nu D_\nu \mathbf{e}_\nu \| + (\| \mathcal{R}_\nu D_\nu \| + |\widehat{\ell}_\nu - \rho_\nu| \, \| \mathcal{R}_\nu \|) \, \| b_\nu - \mathbf{e}_\nu \| = \beta_\nu + \alpha_\nu \| b_\nu - \mathbf{e}_\nu \| \tag{74}$$

On the other hand, from the decomposition $b_\nu = \langle \mathbf{e}_\nu, b_\nu \rangle \mathbf{e}_\nu + P_\nu^\perp b_\nu$, and observing that

$$1 - \langle \mathbf{e}_\nu, b_\nu \rangle = 1 - \sqrt{1 - \| P_\nu^\perp b_\nu \|^2} \leq \| P_\nu^\perp b_\nu \|^2,$$

we also have

$$\| b_\nu - \mathbf{e}_\nu \| \;\leq\; \| P_\nu^\perp b_\nu \| \, (1 + \| P_\nu^\perp \mathbf{e}_\nu \|) = y_\nu(1 + y_\nu) \leq y_\nu(1 + \alpha_\nu)$$

where the last inequality is a result of the fact that from (69) one gets $\| P_\nu^\perp b_\nu \| \leq \alpha_\nu$. Substituting this in (74) we get $y_\nu \leq \beta_\nu + \alpha_\nu(1+\alpha_\nu)y_\nu$ implying that $y_\nu \leq \frac{\beta_\nu}{1-\alpha_\nu(1+\alpha_\nu)}$ whenever $\alpha_\nu < \frac{\sqrt{5}-1}{2}$. Therefore, if $\alpha_\nu < \frac{\sqrt{5}-1}{2}$ then $\| b_\nu - \mathbf{e}_\nu \| \;\leq\; \frac{\beta_\nu(1+\alpha_\nu)}{1-\alpha_\nu(1+\alpha_\nu)}$. Substituting this in the general bound

$$\| r_\nu \| \;\leq\; \| P_\nu^\perp b_\nu \|^2 + \alpha_\nu \| b_\nu - \mathbf{e}_\nu \| \tag{75}$$

and using the last two relationships, we get the first inequality in (72). The second inequality is a trivial consequence of (75).

Next task is to establish that $\beta_\nu = o_P(1)$ and $\alpha_\nu = o_P(1)$. First, consider the following decomposition.

$$\begin{aligned} D_\nu &= (\mathbf{S}_{AA} - \Lambda) + \Lambda^{1/2}\left(T^T\mathcal{M}(\rho_\nu I - \mathcal{M})^{-1}T - \frac{1}{n}\, trace(\mathcal{M}(\rho_\nu I - \mathcal{M})^{-1})I\right)\Lambda^{1/2} \\ &+ \left(\frac{1}{n}\, trace(\mathcal{M}(\rho_\nu I - \mathcal{M})^{-1}) - \gamma\int \frac{x}{\rho_\nu - x} dF_\gamma(x)\right)\Lambda \\ &\quad + (\rho_\nu - \widehat{\ell}_\nu)\Lambda^{1/2}T^T\mathcal{M}(\rho_\nu I - \mathcal{M})^{-1}(\widehat{\ell}_\nu I - \mathcal{M})^{-1}T\Lambda^{1/2} \end{aligned} \tag{76}$$

Since $\widehat{\ell}_\nu \stackrel{a.s.}{\to} \rho_\nu > \kappa_\gamma$ and $\mu_1 \stackrel{a.s.}{\to} \kappa_\gamma$, in view of the analysis carried out in *Section 4*, it is straightforward to see that $\| D_\nu \| \stackrel{a.s.}{\to} 0$. Therefore, $\alpha_\nu \stackrel{a.s.}{\to} 0$ and $\beta_\nu \stackrel{a.s.}{\to} 0$ from the definition (71). However, because of the special structure, we can get a much better bound for β_ν. For that we need to look at the term $\mathcal{R}_\nu D_\nu \mathbf{e}_\nu$ more closely, which we do now.

Define $V^{(i,\nu)} := T^T\mathcal{M}(\rho_\nu I - \mathcal{M})^{-i}T - \frac{1}{n} trace(\mathcal{M}(\rho_\nu I - \mathcal{M})^{-i})I$ for $i = 1, 2$. Expanding D_ν upto second order around ρ_ν, and observing that $\mathcal{R}_\nu \Delta \mathbf{e}_\nu = 0$ for *any* diagonal matrix Δ, we have

$$\begin{aligned} \mathcal{R}_\nu D_\nu \mathbf{e}_\nu &= \mathcal{R}_\nu(\mathbf{S}_{AA} - \Lambda)\mathbf{e}_\nu + \mathcal{R}_\nu\Lambda^{1/2}V^{(1,\nu)}\Lambda^{1/2}\mathbf{e}_\nu + (\rho_\nu - \widehat{\ell}_\nu)\mathcal{R}_\nu\Lambda^{1/2}V^{(2,\nu)}\Lambda^{1/2}\mathbf{e}_\nu \\ &\quad + (\rho_\nu - \widehat{\ell}_\nu)^2\left[\mathcal{R}_\nu\Lambda^{1/2}T^T\mathcal{M}(\rho_\nu I - \mathcal{M})^{-2}(\widehat{\ell}_\nu I - \mathcal{M})^{-1}T\Lambda^{1/2}\mathbf{e}_\nu\right] \\ &= \mathcal{R}_\nu(K(\rho_\nu) - \frac{\rho_\nu}{\ell_\nu}\Lambda)\mathbf{e}_\nu + (\rho_\nu - \widehat{\ell}_\nu)\mathcal{R}_\nu\overline{K}(\rho_\nu)\mathbf{e}_\nu + (\widehat{\ell}_\nu - \rho_\nu)^2\overline{r}_\nu \end{aligned} \tag{77}$$

where $\overline{K}(\rho_\nu) = \Lambda^{1/2}V^{(2,\nu)}\Lambda^{1/2}$ and $\overline{r}_\nu$ is the vector appearing inside square brackets the second line. From this expansion and the observations (i) all except the diagonal of the matrix $\Lambda^{1/2}T^T\mathcal{M}(\rho_\nu I - \mathcal{M})^{-i}T\Lambda^{1/2}\mathbf{e}_\nu$ is $O_P(n^{-1/2})$ for $i = 1, 2$ (obtained through an inequality similar to (38)), (ii) all except the diagonal of $\mathbf{S}_{AA}$ is $O_P(n^{-1/2})$, and (iii) $\mathcal{R}_\nu$ is diagonal with (ν,ν)-th entry equal to 0, it easily follows that

$$\beta_\nu = O_P(n^{-1/2}) + (\widehat{\ell}_\nu - \rho_\nu)^2 O_P(1). \tag{78}$$

Remark : The proof of (78) is somewhat long winded and deliberately so. Note that if we use (i) and (ii) above, the condition $\frac{N}{n} - \gamma = o(n^{-1/2})$, and a decomposition similar to the decomposition of (ν,ν)-th element of $\sqrt{n}(K(\rho_\nu) - \frac{\rho_\nu}{\ell_\nu}\Lambda)$ as $W_n + W_n'$, as in the proof of *Theorem 3*, then by Bai and Silverstein (2004, Theorem 1.1) we immediately get $\| D_\nu \| = O_P(n^{-1/2}) + (\widehat{\ell}_\nu - \rho_\nu)O_P(1)$ by considering a second order expansion of D_ν in the spirit of (76). However, the way we have done it, the bound in (78) is actually a concentration bound and does not depend on the asymptotic limit theorem of Bai and Silverstein (2004).

As a simple consequence of (78), *Lemma 1* and *Theorem 3* we get the following:

Corollary 1: *When $\ell_\nu > 1 + \sqrt{\gamma}$ and of multiplicity one, $b_\nu = \mathbf{e}_\nu + O_P(n^{-1/2})$.*

6.1 Explanation for expansion (60)

RHS of (58) can be written as

$$\mathbf{e}_\nu^T K(\widehat{\ell}_\nu)\mathbf{e}_\nu + 2\mathbf{e}_\nu^T K(\widehat{\ell}_\nu)(b_\nu - \mathbf{e}_\nu) + (b_\nu - \mathbf{e}_\nu)^T K(\widehat{\ell}_\nu)(b_\nu - \mathbf{e}_\nu) \tag{79}$$

First term in (79) is the major contributor in (60), since 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, by (70), (71), (72) and (78),

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

Finally, to check the negligibility of the second term in (79), we observe that by (70),

$$\mathbf{e}_\nu^T K(\widehat{\ell}_\nu)(b_\nu - \mathbf{e}_\nu) = -\mathbf{e}_\nu^T D_\nu \mathcal{R}_\nu D_\nu \mathbf{e}_\nu + \mathbf{e}_\nu^T K(\widehat{\ell}_\nu) r_\nu = -\mathbf{e}_\nu^T D_\nu \mathcal{R}_\nu D_\nu \mathbf{e}_\nu + o_P(n^{-1/2}) + (\widehat{\ell}_\nu - \rho_\nu)^2 o_P(1),$$

where in the last step we used (72) together with (78). Expanding $D_\nu \mathbf{e}_\nu$ as in (77), and using the definition of $\mathcal{R}_\nu$ we get the expression

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

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

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

6.2 Proof of Theorem 4

Part (a) : As a convention we choose $\langle \mathbf{p}_\nu, \widetilde{\mathbf{e}}_\nu \rangle \geq 0$. First note that with $p_{A,\nu}$ as in (8)

$$\langle \mathbf{p}_\nu, \widetilde{\mathbf{e}}_\nu \rangle = \langle p_{A,\nu}, \mathbf{e}_\nu \rangle = \sqrt{1 - R_\nu^2} \; \langle b_\nu, \mathbf{e}_\nu \rangle$$

Since $\beta_\nu \stackrel{a.s.}{\to} 0$, $\alpha_\nu \stackrel{a.s.}{\to} 0$, from (70) and (72), $\langle b_\nu, \mathbf{e}_\nu \rangle \stackrel{a.s.}{\to} 1$. Therefore, from (17), (62), *Theorem 2* and above display, we have

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

from which we get (6) after invoking *Lemma B.2.*

Part (b) : From (17) it is clear that in order that (7) holds, we need either $b_\nu^T \Lambda^{1/2} T^T \mathcal{M}(\widehat{\ell}_\nu I - \mathcal{M})^{-2} T \Lambda^{1/2} b_\nu \stackrel{a.s.}{\to} \infty$ or $\langle b_\nu, \mathbf{e}_\nu \rangle \stackrel{a.s.}{\to} 0$. Clearly, we can no longer use the perturbation analysis argument to study the behaviour of b_ν, since in this case $\widehat{\ell}_\nu \stackrel{a.s.}{\to} \kappa_\gamma$. However we shall 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.* This will prove the result.

Our approach will be to show that given $\epsilon > 0$, we can find a $C_\epsilon > 0$ such that the probability $\mathbb{P}(\lambda_{min}(E) \leq C_\epsilon)$ is summable over n and that $C_\epsilon \to \infty$ as $\epsilon \to 0$.

First, 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

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

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

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

since by *interlacing inequality* $\widehat{\ell}_\nu \geq \mu_\nu$. Thus, in view of *Proposition 5*, we only need to provide a lower bound for the smallest eigenvalue of $\overline{E}_\nu$. However, it will be more convenient to work with the matrix

$$\overline{E} = \sum_{j=1}^{N-M} \frac{\mu_j}{(\kappa_\gamma + \epsilon - \mu_j)^2} \mathbf{t}_j \mathbf{t}_j^T = T^T \mathcal{M}((\kappa_\gamma + \epsilon) I - \mathcal{M})^{-2} T \tag{80}$$

Proving summability of $\mathbb{P}(\lambda_{min}(\overline{E}) \leq C_\epsilon, \overline{J}_{1,\nu})$ suffices because it is easy to see that $\| \overline{E}_\nu - \overline{E} \| \stackrel{a.s.}{\to} 0$ as $n \to \infty$.

By *Proposition 6*, and calculations similar to those in deriving (38), respectively, given $\delta > 0$, such that $\delta < \frac{16(\kappa_\gamma + \epsilon/2)}{\epsilon^2}$, $\exists\, n_5(\delta, \epsilon, \gamma)$ such that for all $n \geq n_5(\delta, \epsilon, \gamma)$,

$$\begin{aligned} &\mathbb{P}(|t_j^T \mathcal{M}((\kappa_\gamma + \epsilon) I - \mathcal{M})^{-2} t_j - \gamma \int \frac{x}{(\kappa_\gamma + \epsilon - x)^2} dF_\gamma(x)| > \delta,\ \overline{J}_{1,\nu}) \\ \leq\ & 2\exp\left(-\frac{n}{N-M} \frac{n(\delta/4)^2 (\epsilon/2)^4}{6(\kappa_\gamma + \epsilon/2)^2}\right) + 2\exp\left(-\frac{n}{n+N-M} \frac{n^2(\delta/4)^2}{2} \frac{(\epsilon/2)^6}{16(\kappa_\gamma+\epsilon)^2(\kappa_\gamma + \epsilon/2)}\right) \\ & + 2\exp\left(-\frac{n}{n+N-M} \frac{n^2(\delta/4)^2}{2} \frac{(\epsilon/2)^4}{4(\kappa_\gamma + \epsilon/2)}\right), \qquad 1 \leq j \leq M, \end{aligned} \tag{81}$$

and

$$\mathbb{P}(|t_j^T \mathcal{M}((\kappa_\gamma + \epsilon) I - \mathcal{M})^{-2} t_k| > \delta,\ \overline{J}_{1,\nu}) \leq 2\exp\left(-\frac{n}{N-M} \frac{n\delta^2 (\epsilon/2)^4}{3(\kappa_\gamma + \epsilon/2)^2}\right), \qquad 1 \leq j \neq k \leq M. \tag{82}$$

If a symmetric matrix A is written as $A = B + C$ where B is a diagonal matrix with the same diagonal as A, then $|\lambda_{min}(A) - \lambda_{min}(B)| \leq \| C \|_{HS}$. If we denote the RHS of (81) and (82) by ε_5 and ε_6, respectively, then similar decomposition of the matrix $\overline{E}$ yields

$$\mathbb{P}(\lambda_{min}(\overline{E}) \leq \int \frac{x}{(\kappa_\gamma + \epsilon - x)^2} dF_\gamma(x) - \delta(1 + \sqrt{M(M-1)}),\ \overline{J}_{1,\nu}) \leq M\varepsilon_5 + \frac{M(M-1)}{2}\varepsilon_6 \quad (83)$$

for $0 < \delta < \frac{16(\kappa_\gamma + \epsilon/2)}{\epsilon^2}$, and for all $n \geq n_5(\delta, \epsilon, \gamma)$.

On the other hand, observe that if $0 < \epsilon < 2\gamma$, then

$$\begin{aligned} \int \frac{x}{(\kappa_\gamma + \epsilon - x)^2} dF_\gamma(x) &> \int_{\kappa_\gamma - \epsilon}^{\kappa_\gamma - \epsilon/2} \frac{x}{(\kappa_\gamma + \epsilon - x)^2} f_\gamma(x) dx \\ = \frac{1}{2\pi\gamma} \int_{\kappa_\gamma - \epsilon}^{\kappa_\gamma - \epsilon/2} \frac{\sqrt{(\kappa_\gamma - x)(x - (1 - \sqrt{\gamma})^2)}}{(\kappa_\gamma + \epsilon - x)^2} dx &> \frac{1}{2\pi\gamma} \left[\frac{1}{(2\epsilon)^2} \sqrt{\frac{\epsilon}{2}} \sqrt{\kappa_\gamma - \epsilon - (1 - \sqrt{\gamma})^2} \right] \frac{\epsilon}{2} \\ = \frac{1}{16\sqrt{2}\pi\gamma} \frac{\sqrt{4\gamma - \epsilon}}{\sqrt{\epsilon}} &> \frac{1}{16\sqrt{\gamma}\pi} \frac{1}{\sqrt{\epsilon}} \end{aligned}$$

Therefore set $\delta = \epsilon(1 + \sqrt{M(M-1)})^{-1}$ and choose ϵ small enough so that $\frac{\sqrt{\gamma}}{16\pi} \frac{1}{\sqrt{\epsilon}} - \epsilon > 0$. Call the last quantity C_ϵ and observe that C_ϵ satisfies the requirement : $C_\epsilon \to \infty$ as $\epsilon \to 0$. By (83) the result follows.

7 Appendix A

7.1 Weak concentration inequalities for random quadratic forms

The following two lemmas will be referred to as *weak concentration inequalities.*

Suppose $C : \mathcal{X} \to \mathbb{R}^{n \times n}$ is a measureable 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 X and Y are i.i.d. $N_n(0, I)$ independent of Z. Then for every $L > 0$ and $0 < \delta < 1$,*

$$\mathbb{P}(\frac{1}{n}|X^T C(Z) Y| > t, \| C(Z) \| \leq L) \leq 2\exp\left(-\frac{(1-\delta)nt^2}{2L^2}\right), \quad \text{for } 0 < t < \frac{\delta}{1-\delta}L \qquad (84)$$

Lemma A.2 : *Suppose X is distributed as $N_n(0, I)$ independent of Z. Also let $C(z) = C^T(z)$ for all $z \in \mathcal{X}$. Let trace(B) denote the trace of a square matrix B. Then, for every $L > 0$ and $0 < \delta < 1$,*

$$\mathbb{P}(\frac{1}{n}|X^T C(Z) X - trace(C(Z))| > t, \| C(Z) \| \leq L) \leq 2\exp\left(-\frac{(1-\delta)nt^2}{4L^2}\right), \quad \text{for } 0 < t < \frac{2\delta}{1-\delta}L \qquad (85)$$

Proof of Lemma A.1 : In the proof for convenience we occassionally write C instead of $C(Z)$. Let $0 < \lambda < \frac{\delta}{L}$. Then if $Z \in D_L$ with $D_L := \{z : \| C(z) \| \leq L\}$,

$$\begin{aligned}
\mathbb{P}(\frac{1}{n}X^T C(Z)Y > t|Z) &\leq e^{-n\lambda t}\mathbb{E}\left[e^{\lambda X^T C(Z)Y}|Z\right] \\
&= e^{-n\lambda t}\mathbb{E}\left[\exp\left(\frac{\lambda^2}{2} \| CY \|^2\right)|Z\right] \\
&= e^{-n\lambda t}(2\pi)^{-n/2}\int_{\mathbb{R}^n}\exp\left(-\frac{1}{2}y^T(I-\lambda^2 C^T C)y\right)dy \\
&= e^{-n\lambda t}\det(I-\lambda^2 C^T C)^{-1/2} \qquad (86)
\end{aligned}$$

The last step is justified by the fact that $\lambda^2 \| C^T C \| \leq \lambda^2 L^2 < 1$ (by choice of λ) so the matrix $I - \lambda^2 C^T C$ is positive definite on D_L. Now, use the fact that $\log\det(I - \lambda^2 C^T C) = \sum_{i=1}^n \log(1 - \lambda^2\sigma_i^2(C))$, where $\sigma_i(C)$ is the i-th largest singular value of C. Thus, since $Z \in D_L$,

$$-\log\det(I-\lambda^2 C^T C) = \sum_{i=1}^n\sum_{k=1}^\infty \frac{1}{k}(\lambda^2\sigma_i^2(C))^k \leq \lambda^2(\sum_{i=1}^n \sigma_i^2(C))\sum_{k=1}^\infty\frac{1}{k}(\lambda L)^{2(k-1)} < n\lambda^2 L^2\sum_{k=0}^\infty(\lambda^2 L^2)^k$$

The geometric series in the last term converges for $\lambda < \frac{1}{L}$ and hence combining with (86) we get, for $0 < \delta < 1$,

$$\mathbb{P}(\frac{1}{n}X^T C(Z)Y > t|Z) \leq \inf_{0<\lambda<\frac{\delta}{L}} e^{-n\lambda t + \frac{1}{2(1-\delta^2)}n\lambda^2 L^2} < \inf_{0<\lambda<\frac{\delta}{L}} e^{-n\lambda t+\frac{1}{2(1-\delta)}n\lambda^2 L^2} \qquad (87)$$

The function $f_t(\lambda) := -n\lambda t + \frac{1}{2(1-\delta)}n\lambda^2 L^2$ achieves its global minimum at $\lambda_t = \frac{t(1-\delta)}{L^2}$. Therefore if $t < \frac{\delta L}{1-\delta}$ then $\lambda_t < \frac{\delta}{L}$ so that we get the upper bound $\exp(f_t(\lambda_t)) = -\frac{(1-\delta)nt^2}{2L^2}$ in (87) for $Z \in D_L$. By symmetry, the same upper bound holds for $\mathbb{P}(\frac{1}{n}X^T C(Z)Y < -t|Z)$ and combining these two and then taking expectation w.r.t. the distribution of Z over the set D_L we get (84).

Proof of Lemma A.2 : As in the proof of *Lemma A.1*, for $Z \in D_L$, $0 < \lambda < \frac{\delta}{L}$, and $t > 0$,

$$\begin{aligned}
\mathbb{P}(\frac{1}{n}(X^T C(Z)X - trace(C(Z))) > t|Z) &\leq e^{-\frac{\lambda}{2}(nt+trace(C(Z)))}\mathbb{E}\left[e^{\frac{\lambda}{2}X^T C(Z)X}|Z\right] \\
&= e^{-\frac{\lambda}{2}(nt+trace(C))}\det(I-\lambda C)^{-1/2}, \qquad (88)
\end{aligned}$$

where in the second step we use the fact that $I - \lambda C$ is positive definite. Denoting the eigenvalues of $C(Z)$ in decreasing order by $\mu_i(C)$, we have (since by assumption $\| C \| \leq L < \frac{1}{\lambda}$)

$$\begin{aligned}
-\log\det(I-\lambda C) - \lambda\, trace(C) &= \sum_{i=1}^n\sum_{k=1}^\infty\frac{1}{k}(\lambda\mu_i(C))^k - \lambda\, trace(C) = \sum_{k=2}^\infty\frac{1}{k}\lambda^k\sum_{i=1}^n(\mu_i(C))^k \\
&\leq \frac{n\lambda^2 L^2}{2}\sum_{k=2}^\infty\frac{2}{k}(\lambda L)^{k-2} \leq \frac{n\lambda^2 L^2}{2}\sum_{k=0}^\infty(\lambda L)^k,
\end{aligned}$$

where the inequality in the third step is by $\| C(Z) \| = \max\{|\mu_1|, |\mu_n|\}$. Now the rest of the proof simply retraces the argument of the proof of *Lemma A.1* and is therefore omitted.

7.2 Concentration inequalities for Lipschitz functionals of random matrices

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

Lemma A.3 : *Suppose* $\mathbf{Y}$ *is an* $m \times n$ *matrix,* $m \leq n$*, with independent (real or complex) entries* Y_{kl} *following law* P_{kl}*,* $1 \leq k \leq m, 1 \leq l \leq n$*. Let* $\mathbf{S}_\Delta = \mathbf{Y}\Delta\mathbf{Y}^*$ *be a generalized Wishart matrix where* Δ *is a diagonal matrix with real, nonnegative diagonal entries and spectral radius* $\phi_\Delta > 0$*. Suppose the family* $\{P_{kl} : 1 \leq k \leq m, 1 \leq l \leq n\}$ *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* $\delta > 0$*,*

$$\mathbb{P}\left(|\frac{1}{m}trace\ f(\frac{1}{m+n}\mathbf{S}_\Delta) - \mathbb{E}(\frac{1}{m}trace\ f(\frac{1}{m+n}\mathbf{S}_\Delta))| > \delta\right) \leq 2\exp\left(-\frac{m^2\delta^2}{2c\phi_\Delta|g|_{\mathcal{L}}^2}\right) \tag{89}$$

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

In order to apply this result to our context we take $m = N - M$, $\mathbf{Y} = \mathbf{Z}_B$ and $\Delta = \frac{m+n}{n}I_n$, and recall that $N(0,1)$ satisfies logarithmic Sobolev inequality with constant $c = 1$ (Bogachev, 1998, Theorem 1.6.1). Then define $f_k(x) = G_k(x;\rho,\gamma,\epsilon)$, $k = 1,2$, where $G_k(x;\rho,\gamma,\epsilon)$ is defined in (90), and $g_k(x) = f_k(x^2)$, and notice that $g_k(x)$ is Lipschitz with $|g_k|_{\mathcal{L}} = \frac{2k(\kappa_\gamma+\epsilon)^{1/2}}{(\rho-\kappa_\gamma-\epsilon)^{k+1}}$. Further, $\phi_\Delta = \frac{m+n}{n}$ and $\mathbf{S}_\Delta = (m+n)\mathbf{S}_{BB}$.

$$G_k(x;\rho,\gamma,\epsilon) = \begin{cases} \frac{1}{(\rho-x)^k} & x \leq \kappa_\gamma + \epsilon \\ \frac{1}{(\rho-\kappa_\gamma-\epsilon)^k} & x > \kappa_\gamma+\epsilon \end{cases} \quad \text{where} \quad \rho > \kappa_\gamma + \epsilon, \quad k = 1,2,\ldots \tag{90}$$

Therefore, applying *Lemma A.3* we get the following :

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

$$\begin{aligned} &\mathbb{P}\left(|\frac{1}{n}trace\ G_k(\mathbf{S}_{BB};\rho,\gamma,\epsilon) - \mathbb{E}(\frac{1}{n}trace\ G_k(\mathbf{S}_{BB};\rho,\gamma,\epsilon))| > \delta\right) \\ \leq\ & 2\exp\left(-\frac{n}{n+N-M}\frac{n^2\delta^2}{2}\frac{(\rho-\kappa_\gamma-\epsilon)^{2(k+1)}}{4k^2(\kappa_\gamma+\epsilon)}\right) \\ =\ & 2\exp\left(-\frac{n^2\delta^2}{2(1+\gamma)}\frac{(\rho-\kappa_\gamma-\epsilon)^{2(k+1)}}{4k^2(\kappa_\gamma+\epsilon)}(1+o(1))\right) \end{aligned} \tag{91}$$

7.3 Proof of Proposition 2

If we denote the singular values of $\mathbf{Z}_B$ by $\sigma_1(\mathbf{Z}_B) > \sigma_2(\mathbf{Z}_B) > \ldots > \sigma_{N-M}(\mathbf{Z}_B)$. Using a concentration inequality for singular values of random matrices (Ledoux, 2001),

$$\mathbb{P}(|\sigma_i(\mathbf{Z}_B) - m(\sigma_i(\mathbf{Z}_B))| > r) \leq 2e^{-\frac{r^2}{4}}, \quad r > 0, \quad 1 \leq i \leq N-M \tag{92}$$

where $m(\sigma_i(\mathbf{Z}_B))$ is a median of $\sigma_i(\mathbf{Z}_B)$. In this case, since $N - M < n$ (at least for sufficiently large n), the distribution of $\sigma_i(\mathbf{Z}_B)$ is continuous so that the median is unique.

Now, since $\mu_i := \lambda_i(\mathbf{S}_{BB}) = \frac{1}{n}(\sigma_i(\mathbf{Z}_B))^2$ and $\sigma_i(\mathbf{Z}_B) \geq 0$ so that

$$m(\mu_i) := m(\lambda_i(\mathbf{S}_{BB})) = m(\sigma_i(\mathbf{S}_{BB})) = \frac{1}{n} m(\sigma_i(\mathbf{Z}_B))^2,$$

it follows that for $r > 0$ and every $i = 1, \ldots, N-M$,

$$2e^{-\frac{nr^2}{4}} \quad \geq \quad \mathbb{P}(|\frac{1}{\sqrt{n}}\sigma_i(\mathbf{Z}_B) - \frac{1}{\sqrt{n}}m(\sigma_i(\mathbf{Z}_B))| > r) \quad \geq \quad \mathbb{P}(|\mu_i - m(\mu_i)| > r(2\sqrt{m(\mu_i)} + r)) \quad (93)$$

The last inequality follows from the fact that for real numbers $x, y \in \mathbb{R}_+$, on the set $|x - y| \leq r$, we have $|x^2 - y^2| = |x - y|(x + y) \leq r(2y + r)$, and then taking $x = \frac{1}{\sqrt{n}}\sigma_i(\mathbf{Z}_B)$ and $y = \frac{1}{\sqrt{n}}m(\sigma_i(\mathbf{Z}_B))$.

Denoting $m(\mu_i)$ by m_i for convenience, set $s = r(2\sqrt{m_i} + r) = (r + \sqrt{m_i})^2 - m_i$. Then solving for r we get for $s > 0$, $r = \sqrt{s + m_i} - \sqrt{m_i}$. Substituting in the last display we get

$$\mathbb{P}(|\mu_i - m_i| > s) \leq 2e^{-\frac{n}{4}(\sqrt{s+m_i} - \sqrt{m_i})^2}, \quad s > 0, \quad i = 1, \ldots, N-M. \quad (94)$$

The next step in the proof of *Proposition 2* is to use the following results on the weak convergence of the largest eigenvalue in the identity covariance case. The limiting distribution $\mathbb{F}_1$ is the so-called Tracy-Widom law of order 1.

Result [Johnstone (2001a)] : *When* $\frac{N}{n} \to \gamma \in (0, 1)$, *under the assumption of normality*

$$\gamma^{-1/2}(1 + \sqrt{\gamma})^{-4/3}(N - M)^{2/3}\left[\mu_1 - \left(1 + \sqrt{\frac{N - M}{n}}\right)^2\right] \Longrightarrow \mathbb{F}_1. \quad (95)$$

By (95) it follows that $\gamma^{-1/2}(1 + \sqrt{\gamma})^{-4/3}N^{2/3}(m_1 - (1 + \sqrt{\frac{N-M}{n}})^2) \to m(\mathbb{F}_1)$ where $m(\mathbb{F}_1)$ is the median of $\mathbb{F}_1$. In particular,

$$m_1 - \kappa_\gamma = O(|\frac{N}{n} - \gamma|) + O(n^{-2/3}) \quad (96)$$

Now to complete the proof of *Proposition 2* observe that

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

for $n \geq n_0(\gamma, \delta)$. Now since $\delta < \kappa_\gamma/2$ implies $\kappa_\gamma + \frac{\delta}{4} < \frac{9}{8}\kappa_\gamma$, for $n \geq n_0(\gamma, \delta)$ we have $\sqrt{s + m_1} - \sqrt{m_1} > \frac{\sqrt{2}s}{3\sqrt{\kappa_\gamma}}$. Therfore, choosing $s = \frac{3\delta}{4}$ and substituting in (94) we get the result after applying the condition $|m_1 - \kappa_\gamma| \leq \frac{\delta}{4}$. Note also that by (96) we also get the rate at which $n_0(\gamma, \delta)$ should grow as $\delta \downarrow 0$.

8 Appendix B

8.1 Expression for ρ_ν

Lemma B.1 : *For* $\ell_\nu \geq 1 + \sqrt{\gamma}$, $\rho_\nu = \ell_\nu\left(1 + \frac{\gamma}{\ell_\nu - 1}\right)$ *solves (3).*

Proof : We prove this by showing that for any $\ell > 1+\sqrt{\gamma}$ we have

$$\int \frac{x}{\rho(\ell)-x} f_\gamma(x)dx = \frac{1}{\ell-1} \tag{97}$$

where $\rho(\ell) = \ell\left(1+\frac{\gamma}{\ell-1}\right)$, and $f_\gamma(x)$ is the density of Marchenko-Pastur law with parameter $\gamma(\leq 1)$ and is given by

$$f_\gamma(x) = \frac{1}{2\pi\gamma x}\sqrt{(b(\gamma)-x)(x-a(\gamma))}\mathbb{I}(a(\gamma)\leq x\leq b(\gamma)), \text{ where } a(\gamma) = (1-\sqrt{\gamma})^2,\ b(\gamma) = (1+\sqrt{\gamma})^2$$

The LHS of (97) is equal to

$$\begin{aligned}
& \frac{1}{2\pi\gamma}\int_{a(\gamma)}^{b(\gamma)} \frac{\sqrt{(b(\gamma)-x)(x-a(\gamma))}}{\rho(\ell)-x}dx \\
= & \frac{1}{2\pi\gamma}\int_{-2\sqrt{\gamma}}^{2\sqrt{\gamma}} \frac{\sqrt{(2\sqrt{\gamma}-y)(y+2\sqrt{\gamma})}}{\rho(\ell)-(1+\gamma)-y}dy, \qquad \text{(setting } y = x-(1+\gamma))
\end{aligned}$$

Since $\rho(\ell)-(1+\gamma) = (\ell-1)+\frac{\gamma}{\ell-1}$, setting $K = \ell-1$ we can rewrite the last expression as

$$\begin{aligned}
& \frac{K}{2\pi\gamma}\int_{-2\sqrt{\gamma}}^{2\sqrt{\gamma}} \frac{\sqrt{4\gamma-y^2}}{K^2+\gamma-Ky}dy \\
= & \frac{2K}{\pi}\int_{-1}^{1} \frac{\sqrt{1-z^2}}{K^2+\gamma-2K\sqrt{\gamma}z}dz, \qquad \text{(setting } z = \frac{y}{2\sqrt{\gamma}}) \\
= & \frac{2K}{\pi}\left[\int_0^1 \sqrt{1-z^2}\left(\frac{1}{K^2+\gamma-2K\sqrt{\gamma}z}+\frac{1}{K^2+\gamma+2K\sqrt{\gamma}z}\right)dz\right] \\
= & \frac{4K(K^2+\gamma)}{\pi}\int_0^1 \frac{\sqrt{1-z^2}}{(K+\gamma)^2-4K^2\gamma z^2}dz \\
= & \frac{4K(K^2+\gamma)}{\pi}\int_0^{\pi/2} \frac{\cos^2\theta}{(K+\gamma)^2-4K^2\gamma\sin^2\theta}d\theta, \qquad \text{(setting } \sin\theta = z) \\
= & \left(\frac{K^2+\gamma}{K\gamma}\right)\frac{1}{\pi}\int_0^{\pi/2} \frac{(K^2+\gamma)^2-4K^2\gamma\sin^2\theta-((K^2+\gamma)^2-4K^2\gamma)}{(K^2+\gamma)^2-4K^2\gamma\sin^2\theta}d\theta
\end{aligned}$$

Substituting the formula for indefinite integral (for $a^2 > b^2$)

$$\int \frac{dx}{a^2-b^2\sin^2 cx} = \frac{1}{ac\sqrt{a^2-b^2}}\tan^{-1}\left(\frac{\sqrt{a^2-b^2}\tan cx}{a}\right) \tag{98}$$

and then using the fact that $\tan 0 = 0$ and $\tan\frac{\pi}{2} = \infty$, the last expression equals

$$\left(\frac{K^2+\gamma}{K\gamma}\right)\frac{1}{\pi}\left[\frac{\pi}{2}-\frac{(K^2-\gamma)^2}{(K^2+\gamma)(K^2-\gamma)}\frac{\pi}{2}\right] = \frac{1}{2}\left(\frac{K^2+\gamma}{K\gamma}\right)\frac{(K^2+\gamma)-(K^2-\gamma)}{K^2+\gamma} = \frac{1}{K}$$

thus completing the proof for the case $\ell > 1+\sqrt{\gamma}$. The case $\ell = 1+\sqrt{\gamma}$ follows from this by applying *Monotone convergence theorem* to the nonnegative functions $\{\frac{x}{\rho(\ell)-x}\mathbb{I}(a(\gamma) < x < b(\gamma)) : \ell \geq 1+\sqrt{\gamma}\}$, since $\rho(\ell)$ is monotonically increasing in $\ell \in [1+\sqrt{\gamma},\infty)$.

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

$$\int \frac{x}{(\rho(\ell)-x)^2} dF_\gamma(x) = \frac{1}{(\ell-1)^2-\gamma} \tag{99}$$

Proof : Just as in the proof of *Lemma B.1*, after substituting $z = (2\sqrt{\gamma})^{-1}(x-(1+\gamma))$, and letting $K = \ell - 1$ we get

$$\begin{aligned}
&\int \frac{x}{(\rho(\ell)-x)^2} dF_\gamma(x) \\
= \;& \frac{2K^2}{\pi}\int_{-1}^{1} \frac{\sqrt{1-z^2}}{(K^2+\gamma-2K\sqrt{\gamma}z)^2} dz \\
= \;& \frac{2K^2}{\pi}\int_{0}^{1} \sqrt{1-z^2}\left(\frac{1}{(K^2+\gamma-2K\sqrt{\gamma}z)^2}+\frac{1}{(K^2+\gamma+2K\sqrt{\gamma}z)^2}\right) dz \\
= \;& \frac{2K^2}{\pi}\int_{0}^{1} \frac{2((K^2+\gamma)^2+4K^2\gamma z^2)\sqrt{1-z^2}}{((K^2+\gamma)^2-4K^2\gamma z^2)^2} dz \\
= \;& \frac{4K^2}{\pi}\int_{0}^{1} \frac{2(K^2+\gamma)^2\sqrt{1-z^2}}{((K^2+\gamma)^2-4K^2\gamma z^2)^2} dz - \frac{4K^2}{\pi}\int_{0}^{1} \frac{\sqrt{1-z^2}}{(K^2+\gamma)^2-4K^2\gamma z^2} dz \\
= \;& \frac{8K^2(K^2+\gamma)^2}{\pi}\int_{0}^{\pi/2} \frac{\cos^2\theta d\theta}{((K^2+\gamma)^2-4K^2\gamma\sin^2\theta)^2} - \frac{1}{K^2+\gamma}, \quad \text{setting } \sin\theta = z \text{ and by (97)} \\
= \;& \frac{8K^2(K^2+\gamma)^2}{\pi}\frac{1}{4K^2\gamma}\int_{0}^{\pi/2} \frac{(K^2+\gamma)^2-4K^2\gamma\sin^2\theta-((K^2+\gamma)^2-4K^2\gamma)}{((K^2+\gamma)^2-4K^2\gamma\sin^2\theta)^2} d\theta - \frac{1}{K^2+\gamma} \\
= \;& \frac{2(K^2+\gamma)^2}{\pi\gamma}\left[\int_{0}^{\pi/2} \frac{d\theta}{(K^2+\gamma)^2-4K^2\gamma\sin^2\theta} - \int_{0}^{\pi/2} \frac{(K^2-\gamma)^2 d\theta}{((K^2+\gamma)^2-4K^2\gamma\sin^2\theta)^2}\right] - \frac{1}{K^2+\gamma} \\
= \;& \frac{2(K^2+\gamma)^2}{\pi\gamma}\frac{1}{(K^2+\gamma)(K^2-\gamma)}\frac{\pi}{2} - \frac{1}{2(K^2+\gamma)} \\
& - \frac{2(K^4-\gamma^2)^2}{\pi\gamma}\int_{0}^{\pi/2} \frac{d\theta}{((K^2+\gamma)^2-4K^2\gamma\sin^2\theta)^2}, \\
= \;& \frac{1}{\gamma}\left(\frac{K^2+\gamma}{K^2-\gamma}\right) - \frac{1}{K^2+\gamma} - \frac{(K^4-\gamma^2)^2}{\pi\gamma}\int_{0}^{\pi} \frac{d\phi}{(K^4+\gamma^2+2K^2\gamma\cos\phi)^2}
\end{aligned} \tag{100}$$

where eighth equality is due to (98) and in the last step we used $\cos 2\theta = 1-2\sin^2\theta$ before setting $\phi = 2\theta$. Since for $a > b$ we have

$$\begin{aligned}
\int \frac{dx}{(a+b\cos x)^2} &= -\frac{b\sin x}{(a^2-b^2)(a+b\cos x)} + \frac{a}{a^2-b^2}\int \frac{dx}{a+b\cos x} \\
&= -\frac{b\sin x}{(a^2-b^2)(a+b\cos x)} + \frac{2a}{(a^2-b^2)^{3/2}}\tan^{-1}\left(\frac{\sqrt{a^2-b^2}\tan\frac{x}{2}}{a+b}\right)
\end{aligned}$$

setting $a = K^4+\gamma^2$ and $b = 2K^2\gamma$ we get

$$\begin{aligned}
\frac{(K^4-\gamma^2)^2}{\pi\gamma}\int_0^{\pi} \frac{d\phi}{(K^4+\gamma^2+2K^2\gamma\cos\phi)^2} &= \frac{(K^4-\gamma^2)^2}{\pi\gamma}\frac{2(K^4+\gamma^2)}{((K^4+\gamma^2)^2-4K^4\gamma^2)^{3/2}}\frac{\pi}{2} \\
&= \frac{1}{\gamma}\left(\frac{K^4+\gamma^2}{K^4-\gamma^2}\right)
\end{aligned}$$

Substituting the last expression in (100) we get

$$\begin{aligned}\int \frac{x}{(\rho(\ell)-x)^2} dF_\gamma(x) &= \frac{1}{\gamma}\left(\frac{K^2+\gamma}{K^2-\gamma}\right) - \frac{1}{K^2+\gamma} - \frac{1}{\gamma}\left(\frac{K^4+\gamma^2}{K^4-\gamma^2}\right) \\ &= \frac{(K^2+\gamma)^2 - \gamma(K^2-\gamma) - (K^4+\gamma^2)}{\gamma(K^2-\gamma)(K^2+\gamma)} = \frac{1}{K^2-\gamma}\end{aligned}$$

8.2 Lower bound on eigenvalues in the general case

In this section we provide a lower bound for the sample eigenvalues $\widehat{\ell}_\nu$ which holds with high probability when $\ell_\nu > 1+\sqrt{\gamma}$. It will be more useful to provide a lower bound for $\widehat{\lambda}_{\nu,\nu} = \lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu})$. We do that using equations (14) and (15) and the observation that for any $\nu \geq 1$

$$\sum_{k=1}^{\nu} \lambda_k(\mathbf{S}_{\overline{\Gamma}_\nu}) = \max_{L \in \mathcal{O}_{\nu,N-M+\nu}} \textit{trace}\ (L^T \mathbf{S}_{\overline{\Gamma}_\nu} L) \tag{101}$$

where $\mathcal{O}_{\nu,N-M+\nu}$ is the set of $(N-M+\nu)\times\nu$ matrices whoses columns are orthonormal. Thus, our approach is to construct an appropriate L_ν for every ν with $\ell_\nu > 1+\sqrt{\gamma}$ such that the lower bound $\textit{trace}\ (L_\nu^T \mathbf{S}_{\overline{\Gamma}_\nu} L_\nu)$ is close to $\sum_{k=1}^{\nu} \rho_k$.

Thereafter by utilizing (21) we have

$$\sum_{k=1}^{\nu-1} \lambda_1(\mathbf{S}_{\Gamma_{k-1}}) + \lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu}) \geq \sum_{k=1}^{\nu-1} \widehat{\ell}_k + \lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu}) \geq \sum_{k=1}^{\nu} \lambda_k(\mathbf{S}_{\overline{\Gamma}_\nu}) \tag{102}$$

We construct the $(N-M+\nu)\times\nu$ matrix L_ν as follows. Let $\widetilde{R}_k, k=1,\dots,\nu$ be numbers between 0 and 1 to be specified. Write

$$L_\nu = \begin{bmatrix} L_{A,\nu} \\ L_{B,\nu} \end{bmatrix} \qquad \text{where} \qquad L_{A,\nu} = \textit{diag}(\sqrt{1-\widetilde{R}_1^2},\dots,\sqrt{1-\widetilde{R}_\nu^2})$$

and $L_{B,\nu} = V\widetilde{\Xi}\widetilde{D}$ where $\widetilde{D} = \textit{diag}(\widetilde{R}_1,\dots,\widetilde{R}_\nu)$, V is as in (13) and the matrix $\widetilde{\Xi} = (\widetilde{\zeta}_1 : \dots : \widetilde{\zeta}_\nu)$ is obtained by Gram-Schmidt orthonormalization of the matrix Ξ whose columns are $\overline{\zeta}_k / \parallel \overline{\zeta}_k \parallel$ where

$$\overline{\zeta}_k = \sqrt{\ell_k}\mathcal{M}^{1/2}(\rho_k I - \mathcal{M})^{-1} t_k, \qquad k=1,\dots,\nu \tag{103}$$

To be specific, we set $\widetilde{\zeta}_\nu = \overline{\zeta}_\nu / \parallel \overline{\zeta}_\nu \parallel$ and assume that the orthonormalization is carried out backwards (w.r.t. the columns of Ξ). First thing to notice is that

$$\parallel \overline{\zeta}_k \parallel^2 = \ell_k t_k^T \mathcal{M}(\rho_k I - \mathcal{M})^{-2} t_k$$

and

$$\overline{\zeta}_j^T \overline{\zeta}_k = \sqrt{\ell_j \ell_k} t_j^T \mathcal{M}(\rho_k I - \mathcal{M})^{-1}(\rho_j I - \mathcal{M})^{-1} t_k, \qquad \text{for } \ 1 \leq j \neq k \leq \nu$$

With $J_\gamma(\epsilon) = \{\mu_1 < \kappa_\gamma + \epsilon\}$, we have, due to *Lemma A.1* (taking $\delta = \frac{1}{3}$ in the lemma),

$$
\begin{aligned}
& \mathbb{P}(\overline{\zeta}_j^T \overline{\zeta}_k \geq \sqrt{\ell_j \ell_k}\delta_0, J_\gamma(\epsilon/2)) \\
= \quad & 2\exp\left(-\frac{n}{N-M}\frac{(\rho_j - \epsilon/2 - \kappa_\gamma)^2(\rho_k - \epsilon/2 - \kappa_\gamma)^2 n\delta_0^2}{12(\kappa_\gamma + \epsilon/2)^2}\right) \\
= \quad & 2\exp\left(-\frac{1}{\gamma}\frac{(\rho_j - \epsilon/2 - \kappa_\gamma)^2(\rho_k - \epsilon/2 - \kappa_\gamma)^2 n\delta_0^2}{12(\kappa_\gamma + \epsilon/2)^2}(1 + o(1))\right), \\
& \qquad \text{for } 0 < \delta_0 < \frac{2(\kappa_\gamma + \epsilon/2)}{(\rho_j - \epsilon/2 - \kappa_\gamma)(\rho_k - \epsilon/2 - \kappa_\gamma)} \quad \text{for } 1 \leq j \neq k \leq \nu
\end{aligned} \tag{104}
$$

We now choose $\widetilde{R}_k$ as

$$
\widetilde{R}_k = \frac{\parallel \overline{\zeta}_k \parallel}{\sqrt{1 + \parallel \overline{\zeta}_k \parallel^2}} \qquad \text{or} \qquad \sqrt{1 - \widetilde{R}_k^2} = \frac{1}{\sqrt{1 + \parallel \overline{\zeta}_k \parallel^2}} \tag{105}
$$

Our aim is to prove the following proposition.

Proposition B.1 : *With this choice of L_ν, given $\epsilon > 0$, $\exists\ n_6(\epsilon, \Lambda, \gamma)$ such that for $n \geq n_6(\epsilon, \Lambda, \gamma)$,*

$$
\mathbb{P}(trace(L_\nu^T \mathbf{S}_{\overline{\Gamma}_\nu} L_\nu) \leq \sum_{k=1}^{\nu} \rho_k - \epsilon/2,\ \mu_1 < \kappa_\gamma + \epsilon/2) \leq \varepsilon_7(n, \epsilon, \Lambda, \gamma) \tag{106}
$$

where $\sum_{n=n_6(\epsilon,\Lambda,\gamma)}^{\infty} \varepsilon_7(n, \epsilon, \Lambda, \gamma) < \infty$.

Once we have this result, we apply *Proposition 3* (with $\rho_\nu + \epsilon$ replaced by $\rho_k + \epsilon/(2\nu)$ and $\widehat{\ell}_\nu$ replaced by $\lambda_1(S_{\Gamma_{k-1}})$, $k = 1, \ldots, \nu - 1$; the validity of this is readily cheked by following the first step of the proof), utilize (101), (102), and the inequality $\widehat{\ell}_\nu \geq \lambda_\nu(\mathbf{S}_{\overline{\Gamma}_\nu})$, in combination with (106) to prove the following.

Proposition B.2 : *Given $\epsilon > 0$, $\exists\ n_7(\epsilon, \Lambda, \gamma)$ such that for $n \geq n_7(\epsilon, \Lambda, \gamma)$,*

$$
\mathbb{P}(\widehat{\ell}_\nu \leq \rho_\nu - \epsilon,\ \mu_1 < \kappa_\gamma + \epsilon/2) \leq \varepsilon_8(n, \epsilon, \Lambda, \gamma)
$$

where $\sum_{n=n_7(\epsilon,\Lambda,\gamma)}^{\infty} \varepsilon_8(n, \epsilon, \Lambda, \gamma) < \infty$, provided ϵ is small enough so that $\rho_\nu > \kappa_\gamma + 2\epsilon$.

The rest of the section is devoted to giving an outline of the proof of *Proposition B.1.* First step in that direction is to express $trace(L_\nu^T \mathbf{S}_{\overline{\Gamma}_\nu} L_\nu)$ as

$$
trace(L_\nu^T \mathbf{S}_{\overline{\Gamma}_\nu} L_\nu) = trace(L_{A,\nu}^T \overline{\mathbf{S}}_{AA,\nu} L_{A,\nu}) + 2\ trace(L_{A,\nu}^T \overline{\mathbf{S}}_{AB,\nu} L_{B,\nu}) + trace(L_{B,\nu}^T \mathbf{S}_{BB} L_{B,\nu}) \tag{107}
$$

Here, as before, $\overline{\mathbf{S}}_{AA,\nu}$ denotes the submatrix of $\mathbf{S}_{AA}$ consisting of first ν rows and first ν columns. $\overline{\mathbf{S}}_{AB,\nu}$ is analogously defined. By definition of $L_{A,\nu}$,

$$
trace(L_{A,\nu}^T \overline{\mathbf{S}}_{AA,\nu} L_{A,\nu}) = \sum_{k=1}^{\nu}(1 - \widetilde{R}_k^2) s_{kk} = \sum_{k=1}^{\nu}(1 - \widetilde{R}_k^2)\ell_k \frac{1}{n} \parallel Z_{A,k} \parallel^2 \tag{108}
$$

Next,

$$trace(L_{A,\nu}^T \overline{\mathbf{S}}_{AB,\nu} L_{B,\nu}) = \; trace(L_{A,\nu}^T \Lambda_\nu^{1/2} T_\nu^T \mathcal{M}^{1/2} \widetilde{\Xi} \widetilde{D}) = \sum_{k=1}^{\nu} \widetilde{R}_k \sqrt{1-\widetilde{R}_k^2} \sqrt{\ell_k} t_k^T \mathcal{M}^{1/2} \widetilde{\zeta}_k \tag{109}$$

Finally,

$$trace(L_{B,\nu}^T \mathbf{S}_{BB} L_{B,\nu}) = \; trace(\widetilde{D} \widetilde{\Xi}^T \mathcal{M} \widetilde{\Xi} \widetilde{D}) = \sum_{k=1}^{\nu} \widetilde{R}_k^2 \widetilde{\zeta}_k^T \mathcal{M} \widetilde{\zeta}_k \tag{110}$$

Now let us find an expression for $\widetilde{\zeta}_k$. By definition, $\widetilde{\zeta}_\nu = \overline{\zeta}_\nu / \parallel \overline{\zeta}_\nu \parallel$ and

$$\widetilde{\zeta}_j = \left(\frac{\overline{\zeta}_j}{\parallel \overline{\zeta}_j \parallel} - \sum_{k=j+1}^{\nu} c_{jk} \widetilde{\zeta}_k \right) / \parallel \frac{\overline{\zeta}_j}{\parallel \overline{\zeta}_j \parallel} - \sum_{k=j+1}^{\nu} c_{jk} \widetilde{\zeta}_k \parallel, \qquad j = \nu-1, \nu-2, \ldots, 1, \tag{111}$$

where c_{jk} are determined from the orthogonality relations. Therefore,

$$c_{jk} = \frac{\langle \overline{\zeta}_j, \widetilde{\zeta}_k \rangle}{\parallel \overline{\zeta}_j \parallel}, \qquad \text{for} \;\; \nu \geq k > j.$$

Thus we can express $\widetilde{\Xi}$ as $\Xi\Delta$ where Δ is a lower triangular matrix whose entries are given as follows:

$$\Delta_{jk} = \begin{cases} 0 & \text{if } 1 \leq k \leq j-1 \\ -c_{jk}\Delta_{jj} & \text{if } j+1 \leq k \leq \nu \end{cases} \qquad \text{with} \;\; \Delta_{jj} = \left(\parallel \frac{\overline{\zeta}_j}{\parallel \overline{\zeta}_j \parallel} - \sum_{k=j+1}^{\nu} c_{jk} \widetilde{\zeta}_k \parallel \right)^{-1}$$

Note that

$$\Delta_{jj}^{-2} = \frac{\parallel \overline{\zeta}_k \parallel^2}{\parallel \overline{\zeta}_k \parallel^2} - 2 \sum_{k>j} c_{jk} \frac{\langle \overline{\zeta}_j, \widetilde{\zeta}_k \rangle}{\parallel \overline{\zeta}_j \parallel} + \sum_{k>j} c_{jk}^2 = 1 - \sum_{k>j} c_{jk}^2$$

This implies that for $k > j$,

$$\begin{aligned} c_{jk} = \frac{\langle \overline{\zeta}_j, \widetilde{\zeta}_k \rangle}{\parallel \overline{\zeta}_j \parallel} \;\; &= \;\; \sum_{k' \geq k} \frac{\langle \overline{\zeta}_j, \overline{\zeta}_{k'} \rangle}{\parallel \overline{\zeta}_j \parallel \cdot \parallel \overline{\zeta}_{k'} \parallel} \Delta_{kk'} \\ &= \;\; \Delta_{kk} \left(\frac{\langle \overline{\zeta}_j, \overline{\zeta}_k \rangle}{\parallel \overline{\zeta}_j \parallel \cdot \parallel \overline{\zeta}_k \parallel} - \sum_{k'>k} \frac{\langle \overline{\zeta}_j, \overline{\zeta}_{k'} \rangle}{\parallel \overline{\zeta}_j \parallel \cdot \parallel \overline{\zeta}_{k'} \parallel} c_{kk'} \right) \\ &= \;\; \Delta_{kk} (\tau_{jk} - \sum_{k'>k} \tau_{jk'} c_{kk'}) \end{aligned} \tag{112}$$

where $\tau_{jk} = \frac{\langle \overline{\zeta}_j, \overline{\zeta}_k \rangle}{\parallel \overline{\zeta}_j \parallel \cdot \parallel \overline{\zeta}_k \parallel}$. Moreover,

$$|1 - \Delta_{kk}| = \left| 1 - \frac{1}{\sqrt{1 - \sum_{k'>k} c_{kk'}^2}} \right| \leq \frac{\sum_{k'>k} c_{kk'}^2}{1 - \sum_{k'>k} c_{kk'}^2} \quad \Longrightarrow \quad \Delta_{kk} \leq \frac{1}{1 - \sum_{k'>k} c_{kk'}^2} \tag{113}$$

We aim to show that $\widetilde{\zeta}_j$ is close to $\frac{\overline{\zeta}_j}{\|\overline{\zeta}_j\|}$. The following lemma helps us make such a statement.

Lemma B.3 : *Let $A > 1$ be arbitrary. Suppose $\delta > 0$ is such that $\delta < \frac{A-1}{\nu A^2}$. If $|\tau_{jk}| \leq \delta$ for all $\nu \geq k > j \geq 1$, then*

$$|c_{jk}| \leq A\delta, \qquad \text{and} \qquad \Delta_{jj} \leq \frac{1}{1-(\nu-j)A^2\delta^2}, \qquad \nu \geq k > j \geq 1. \tag{114}$$

Proof : We prove the result by backward induction on j, k. First note that $c_{j\nu} = \tau_{j\nu}$ for $j = 1, \ldots, \nu-1$. Thus $|c_{j\nu}| \leq \delta$ for $j = 1, \ldots, \nu-1$. And by (113), $\Delta_{\nu-1,\nu-1} \leq \frac{1}{1-\delta^2}$. So the induction hypothesis is satisfied for $j = \nu-1$, $k = \nu$. Suppose the hypothesis holds for all $\nu \geq k > j \geq J+1$ where $J \geq 1$. Want to show that the same holds for $j = J$. Evidently $|c_{J\nu}| \leq \delta$. From (112) we have, for $k = J+1, \ldots, \nu-1$

$$\begin{aligned}
|c_{Jk}| &\leq \Delta_{kk}(|\tau_{Jk}| + \sum_{k'>k} |\tau_{Jk}||c_{kk'}|) \leq \frac{1}{1-(\nu-k)A^2\delta^2}(\delta + \sum_{k'=k+1}^{\nu} \delta \cdot A\delta) \qquad \text{(by hypothesis)} \\
&= \frac{\delta(1+(\nu-k)A\delta)}{1-(\nu-k)A^2\delta^2} \leq \frac{\delta(1+(\nu-k)A\delta)}{1-(\nu-k)^2A^2\delta^2} = \frac{\delta}{1-(\nu-k)A\delta} \leq A\delta
\end{aligned}$$

Here the last inequality follows from the fact

$$\frac{1}{1-(\nu-k)A\delta} \leq A \iff 1 - \frac{1}{A} \geq (\nu-k)A\delta \iff \delta \leq \frac{A-1}{(\nu-k)A^2}$$

and the last condition holds since $\delta < \frac{A-1}{\nu A^2}$. The assertion about Δ_{JJ} follows easily from this and (113).

We are now in a position to finish the proof of *Proposition B.1.* We avoid all the messy details since most of it is mere repitition of the analysis we carried out in *Section 4.* We just show how the three terms behave asymptotically as $n \to \infty$. *Proposition 6* shows that for large n, $\| \overline{\zeta}_k \|^2$ concentrates around $\ell_k\gamma \int \frac{x}{(\rho_k - x)^2} dF_\gamma(x) = \frac{\ell_k\gamma}{(\ell_k-1)^2-1}$. This and (104) imply that for every pair $j \neq k$, τ_{jk} concentrates about 0. Therefore by *Lemma B.3* we see that $\widetilde{\zeta}_j - \overline{\zeta}_j / \| \overline{\zeta}_j \|$ is a vector whose norm concentrates around zero. With this piece of information, we can strip off the insignificant terms in (109) and (110) to claim that

$$\begin{aligned}
trace(L_{A,\nu}^T \overline{\mathbf{S}}_{AB,\nu} L_{B,\nu}) &\sim \sum_{k=1}^{\nu} \frac{1}{\| \overline{\zeta}_k \|} \widetilde{R}_k \sqrt{1-\widetilde{R}_k^2} \sqrt{\ell_k} t_k^T \mathcal{M}^{1/2} \overline{\zeta}_k \\
\text{and} \qquad trace(L_{B,\nu}^T \mathbf{S}_{BB} L_{B,\nu}) &\sim \sum_{k=1}^{\nu} \frac{1}{\| \overline{\zeta}_k \|^2} \widetilde{R}_k^2 \overline{\zeta}_k^T \mathcal{M} \overline{\zeta}_k
\end{aligned}$$

where $\sim$ means that the difference between the LHS and RHS concentrates around 0 as $n \to \infty$. Recalling (103) and (105), and using *Proposition 6* and parts of the proof of *Proposition 3*, it is easy to show that

$$trace(L_{A,\nu}^T \overline{\mathbf{S}}_{AA,\nu} L_{A,\nu}) \sim \sum_{k=1}^{\nu} \frac{1}{1+\| \overline{\zeta}_k \|^2} \ell_k \sim \sum_{k=1}^{\nu} \ell_k \left(1 + \frac{\ell_k\gamma}{(\ell_k-1)^2-\gamma}\right)^{-1} \tag{115}$$

$$
\begin{aligned}
trace(L_{A,\nu}^T \overline{\mathbf{S}}_{AB,\nu} L_{B,\nu}) &\sim \sum_{k=1}^{\nu} \frac{1}{1+\parallel \bar{\zeta}_k \parallel^2} \ell_k t_k^T \mathcal{M}(\rho_k I - \mathcal{M})^{-1} t_k \\
&\sim \sum_{k=1}^{\nu} \frac{\ell_k \gamma}{\ell_k - 1} \left(1 + \frac{\ell_k \gamma}{(\ell_k - 1)^2 - \gamma}\right)^{-1} \qquad (116)
\end{aligned}
$$

$$
\begin{aligned}
trace(L_{A,\nu}^T \overline{\mathbf{S}}_{AB,\nu} L_{B,\nu}) &\sim \sum_{k=1}^{\nu} \frac{1}{1+\parallel \bar{\zeta}_k \parallel^2} \ell_k t_k^T \mathcal{M}^2(\rho_k I - \mathcal{M})^{-2} t_k \\
&\sim \sum_{k=1}^{\nu} \ell_k \gamma \left(1 + \frac{\ell_k \gamma}{(\ell_k - 1)^2 - \gamma}\right)^{-1} \left[\frac{\rho_k}{(\ell_k - 1)^2 - 1} - \frac{1}{\ell_k - 1}\right] \qquad (117)
\end{aligned}
$$

Substituing (115), (116) and (117) in (107), after some simplification we get

$$
trace(L_\nu^T \mathbf{S}_{\overline{\Gamma}_\nu} L_\nu) \sim \sum_{k=1}^{\nu} \rho_k
$$

Formalizing this argument we can prove (106).

8.3 Proof of Proposition 6

We simply give an outline. First consider the expansion

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

For the first square-bracketed term use *Lemma A.1* restricting to the set $\{\mu_1 < \kappa_\gamma + \epsilon/2\}$. Subdivide the second term further as

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

Bounds for I and II on the set $\{\mu_1 < \kappa_\gamma + \epsilon/2\}$ are derived by imitating the arguments leading to (47) and (45). Only notable difference is that for I we need to use the function $G_2(\cdot; \rho, \gamma, \epsilon/2)$ instead of $G_1(\cdot; \rho, \gamma, \epsilon/2)$. Keeping track of the constants we derive the upper bound in the statement of *Proposition 6*.

Acknowledgement

I wish to thank Professor Iain Johnstone, who is my thesis advisor, for his support and thoughtful discussions, and also for bringing to my notice two important references. I would also like to thank Dr. Jinho Baik for some helpful communications.

Reference

1. Anderson, T. W. (1963) : Asymptotic theory of principal component analysis, *Annals of Mathematical Statistics*, **34**, p. 122-148.

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

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

4. Bai, Z. D. & Silvertein, J. W. (1999) : Exact separation of eigenvalues of large dimensional sample covariance matrices, *Annals of Probability*, **27**, 1536-1555.

5. Bai, Z. D. & Silvertein, J. W. (2004) : CLT for linear spectral statistics of large dimensional sample covariance matrix, (to appear in the *Annals of Probability*).

6. Bai, Z. D. & Yin, Y. Q. (1993) : Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, *Annals of Probability*, **21**, 1275-1294.

7. Baik, J., Ben Arous, G. & Péché (2004) : Phase transition of the largest eigenvalue for non-null complex covariance matrices, *arXiv:math.PR/0403022 v1.*

8. Baik, J. & Silverstein, J. W. (2004) : Eigenvalues of large sample covariance matrices of spiked population models, *arXiv:math.ST/048165 v1.*

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

10. Buja, A., Hastie, T. & Tibshirani, R. (1995) : Penalized discriminant analysis, *Annals of Statistics*, **23**, 73-102.

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

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

13. Hoyle, D. & Rattray, M. (2004) : Principal component analysis eigenvalue spectra from data with symmetry breaking structure, *Physical Review E*, **69**, 026124.

14. Johnstone, I. M. (2001) : On the distribution of the largest principal component, *Annals of Statistics*, **29**, 295-327.

15. Johnstone, I. M. & Lu, A. Y. (2004) : Sparse principal component analysis, (to appear in *Journal of American Statistical Association*).

16. Kneip, A. & Utikal, K. J. (2001) : Inference for density families using functional principal component analysis. *Journal of American Statistical Association*, **96**, 519-542.

17. Kato, T. (1980) : *Perturbation theory of linear operators*, Springer-Verlag.

18. Ledoit, O. & Wolf, M. (2002) : Some hypothesis test for the covariance matrix when the dimension is large compared to the sample size, *Annals of Statistics*, **30**, 1081-1102.

19. Ledoux, M. (2001) : *The concentration of measure phenomenon*, Mathematical Surveys and Monographs **89**, American Mathematical Society.

20. Muirhead, R. J. (1982) : *Aspects of multivariate statistical theory*, John Wiley & Sons, Inc.

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

22. Rao, C. R. (1973) : *Linear statistical inference and its applications*, Wiley Eastern.

23. Reimann, P., Van den Broeck, C. & Bex, G. J. (1996) : A Gaussian scenario for unsupervised learning, *Journal of Physics A : Mathematical and General*, **29**, 3521-3535.

24. Roy, S. N. (1953) : On a heuristic method of test construction and its use in Multivariate analysis, *Annals of Mathematical Statistics*, **24**, 220-238.

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

26. Soshnikov, A. (2002) : A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, *Journal of Statistical Physics*, **108**(5/6), 1033-1056. (Also *arXiv:math.PR/0104113 v2*).

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

28. Tracy, C. & Widom, H. (1994) : Level-spacing distribution and Airy kernel, *Communications in Mathematical Physics*, **159**, 151-174.

29. Tracy, C. & Widom, H. (1996) : On orthogonal and symplectic matrix ensembles, *Communications in Mathematical Physics*, **177**, 727-754.

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

31. Wachter, K. W. (1976) : Probability plotting points for principal components, *Ninth Interface Symposium Computer Science and Statistics* (D. Hoaglin and R. Welsch eds.) 299-308, Prindle, Weber and Schmidt, Boston.