ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

High Dimensional Correlation Matrices: CLT and Its Applications

Jiti Gao, Xiao Han, Guangming Pan and Yanrong Yang

November 2014

Working Paper 26/14

High Dimensional Correlation Matrices: CLT and Its Applications

Jiti Gao*, Xiao Han†, Guangming Pan‡, and Yanrong Yang§

November 25, 2014

Abstract

Statistical inferences for sample correlation matrices are important in high dimensional data analysis. Motivated by this, this paper establishes a new central limit theorem (CLT) for a linear spectral statistic (LSS) of high dimensional sample correlation matrices for the case where the dimension p and the sample size n are comparable. This result is of independent interest in large dimensional random matrix theory. Meanwhile, we apply the linear spectral statistic to an independence test for p random variables, and then an equivalence test for p factor loadings and n factors in a factor model. The finite sample performance of the proposed test shows its applicability and effectiveness in practice. An empirical application to test the independence of household incomes from different cities in China is also conducted.

Keywords: Central limit theorem; equivalence test; high dimensional correlation matrix; independence test; linear spectral statistics.

JEL Classification: C21, C32

*Jiti Gao, Department of Econometrics and Business Statistics, Monash University, VIC 3145, Australia(Email: jiti.gao@monash.edu).

†Xiao Han, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, 637371(Email: xhan011@e.ntu.edu.sg).

‡Guangming Pan, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, 637371(Email: GMPAN@ntu.edu.sg).

§Yanrong Yang, Department of Econometrics and Business Statistics, Monash University, VIC 3145, Australia(Email: yanrong.yang@monash.edu).

1 Introduction

Big data issues arising in various fields bring great challenges to classical statistical inferences. High dimensionality and large sample size are two critical features of big data. In statistical inferences, there are serious problems, such as, noise accumulation, spurious correlations, and incidental homogeneity, arisen by high dimensionality. In view of this, the development of new statistical models and methods is necessary for big data research. Thus, our task in this paper is to analyze the correlation matrix of a p-dimensional random vector $\mathbf{x} = (X_1, X_2, \ldots, X_p)^*$, with available samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, where $\mathbf{x}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})^*$, where $*$ denotes the conventional conjugate transpose. We consider the setting of the dimensionality p and the sample size n being in the same order.

Correlation matrices are commonly used in statistics to investigate relationships among different variables in a group. It is well known that the sample correlation matrix is not a 'good' estimator of its corresponding population version when the number p of random variables under investigation is comparable to the sample size n. Thus, it is of great interest to understand and investigate the asymptotic behaviour of the sample correlation matrices of high dimensional data. Sample correlation matrices have appeared in some classical statistics for hypothesis tests. Schott (2005) utilized sample correlation matrices to test independence for a large number of random variables having a multivariate normal distribution. Concerning statistical inference for high dimensional data, furthermore, there are many available research methods based on sample covariance matrices, for example, Johnstone (2001), Cai, Zhang and Zhou (2010). As the population mean and variance of the original data are usually unknown, sample covariance matrices cannot provide us with sufficient and correct information about the data. To illustrate this point, a simple example is that we will make an incorrect conclusion in an independence test if the variance of the data under investigation is not identical to one while the statistics based on sample covariance matrices require the variance to be one. Moreover, the main advantage of using sample correlation matrices over sample covariance matrices is that it does not require the first two population moments of the elements of $\mathbf{x}$ to be known. This point makes the linear spectral statistics based on sample correlation matrices more practical in applications. By contrast, linear spectral statistics for sample covariances involve unknown moments, and are therefore practically infeasible.

Large dimensional random matrix theory provides us with a powerful tool to establish asymptotic theory for high dimensional sample covariance matrices. Bai and Silverstein (2004) contributed to the establishment of asymptotic theory for linear spectral statistics based on high dimensional sample covariance matrices. Meanwhile, there are few results available in the literature for investigating high dimensional sample correlation matrices. Jiang (2004), among one of the first, established a limiting spectral distribution for sample correlation matrices. Cai and Jiang (2011) developed some limiting laws of coherence for sample correlation matrices. In addition, both Bao, Pan and Zhou (2012) and Pillai and Yin (2012) established asymptotic distributions for the extreme eigenvalues of the sample correlation matrices under study. By moving one step further, this paper develops a new central limit theorem for a linear spectral statistic (LSS), which is based on the empirical spectral distribution (ESD) of the sample correlation matrix of $\mathbf{x}$. LSS is a general class of statistics in the sense of being able to cover a lot of commonly used statistics. This new CLT is also of independent interest in large dimensional random matrix theory.

In addition to the establishment of a new CLT, we discuss two relevant statistical applications of both the linear spectral statistic of the sample correlation matrix and the resulting asymptotic theory. The first one is an independence test for p random variables included in the vector $\mathbf{x}$. A related study is Schott (2005), who discussed this kind of independence test for p normal random variables. The second application is to test the equivalence of factor loadings or factors in a factor model. As we discuss in Section 3 below, sample correlation matrices can be used directly for testing purposes without estimating factor loadings and factors first.

The rest of the paper is organized as follows. Section 2 introduces a class of linear spectral statistics. An asymptotic theory is established in Section 3.1 and its applications are established in Section 3.2. The finite sample performance of the proposed test is reported and discussed in Section 4. An empirical application to test independence for household incomes from different cities in China is provided in Section 5. Section 6 concludes the main discussion of this paper. The proofs of the main theory stated in Section 3.1 is given in Section 7. The proofs of some necessary lemmas are provided in Section 8.

2 Linear Spectral Statistics

Given a p-dimensional random vector $\mathbf{x} = (X_1, X_2, \ldots, X_p)^*$ with n random samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, where $\mathbf{x}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})^*$, $i = 1, 2, \ldots, n$. Let $\mathbf{X}_n = (\mathbf{y}_1 - \bar{\mathbf{y}}_1, \mathbf{y}_2 - \bar{\mathbf{y}}_2, \ldots, \mathbf{y}_p - \bar{\mathbf{y}}_p)^*$, where $\mathbf{y}_i = (X_{i1}, X_{i2}, \ldots, X_{in})^T$ for $i = 1, 2, \ldots, p$ and $\bar{\mathbf{y}}_i = \frac{1}{n}\sum_{j=1}^n X_{ij}\mathbf{e}$ with $\mathbf{e}$ being a p-dimensional vector whose elements are all 1, in which T denotes the transpose of a matrix or a vector.

Consider the sample correlation matrix $\mathbf{B}_n = (\rho_{ik})_{p\times p}$ with

$$\rho_{ik} = \frac{(\mathbf{y}_i - \bar{\mathbf{y}}_i)^*(\mathbf{y}_k - \bar{\mathbf{y}}_k)}{||\mathbf{y}_i - \bar{\mathbf{y}}_i|| \cdot ||\mathbf{y}_k - \bar{\mathbf{y}}_k||},$$

where $||\cdot||$ is the usual Euclidean norm. $\mathbf{B}_n$ can also be written as

$$\mathbf{B}_n = \mathbf{Y}_n^*\mathbf{Y}_n = \mathbf{D}_n\mathbf{X}_n^*\mathbf{X}_n\mathbf{D}_n,$$

with

$$\mathbf{Y}_n = \Big(\frac{\mathbf{y}_1 - \bar{\mathbf{y}}_1}{||\mathbf{y}_1 - \bar{\mathbf{y}}_1||}, \frac{\mathbf{y}_2 - \bar{\mathbf{y}}_2}{||\mathbf{y}_2 - \bar{\mathbf{y}}_2||}, \ldots, \frac{\mathbf{y}_p - \bar{\mathbf{y}}_p}{||\mathbf{y}_p - \bar{\mathbf{y}}_p||}\Big)$$

and $\mathbf{D}_n = \text{diag}\left(\frac{1}{\|\mathbf{y}_i - \bar{\mathbf{y}}_i\|}\right)_{p\times p}$ is a diagonal matrix.

Let us consider a class of statistics related to the eigenvalues of $\mathbf{B}_n$. To this end, define the empirical spectral distribution (ESD) of the sample correlation matrix $\mathbf{B}_n$ by $F^{\mathbf{B}_n}(x) = \frac{1}{p}\sum_{i=1}^{p} I(\lambda_i \le x)$, where $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_p$ are the eigenvalues of $\mathbf{B}_n$ and $I(\cdot)$ is an indicator function.

If $X_1, X_2, \ldots, X_p$ are independent, $F^{\mathbf{B}_n}(x)$ converges with probability one to the Marcenko-Pastur (simply called M-P) law $F_c(x)$ with $c = \lim_{n\to\infty} p/n$ (see Jiang (2004)), whose density has an explicit expression of the form

$$f_c(x) = \begin{cases} \frac{1}{2\pi x c}\sqrt{(b-x)(x-a)}, & a \le x \le b; \\ 0, & otherwise; \end{cases}$$

and a point mass $1 - 1/c$ at the origin if $c > 1$, where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$.

Linear spectral statistics of the sample correlation matrix are of the form:

$$\frac{1}{p}\sum_{j=1}^{p} f(\lambda_j) = \int f(x) dF^{\mathbf{B}_n}(x),$$

where f is an analytic function on $[0, \infty)$.

We then consider a normalized and scaled linear spectral statistic of the form:

$$T_n(f) = \int f(x) dG_n(x), \tag{2.1}$$

where $G_n(x) = p[F^{\mathbf{B}_n}(x) - F_{c_n}(x)]$.

The test statistic $T_n(f)$ is a general statistic in the sense that it covers many classical statistics as special cases. For example,

1. Schott's Statistic (Schott (2005)):

$$f_1(x) = x^2 - x: \quad T_n(f_1) = tr(\mathbf{B}_n^2) - p - p\int (x^2 + x) dF_{c_n}(x).$$

2. The Likelihood Ratio Test Statistic (Morrison (2005)):

$$f_2(x) = \log(x): \quad T_n(f_2) = \sum_{i=1}^{p} \log(\lambda_i) - p\int \log(x) dF_{c_n}(x),$$

where $\lambda_i : i = 1, 2, \ldots, p$ are eigenvalues of $\mathbf{B}_n$.

One important tool used in developing an asymptotic distribution for $T_n(f)$ is the Stieltjes transform. The Stieltjes transform m_G for any c.d.f G is defined by

$$m_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \quad \Im(z) > 0.$$

The Stieltjes transform $m_G(z)$ and the corresponding distribution $G(x)$ satisfy the following relation:

$$G([x_1, x_2]) = \frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{x_1}^{x_2} \Im\big(m_G(x + i\varepsilon)\big) dx,$$

where x_1 and x_2 are continuity points of G.

Furthermore, the linear spectral statistic can be expressed via the Stieltjes transform of ESD of $\mathbf{B}_n$ as follows:

$$\int f(x) dF^{\mathbf{B}_n}(x) = -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) m_{F^{\mathbf{B}_n}}(z) dz, \tag{2.2}$$

where the contour $\mathcal{C}$ contains the support of $F^{\mathbf{B}_n}$ with probability one.

3 Asymptotic Theory and Two Applications

First, we establish a new central lint theorem for the linear statistic (2.1) in Theorem 1. Second, we show how to apply the linear statistic and its resulting limiting distribution for an independence test for p random variables and then an equivalence test for factor loadings or factors respectively.

3.1 Asymptotic Theory

Before we establish our main theorem, we introduce some notion. Let $\underline{\mathbf{B}}_n = \mathbf{Y}_n \mathbf{Y}_n^*$. The Stieltjes transforms of ESD and LSD for $\underline{\mathbf{B}}_n$ are denoted by $\underline{m}_n(z)$ and $\underline{m}_c(z)$, respectively. Their analogues for $\mathbf{B}_n$ are denoted by $m_n(z)$ and $m_c(z)$, respectively. Moreover, $\underline{m}_{c_n}(z)$ and $m_{c_n}(z)$ become $\underline{m}_c(z)$ and $m_c(z)$, respectively, when c is replaced by c_n. For ease of notation, we denote $m_c(z)$ and $\underline{m}_c(z)$ by $m(z)$ and $\underline{m}(z)$, respectively with omitting the subscript c. Moreover, let $\kappa = \lim_{p \to \infty} \frac{1}{p} \sum_{i=1}^{p} \frac{\mathbb{E}|X_{i1} - \mathbb{E}X_{i1}|^4}{(\mathbb{E}|X_{i1} - \mathbb{E}X_{i1}|^2)^2}$, and $m^{'}(z)$ denote the first derivative of $m(z)$ with respect to z, throughout the rest of this paper.

The following theorem is to establish a joint central limit theorem for the linear spectral statistic of the correlation matrix $\mathbf{B}_n$.

Theorem 1. *Let* $\{X_{ij} : i = 1, 2, \ldots, p; j = 1, 2, \ldots, n\}$ *be independent with* $\sup_{1 \le i \le p} \mathbb{E}|X_{i1}|^4 < \infty$. *Let* $p/n \to c \in (0, +\infty)$ *as* $n \to \infty$. *Let* $f_1, f_2, \ldots, f_r$ *be functions on* $\mathbb{R}$ *and analytic on an open interval containing*

$$\big[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2\big].$$

Then, the random vector $\Big(\int f_1(x)dG_n(x), \ldots, \int f_r(x)dG_n(x)\Big)$ *converges weakly to a Gaussian vector* $(X_{f_1}, \ldots, X_{f_r})$.

When X_{ij} *are real random variables, the asymptotic mean is*

$$
\begin{aligned}
E_r\left[X_{f_j}\right] &= \frac{\kappa-1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\underline{m}(z)\big(z(1+\underline{m}(z))+1-c\big)}{\big((z(1+\underline{m}(z))-c\big)^2-c\big)\big(z(1+\underline{m}(z))-c\big)}dz \\
&-\frac{\kappa-|\psi|^2-2}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{cz\underline{m}(z)m^2(z)\big(1+\underline{m}(z)\big)\big(z(1+\underline{m}(z))+1-c\big)}{\big((z(1+\underline{m}(z))-c)^2-c\big)\big(1+cm(z)\big)}dz \\
&-\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\underline{m}'(z)\Big(z(1+\underline{m}(z))+1-c\Big)}{\underline{m}(z)\big(z+z\underline{m}(z)-c\big)\Big(\big(z(1+\underline{m}(z))-c\big)^2-c\Big)}dz \\
&+\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\Big(1+z\underline{m}(z)-zm(z)\underline{m}(z)-z^2m(z)\underline{m}^2(z)\Big)\Big(1+\underline{m}(z)\Big)\Big(z(1+\underline{m}(z))+1-c\Big)}{z(1+cm(z))\big(z(1+\underline{m}(z))-c)^2-c\big)}dz \\
&+\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\Big(\frac{cm(z)}{z}-czm(z)\underline{m}'(z)\Big)dz
\end{aligned}
$$

and the asymptotic covariance function

$$
\begin{aligned}
& Cov_r(X_{f_j}, X_{f_r}) \\
= & -\frac{1}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f_j(z_1)f_r(z_2)\frac{cm'(z_1)m'(z_2)}{\big(1+c(m(z_1)+m(z_2))+c(c-1)m(z_1)m(z_2)\big)^2}dz_1dz_2 \\
& + \frac{\kappa-1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f_j(z_1)f_r(z_2)\frac{c\underline{m}'(z_1)\underline{m}'(z_2)}{(1+\underline{m}(z_1))^2(1+\underline{m}(z_2))^2}dz_1dz_2 \\
& - \frac{\kappa-|\psi|^2-2}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f_j(z_1)f_r(z_2)V(c, m(z_1), m(z_2))dz_1dz_2,
\end{aligned}
$$

in which $\psi = \frac{\mathbb{E}(X_{i1}-\mathbb{E}X_{i1})^2}{\mathbb{E}|X_{i1}-\mathbb{E}X_{i1}|^2} \equiv 1$ *under the real case,*

$$
\begin{aligned}
V(c, m(z_1), m(z_2)) &= c\Big(m(z_1)\underline{m}(z_1)+z_1m(z_1)\underline{m}'(z_1)+z_1m'(z_1)\underline{m}(z_1)\Big) \\
&\times\Big(m(z_2)\underline{m}(z_2)+z_2m(z_2)\underline{m}'(z_2)+z_2m'(z_2)\underline{m}(z_2)\Big)
\end{aligned}
$$

for $j, k = 1, 2, \ldots, r$, *and the contour* $\oint_{\mathcal{C}}$ *is closed and taken in the positive direction in the complex plane, each enclosing the support of* $F_c(\cdot)$.

When $\{X_{ij}\}$ *are complex variables, assuming that* $\psi = \frac{\mathbb{E}(X_{i1}-\mathbb{E}X_{i1})^2}{\mathbb{E}|X_{i1}-\mathbb{E}X_{i1}|^2}$ *are the same for i=1,2,...,p, the asymptotic mean is*

$$
\begin{aligned}
& E_c\left[X_{f_j}\right] = E_r\left[X_{f_j}\right] \\
& - \frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\Big(\frac{z\underline{m}'(z)}{(1+\underline{m}(z))(z+z\underline{m}(z)-c)} - \frac{c|\psi|^2m^2(z)}{(1+cm(z))[(1+cm(z))^2-c|\psi|^2m^2(z)]}\Big) \\
& \times\Big(-\frac{c(1+\underline{m}(z))\big(z(1+\underline{m}(z))+1-c\big)}{z\underline{m}(z)\Big(\big(z(1+\underline{m}(z))-c\big)^2-c\Big)}\Big)dz;
\end{aligned}
$$

and the asymptotic variance is

$$\begin{aligned}
&Cov_c(X_{f_j}, X_{f_r}) = Cov_r(X_{f_j}, X_{f_r}) \\
&+ \frac{1}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{f_j(z_1) f_r(z_2) c m'(z_1) m'(z_2) dz_1 dz_2}{\left(1 + c(m(z_1) + m(z_2)) + c(c-1) m(z_1) m(z_2)\right)^2} \\
&- \frac{|\psi|^2}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{f_j(z_1) f_r(z_2) c m'(z_1) m'(z_2) dz_1 dz_2}{[(1 + c m(z_1))(1 + c m(z_2)) - c|\psi|^2 m(z_1) m(z_2)]^2}.
\end{aligned}$$

Remark 1. *Especially, when $X_{ij} \sim \mathcal{N}(\mu_i, \sigma_i^2)$, i=1,2,...,p; j=1,2,...,n, we have $\kappa \equiv 3$. Although the asymptotic means and variances given above look complicated, they are easy to calculate in practice. In fact, the LSD's $m(z)$ and $\underline{m}(z)$ can be estimated by $\frac{1}{p} tr(\mathbf{B}_n - z\mathbf{I}_p)^{-1}$ and $\frac{1}{n} tr(\underline{\mathbf{B}}_n - z\mathbf{I}_n)^{-1}$ respectively. Moreover, asymptotic distributions are still the same after plugging in such estimators due to Slutsky's theorem. The integrals involved in Theorem 3.1 may be calculated by the function 'quad' or 'dblquad' in MATLAB.*

3.2 Two Applications

In this section, we provide two statistical applications of linear spectral statistics for sample correlation matrices. They are an independence test for high dimensional random vector and an equivalence test for factor loadings or factors in a factor model.

3.2.1 Independence Test

For the p random variables grouped in the vector $\mathbf{y}$, our goal is to test the following hypotheses:

$$\mathbb{H}_{10} : X_1, \ldots, X_p \textit{ are independent}; \ \textit{vs } \mathbb{H}_{1a} : X_1, \ldots, X_p \textit{ are dependent}. \tag{3.1}$$

For this independence test, we make the best use of the linear spectral statistic (2.1) based on the sample correlation matrix of $\mathbf{x}$ with the available n samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. As stated in the last section, under the null hypothesis, the limit spectral distribution of $\mathbf{B}_n$ is the M-P law. We use this point to imply independence when applying linear spectral statistics. For simplicity, we choose $f(x) = x^2$ in (2.1).

3.2.2 Test for Equivalence of Factor Loadings or Factors

Since it is difficult to find consistent estimates for unknown factors and loadings, this section proposes to use the proposed linear spectral statistic of the sample correlation matrix for directly testing equivalence for either the factor or the loading without requiring consistent estimates.

Consider the factor model

$$X_{it} = \boldsymbol{\lambda}_i^T \mathbf{F}_t + \varepsilon_{it}, \quad i = 1, 2, \ldots, p; \ t = 1, 2, \ldots, n, \tag{3.2}$$

where $\boldsymbol{\lambda}_i$ is an r-dimensional factor loading, $\mathbf{F}_t$ is the corresponding r-dimensional common factor, $\{\varepsilon_{it} : i = 1, 2, \dots, p; t = 1, 2, \dots, n\}$ are the idiosyncratic components and they are independent for $i = 1, 2, \dots, p$ and $t = 1, 2, \dots, n$.

One goal is to test

$$\mathbb{H}_{20} : \boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 = \ldots = \boldsymbol{\lambda}_p. \tag{3.3}$$

The proposed statistic is the linear spectral statistic based on the sample correlation matrix $\mathbf{B}_n$. Under $\mathbb{H}_{20}$, model (3.2) reduces to

$$X_{it} = \boldsymbol{\lambda}^T \mathbf{F}_t + \varepsilon_{it}. \tag{3.4}$$

From (3.4), we have

$$X_{it} - \bar{X}_t = \varepsilon_{it} - \bar{\varepsilon}_t,$$

where $\bar{X}_t = \frac{1}{N} \sum_{i=1}^{N} X_{it}$ and $\bar{\varepsilon}_t = \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{it}$.

In view of this, under the null hypothesis $\mathbb{H}_{20}$, the sample correlation matrix of $\mathbf{x} = (X_{i1}, X_{i2}, \dots, X_{in})^T$ is the same as that of $\boldsymbol{\varepsilon} = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in})^T$. Since the components of $\boldsymbol{\varepsilon}$ are independent, the linear spectral statistic (2.1) follows the asymptotic distribution in Theorem 1. This is the reason why the proposed statistic works in this case.

Another goal is to test

$$\mathbb{H}_{30} : \mathbf{F}_1 = \mathbf{F}_2 = \ldots = \mathbf{F}_n. \tag{3.5}$$

Similarly, we also propose the linear spectral statistic based on the sample correlation matrix $\mathbf{B}_n$. Under $\mathbb{H}_{30}$, model (3.2) reduces to

$$X_{it} = \boldsymbol{\lambda}_i^T \mathbf{F} + \varepsilon_{it}, \tag{3.6}$$

From (3.6), we have

$$X_{it} - \bar{X}_i = \varepsilon_{it} - \bar{\varepsilon}_i,$$

where $\bar{X}_i = \frac{1}{n} \sum_{t=1}^{n} X_{it}$ and $\bar{\varepsilon}_i = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{it}$.

Then under the null hypothesis $\mathbb{H}_{30}$, the sample correlation matrix of $\widetilde{\mathbf{x}} = (X_{1t}, X_{2t}, \dots, X_{pt})^T$ is the same as that of $\widetilde{\boldsymbol{\varepsilon}} = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{pt})^T$. This point makes the proposed statistic (2.1) applicable and useful in this situation.

Remark 2. *We consider a special example of interactive factor model (3.2) of the form:*

$$X_{it} = \alpha_i + f_t + \varepsilon_{it}, \quad i = 1, 2, \dots, p; \; t = 1, 2, \dots, n, \tag{3.7}$$

where α_i is the specific fixed effects corresponding to section i for $i = 1, 2, \ldots, n$, $f_t = f(\frac{t}{T})$ is a trend function, $\{\varepsilon_{it} : i = 1, 2, \ldots, p; t = 1, 2, \ldots, n\}$ are the idiosyncratic components and they are independent for $i = 1, 2, \ldots, p$ and $t = 1, 2, \ldots, n$.

For model (3.7), we consider the null hypothesis test

$$\mathbb{H}_{40} : \alpha_1 = \alpha_2 = \cdots = \alpha_p. \tag{3.8}$$

We may propose the same statistic as that for (3.3).

4 Finite sample analysis

The finite sample performance of the proposed linear spectral statistic in the two applications are being investigated. We present the empirical sizes and powers of the proposed test.

4.1 Empirical sizes and powers

First, we introduce the method of calculating the empirical sizes and powers. Since the asymptotic distribution of the proposed test statistic R_n is a standard normal distribution, it is not difficult to compute the empirical sizes and powers. Let $z_{1-\frac{1}{2}\alpha}$ and $z_{\frac{1}{2}\alpha}$ be the $100(1-\frac{1}{2}\alpha)\%$ and $\frac{1}{2}\alpha$ quantiles of the standard normal distribution. With K replications of the data set simulated under the null hypothesis, we calculate the empirical size as

$$\hat{\alpha} = \frac{\{\sharp \ of \ R_n^H \geq z_{1-\frac{1}{2}\alpha} or \ R_n^H \leq z_{\frac{1}{2}\alpha}\}}{K}, \tag{4.1}$$

where R_n^H represents the value of the test statistic R_n based on the data simulated under the null hypothesis.

In our simulation, we choose $K = 1000$ as the number of the replications. The significance level is $\alpha = 0.05$. Similarly, the empirical power is calculated as

$$\hat{\beta} = \frac{\{\sharp \ of \ R_n^A \geq z_{1-\frac{1}{2}\alpha} or \ R_n^A \leq z_{\frac{1}{2}\alpha}\}}{K}, \tag{4.2}$$

where R_n^A represents the value of the test statistic R_n based on the data simulated under the alternative hypothesis.

4.2 Independence Test

First, we generate the data $\mathbf{x} = (X_1, X_2, \ldots, X_p)$ with n random samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in the following data generating process. Let $\mathbf{x}_i = \mathbf{T}\mathbf{z}_i$, where $\mathbf{z}_i = (Z_{1i}, Z_{2i}, \ldots, Z_{pi})^T$ with the first $[p/2]$ components $(Z_{1i}, Z_{2i}, \ldots, Z_{[p/2]i})$ being generated from the standard normal distribution and the rest of the components $(Z_{[p/2]+1,i}, Z_{[p/2]+2,i}, \ldots, Z_{pi})$ being generated from Gamma(1,1), in which $[m] \leq m$ denotes the largest integer of m. The $p \times p$ deterministic matrix $\mathbf{T}$ is generated in the following scenarios:

1. Independent case: $\mathbf{T} = \mathbf{I}_p$, where $\mathbf{I}_p$ is an identity matrix;
2. Dependent case(1): $\mathbf{T} = \mathbf{I}_p + \frac{1}{\sqrt{n}}\mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ and $\mathbf{v}$ are $p \times 1$ random vectors whose elements are generated from the standard normal distribution;
3. Dependent case(2): $\mathbf{T} = \mathbf{I}_p + \mathbf{d}\mathbf{e}^T + \mathbf{e}\mathbf{d}^T$, where $\mathbf{d} = (0.5, 0, 0, \dots, 0)^T$ is $p \times 1$ vector with the first element being 0.5 and the rest of the elements being 0, and $\mathbf{e}$ is a $p \times 1$ vector whose elements are all 1.

The empirical sizes corresponding to the independent case are listed in Table 1. The table shows that, as the pair (n, p) increases jointly, the sizes are close to the true value 0.05. The empirical powers under the two dependent cases above are presented in Table 2 and Table 3 respectively. The tendency of the powers going to 1, as (n, p) increases, illustrates both the finite–sample applicability and the effectiveness of the proposed test statistic.

Table 1: Independent test: size(half gamma)

n\c	0.2	0.4	0.6	0.8	1
20	0.0248	0.0310	0.0376	0.0366	0.0374
30	0.0360	0.0376	0.0440	0.0400	0.0416
40	0.0360	0.0424	0.0446	0.0452	0.0436
50	0.0410	0.0482	0.0484	0.0512	0.0440
60	0.0428	0.0486	0.0448	0.0482	0.0516

Table 2: Independent test: power(I+$\frac{1}{\sqrt{n}}u_p v_p^*$)

n\c	0.2	0.4	0.6	0.8	1.0
10	0.1640	0.2902	0.4704	0.6404	0.7682
20	0.4092	0.7342	0.9114	0.9816	0.9952
30	0.6244	0.9384	0.9942	0.9998	1.0000
40	0.8076	0.9890	0.9994	1.0000	1.0000
50	0.9022	0.9986	1.0000	1.0000	1.0000

4.3 Equivalence Tests for Factor Loadings or Factors

As for the equivalence test (3.3) for factor loadings, we generate data for factors and idiosyncratic components as follows. The idiosyncratic components $\{\varepsilon_{it} : i = 1, 2, \dots, p; t = 1, 2, \dots, n\}$ are

Table 3: Independent test: power(a=0.5)

(n,c)\d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
(20,0.4)	0.2916	0.6368	0.6310	0.8082	0.8930	0.9318	0.9506	0.9534
(20,0.8)	0.2416	0.3700	0.5806	0.6662	0.7808	0.8452	0.8692	0.9066
(30,0.4)	0.3102	0.6916	0.9326	0.9668	0.9784	0.9884	0.9892	0.9928
(30,0.8)	0.2580	0.6384	0.7828	0.9048	0.9444	0.9622	0.9836	0.9902
(40,0.4)	0.7000	0.8826	0.9762	0.9874	0.9974	0.9976	0.9988	0.9996
(40,0.8)	0.4080	0.7628	0.9284	0.9730	0.9870	0.9944	0.9984	0.9994

generated from the standard normal distribution and the factors $\mathbf{F}_t$ is $AR(1)$, i.e.

$$\mathbf{F}_t = a\mathbf{F}_{t-1} + \boldsymbol{\eta}_t, \quad t = 1, 2, \ldots, n,$$

where $a = 0.2$ and $\{\boldsymbol{\eta}_t\}$ is generated independently from the standard normal distribution. The initial value $\mathbf{F}_0 = \mathbf{0}$. The number of factors takes values of 2 and 3, respectively, in the simulation.

Factor loadings are generated in the following two scenarios.

1. DGP(1): $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}$ for $i = 1, 2, \ldots, p$, where $\boldsymbol{\lambda}$ is generated from the standard normal distribution.

2. DGP(2): $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}$ for $i = 1, 2, \ldots, [d \cdot p]$, where $d = 0.1$; $\boldsymbol{\lambda}_j$ is generated independently from the standard normal distribution for each $j = [d \cdot p], [d \cdot p] + 1, \ldots, p$.

For this test, the empirical sizes under DGP(1) are shown in Table 4 while the empirical powers under DGP(2) are given in Table 5 and Table 6. As (n, p) increases jointly, the empirical sizes tend to the nominal level of 5%. The powers show that our proposed test statistic can capture some local alternatives effectively. As $p = 30$, there are 3 different factor loadings under the alternative hypothesis which can be distinguished by the proposed test statistic.

Table 4: Factor loading test: size

n\c	0.2	0.4	0.6	0.8	1
20	0.0234	0.0320	0.0348	0.0324	0.0346
30	0.0328	0.0374	0.0376	0.0386	0.0404
40	0.0338	0.0386	0.0462	0.0444	0.0454
50	0.0348	0.0440	0.0456	0.0460	0.0424

Table 5: Factor loading test: power(r=2, different factor loadings are at n-direction)

(n,p)\d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(10,10)	0.0690	0.1096	0.1446	0.1812	0.2070	0.2256	0.2486	0.2526	0.2394
(20,10)	0.0726	0.1100	0.1536	0.1886	0.2180	0.2392	0.2646	0.2682	0.2700
(30,10)	0.0742	0.1134	0.1624	0.1964	0.2214	0.2432	0.2586	0.2634	0.2782
(20,20)	0.1100	0.2070	0.3068	0.3964	0.4616	0.5216	0.5578	0.6092	0.6264
(30,20)	0.1010	0.1830	0.2884	0.3744	0.4464	0.4954	0.5486	0.6062	0.6126
(30,30)	0.1412	0.2624	0.4088	0.5266	0.6172	0.7004	0.7464	0.8050	0.8368

Table 6: Factor loading test: power(r=3, different factor loadings are at n-direction)

(n,p)\d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(10,10)	0.0942	0.1648	0.2074	0.2418	0.2854	0.2956	0.2834	0.2964	0.2920
(20,10)	0.1618	0.2970	0.4078	0.4862	0.5642	0.6044	0.6466	0.6854	0.6826
(30,10)	0.2202	0.4230	0.5646	0.6578	0.7318	0.8028	0.8342	0.8612	0.8766
(20,20)	0.1692	0.2816	0.4252	0.5226	0.5998	0.6518	0.7026	0.7406	0.7438
(30,20)	0.2068	0.4228	0.5774	0.7024	0.7808	0.8478	0.8812	0.9074	0.9348
(30,30)	0.1954	0.4052	0.5770	0.6918	0.7768	0.8372	0.8848	0.9092	0.9320

Similarly, for the equivalence test (3.5) for factors, the idiosyncratic components are generated in the same way as the test above. The factor loading $\{\boldsymbol{\lambda}_i\}$ is generated independently from the standard normal distribution.

Factors are generated in the following two scenarios.

1. DGP(3): $\mathbf{F}_t = \mathbf{F}$ for $t = 1, 2, \ldots, n$, where $\mathbf{F}$ is generated independently from the standard normal distribution.

2. DGP(4): $\mathbf{F}_t = \mathbf{F}$ for $i = 1, 2, \ldots, [d \cdot n]$, where $d = 0.1$; $\mathbf{F}_t$ is generated independently from the standard normal distribution for $t = [d \cdot n], [d \cdot n] + 1, \ldots, n$.

The empirical sizes under DGP(3) are shown in Table 7 while the empirical powers under DGP(4) are given in Table 8 and Table 9. The behaviours of the sizes and powers are similar to those discussed in the factor loading test.

Another equivalence test (3.8) is also analyzed. The idiosyncratic components $\{\varepsilon_{it} : i = 1, 2, \ldots, p; t = 1, 2, \ldots, n\}$ are generated independently from the standard normal distribution, and the trend function $f_t = t/n$.

The specific character α_i for each section $i = 1, 2, \ldots, p$ is generated in the following two scenarios.

Table 7: Factor test: size

n\c	0.2	0.4	0.6	0.8	1
20	0.0286	0.0330	0.0348	0.0384	0.0390
30	0.0322	0.0352	0.0396	0.0398	0.0412
40	0.0322	0.0362	0.0410	0.0420	0.0414
50	0.0360	0.0442	0.0462	0.0456	0.0440

Table 8: Factors test: power(r=2, different factors are at n-direction)

(n,p)\d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(10,10)	0.0696	0.1170	0.1528	0.1822	0.1994	0.2248	0.2272	0.2530	0.2474
(20,10)	0.1146	0.2016	0.3024	0.3684	0.4386	0.4850	0.5316	0.5606	0.5710
(30,10)	0.1582	0.2970	0.4260	0.5338	0.6088	0.6850	0.7192	0.7564	0.7734
(20,20)	0.1024	0.2038	0.3002	0.3918	0.4612	0.5214	0.5548	0.5988	0.6158
(30,20)	0.1354	0.2896	0.4130	0.5492	0.6340	0.7116	0.7574	0.8096	0.8310
(30,30)	0.1358	0.2810	0.4058	0.5304	0.6268	0.6988	0.7594	0.8094	0.8302

Table 9: Factors test: power(r=3, different factors are at n-direction)

(n,p)\ d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(10,10)	0.0996	0.1590	0.2088	0.2566	0.2688	0.2910	0.3004	0.2960	0.2930
(20,10)	0.1606	0.2996	0.3968	0.4984	0.5556	0.6016	0.6298	0.6632	0.6784
(30,10)	0.2272	0.4100	0.5502	0.6568	0.7334	0.7912	0.8298	0.8620	0.8748
(20,20)	0.1554	0.2988	0.4358	0.5252	0.5906	0.6592	0.7010	0.7336	0.7584
(30,20)	0.2138	0.4166	0.5762	0.7024	0.7880	0.8506	0.8826	0.9120	0.9256
(30,30)	0.2074	0.4028	0.5660	0.6960	0.7850	0.8362	0.8842	0.9210	0.9304

1. DGP(1): $\alpha_i = \alpha$ with $i = 1, 2, \ldots, p$ where α is generated from standard normal distribution.

2. DGP(2): $\alpha_i = \alpha$ with $i = 1, 2, \ldots, [d \cdot p]$ where $d = 0.1$; α_j is generated from standard normal distribution independently for each $j = [d \cdot p], [d \cdot p] + 1, \ldots, p$.

The empirical sizes and powers are illustrated in Table 10 and Table 11 respectively. In contrast with the powers in the factor loading test, the powers are relatively lower. It is reasonable because the specific characteristic α_i is not affected by the common factors. In summary, the proposed statistic still works well numerically in this case.

Table 10: Specific characteristic test: size

(n,p)\d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(10,10)	0.0286	0.0302	0.0292	0.0318	0.0276	0.0348	0.0324	0.0344	0.0328
(20,10)	0.0364	0.0334	0.0350	0.0392	0.0366	0.0400	0.0360	0.0350	0.0334
(30,10)	0.0360	0.0424	0.0334	0.0338	0.0386	0.0400	0.0398	0.0360	0.0360
(20,20)	0.0372	0.0344	0.0392	0.0388	0.0402	0.0378	0.0386	0.0414	0.0392
(30,20)	0.0390	0.0408	0.0388	0.0356	0.0432	0.0418	0.0418	0.0390	0.0382
(30,30)	0.0440	0.0420	0.0434	0.0412	0.0432	0.0396	0.0434	0.0442	0.0436

Table 11: Specific characteristic test: power

(n,p)\d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(10,10)	0.0560	0.0892	0.1298	0.1644	0.1998	0.2292	0.2544	0.2726	0.2822
(20,10)	0.0638	0.0940	0.1266	0.1670	0.1970	0.2256	0.2368	0.2664	0.2562
(30,10)	0.0532	0.0864	0.1158	0.1572	0.1768	0.2058	0.2150	0.2480	0.2392
(20,20)	0.0758	0.1534	0.2258	0.3076	0.3756	0.4428	0.5078	0.5430	0.5826
(30,20)	0.0644	0.1434	0.2056	0.2738	0.3404	0.4106	0.4608	0.5094	0.5396
(30,30)	0.0912	0.1852	0.2924	0.3946	0.4972	0.5766	0.6544	0.7102	0.7638

5 Empirical Application

In this section, we analyze the relationship of the household incomes among different cities for rural China. The main goal is to test whether they are independent or not.

The data set is drawn from the 'Rural Household Income and Expenditure Survey' conducted by the State Statistics Bureau of China (SSB) and the Chinese Academy of Social

Science (CASS). The data set was collected in 1995 and provides useful information about 7998 households in rural areas of 19 Chinese provinces.

In this study, we focus on testing independence of the household incomes among different cities. After deleting observations with missing or implausible values of the household income variables, a sample of 96 households is retained for 69 different cities.

Table 12: P-values of independence test for household incomes from different cities

(p,n)	(5, 10)	(15, 20)	(40, 50)	(50, 60)	(60, 70)	(69,80)	(69, 96)
$p - values$	0.5260	0.4430	0.5620	0.5290	0.0890	0.0680	0.0540

The proposed linear spectral statistic is applied to this independence test. Different number of cities and various number of households are considered. The p-values of the proposed test are reported in Table 12. The p-values decrease as the number of cities increases. This phenomenon makes sense since the possibility of the dependence becomes larger as the number of cities becomes bigger. Since the p-values are all greater than 0.01, we conclude that the household incomes from different cities are independent.

6 Conclusions

In this paper, we have established a new central limit theorem for a linear spectral statistic of sample correlation matrices for the case where the dimensionality p and the sample size n are comparable. Two useful statistical applications are considered. The first one is an independence test for p random variables while the second one is an equivalence test in factor models. The advantage of using the linear spectral statistic based on sample correlation matrices over sample covariance matrices is that we do not require the knowledge of the first two moments or the underlying distribution of the p random variables under investigation. The finite sample performance of the proposed test is evaluated. An empirical application to test cross-section independence for the household income in different cities of China is discussed.

7 Appendix: Proof of the main theorem

We start by listing some necessary lemmas.

7.1 Lemmas

Lemma 1 (Jiang (2004); Xiao and Zhou (2010)). *Suppose* $p/n \to c \in (0, +\infty)$. *If* $\mathbb{E}|X_{11}|^4 < \infty$ *and* $\mathbb{E}X_{11} = 0$, *then* $\lambda_{\max}(\mathbf{B}_n) \overset{a.s.}{\to} (1+\sqrt{c})^2$ *and* $\lambda_{\min}(\mathbf{B}_n) \overset{a.s.}{\to} (1-\sqrt{c})^2$.

Lemma 2 (Corollary 7.38 of Horn and Johnson (1999)). *Let* A *and* B *be two complex* $p \times n$ *matrices. Define* r=min$\{p, n\}$. *If* $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r$ *are the first* r *largest eigenvalues of* A^*A *and* $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_r$ *are the first* r *largest eigenvalues of* B^*B, *then*

$$\max_{1\leq i\leq r} |\sqrt{\sigma_i} - \sqrt{\lambda_i}| \leq \|A - B\|,$$

where $\|A - B\|$ *denotes the largest eigenvalue of* $(A - B)^*(A - B)$.

Lemma 3 (Burkholder (1973)). *Let* $\{X_k\}$ *be a complex martingale difference sequence with respect to the increasing* $\sigma - field\{\mathcal{F}_k\}$. *Then for* q> 1,

$$\mathbb{E}|\sum X_k|^q \leq K_q \left(\mathbb{E}\left(\sum \mathbb{E}_{k-1}|X_k|^2\right)^{q/2} + \mathbb{E}\left[\sum |X_k|^q\right]\right).$$

Lemma 4 (Theorem 35.12 of Billingsley (1995)). *Suppose for each* n, $Y_{n1}, Y_{n2}, \ldots, Y_{nr_n}$ *is a real martingale difference sequence with respect to the increasing* σ*-field* $\{\mathcal{F}_{nj}\}$ *having second moments. If as* $n \to \infty$,

$$\sum_{j=1}^{r_n} \mathbb{E}(Y_{nj}^2|\mathcal{F}_{n,j-1}) \overset{i.p.}{\to} \sigma^2, \tag{7.1}$$

where σ^2 *is positive constant, and for each* $\varepsilon > 0$,

$$\sum_{j=1}^{r_n} \mathbb{E}(Y_{nj}^2 I_{|Y_{nj}|\geq\varepsilon}) \to 0, \tag{7.2}$$

then $\sum_{j=1}^{r_n} Y_{nj} \overset{D}{\to} \mathcal{N}(0, \sigma^2)$.

The proofs of Lemmas 5-7 below are given in the supplementary document.

Lemma 5. *Suppose that* $\{X_i\}_{i=1}^n$ *are i.i.d. random variables with* $EX_1 = 0$ *and* $E|X_1|^2 = 1$. *Let* $\mathbf{y} = (X_1, ..., X_n)^T$ *and* $\bar{\mathbf{y}} = \frac{\sum_{i=1}^n X_i}{n}\mathbf{e}$, *where* $\mathbf{e} = (1, 1, ..., 1)^T$ *is an* n*-dimensional vector. Assuming that* $\mathbf{A}$ *is a deterministic complex matrix, then for any given* $q \geq 2$, *there is a positive constant* K_q *depending on* q *such that*

$$E\left|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right|^q \leq K_q\left\{n^{-q}(v_{2q}tr(\mathbf{A}\mathbf{A}^*)^q + (v_4 tr(\mathbf{A}\mathbf{A}^*))^{q/2}) + \mathbb{P}(B_n^c(\epsilon))\|\mathbf{A}\|^q\right\}, \tag{7.3}$$

where $B_n(\epsilon) = \left\{\mathbf{y} : |\frac{\|\mathbf{y}-\bar{\mathbf{y}}\|^2}{n} - 1| \leq \epsilon\right\}$ *and* $\boldsymbol{\alpha} = \frac{(\mathbf{y}-\bar{\mathbf{y}})^T}{\|\mathbf{y}-\bar{\mathbf{y}}\|}$, *in which* $\epsilon > 0$ *is a constant.*

Remark 3. *Note that* $\mathbb{P}(B_n^c(\epsilon)) = O(n^{-q/2}v_4^{q/2} + n^{-q+1}v_{2q})$*. If* $\|\mathbf{A}\| \le K$ *and* $|X_i| \le \sqrt{n}\delta_n$*, we have*

$$\mathbb{E}\left|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right|^q \le K_q n^{-1}\delta_n^{2q-4}. \tag{7.4}$$

Remark 4. *Similar to Lemma 5, one can prove that under the same conditions of Lemma 5 (replacing* $\boldsymbol{\alpha}*$ *by* $\boldsymbol{\alpha}^T$*), we have*

$$\mathbb{E}\left|\boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha} - \frac{\mathbb{E}X_1^2}{n}tr\mathbf{A}\right|^q \le K_q\left\{n^{-q}(v_{2q}tr(\mathbf{A}\mathbf{A}^*)^q + (v_4tr(\mathbf{A}\mathbf{A}^*))^{q/2}) + \mathbb{P}(B_n^c(\epsilon))\|\mathbf{A}\|^q\right\}. \tag{7.5}$$

Lemma 6. *In addition to the condition of Lemma 5, if* $\mathbb{E}|X_1|^4 < \infty$*,* $\|\mathbf{A}\| \le K$ *and* $\|\mathbf{B}\| \le K$*, then*

$$\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A})(\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{B}) = \sum_{i=1}^n \frac{1}{n^2}(\mathbb{E}|X_1|^4 - |\mathbb{E}(X_1^2)|^2 - 2)\mathbf{A}_{ii}\mathbf{B}_{ii}$$
$$+ \frac{|\mathbb{E}X_1^2|^2}{n^2}tr(\mathbf{A}\mathbf{B}^T) + \frac{1}{n^2}tr(\mathbf{A}\mathbf{B}) + \frac{1 - \mathbb{E}|X_1|^4}{n^3}tr\mathbf{A}tr\mathbf{B} + o\left(\frac{1}{n}\right).$$

In the sequel, we assume that $\{X_{ij}\}$ satisfies

$$|X_{ij}| < \delta_n\sqrt{n},\ \mathbb{E}X_{ij} = 0,\ \mathbb{E}|X_{ij}|^2 = 1,\ \mathbb{E}|X_{ij}|^4 < \infty,\ \text{and} \quad \kappa = \lim_{p\to\infty}\frac{1}{p}\sum_{i=1}^p \mathbb{E}|X_{i1}|^4. \tag{7.6}$$

Lemma 7. *For any* $l \in \mathbf{N}^+$*,* $\mu_1 > (1+\sqrt{c})^2$ *and* $0 < \mu_2 < \mathbf{I}_{(0,1)}(c)(1-\sqrt{c})^2$*, under condition (7.6), we have*

$$P(\|\mathbf{B}_n\| \ge \mu_1) = o(n^{-l}) \tag{7.7}$$

and

$$P(\lambda_{\min}^{\mathbf{B}_n} \le \mu_2) = o(n^{-l}). \tag{7.8}$$

7.2 Proof of Theorem 1

The overall strategy of our proof is similar to that in Bai and Silverstein (2004). Since many tools proposed in Bai and Silverstein (2004) can not be utilized for the sample correlation matrix case, we therefore develop a number of new techniques for the proof of Theorem 1. Among them, to apply the Cauchy integral formula in (7.9) below and prove tightness, we develop Lemma 7 to make sure that the extreme eigenvalues of $\mathbf{B}_n$ are highly concentrated around two edges of the support. To convert random quadratic forms into the corresponding traces, we establish a moment inequality for random quadratic forms in Lemma 5. Lemma 6 also establishes a precise estimator for the expectation of the product of two random quadratic forms before we may apply central limit theorems for martingale differences. Moreover, we find out the limit of the quadratic form $\mathbf{1}^T\mathbb{E}(\underline{\mathbf{B}}_n - z\mathbf{I})^{-1}\mathbf{1}/n$ is independent of $\underline{m}(z)$, which is quite different from what may be obtained in the case of covariance matrices (here all entries of the vector $\mathbf{1}$ are one). One can refer to Lemma 8 in the supplementary document for detail.

By the Cauchy integral formula, we have

$$\int f(x)dG(x) = -\frac{1}{2\pi i}\int_{\mathcal{C}} f(z)m_G(z)dz \tag{7.9}$$

valid for any c.d.f G and any function f analytic on an open set containing the support of G. In our case, $G(x) := G_n(x) = p(F^{\mathbf{B}_n}(x) - F_{c_n}(x))$.

Note that the support of $G_n(x)$ is random. Fortunately, it is well known that the extreme eigenvalues of $\mathbf{B}_n$ are highly concentrated around two edges of the support of the limiting M-P law $F_c(x)$ (see Lemma 7). Then the contour $\mathcal{C}$ can be appropriately chosen. Moreover, as in Bai and Silverstein (2004), by Lemma 7, we can replace the process $\{M_n(z), \mathcal{C}\}$ by a slightly modified process $\{\widehat{M}_n(z), \mathcal{C}\}$. Below we present the definitions of the contour $\mathcal{C}$ and the modified process $\widehat{M}_n(z)$. Let x_r be any number greater than $(1+\sqrt{c})^2$. Let x_l be any negative number if $c \geq 1$. Otherwise we choose $x_l \in (0, (1-\sqrt{c})^2)$. Now let $\mathcal{C}_u = \{x + iv_0 : x \in [x_l, x_r]\}$.

Then we define $\mathcal{C}^+ \equiv \{x_l + iv : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + iv : v \in [0, v_0]\}$, and $\mathcal{C} = \mathcal{C}^+ \cup \overline{\mathcal{C}^+}$. Now we define the subsets $\mathcal{C}_n$ of $\mathcal{C}$ on which $M_n(\cdot)$ equals to $\widehat{M}_n(\cdot)$. Let $\{\varepsilon_n\}$ be a sequence decreasing to zero satisfying for some $\alpha \in (0,1)$, $\varepsilon_n \geq n^{-\alpha}$.

Now we set

$$\mathcal{C}_l = \begin{cases} \{x_l + iv : v \in [n^{-1}\varepsilon_n, v_0]\} & \text{if } x_l > 0, \\ \{x_l + iv : v \in [0, v_0]\} & \text{if } x_l < 0, \end{cases}$$

and $\mathcal{C}_r = \{x_r + iv : v \in [n^{-1}\varepsilon, v_0]\}$.

Then we define $\mathcal{C}_n = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_r$. The process $\widehat{M}_n(z)$ is defined as

$$\widehat{M}_n(z) = \begin{cases} M_n(z) & \text{for } z \in \mathcal{C}_n, \\ M_n(x_r + in^{-1}\varepsilon_n) & \text{for } x = x_r, v \in [0, n^{-1}\varepsilon_n], \\ M_n(x_l + in^{-1}\varepsilon_n) & \text{for } x = x_l, v \in [0, n^{-1}\varepsilon_n]. \end{cases}$$

To prove Theorem 1, as in Bai and Silverstein (2004), it suffices to prove the CLT for $\widehat{M}_n(z)$ with $z \in \mathcal{C}$. We state the result in the following proposition and then prove it.

Proposition 1. *Under the conditions of Theorem 1, $\{\widehat{M}_n(\cdot)\}$ forms a tight sequence on $\mathcal{C}^+$. And $\{\widehat{M}_n(\cdot)\}$ converges weakly to a two-dimensional Gaussian process $\{M(\cdot)\}$ satisfying for $z \in \mathcal{C}^+$.*

Under the real random variable case,

$$\begin{aligned}\mathbb{E}M(z) &= -\Big(\kappa-1\Big)\frac{c\underline{m}(z)\big(z(1+\underline{m}(z))+1-c\big)}{\big((z(1+\underline{m}(z)))-c\big)^2-c\big)\big(z(1+\underline{m}(z))-c\big)}\\
&+\Big(\kappa-|\psi|^2-2\Big)\frac{cz\underline{m}(z)m^2(z)\big(1+\underline{m}(z)\big)\big(z(1+\underline{m}(z))+1-c\big)}{\big((z(1+\underline{m}(z))-c)^2-c\big)\big(1+cm(z)\big)}\\
&+\frac{c\underline{m}'(z)\Big(z(1+\underline{m}(z))+1-c\Big)}{\underline{m}(z)\big(z+z\underline{m}(z)-c\big)\Big(\big(z(1+\underline{m}(z))-c\big)^2-c\Big)}\\
&-\frac{c\Big(1+z\underline{m}(z)-zm(z)\underline{m}(z)-z^2m(z)\underline{m}^2(z)\Big)\Big(1+\underline{m}(z)\Big)\Big(z(1+\underline{m}(z))+1-c\Big)}{z(1+cm(z))\big(z(1+\underline{m}(z))-c)^2-c\big)}\\
&-\frac{cm(z)}{z}+czm(z)\underline{m}'(z)
\end{aligned} \tag{7.10}$$

and for $z_i, z_j \in \mathcal{C}$

$$\begin{aligned}\mathrm{Cov}(M(z_i),M(z_j)) &= 2\frac{cm'(z_1)m'(z_2)}{\big(1+c(m(z_1)+m(z_2))+(c^2-c)m(z_1)m(z_2)\big)^2}\\
&-\Big(\kappa-1\Big)\frac{c\underline{m}'(z_1)\underline{m}'(z_2)}{(1+\underline{m}(z_1))^2(1+\underline{m}(z_2))^2}\\
&+\Big(\kappa-|\psi|^2-2\Big)V(c,m(z_1),m(z_2)),
\end{aligned} \tag{7.11}$$

where $V(c, m(z_1), m(z_2))$ is defined in Theorem 3.1.

When $\{X_{ij}\}$ are complex variables, assuming that $\psi = \frac{\mathbb{E}(X_{i1}-\mathbb{E}X_{i1})^2}{\mathbb{E}|X_{i1}-\mathbb{E}X_{i1}|^2}$ are the same for i=1,2,...,p, the asymptotic mean is

$$\begin{aligned}(7.10)+\Big(&\frac{z\underline{m}'(z)}{(1+\underline{m}(z))(z+z\underline{m}(z)-c)}-\frac{c|\psi|^2m^2(z)}{(1+cm(z))[(1+cm(z))^2-c|\psi|^2m^2(z)]}\Big)\\
&\cdot\Big(-\frac{c(1+\underline{m}(z))\big(z(1+\underline{m}(z))+1-c\big)}{z\underline{m}(z)\Big(\big(z(1+\underline{m}(z))-c\big)^2-c\Big)}\Big);
\end{aligned} \tag{7.12}$$

and the asymptotic variance is

$$\begin{aligned}(7.11) \quad &- \quad \frac{cm'(z_1)m'(z_2)}{\big(1+c(m(z_1)+m(z_2))+c(c-1)m(z_1)m(z_2)\big)^2}\\
&+ \quad |\psi|^2\frac{cm'(z_1)m'(z_2)}{[(1+cm(z_1))(1+cm(z_2))-c|\psi|^2m(z_1)m(z_2)]^2}.
\end{aligned} \tag{7.13}$$

By the discussions in Bai and Silverstein (2004), we see that Theorem 1 holds if Proposition 1 is proved. Thus the rest of the work will be devoted to the proof of Proposition 1.

Before proving Proposition 1, we need to truncate the elements of $\mathbf{X}_n$ as follows.

7.2.1 Truncation, Centralization and Rescaling

By the same method as that in page 559 of Bai and Silverstein (2004), we can choose a positive sequence of $\{\delta_n\}$ such that

$$\delta_n \to 0,\ \delta_n n^{1/4} \to \infty,\ \delta_n^{-4} E X_{11}^4 I(|X_{11}| \geq \delta_n \sqrt{n}) \to 0.$$

Let $\hat{\mathbf{B}}_n = \hat{\mathbf{D}}_n \hat{\mathbf{X}}_n \hat{\mathbf{X}}_n^* \hat{\mathbf{D}}_n$, where $\hat{\mathbf{X}}_n$ is $p \times n$ matrix having $(i,j)th$ entry $\hat{X}_{ij} - \frac{1}{n}\sum_{k=1}^n \hat{X}_{ik}$, $\hat{X}_{ij} = X_{ij} I_{\{|X_{ij}| < \delta_n \sqrt{n}\}}$ and $\hat{\mathbf{D}}_n$ is $\mathbf{D}_n$ with $\mathbf{X}_n$ replaced by $\hat{\mathbf{X}}_n$. We then have

$$P(\mathbf{B}_n \neq \hat{\mathbf{B}}_n) \leq np \cdot P(|X_{11}| \geq \delta_n \sqrt{n}) \leq K \delta_n^{-4} \int_{\{|X_{11}| \geq \delta_n \sqrt{n}\}} |X_{11}|^4 = o(1).$$

Define $\widetilde{\mathbf{B}}_n = \frac{1}{n} \widetilde{\mathbf{D}}_n \widetilde{\mathbf{X}}_n \widetilde{\mathbf{X}}_n^* \widetilde{\mathbf{D}}_n$, where $\widetilde{\mathbf{X}}_n$ is $p \times n$ matrix having $(i,j)th$ entry $\widetilde{X}_{ij} - \frac{1}{n}\sum_{k=1}^n \tilde{X}_{ik}$, $\widetilde{X}_{ij} = (\hat{X}_{ij} - \mathbb{E}\hat{X}_{ij})/\sigma_n$ with $\sigma_n^2 = \mathbb{E}|\hat{X}_{ij} - \mathbb{E}\hat{X}_{ij}|^2$; and $\widetilde{\mathbf{D}}_n$ is $\mathbf{D}_n$ with X_{ij} replaced by $\widetilde{X}_{ij}$. Throughout this paper, we use M and K to denote a constant which can represent different constants at difference appearance.

From Lemma 1 and Yin, Bai and Krishnaiah (1988), we see that

$$|\limsup_n \lambda_{\max}^{\hat{\mathbf{B}}_n}| \leq M(1+\sqrt{c})^2, \quad |\limsup_n \lambda_{\max}^{\widetilde{\mathbf{B}}_n}| \leq M(1+\sqrt{c})^2.$$

Let $\hat{G}_n(x)$ and $\widetilde{G}_n(x)$ be $G_n(x)$ with $\mathbf{B}_n$ replaced by $\hat{\mathbf{B}}_n$ and $\widetilde{\mathbf{B}}_n$ respectively. Then for each $j = 1, 2, \ldots, k$,

$$\begin{aligned} &|\int f_j(x) d\hat{G}_n(x) - \int f_j(x) d\widetilde{G}_n(x)| \leq M \sum_{k=1}^p |\lambda_k^{\hat{\mathbf{B}}_n} - \lambda_k^{\widetilde{\mathbf{B}}_n}| \\ \leq \ & M\Big(\sum_{k=1}^p (\sqrt{\lambda_k^{\hat{\mathbf{B}}_n}} - \sqrt{\lambda_k^{\widetilde{\mathbf{B}}_n}})^2\Big)^{1/2} \Big(\sum_{k=1}^p (\lambda_k^{\hat{\mathbf{B}}_n} + \lambda_k^{\widetilde{\mathbf{B}}_n})\Big)^{1/2}. \end{aligned}$$

Moreover, similar to Bai and Silverstein (2004) (page 560), we have

$$\begin{aligned} &\sum_{k=1}^p (\sqrt{\lambda_k^{\hat{\mathbf{B}}_n}} - \sqrt{\lambda_k^{\widetilde{\mathbf{B}}_n}})^2 \\ \leq \ & M\Big(\frac{1}{n} tr(\hat{\mathbf{D}}_n \hat{\mathbf{X}}_n - \widetilde{\mathbf{D}}_n \widetilde{\mathbf{X}}_n)(\hat{\mathbf{D}}_n \hat{\mathbf{X}}_n - \widetilde{\mathbf{D}}_n \widetilde{\mathbf{X}}_n)^*\Big)^{1/2} \big(p(\lambda_{\max}^{\hat{\mathbf{B}}_n} + \lambda_{\max}^{\widetilde{\mathbf{B}}_n})\big)^{1/2}, \end{aligned}$$

where K_j is a bound in $|f_j'(z)|$.

Meanwhile, we have

$$\begin{aligned} |\sigma_n^2 - 1| &\leq M \int_{\{|X_{11}| \geq 2\delta_n \sqrt{n}\}} |X_{11}|^2 \\ &\leq 2\delta_n^{-2} n^{-1} \int_{\{|X_{11}| \geq \delta_n \sqrt{n}\}} |X_{11}|^4 = o(\delta_n^2 n^{-1}) \end{aligned} \quad (7.14)$$

and $|\mathbb{E}\hat{X}_{11}| \leq 2\big|\int_{\{|X_{11}| \geq \delta_n \sqrt{n}\}} X_{11}\big| = o(\delta_n n^{-3/2})$.

We therefore obtain

$$
\begin{aligned}
& \Big(\frac{1}{n}tr\big[(\hat{\mathbf{D}}_n\hat{\mathbf{X}}_n-\widetilde{\mathbf{D}}_n\widetilde{\mathbf{X}}_n)(\hat{\mathbf{D}}_n\hat{\mathbf{X}}_n-\widetilde{\mathbf{D}}_n\widetilde{\mathbf{X}}_n)^*\big]\Big)^{1/2} \\
= & \Big(\frac{1}{n}tr\big[\big(\hat{\mathbf{D}}_n(\hat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n)+(\hat{\mathbf{D}}_n-\widetilde{\mathbf{D}}_n)\widetilde{\mathbf{X}}_n\big)\big(\hat{\mathbf{D}}_n(\hat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n)+(\hat{\mathbf{D}}_n-\widetilde{\mathbf{D}}_n)\widetilde{\mathbf{X}}_n\big)^*\big]\Big)^{1/2} \\
= & \Big(\frac{1}{n}tr\big[\big(\hat{\mathbf{D}}_n(\hat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n)(\hat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n^*)\hat{\mathbf{D}}_n+\hat{\mathbf{D}}_n(\hat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n)\widetilde{\mathbf{X}}_n^*(\hat{\mathbf{D}}_n-\widetilde{\mathbf{D}}_n) \\
+ & (\hat{\mathbf{D}}_n-\widetilde{\mathbf{D}}_n)\widetilde{\mathbf{X}}_n(\hat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n^*)\hat{\mathbf{D}}_n+(\hat{\mathbf{D}}_n-\widetilde{\mathbf{D}}_n)\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^*(\hat{\mathbf{D}}_n-\widetilde{\mathbf{D}}_n)\big]\Big)^{1/2}.
\end{aligned} \tag{7.15}
$$

For the first term on the right hand side above, under the condition of Theorem 1, By (7.14), we can see

$$
\begin{aligned}
& \frac{1}{n}\mathbb{E}tr\big[\hat{\mathbf{D}}_n(\hat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n)(\hat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n)^*\hat{\mathbf{D}}_n^*\big] \\
= & \frac{1}{n}\sum_{i,j}\mathbb{E}\frac{[\hat{X}_{ij}-\tilde{X}_{ij}-\frac{1}{n}\sum_{k=1}^n(\hat{X}_{ik}-\tilde{X}_{ik})]^2}{\sum_{k=1}^n(\hat{X}_{ik}-\frac{1}{n}\sum_{l=1}^n\hat{X}_{il})^2} \\
= & \frac{(1-\frac{1}{\sigma_n})^2}{n}\sum_{i,j}\mathbb{E}\frac{(\hat{X}_{ij}-\frac{1}{n}\sum_{l=1}^n\hat{X}_{il})^2}{\sum_{k=1}^n(\hat{X}_{ik}-\frac{1}{n}\sum_{l=1}^n\hat{X}_{il})^2}=\frac{p}{n}(1-\frac{1}{\sigma_n})^2=o(\delta_n^4n^{-2}).
\end{aligned} \tag{7.16}
$$

The remaining terms of (7.15) can be similarly verified to have an order of o(1/n) and so (7.15) $=o(n^{-1/2})$. In view of above, we obtain $\int f_j(x)dG_n(x)=\int f_j(x)d\widetilde{G}_n(x)+o_p(1)$. Since $\mathbb{E}|\tilde{X}_{ij}|^4=\mathbb{E}|X_{ij}|^4+O(n^{-1})$, it will not affect $\kappa=\lim_{p\to\infty}\frac{1}{p}\sum_{i=1}^p\mathbb{E}|X_{i1}|^4$.

7.2.2 Convergence of $M_n(z)$

Let $\underline{\mathbf{B}}_n=\mathbf{Y}_n\mathbf{Y}_n^*$. The Stieltjes transforms of ESD and LSD for $\underline{\mathbf{B}}_n$ are denoted by $\underline{m}_n(z)$ and $\underline{m}_c(z)$ respectively. Their analogues for $\mathbf{B}_n$ are denoted by $m_n(z)$ and $m_c(z)$ respectively. Moreover, $\underline{m}_{c_n}(z)$ and $m_{c_n}(z)$ are $\underline{m}_c(z)$ and $m_c(z)$ respectively with c replaced by c_n. For ease of notation, we also denote $m_c(z)$ and $\underline{m}_c(z)$ by $m(z)$ and $\underline{m}(z)$ respectively with omitting the subscript c.

Since $M_n(z)=p\big[m_n(z)-m_{c_n}(z)\big]=n\big[\underline{m}_n(z)-\underline{m}_{c_n}(z)\big]$, we write for $z\in\mathcal{C}_n$, $M_n(z)=M_n^{(1)}(z)+M_n^{(2)}(z)$, where $M_n^{(1)}(z)=n\big[\underline{m}_n(z)-\mathbb{E}\underline{m}_n(z)\big]$ and $M_n^{(2)}(z)=n\big[\mathbb{E}\underline{m}_n(z)-\underline{m}_{c_n}(z)\big]$.

Following the steps in Bai and Silverstein (2004), it suffices to show the following four statements:

1. *Finite-dimensional convergence of* $M_n^{(1)}(z)$ *in distribution on* $\mathcal{C}_n$;

2. $M_n^{(1)}(z)$ *is tight on* $\mathcal{C}_n$;

3. $\{M_n^{(2)}(z)\}$ *for* $z\in\mathcal{C}_n$ *is bounded and equicontinuous*;

4. $M_n^{(2)}(z)$ *converges to a constant and find its limit.*

Step 1:

First, we introduce some notations. In the following proof, we assume that $v = \Im z \geq v_0 > 0$. Moreover,

$$\mathbf{r}_j = \frac{\mathbf{y}_j - \bar{\mathbf{y}}_j}{||\mathbf{y}_j - \bar{\mathbf{y}}_j||}, \forall j = 1, 2, \ldots, p; \quad \underline{\mathbf{B}}_j^{(n)} = \underline{\mathbf{B}}_n - \mathbf{r}_j\mathbf{r}_j^*, \quad \mathbf{D}(z) = \underline{\mathbf{B}}_n - z\mathbf{I}_n,$$
$$\mathbf{D}_j(z) = \mathbf{D}(z) - \mathbf{r}_j\mathbf{r}_j^*, \quad \beta_j(z) = \frac{1}{1 + \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j^*}, \quad \widetilde{\beta}_j(z) = \frac{1}{1 + \frac{1}{n}tr\mathbf{D}_j^{-1}(z)},$$
$$b_n(z) = \frac{1}{1 + \frac{1}{n}\mathbb{E}tr\mathbf{D}_1^{-1}(z)}, \quad \varepsilon_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{n}tr\mathbf{D}_j^{-1}(z),$$

and $\delta_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j - \frac{1}{n}tr\mathbf{D}_j^{-2}(z)$. By Lemma 5, we have for any $r \geq 2$

$$\mathbb{E}|\varepsilon_j(z)|^r \leq \frac{M}{v^{2r}}n^{-1}\delta_n^{2r-4} \tag{7.17}$$

and

$$\mathbb{E}|\delta_j(z)|^r \leq \frac{M}{v^{2r}}n^{-1}\delta_n^{2r-4}. \tag{7.18}$$

It is easy to see that

$$\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z) = -\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\beta_j(z), \tag{7.19}$$

where we use the formula that $\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1} = \mathbf{A}_2^{-1}(\mathbf{A}_2 - \mathbf{A}_1)\mathbf{A}_1^{-1}$ holds for any two invertible matrices $\mathbf{A}_1$ and $\mathbf{A}_2$. Note that $|\beta_j(z)|$, $|\widetilde{\beta}_j(z)|$ and $|b_n(z)|$ are bounded by $\frac{|z|}{v}$.

Let $\mathbb{E}_0(\cdot)$ denote expectation and $\mathbb{E}_j(\cdot)$ denote conditional expectation with respect to the σ-field generated by $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_j$, where $j = 1, 2, \ldots, p$. Next, we write $M_n^{(1)}(z)$ as a sum of martingale difference sequences(MDS), and then utilize the CLT of MDS which is listed in Lemma 4 to derive the asymptotic distribution of $M_n^{(1)}(z)$, which can be written as

$$\begin{aligned}
M_n^{(1)}(z) &= n[\underline{m}_n(z) - \mathbb{E}\underline{m}_n(z)] = tr[\mathbf{D}^{-1}(z) - \mathbb{E}\mathbf{D}^{-1}(z)] \\
&= \sum_{j=1}^{p}[tr\mathbb{E}_j\mathbf{D}^{-1}(z) - tr\mathbb{E}_{j-1}\mathbf{D}^{-1}(z)] \\
&= \sum_{j=1}^{p}\Big(tr\mathbb{E}_j[\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)] - tr\mathbb{E}_{j-1}[\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)]\Big) \\
&= -\sum_{j=1}^{p}(\mathbb{E}_j - \mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j.
\end{aligned} \tag{7.20}$$

Evidently, $\beta_j(z)$ can be written as

$$\beta_j(z) = \widetilde{\beta}_j(z) - \beta_j(z)\widetilde{\beta}_j(z)\varepsilon_j(z) = \widetilde{\beta}_j(z) - \widetilde{\beta}_j^2(z)\varepsilon_j(z) + \widetilde{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z).$$

From this and the definition of $\delta_j(z)$, (7.20) has the following expression

$$\begin{aligned}
&(\mathbb{E}_j-\mathbb{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j=(\mathbb{E}_j-\mathbb{E}_{j-1})\big[\big(\widetilde{\beta}_j(z)-\widetilde{\beta}_j^2(z)\varepsilon_j(z)\\
&+\widetilde{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z)\big)\big(\delta_j(z)+\frac{1}{n}tr\mathbf{D}_j^{-2}(z)\big)\big]=(\mathbb{E}_j-\mathbb{E}_{j-1})\big[\widetilde{\beta}_j(z)\delta_j(z)\\
&-\widetilde{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)-\widetilde{\beta}_j^2(z)\varepsilon_j(z)\frac{1}{n}tr\mathbf{D}_j^{-2}(z)+\widetilde{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j\big]\\
&=\ \mathbb{E}_j\big[\widetilde{\beta}_j(z)\delta_j(z)-\widetilde{\beta}_j^2(z)\varepsilon_j(z)\frac{1}{n}tr\mathbf{D}_j^{-2}(z)\big]\\
&-(\mathbb{E}_j-\mathbb{E}_{j-1})\big[\widetilde{\beta}_j^2(z)\big(\varepsilon_j(z)\delta_j(z)-\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j\varepsilon_j^2(z)\big)\big]-\mathbb{E}_{j-1}[\widetilde{\beta}_j(z)\delta_j(z)],
\end{aligned} \tag{7.21}$$

where the second equality uses the fact that $(\mathbb{E}_j-\mathbb{E}_{j-1})\widetilde{\beta}_j(z)\frac{1}{n}tr\mathbf{D}_j^{-2}(z)=0$.

By making a minor change to Lemma 8 in the supplementary (i.e. Replace $\mathbf{D}^{-1}(z)$ by $\mathbf{D}_j^{-1}(z)$), we have $\mathbb{E}\Big|\frac{1}{n}\mathbf{1}^T\mathbf{D}_j^{-1}(z)\mathbf{1}+\frac{1}{z}\Big|^2\to 0$. Thus

$$\begin{aligned}
-\sum_{j=1}^p\mathbb{E}_{j-1}\delta_j(z)&=\frac{1}{n(n-1)}\sum_{j=1}^p\sum_{k\neq\ell}\mathbb{E}_{j-1}\big(\mathbf{D}_j^{-2}(z)\big)_{k\ell}\\
&=\ \frac{1}{n(n-1)}\sum_{j=1}^p\mathbb{E}_{j-1}\big(\mathbf{1}^T\mathbf{D}_j^{-2}(z)\mathbf{1}-tr\mathbf{D}_j^{-2}(z)\big)\overset{i.p.}{\to}\ \frac{c}{z^2}-c\underline{m}'(z),
\end{aligned} \tag{7.22}$$

where the last step uses the fact that $\mathbf{1}^T\mathbf{D}_j^{-2}(z)\mathbf{1}=(\mathbf{1}^T\mathbf{D}_j^{-1}(z)\mathbf{1})'\to\frac{1}{z^2},\ \ \frac{1}{n}tr\mathbf{D}_j^{-2}(z)\to\underline{m}'(z)$ in L_2 by Lemma 2.3 of Bai and Silverstein (2004).

It follows from (7.22) that

$$\sum_{j=1}^p\mathbb{E}_{j-1}\widetilde{\beta}_j(z)\delta_j(z)\overset{i.p.}{\to}\frac{cm(z)}{z}-czm(z)\underline{m}'(z), \tag{7.23}$$

where we use the fact that

$$\begin{aligned}
&\mathbb{E}\Big|\sum_{j=1}^p\mathbb{E}_{j-1}\Big(\hat{\beta}_j(z)-b_1(z)\Big)\delta_j(z)\Big|\\
&\leq\sum_{j=1}^p\frac{1}{n(n-1)}\mathbb{E}\Big|\hat{\beta}_j(z)-b_1(z)\Big|\cdot\Big|\mathbf{1}^T\mathbf{D}_j^{-1}(z)\mathbf{1}-tr\mathbf{D}_j^{-1}(z)\Big|\to 0.
\end{aligned}$$

By (7.17) and (7.18), we have

$$\begin{aligned}
E\Big|\sum_{j=1}^p(\mathbb{E}_j-\mathbb{E}_{j-1})\widetilde{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)\Big|^2\ &=\ \sum_{j=1}^p\mathbb{E}\Big|(\mathbb{E}_j-\mathbb{E}_{j-1})\widetilde{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)\Big|^2\\
&\leq\ 4\sum_{j=1}^p\mathbb{E}\Big|\widetilde{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)\Big|^2=o(1),
\end{aligned} \tag{7.24}$$

where the first equality uses the fact that $(\mathbb{E}_j-\mathbb{E}_{j-1})\widetilde{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)$ is a martingale difference sequence. Therefore, $\sum_{j=1}^p(\mathbb{E}_j-\mathbb{E}_{j-1})\widetilde{\beta}_j^2(z)\varepsilon_j(z)\delta_j(z)$ converges to 0 in probability. By the same argument, we have

$$\sum_{j=1}^{p}(\mathbb{E}_j-\mathbb{E}_{j-1})\widetilde{\beta}_j^2(z)\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j\varepsilon_j^2(z)\overset{i.p.}{\to}0. \tag{7.25}$$

With (7.20)–(7.25), we need to consider the limit of the following term:

$$\sum_{i=1}^{r}\alpha_i\sum_{j=1}^{p}Y_j(z_i)=\sum_{j=1}^{p}\sum_{i=1}^{r}\alpha_iY_j(z_i), \tag{7.26}$$

where $\Im(z_i)\neq 0$, $\{\alpha_i : i=1,2,\ldots,r\}$ are constants, and

$$Y_j(z)=-\mathbb{E}_j\Big(\widetilde{\beta}_j(z)\delta_j(z)-\widetilde{\beta}_j^2(z)\varepsilon_j(z)\frac{1}{n}tr\mathbf{D}_j^{-2}(z)\Big)=-\mathbb{E}_j\frac{d}{dz}\big(\widetilde{\beta}_j(z)\varepsilon_j(z)\big).$$

By Lemma 5, we obtain,

$$\mathbb{E}|Y_j(z)|^4\le K\mathbb{E}|\varepsilon_j(z)|^4=o\left(\frac{1}{p}\right). \tag{7.27}$$

It follows from (7.27) that

$$\sum_{j=1}^{p}\mathbb{E}\left(\left|\sum_{i=1}^{r}\alpha_iY_j(z_i)\right|^2 I_{(|\sum_{i=1}^r\alpha_iY_j(z_i)|\ge\varepsilon)}\right)\le\frac{1}{\varepsilon^2}\sum_{j=1}^{p}\mathbb{E}\left|\sum_{i=1}^{r}\alpha_iY_j(z_i)\right|^4\to 0.$$

From Lemma 4, it suffices to prove that

$$\sum_{j=1}^{p}\mathbb{E}_{j-1}[Y_j(z_1)Y_j(z_2)] \tag{7.28}$$

converges in probability to a constant. Once it is proved, we can conclude that $M_n^{(1)}(z)$ converges in finite dimension to a normal distribution.

Since

$$\frac{\partial^2}{\partial z_1\partial z_2}\Big(\sum_{j=1}^{p}\mathbb{E}_{j-1}\big[\mathbb{E}_j\big(\widetilde{\beta}_j(z_1)\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\widetilde{\beta}_j(z_2)\varepsilon_j(z_2)\big)\big]\Big)=(7.28),$$

and by the same arguments as those on page 571 of Bai and Silverstein (2004), it is enough to consider the limit of

$$\sum_{j=1}^{p}\mathbb{E}_{j-1}\big[\mathbb{E}_j\big(\widetilde{\beta}_j(z_1)\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\widetilde{\beta}_j(z_2)\varepsilon_j(z_2)\big)\big]. \tag{7.29}$$

By the fact that

$$\lim_{n\to}\mathbb{E}\left|\frac{1}{n}tr\mathbb{E}_j\Big(\mathbf{D}_j^{-1}(z)\Big)-\underline{m}(z)\right|^2=0, \tag{7.30}$$

$|\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j| \le \frac{|z|}{v^2}$ and Lemma 3, we obtain

$$\mathbb{E}\big|\widetilde{\beta}_j(z_i) - b_n(z_i)\big|^2 \le \frac{K}{n}. \tag{7.31}$$

By (7.31), we have

$$\mathbb{E}\big|\mathbb{E}_{j-1}[\mathbb{E}_j\big(\widetilde{\beta}_j(z_1)\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\widetilde{\beta}_j(z_2)\varepsilon_j(z_2)\big)] - \mathbb{E}_{j-1}[\mathbb{E}_j\big(b_n(z_1)\varepsilon_j(z_1)\big)\mathbb{E}_j\big(b_n(z_2)\varepsilon_j(z_2)\big)]\big| = O(n^{-3/2}).$$

From this, it follows that

$$\sum_{j=1}^p \mathbb{E}_{j-1}\big[\mathbb{E}_j\big(\widetilde{\beta}_j(z_1)\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\widetilde{\beta}_j(z_2)\varepsilon_j(z_2)\big)\big] - b_n(z_1)b_n(z_2)\sum_{j=1}^p \mathbb{E}_{j-1}\big[\mathbb{E}_j\big(\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\varepsilon_j(z_2)\big)\big] \stackrel{i.p.}{\to} 0.$$

Then it is enough to prove that

$$b_n(z_1)b_n(z_2)\sum_{j=1}^p \mathbb{E}_{j-1}\big[\mathbb{E}_j\big(\varepsilon_j(z_1)\big)\mathbb{E}_j\big(\varepsilon_j(z_2)\big)\big] \tag{7.32}$$

converges to a constant in probability, which further gives the limit of (7.26).

By Lemma 6, (7.32) becomes

$$(7.32) = \begin{cases} J_1 + 2J_2 + J_3 + o_P(1), & \textit{under the real case;} \\ J_1 + J_2 + J_3 + J_4 + o_P(1), & \textit{under the complex case,} \end{cases} \tag{7.33}$$

where

$$J_1 = \frac{1}{n^3}b_n(z_1)b_n(z_2)\Big[\sum_{j=1}^p(1-\mathbb{E}|X_{j1}|^4)tr\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)tr\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big)\Big];$$
$$J_2 = \frac{1}{n^2}b_n(z_1)b_n(z_2)\Big[\mathbb{E}\sum_{j=1}^p tr\big[\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big)\big];$$
$$J_3 = \frac{1}{n^2}b_n(z_1)b_n(z_2)\Big[\sum_{j=1}^p(\mathbb{E}|X_{j1}|^4 - 2 - |\mathbb{E}X_{j1}^2|^2)\sum_{k=1}^n \mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)_{kk}\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big)_{kk}\Big];$$
$$J_4 = b_n(z_1)b_n(z_2)\frac{1}{n^2}\Big[\sum_{j=1}^p |\mathbb{E}X_{j1}^2|^2\big[\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)^T\big)\big]\Big].$$

Next, we study the limit of the term J_2. Let $\mathbf{D}_{ij}(z) = \mathbf{D}(z) - \mathbf{r}_i\mathbf{r}_i^* - \mathbf{r}_j\mathbf{r}_j^*$, $b_1(z) = \frac{1}{1+\frac{1}{n}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)}$ and $\beta_{ij}(z) = \frac{1}{1+\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i}$.

We have the equality $\mathbf{D}_j(z_1) + z_1\mathbf{I}_n - \frac{p-1}{n}b_1(z_1)\mathbf{I}_n = \sum_{i\ne j}^p \mathbf{r}_i\mathbf{r}_i^* - \frac{p-1}{n}b_1(z_1)\mathbf{I}_n$. Multiplying by $(z_1\mathbf{I}_n - \frac{p-1}{n}b_1(z)\mathbf{I}_n)^{-1}$ on the left-hand side and $\mathbf{D}_j^{-1}(z_1)$ on the right-hand side, and using

$$\mathbf{r}_i^*\mathbf{D}_j^{-1}(z_1) = \beta_{ij}(z_1)\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1), \tag{7.34}$$

we get

$$\begin{aligned}\mathbf{D}_j^{-1}(z_1) &= -\mathbf{H}_n(z_1) + \sum_{i\neq j}^{p} \beta_{ij}(z_1)\mathbf{H}_n(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1) - \frac{p-1}{n} b_1(z_1)\mathbf{H}_n(z_1)\mathbf{D}_j^{-1}(z_1) \\ &= -\mathbf{H}_n(z_1) + b_1(z_1)A(z_1) + B(z_1) + C(z_1), \end{aligned} \tag{7.35}$$

where $\mathbf{H}_n(z_1) = \big(z_1\mathbf{I}_n - \frac{p-1}{n} b_1(z_1)\mathbf{I}_n\big)^{-1}$, $A(z_1) = \sum_{i\neq j}^{p} \mathbf{H}_n(z_1)\big(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{I}_n\big)\mathbf{D}_{ij}^{-1}(z_1)$, $B(z_1) = \sum_{i\neq j}^{p} \big(\beta_{ij}(z_1) - b_1(z_1)\big)\mathbf{H}_n(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)$ and

$$C(z_1) = n^{-1} b_1(z_1)\mathbf{H}_n(z_1)\sum_{i\neq j}^{p}\big(\mathbf{D}_{ij}^{-1}(z_1) - \mathbf{D}_j^{-1}(z_1)\big).$$

For any real t, $\Big|1 - \frac{t}{z(1+n^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z))}\Big|^{-1} \leq \frac{\big|z\big(1+n^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z)\big)\big|}{\Im\big(z(1+n^{-1}\mathbb{E}tr\mathbf{D}_{12}^{-1}(z))\big)} \leq \frac{|z|(1+1/v_0)}{v_0}$.

Thus,

$$||\mathbf{H}_n(z_1)|| \leq \frac{1+1/v_0}{v_0}. \tag{7.36}$$

For any random matrix $\mathbf{M}$, denote its nonrandom bound by $|||\mathbf{M}|||$. From (7.31), Lemma 5 and (7.36), we get

$$\begin{aligned}\mathbb{E}\big|tr\mathbf{B}(z_1)\mathbf{M}\big| &\leq p\mathbb{E}^{1/2}\big(\big|\beta_{12}(z_1) - b_1(z_1)\big|^2\big)\cdot\mathbb{E}^{1/2}\big(\big|\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{M}H_n(z_1)\mathbf{r}_i\big|^2\big) \\ &\leq K|||\mathbf{M}|||\frac{|z_1|^2(1+1/v_0)}{v_0^5}n^{1/2}. \end{aligned} \tag{7.37}$$

For any $n\times n$ matrix $\mathbf{A}$, we have

$$\big|tr\big(\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)\big)\mathbf{A}\big| \leq \frac{||\mathbf{A}||}{\Im(z)}, \tag{7.38}$$

With (7.38), we obtain

$$\big|tr\mathbf{C}(z_1)\mathbf{M}\big| \leq |||\mathbf{M}|||\frac{|z_1|(1+1/v_0)}{v_0^3}. \tag{7.39}$$

For any nonrandom $\mathbf{M}$, it follows from Lemma 5 and (7.36) that

$$\begin{aligned}\mathbb{E}\big|tr\mathbf{A}(z_1)\mathbf{M}\big| &\leq K\mathbb{E}^{1/2}\big(tr\mathbf{D}_{ij}^{-1}(z_1)\mathbf{M}\mathbf{H}_n(z_1) \\ &\qquad \cdot\mathbf{H}_n(\bar{z}_1)\mathbf{M}^*\mathbf{D}_{ij}^{-1}(\bar{z}_1)\big) \\ &\leq K||\mathbf{M}||\frac{1+1/v_0}{v_0^2}n^{1/2}. \end{aligned} \tag{7.40}$$

By using (7.19), we can write $tr\mathbb{E}_j\big(\mathbf{A}(z_1)\big)\mathbf{D}_j^{-1}(z_2) = A_1(z_1,z_2) + A_2(z_1,z_2) + A_3(z_1,z_2) + R(z_1,z_2)$, where

$$
\begin{aligned}
A_1(z_1,z_2) &= -tr\sum_{i<j}^{p}\mathbf{H}_n(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\beta_{ij}(z_2)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_2)\\
&= -\sum_{i<j}^{p}\beta_{ij}(z_2)\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_2)\mathbf{H}_n(z_1)\mathbf{r}_i;\\
A_2(z_1,z_2) &= -tr\sum_{i<j}^{p}\mathbf{H}_n(z_1)n^{-1}\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2)-\mathbf{D}_{ij}^{-1}(z_2)\big);\\
A_3(z_1,z_2) &= tr\sum_{i<j}^{p}\mathbf{H}_n(z_1)\big(\mathbf{r}_i\mathbf{r}_i^*-n^{-1}\mathbf{I}_n\big)\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2);\\
R(z_1,z_2) &= tr\mathbb{E}_j\sum_{i>j}\mathbf{H}_n(z_1)\Big(-\frac{1}{n(n-1)}\mathbf{e}\mathbf{e}^*+\frac{1}{n(n-1)}\mathbf{I}_n\Big)\mathbf{D}_{ij}^{-1}(z_1)\mathbf{D}_j^{-1}(z_2),
\end{aligned}
$$

where $\mathbf{e}$ is an n-dimensional vector with all elements being 1 and $\mathbb{E}\bar{\mathbf{r}}_{ik}\mathbf{r}_{ij} = -\frac{1}{n(n-1)}$, $k \neq j$(see (1.16) in the supplementary).

It is easy to see that $R(z_1,z_2) = O(1)$. We get from (7.38) and (7.36) that $|A_2(z_1,z_2)| \leq \frac{1+1/v_0}{v_0^2}$. Similar to (7.37), we have $\mathbb{E}|A_3(z_1,z_2)| \leq \frac{1+1/v_0}{v_0^3}n^{1/2}$. Using Lemma 5 and (7.31), we have, for $i<j$,

$$
\begin{aligned}
&\mathbb{E}\big|\beta_{ij}(z_2)\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_2)\mathbf{H}_n(z_1)\mathbf{r}_i\\
&\quad -b_1(z_2)n^{-2}tr\big(\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2)\big)tr\big(\mathbf{D}_{ij}^{-1}(z_2)\mathbf{H}_n(z_1)\big)\big| \leq Kn^{-1/2}. \qquad (7.41)
\end{aligned}
$$

By (7.38), we have

$$
\begin{aligned}
&\big|tr\big(\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\mathbf{D}_{ij}^{-1}(z_2)\big)tr\big(\mathbf{D}_{ij}^{-1}(z_2)\mathbf{H}_n(z_1)\big)\\
&\quad -tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\big)tr\big(\mathbf{D}_j^{-1}(z_2)\mathbf{H}_n(z_1)\big)\big| \leq Kn. \qquad (7.42)
\end{aligned}
$$

It follows from (7.41) and (7.42) that

$$
\mathbb{E}\Big|A_1(z_1,z_2)+\frac{j-1}{n^2}b_1(z_2)tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\big)tr\big(\mathbf{D}_j^{-1}(z_2)\mathbf{H}_n(z_1)\big)\Big| \leq Kn^{1/2}.
$$

Therefore, by (7.35)–(7.42), we obtain that

$$
\begin{aligned}
&tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\big)\Big[1+\frac{j-1}{n^2}b_1(z_1)b_1(z_2)tr\big(\mathbf{D}_j^{-1}(z_2)\mathbf{H}_n(z_1)\Big]\\
=\ & -tr\big(\mathbf{H}_n(z_1)\mathbf{D}_j^{-1}(z_2)\big)+A_4(z_1,z_2),
\end{aligned}
$$

where $\mathbb{E}|A_4(z_1,z_2)| \leq Kn^{1/2}$.

By (7.35) for $\mathbf{D}_j^{-1}(z_2)$ and (7.37)-(7.40), we have

$$
\begin{aligned}
&tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\big)\Big[1-\frac{j-1}{n^2}b_1(z_1)b_1(z_2)tr\big(\mathbf{H}_n(z_2)\mathbf{H}_n(z_1)\big)\Big] \qquad (7.43)\\
=\ & tr\Big(\mathbf{H}_n(z_2)\mathbf{H}_n(z_1)\Big)+A_5(z_1,z_2),
\end{aligned}
$$

where $\mathbb{E}|A_5(z_1,z_2)| \le Kn^{1/2}$.

From (7.38), we have $|b_1(z) - b_n(z)| \le Kn^{-1}$. Using $\mathbb{E}\big|\frac{1}{n}tr\mathbf{D}^{-1}(z) - \mathbb{E}\frac{1}{n}tr\mathbf{D}^{-1}(z)\big|^k \le C_k n^{-k/2}$, we have

$$|b_n(z) - \mathbb{E}\beta_1(z)| \le Kn^{-1/2}. \tag{7.44}$$

As in (2.2) of Silverstein (1995), one may verify that

$$m_n(z) = -\frac{1}{pz}\sum_{j=1}^{p}\beta_j(z). \tag{7.45}$$

It follows from (7.45) that

$$\mathbb{E}\beta_1(z) = -z\mathbb{E}m_n(z). \tag{7.46}$$

From (7.44), (7.46) and Lemma 5, we have

$$|b_1(z) + zm_{c_n}(z)| \le Kn^{-1/2}. \tag{7.47}$$

Let $\mathbf{Q}_n(z) = \big(\mathbf{I}_n + \frac{p-1}{n}m_{c_n}(z)\mathbf{I}_n\big)^{-1}$. So by (7.43), we get

$$\begin{aligned}&tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\big)\big[1 - \frac{j-1}{n^2}m_{c_n}(z_1)m_{c_n}(z_2)tr\mathbf{Q}_n(z_2)\mathbf{Q}_n(z_1)\big] \\ &= \frac{1}{z_1z_2}tr\big(\mathbf{Q}_n(z_2)\mathbf{Q}_n(z_1)\big) + A_6(z_1,z_2),\end{aligned} \tag{7.48}$$

where $\mathbb{E}|A_6(z_1,z_2)| \le Kn^{1/2}$.

Rewrite (7.48) as

$$\begin{aligned}&tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbf{D}_j^{-1}(z_2)\big)\big[1 - \frac{j-1}{n}\frac{m_{c_n}(z_1)m_{c_n}(z_2)}{\big(1+\frac{p-1}{n}m_{c_n}(z_2)\big)\big(1+\frac{p-1}{n}m_{c_n}(z_1)\big)}\big] \\ &= \frac{n}{z_1z_2}\frac{1}{\big(1+\frac{p-1}{n}m_{c_n}(z_1)\big)\big(1+\frac{p-1}{n}m_{c_n}(z_2)\big)} + A_6(z_1,z_2).\end{aligned} \tag{7.49}$$

Then J_2 can be written as $J_2 = a_n(z_1,z_2)\frac{1}{p}\sum_{j=1}^{p}\frac{1}{1-\frac{j-1}{p}a_n(z_1,z_2)} + A_7(z_1,z_2)$, where $a_n(z_1,z_2) = \frac{c_n m_{c_n}(z_1)m_{c_n}(z_2)}{\big(1+\frac{p-1}{n}m_{c_n}(z_1)\big)\big(1+\frac{p-1}{n}m_{c_n}(z_2)\big)}$ and $\mathbb{E}|A_7(z_1,z_2)| \le Kn^{-1/2}$.

Note that the limit of $a_n(z_1,z_2)$ is $a(z_1,z_2) = \frac{cm(z_1)m(z_2)}{\big(1+cm(z_1)\big)\big(1+cm(z_2)\big)}$. Thus by (7.49), the i.p. limit of $\frac{\partial^2}{\partial z_2\partial z_1}J_2$ is

$$\begin{aligned}&\frac{\partial^2}{\partial z_2\partial z_1}\int_0^{a(z_1,z_2)}\frac{1}{1-z}dz = \frac{\partial}{\partial z_2}\big(\frac{\partial a(z_1,z_2)/\partial z_1}{1-a(z_1,z_2)}\big) \\ =\;& \frac{\partial}{\partial z_2}\Big(\frac{cm(z_2)m'(z_1)}{\Big(1+cm(z_1)\Big)\Big(1+c\big(m(z_1)+m(z_2)\big)+c(c-1)m(z_1)m(z_2)\Big)}\Big) \\ =\;& \frac{cm'(z_1)m'(z_2)}{\Big(1+c\big(m(z_1)+m(z_2)\big)+c(c-1)m(z_1)m(z_2)\Big)^2}.\end{aligned}$$

For $\frac{\partial^2}{\partial z_1 \partial z_2} J_1$ in (7.33), similar to (7.22), by (7.47) and (7.30), we have

$$\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{p}\mathbb{E}|X_{j1}|^4\frac{1}{n}tr\mathbb{E}_j\mathbf{D}_j^{-1}(z_1)\frac{1}{n}tr\mathbb{E}_j\mathbf{D}_j^{-1}(z_2)-\frac{\underline{m}_{c_n}(z_1)\underline{m}_{c_n}(z_2)}{n}\sum_{j=1}^{p}\mathbb{E}|X_{j1}|^4\right|=o(1)$$

So we can conclude that

$$J_1\overset{i.p.}{\to}\Big(1-\kappa\Big)cz_1m(z_1)z_2m(z_2)\underline{m}(z_1)\underline{m}(z_2)=\Big(1-\kappa\Big)\frac{c\underline{m}(z_1)\underline{m}(z_2)}{\big(1+\underline{m}(z_1)\big)\big(1+\underline{m}(z_2)\big)},$$

where the equality above uses the relation between $m(z)$ and $\underline{m}(z)$: $m(z)=-\frac{1}{z\left(1+\underline{m}(z)\right)}$. Then the second derivative of J_1 with respect to z_1 and z_2 is

$$\frac{\partial^2}{\partial z_1\partial z_2}J_1\overset{i.p.}{\to}\Big(1-\kappa\Big)\frac{c\underline{m}^{'}(z)\underline{m}^{'}(z_2)}{\big(1+\underline{m}(z_1)\big)^2\big(1+\underline{m}(z_2)\big)^2}.$$

The next aim is to establish the limit of $\frac{\partial^2}{\partial z_1 \partial z_2} J_3$ in (7.33). It is enough to find the limit of $\frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{p}(\mathbb{E}|X_{j1}|^4-2-|\mathbb{E}X_{j1}^2|^2)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)_{kk}\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big)_{kk}$.

First, we claim that

$$\frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{p}(\mathbb{E}|X_{j1}|^4-2-|\mathbb{E}X_{j1}^2|^2)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)-\mathbb{E}\mathbf{D}_j^{-1}(z_1)\big)_{kk}\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big)_{kk}=O_p(n^{-1/2}). \tag{7.50}$$

Actually,

$$\begin{aligned}&\mathbb{E}|\frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{p}(\mathbb{E}|X_{j1}|^4-2-|\mathbb{E}X_{j1}^2|^2)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)-\mathbb{E}\mathbf{D}_j^{-1}(z_1)\big)_{kk}\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)\big)_{kk}|\\ \leq\ &\frac{pK}{n^2v_0}\sum_{k=1}^{n}\mathbb{E}|\mathbf{e}_k'(\mathbf{D}_1^{-1}(z_1)-\mathbb{E}\mathbf{D}_1^{-1}(z_1))\mathbf{e}_k|\leq Kn^{-1/2},\end{aligned} \tag{7.51}$$

where the last inequality follows from (1.48) in the supplementary (replacing $\mathbf{D}$ by $\mathbf{D}_1$). With (7.51), it remains to find the limit of

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\big(\mathbf{D}_j^{-1}(z_1)\big)_{kk}\mathbb{E}\big(\mathbf{D}_j^{-1}(z_2)\big)_{kk}. \tag{7.52}$$

It is easy to see that the sum of expectations in (7.52) is exactly the same for any j. Moreover, we have

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\big(\mathbf{D}_j^{-1}(z_1)\big)_{kk}\mathbb{E}\big(\mathbf{D}_j^{-1}(z_2)\big)_{kk}\overset{i.p.}{\to}\underline{m}(z_1)\underline{m}(z_2). \tag{7.53}$$

By (7.6), (7.47) and (7.53), we get $J_3 \xrightarrow{i.p.} \Big(\kappa - 2 - |\mathbb{E}X_{11}^2|^2\Big) c z_1 z_2 m(z_1)\underline{m}(z_1)m(z_2)\underline{m}(z_2)$. Thus the limit of $\frac{\partial^2}{\partial z_1 \partial z_2} J_3$ is

$$\begin{aligned}\frac{\partial^2}{\partial z_1 \partial z_2} J_3 &\xrightarrow{i.p.} \Big(\kappa - |\mathbb{E}X_{11}^2|^2 - 2\Big) \cdot c(m(z_1)\underline{m}(z_1) + z_1 m(z_1)\underline{m}'(z_1) + z_1 m'(z_1)\underline{m}(z_1)) \\ &\times (m(z_2)\underline{m}(z_2) + z_2 m(z_2)\underline{m}'(z_2) + z_2 m'(z_2)\underline{m}(z_2)).\end{aligned} \tag{7.54}$$

For the complex case, the limit of $\frac{\partial^2}{\partial z_1 \partial z_2} J_4$ is derived in Lemma 9 in the supplementary.

Step 2:

The tightness of $M_n^{(1)}(z)$ is similar to that provided in Bai and Silverstein (2004). It is sufficient to prove the moment condition (12.51) of Billingsley (1968), i.e.

$$\sup_{n; z_1, z_2 \in \mathcal{C}_n} \frac{\mathbb{E}|M_n^{(1)}(z_1) - M_n^{(1)}(z_2)|^2}{|z_1 - z_2|^2} \tag{7.55}$$

is finite.

Before proceeding, we provide some results needed in the proof later. First, moments of $||\mathbf{D}^{-1}(z)||$, $||\mathbf{D}_j^{-1}(z)||$ and $||\mathbf{D}_{ij}^{-1}(z)||$ are bounded in n and $z \in \mathcal{C}_n$. It is easy to see that it is true for $z \in \mathcal{C}_u$ and for $z \in \mathcal{C}_\ell$ if $x_\ell < 0$. For $z \in \mathcal{C}_r$ or, if $x_\ell > 0$, $z \in \mathcal{C}_\ell$, we have from Lemma 7 that

$$\begin{aligned}\mathbb{E}||\mathbf{D}_j^{-1}(z)||^m &\leq K_1 + v^{-m} P(||\mathbf{B}_j|| \geq \eta_r \ or \ \lambda_{\min}^{\mathbf{B}_j} \leq \eta_\ell) \\ &\leq K_1 + K_2 n^m \varepsilon^{-m} n^{-\ell} \leq K\end{aligned}$$

for large ℓ. Here η_r is any number between $(1+\sqrt{c})^2$ and x_r; if $x_\ell > 0$, η_ℓ is any number between x_ℓ and $(1-\sqrt{c})^2$ and if $x_\ell < 0$, η_ℓ can be any negative number. So for any positive integer m,

$$\max\Big(\mathbb{E}||\mathbf{D}^{-1}(z)||^m, \mathbb{E}||\mathbf{D}_j^{-1}(z)||^m, \mathbb{E}||\mathbf{D}_{ij}^{-1}(z)||^m\Big) \leq K. \tag{7.56}$$

By the argument above, we can extend (7.4) in the remark of Lemma 5 and get

$$\Big|\mathbb{E}\Big(a(v) \prod_{\ell=1}^{q} \big(\mathbf{r}_1^* \mathbf{B}_\ell(v) \mathbf{r}_1 - n^{-1} tr \mathbf{B}_\ell(v)\big)\Big)\Big| \leq K n^{-1} \delta_n^{(2q-4)}, \tag{7.57}$$

where $\mathbf{B}_\ell(v)$ is independent of $\mathbf{r}_1$ and

$$\max(|a(v)|, ||\mathbf{B}_\ell(v)||) \leq K(1 + n^s I(||\mathbf{B}_n|| \geq \eta_r \ or \ \lambda_{\min}^{\widetilde{\mathbf{B}}} \leq \eta_\ell)),$$

with $\widetilde{\mathbf{B}}$ being $\underline{\mathbf{B}}_j^{(n)}$ or $\underline{\mathbf{B}}_n$.

By (7.57), we have

$$\mathbb{E}|\varepsilon_j(z)|^m \leq K_m n^{-1} \delta_n^{2m-4}. \tag{7.58}$$

Let $\gamma_j(z) = \mathbf{r}_j^* \mathbf{D}_j^{-1}(z)\mathbf{r}_j - n^{-1}\mathbb{E}tr\mathbf{D}_j^{-1}(z)$. By Lemma 3, (7.57) and Hölder's inequality, with similar derivation on page 580 of Bai and Silverstein (2004), we have

$$\mathbb{E}|\gamma_j(z) - \varepsilon_j(z)|^m \le \frac{K_m}{n^{m/2}}. \tag{7.59}$$

It follows from (7.58) and (7.59) that

$$\mathbb{E}|\gamma_j|^m \le K_m n^{-1} \delta_n^{2m-4}, \quad m \ge 2. \tag{7.60}$$

Next, we prove that $b_n(z)$ is bounded. With (7.57), we have for any $m \ge 1$,

$$\mathbb{E}|\beta_1(z)|^m \le K_m. \tag{7.61}$$

Since $b_n(z) = \beta_1(z) + \beta_1(z) b_n(z)\gamma_1(z)$, it is derived from (7.60) and (7.61) that $|b_n(z)| \le K_1 + K_2|b_n(z)|n^{-1/2}$.

Then

$$|b_n(z)| \le \frac{K_1}{1 - K_2 n^{-1/2}} \le K. \tag{7.62}$$

With (7.57)–(7.62) and the same approach on page 581-583 of Bai and Silverstein (2004), we can obtain that (7.55) is finite.

Steps 3 and 4:

First, we list some results which are used later in this part. The derivations of these results are similar to those for sample covariance matrices in Bai and Silverstein (2004):

$$\sup_{z \in \mathcal{C}_n} |\underline{m}_n(z) - \underline{m}(z)| \to 0, \quad as \quad n \to \infty.$$

$$\sup_{n; z \in \mathcal{C}_n} \left|\left| \left(c_n \mathbb{E}\underline{m}_n(z)\mathbf{I}_n + \mathbf{I}_n\right)^{-1} \right|\right| < \infty. \tag{7.63}$$

$$\sup_{z \in \mathcal{C}_n} \left| \frac{\mathbb{E}\left(\underline{m}_n^2(z)\right)}{\left(1 + c_n \mathbb{E}\underline{m}_n(z)\right)^2} \right| < \xi, \quad \xi \in (0,1).$$

$$\mathbb{E}\left| tr\mathbf{D}_1^{-1}(z)\mathbf{M} - \mathbb{E}tr\mathbf{D}_1^{-1}(z)\mathbf{M}\right|^2 \le K||\mathbf{M}||^2. \tag{7.64}$$

Next, we derive an identity. Let

$$\mathbf{G}_n(z) = c_n \mathbb{E}m_n(z)\mathbf{I}_n + \mathbf{I}_n. \tag{7.65}$$

Write $\underline{\mathbf{B}}_n - z\mathbf{I}_n - \big(-z\mathbf{G}_n(z)\big)$ as $\sum_{j=1}^p \mathbf{r}_j\mathbf{r}_j^* - \big(-zc_n\mathbb{E}m_n(z)\big)\mathbf{I}_n$. Taking first inverse and then expected value, we get

$$\begin{aligned}
&\big(-z\mathbf{G}_n(z)\big)^{-1} - \mathbb{E}\big(\underline{\mathbf{B}}_n - z\mathbf{I}_n\big)^{-1} \\
=\;& \big(-z\mathbf{G}_n(z)\big)^{-1}\mathbb{E}\Big[\Big(\sum_{j=1}^p \mathbf{r}_j\mathbf{r}_j^* - (-zc_n\mathbb{E}m_n(z)\mathbf{I}_n)\Big)\big(\underline{\mathbf{B}}_n - z\mathbf{I}_n\big)^{-1}\Big] \\
=\;& -z^{-1}\sum_{j=1}^p \mathbb{E}\beta_j(z)\big[\mathbf{G}_n^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\big(\underline{\mathbf{B}}_{(j)}^n - z\mathbf{I}_n\big)^{-1}\big] \\
&+z^{-1}\mathbb{E}\big[\mathbf{G}_n^{-1}(z)\big(-zc_n\mathbb{E}m_n(z)\big)\mathbf{I}_n\big(\underline{\mathbf{B}}_n - z\mathbf{I}_n\big)^{-1}\big] \\
=\;& -z^{-1}\sum_{j=1}^p \mathbb{E}\beta_j(z)\big[\mathbf{G}_n^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\big(\underline{\mathbf{B}}_{(j)}^n - z\mathbf{I}_n\big)^{-1} - \frac{1}{n}\mathbf{G}_n^{-1}(z)\mathbb{E}\big(\underline{\mathbf{B}}_n - z\mathbf{I}_n\big)^{-1}\big] \\
=\;& -z^{-1}\sum_{j=1}^p \mathbb{E}\beta_j(z)\big[\mathbf{G}_n^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z) - \frac{1}{n}\mathbf{G}_n^{-1}(z)\mathbb{E}\mathbf{D}^{-1}(z)\big],
\end{aligned}$$

where the second equality uses (7.34) and the third equality uses (7.46).

Taking trace on both sides and dividing by $-\frac{p}{z}$, we get

$$\begin{aligned}
&\frac{1}{c_n\big(1+c_n\mathbb{E}m_n(z)\big)} + \frac{z}{c_n}\mathbb{E}\underline{m}_n(z) \\
=\;& \frac{1}{p}\sum_{j=1}^p \mathbb{E}\beta_j(z)\big[\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_j - \frac{1}{n}\mathbb{E}tr\left(\mathbf{G}_n^{-1}(z)\mathbf{D}^{-1}(z)\right)\big],
\end{aligned}$$

Next, we investigate the limit of

$$\begin{aligned}
&n\Big(\frac{1}{c_n\big(1+c_n\mathbb{E}m_n(z)\big)} + \frac{z}{c_n}\mathbb{E}\underline{m}_n(z)\Big) \\
=\;& \frac{n}{p}\sum_{j=1}^p \mathbb{E}\beta_j(z)\big[\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_j - \frac{1}{n}\mathbb{E}tr\left(\mathbf{G}_n^{-1}(z)\mathbf{D}^{-1}(z)\right)\big].
\end{aligned}\tag{7.66}$$

We need only to calculate the limit of $\mathbb{E}\beta_1(z)\big[\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 - \frac{1}{n}\mathbb{E}tr\big(\mathbf{G}_n^{-1}(z)\mathbf{D}^{-1}(z)\big)\big]$. By similar arguments to Steps 1 and 2, we can get the limit of (7.66). Let $\gamma_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{n}\mathbb{E}tr\mathbf{D}_j^{-1}(z)$. By (7.34) and the fact that $\beta_1(z) = b_n(z)\big(1-\beta_1(z)\gamma_1(z)\big)$, we have

$$\begin{aligned}
&\mathbb{E}tr\big(\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z)\big) - \mathbb{E}tr\big(\mathbf{G}_n^{-1}(z)\mathbf{D}^{-1}(z)\big) \\
=\;& \mathbb{E}\beta_1(z)tr\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z) \\
=\;& b_n(z)\mathbb{E}\left[(1-\beta_1(z)\gamma_1(z))\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\right].
\end{aligned}\tag{7.67}$$

From Lemma 5 and (7.63), we get

$$\big|\mathbb{E}\beta_1(z)\gamma_1(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z)\mathbf{r}_1\big| \le Kn^{-1}.$$

Therefore,

$$\Big|(7.67) - n^{-1}b_n(z)\mathbb{E}tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z)\Big| \le Kn^{-1}.$$

Since $\beta_1(z) = b_n(z) - b_n^2(z)\gamma_1(z) + \beta_1(z)b_n^2(z)\gamma_1^2(z)$, we have

$$\begin{aligned}
& n\mathbb{E}\beta_1(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 - \mathbb{E}\beta_1(z)\mathbb{E}tr\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z) \\
= \; & \frac{nb_n(z)}{c_n\mathbb{E}m_n(z)+1}\mathbb{E}\gamma_1(z) + b_n^2(z)\frac{tr\mathbb{E}\mathbf{D}_1^{-1}(z)}{c_n\mathbb{E}m_n(z)+1}\mathbb{E}\gamma_1(z) - b_n^2(z)n\mathbb{E}\gamma_1(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 \\
& +b_n^2(z)\Big(n\mathbb{E}\beta_1(z)\gamma_1^2(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 - \big(\mathbb{E}\beta_1(z)\gamma_1^2(z)\big)\mathbb{E}tr\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z)\Big) \\
& +nb_1(z)\Big(\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\big[\mathbf{G}_n(z)\big]^{-1}\mathbf{r}_1 - \mathbb{E}tr\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z)\Big) + o(1) \\
= \; & \frac{nb_n(z)}{c_n\mathbb{E}m_n(z)+1}\mathbb{E}\gamma_1(z) + b_n^2(z)\frac{tr\mathbb{E}\mathbf{D}_1^{-1}(z)}{c_n\mathbb{E}m_n(z)+1}\mathbb{E}\gamma_1(z) - b_n^2(z)n\mathbb{E}\gamma_1(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 \\
& +b_n^2\Big(\mathbb{E}\big[n\beta_1(z)\gamma_1^2(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 - \beta_1(z)\gamma_1^2(z)tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\big]\Big) \\
& +b_n^2(z)Cov\big(\beta_1(z)\gamma_1^2(z), tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\big) + o(1).
\end{aligned} \tag{7.68}$$

By (7.60) and (7.67), we have

$$\Big|\mathbb{E}\big[n\beta_1(z)\gamma_1^2(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 - \beta_1(z)\gamma_1^2(z)tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\big]\Big| \le K\delta_n^2.$$

Using (7.60), (7.67), (7.64) and (7.61), we have

$$\begin{aligned}
& \Big|Cov\big(\beta_1(z)\gamma_1^2(z), tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\big)\Big| \\
\le \; & \big(\mathbb{E}|\beta_1(z)|^4\big)^{1/4}\big(c_n\mathbb{E}|\gamma_1(z)|^8\big)^{1/4}\Big(\mathbb{E}\big|tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z) - \mathbb{E}tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\big|^2\Big)^{1/2} \\
\le \; & K\delta_n^3 n^{-1/4}.
\end{aligned}$$

Since $\beta_1(z) = b_n(z) - b_n(z)\beta_1(z)\gamma_1(z)$, from (7.60) and (7.61), it follows that $\mathbb{E}\beta_1(z) = b_n(z) + O(n^{-1/2})$. Write

$$\begin{aligned}
& \mathbb{E}n\gamma_1(z)\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 \\
= \; & n\mathbb{E}\big[\big(\mathbf{r}_1^*\mathbf{D}_1^{-1}\mathbf{r}_1 - n^{-1}tr\mathbf{D}_1^{-1}(z)\big)\big(\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 - n^{-1}tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\big)\big] \\
& +n^{-1}Cov\big(tr\mathbf{D}_1^{-1}(z), tr\mathbf{D}_1^{-1}(z)\mathbf{G}_n^{-1}(z)\big).
\end{aligned}$$

From (7.64), we see that the second term above is $O(n^{-1})$. For the other term $\mathbb{E}\beta_j(z)\big[\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_j - \frac{1}{n}\mathbb{E}tr\mathbf{G}_n^{-1}(z)\mathbf{D}_j^{-1}(z)$, repeat the same steps above, we can get a similar result by replacing the subscript 1 by j. By (7.66) and (7.68), we arrive at

$$\begin{aligned}
& n\Big(\frac{1}{c_n\big(1+c_n\mathbb{E}m_n(z)\big)} + \frac{z}{c_n}\mathbb{E}m_n(z)\Big) \qquad (7.69) \\
= \; & \frac{1}{p}\sum_{j=1}^{p}(W_1^{(j)} + W_2^{(j)} + W_3^{(j)}) + o(1),
\end{aligned}$$

where $W_1^{(j)} = b_n^2(z)n^{-1}\mathbb{E}tr\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{D}_j^{-1}(z)$,
$W_2^{(j)} = -b_n^2(z)n\mathbb{E}\big[\big(\mathbf{r}_1^*\mathbf{D}_j^{-1}(z)\mathbf{r}_1 - n^{-1}tr\mathbf{D}_j^{-1}(z)\big)\big(\mathbf{r}_1^*\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\mathbf{r}_1 - n^{-1}tr\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\big)\big]$ and $W_3^{(j)} = \frac{nb_n(z)}{c_n\mathbb{E}m_n(z)+1}\mathbb{E}\gamma_j(z) + b_n^2(z)\frac{tr\mathbb{E}\mathbf{D}_j^{-1}(z)}{c_n\mathbb{E}m_n(z)+1}\mathbb{E}\gamma_j(z)$.

To calculate the limit of $W_1^{(j)}$, we need to expand $\mathbf{D}_j^{-1}(z)$ to the form like (7.35). Similar to Bai and Silverstein (2004), we recalculate (7.37) and (7.18) by (7.58)–(7.60). We omit the details here. After these steps, we have

$$\lim_{n\to\infty} W_1^{(j)} = \frac{z^2m^2(z)\underline{m}'(z)}{cm(z)+1}. \tag{7.70}$$

For W_2, using Lemma 6 on W_2, we have

$$W_2^{(j)} = \begin{cases} W_{2,1}^{(j)} + W_{2,2}^{(j)} + W_{2,3}^{(j)} + W_{2,4}^{(j)} + W_{2,5}^{(j)}, & \textit{under the complex case};\\ W_{2,1}^{(j)} + 2W_{2,2}^{(j)} + W_{2,4}^{(j)} + W_{2,5}^{(j)}, & \textit{under the real case}, \end{cases} \tag{7.71}$$

where

$$W_{2,1}^{(j)} = -\frac{1-\mathbb{E}|X_{j1}|^4}{n^2}b_n^2(z)\mathbb{E}tr\mathbf{D}_j^{-1}(z)\mathbb{E}tr\big[\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\big],$$

$$W_{2,2}^{(j)} = -\frac{1}{n-1}b_n^2(z)\mathbb{E}tr\Big[\mathbf{D}_j^{-2}(z)\mathbf{G}_n^{-1}(z)\Big],$$

$$W_{2,3}^{(j)} = -\frac{|\mathbb{E}X_{j1}^2|^2}{n}b_n^2(z)\mathbb{E}tr\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\big(\mathbf{D}_j^{-1}(z)\big)^*,$$

$$W_{2,4}^{(j)} = -\Big[\frac{\mathbb{E}|X_{j1}|^4}{n} - \frac{2}{n} - n\mathbb{E}\big(\frac{(X_{j1}^*)^2X_{j2}^2}{n^2}\big)\Big]b_n^2(z)\sum_{k=1}^n\mathbb{E}\big(\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\big)_{kk}\big(\mathbf{D}_j^{-1}(z)\big)_{kk},$$

and $W_{2,5}^{(j)} = o_{L_1}(1)$ uniformly for j.

The limits of $W_{2,1}^{(j)}$, $W_{2,2}^{(j)}$ and $W_{2,3}^{(j)}$ can be easily obtained as

$$\lim_{n\to\infty} W_{2,1}^{(j)} = \Big(\mathbb{E}|X_{j1}|^4 - 1\Big)\frac{z\underline{m}^2(z)}{\big(1+\underline{m}(z)\big)\big(z(1+\underline{m}(z))-c\big)},$$

$$\lim_{n\to\infty} W_{2,2}^{(j)} = -\frac{z\underline{m}'(z)}{\big(1+\underline{m}(z)\big)\big(z+z\underline{m}(z)-c\big)},$$

$$\lim_{n\to\infty} W_{2,3}^{(j)} = -|\mathbb{E}X_{11}^2|^2\frac{cm^2(z)}{1+cm(z)}\cdot\frac{1}{(1+cm(z))^2 - c|\mathbb{E}X_{11}^2|^2m^2(z)},$$

where the last limit is derived similarly to Lemma 9 in the supplementary document.

For $W_{2,4}^{(j)}$, similar to deriving the limit of $\frac{\partial^2}{\partial z_1\partial z_2}J_3$ in (7.54), we have

$$\frac{1}{n}\sum_{k=1}^n\mathbb{E}\big[\big(\mathbf{D}_j^{-1}(z)\big)_{kk}\big]\mathbb{E}\big[\big(\mathbf{D}_j^{-1}(z)\mathbf{G}_n^{-1}(z)\big)_{kk}\big] \xrightarrow{i.p.} \frac{\underline{m}^2(z)}{cm(z)+1}. \tag{7.72}$$

By (7.72) and a simple calculation, we get $\lim_{n\to\infty} W_{2,4}^{(j)} = -\Big(\mathbb{E}|X_{j1}|^4 - 3\Big)\frac{z^2m^2(z)\underline{m}^2(z)}{cm(z)+1}$.

For W_3, by lemma 8 in the supplementary document, we have

$$
\begin{aligned}
n\mathbb{E}\gamma_j(z) &= n\mathbb{E}\Big(\frac{1}{n(n-1)}\sum_{k\neq\ell}\big(\mathbf{D}_j^{-1}(z)\big)_{k\ell}\Big) \\
&= \frac{1}{n-1}\mathbb{E}\mathbf{1}^T\mathbf{D}_j^{-1}(z)\mathbf{1} - \frac{1}{n-1}\mathbb{E}tr\mathbf{D}_j^{-1}(z) \stackrel{i.p.}{\to} -\frac{1}{z} - \underline{m}(z). \qquad (7.73)
\end{aligned}
$$

Then it follows from (7.73) and (7.47) that

$$
W_3^{(j)} \stackrel{i.p.}{\to} \frac{m(z)+zm(z)\underline{m}(z)-zm^2(z)\underline{m}(z)-z^2m^2(z)\underline{m}^2(z)}{1+cm(z)}.
$$

Therefore, it follows from (7.69) and calculations above, we can obtain

$$
\begin{aligned}
&\lim_{n\to\infty} n\Big(\frac{1}{c_n\big(1+c_n\mathbb{E}m_n(z)\big)} + \frac{z}{c_n}\mathbb{E}m_n(z)\Big) \\
&= \lim_{n\to\infty}\frac{1}{p}\sum_{j=1}^{p}(W_1^{(j)} + W_{2,1}^{(j)} + W_{2,2}^{(j)} + W_{2,3}^{(j)} + W_{2,4}^{(j)} + W_3^{(j)}), \\
&\text{ under the complex case, or;} \\
&= \lim_{n\to\infty}\frac{1}{p}\sum_{j=1}^{p}(W_1^{(j)} + W_{2,1}^{(j)} + 2W_{2,2}^{(j)} + W_{2,4}^{(j)} + W_3^{(j)}), \\
&\text{under the real case.} \qquad (7.74)
\end{aligned}
$$

The goal is to find the limit of $M_n^{(2)}(z) = n\Big(\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z)\Big)$. It has a relation to the limit in (7.74). We illustrate this point below.

Recall that $m_{c_n}(z)$ and $\underline{m}_{c_n}(z)$ satisfy the following equations

$$
m_{c_n}(z) = \frac{1}{1-c_n-c_nzm_{c_n}(z)-z}, \qquad (7.75)
$$

$$
\underline{m}_{c_n}(z) = -\big(z - \frac{c_n}{1+\underline{m}_{c_n}(z)}\big)^{-1}. \qquad (7.76)
$$

Let $A_n(z) = \frac{1}{c_n\big(1+c_n\mathbb{E}m_n(z)\big)} + \frac{z}{c_n}\mathbb{E}\underline{m}_n(z)$. Since

$$
\mathbb{E}\underline{m}_n(z) = -\frac{1-c_n}{z} + c_n\mathbb{E}m_n(z), \qquad (7.77)
$$

With (7.77), we have

$$
\begin{aligned}
A_n(z) &= \frac{1}{c_n + c_n\mathbb{E}\underline{m}_n(z) + c_n\frac{1-c_n}{z}} + \frac{z}{c_n}\mathbb{E}\underline{m}_n(z) \\
&= \mathbb{E}\underline{m}_n(z)\Big[\frac{z}{c_n} + \frac{1}{c_n + c_n(1-c_n)/z + (c_n-1)\mathbb{E}\underline{m}_n(z)}\Big(\frac{1}{\mathbb{E}\underline{m}_n(z)} - \frac{1}{c_n(1+\mathbb{E}\underline{m}_n(z)+\frac{1-c_n}{z})}\Big)\Big].
\end{aligned}
$$

Then it follows that

$$\mathbb{E}\underline{m}_n(z) = \Big[-\frac{z}{c_n}\Big(c_n + \frac{c_n(1-c_n)}{z} + (c_n-1)\mathbb{E}\underline{m}_n(z)\Big) \\ + \frac{1}{c_n(1+\mathbb{E}\underline{m}_n(z)+\frac{1-c_n}{z})} + \Big(c_n + \frac{c_n(1-c_n)}{z} + (c_n-1)\mathbb{E}\underline{m}_n(z)\Big)\frac{A_n(z)}{\mathbb{E}\underline{m}_n(z)}\Big]^{-1}. \tag{7.78}$$

By (7.76) and (7.78), we have

$$\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z) = A^{-1}B^{-1}(B+A), \tag{7.79}$$

where $A = z - \frac{c_n}{1+\underline{m}_{c_n}(z)}$ and

$$B = -\frac{z}{c_n}\Big(c_n + \frac{c_n(1-c_n)}{z} + (c_n-1)\mathbb{E}\underline{m}_n(z)\Big) + \frac{1}{c_n(1+\mathbb{E}\underline{m}_n(z)+\frac{1-c_n}{z})} \\ + \Big(c_n + \frac{c_n(1-c_n)}{z} + (c_n-1)\mathbb{E}\underline{m}_n(z)\Big)\frac{A_n(z)}{\mathbb{E}\underline{m}_n(z)}$$

By the definition of $A_n(z)$ and (7.77), we know

$$\frac{1}{c_n\big(1+\mathbb{E}\underline{m}_n(z)+\frac{1-c_n}{z}\big)} = \frac{1}{c_n(1+c_n\mathbb{E}m_n(z))} = A_n(z) - \frac{z}{c_n}\mathbb{E}\underline{m}_n(z). \tag{7.80}$$

Then it follows from (7.80) that

$$\begin{aligned} B + A = & -\frac{1}{\underline{m}_{c_n}(z)} - z - (1-c_n) - z\mathbb{E}\underline{m}_n(z) \\ & + A_n(z)\Big(1 + \frac{1}{\mathbb{E}\underline{m}_n(z)}(c_n + \frac{c_n(1-c_n)}{z}) + (c_n-1)\Big) \\ = & \ c_n z\Big(m_{c_n}(z) - \mathbb{E}m_n(z)\Big) + A_n(z)\Big(1 + \frac{1}{\mathbb{E}\underline{m}_n(z)}(c_n + \frac{c_n(1-c_n)}{z}) + (c_n-1)\Big), \end{aligned} \tag{7.81}$$

where the last equality uses (7.75) and (7.77).

From (7.79), (7.81) and the fact that $n\Big(\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z)\Big) = p\Big(\mathbb{E}m_n(z) - m_{c_n}(z)\Big)$, we have

$$n\Big(\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z)\Big) = \frac{nA_n(z)\Big(1 + \frac{U_n}{\mathbb{E}\underline{m}_n(z)}A^{-1}B^{-1}\Big)}{1 + A^{-1}B^{-1}z}, \tag{7.82}$$

where $U_n = c_n + \frac{c_n(1-c_n)}{z} + (c_n-1)\mathbb{E}\underline{m}_n(z)$.

With tedious but simple calculations, we obtain the limit of each part on the right hand side of (7.82) as follows.

$$\lim_{n\to\infty} A = z - \frac{c}{1+\underline{m}(z)}, \quad \lim_{n\to\infty} B = -z - (1-c) - z\underline{m}(z), \\ \lim_{n\to}\Big(1 + \frac{U_n}{\mathbb{E}\underline{m}_n(z)}\Big) = c + \frac{zc + c(1-c)}{z\underline{m}(z)}. \tag{7.83}$$

It follows from (7.83) that

$$\lim_{n\to} \frac{1+\frac{U_n}{\mathbb{E}\underline{m}_n(z)}A^{-1}B^{-1}}{1+A^{-1}B^{-1}z} = -\frac{c(1+\underline{m}(z))\big(z(1+\underline{m}(z))+1-c\big)}{z\underline{m}(z)\Big(\big(z(1+\underline{m}(z))-c\big)^2-c\Big)}. \tag{7.84}$$

Thus, it follows from (7.74) and (7.83) that

$$\begin{aligned}
& n\Big(\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z)\Big) \\
&= \lim_{p\to\infty} \frac{1}{p}\sum_{j=1}^{p}\Big(W_1^{(j)} + W_{2,1}^{(j)} + W_{2,2}^{(j)} + W_{2,3}^{(j)} + W_{2,4}^{(j)} + W_3^{(j)}\Big) \\
&\times \left(-\frac{c(1+\underline{m}(z))\big(z(1+\underline{m}(z))+1-c\big)}{z\underline{m}(z)\Big(\big(z(1+\underline{m}(z))-c\big)^2-c\Big)}\right),
\end{aligned}$$

under the complex random variable case, or

$$\begin{aligned}
&= \lim_{n\to\infty} \frac{1}{p}\sum_{j=1}^{p}\Big(W_1^{(j)} + W_{2,1}^{(j)} + 2W_{2,2}^{(j)} + W_{2,4}^{(j)} + W_3^{(j)}\Big) \\
&\times \left(-\frac{c(1+\underline{m}(z))\big(z(1+\underline{m}(z))+1-c\big)}{z\underline{m}(z)\Big(\big(z(1+\underline{m}(z))-c\big)^2-c\Big)}\right),
\end{aligned}$$

under the real random variable case.

8 Some Lemmas and Proofs

8.1 Lemma 5

Suppose that $\{X_i\}_{i=1}^n$ are i.i.d. random variables with $EX_1 = 0$ and $E|X_1|^2 = 1$. Let $\mathbf{y} = (X_1, ..., X_n)^T$ and $\bar{\mathbf{y}} = \frac{\sum_{i=1}^n X_i}{n}\mathbf{e}$, where $\mathbf{e} = (1, 1, ..., 1)^T$ is an n-dimensional vector. Assuming that $\mathbf{A}$ is a deterministic complex matrix, then for any given $q \geq 2$, there is a positive constant K_q depending on q such that

$$\begin{aligned}
& E\left|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right|^q \\
& \leq K_q\left\{n^{-q}(v_{2q}tr(\mathbf{A}\mathbf{A}^*)^q + (v_4 tr(\mathbf{A}\mathbf{A}^*))^{q/2}) + \mathbb{P}(B_n^c(\epsilon))\|\mathbf{A}\|^q\right\},
\end{aligned} \tag{8.1}$$

where $B_n(\epsilon) = \left\{\mathbf{y} : |\frac{\|\mathbf{y}-\bar{\mathbf{y}}\|^2}{n} - 1| \leq \epsilon\right\}$ and $\boldsymbol{\alpha} = \frac{(\mathbf{y}-\bar{\mathbf{y}})^T}{\|\mathbf{y}-\bar{\mathbf{y}}\|}$ for some $\epsilon > 0$.

Remark 5. *Note that* $\mathbb{P}(B_n^c(\epsilon)) = O(n^{-q/2}v_4^{q/2} + n^{-q+1}v_{2q})$. *If* $\|\mathbf{A}\| \leq K$ *and* $|X_i| \leq \sqrt{n}\delta_n$, *we have*

$$\mathbb{E}\left|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right|^q \leq K_q n^{-1}\delta_n^{2q-4}. \tag{8.2}$$

Remark 6. *Similar to Lemma 5, one can prove that under the same conditions of Lemma 5 (replacing* $\boldsymbol{\alpha}*$ *by* $\boldsymbol{\alpha}^T$*), we have*

$$\mathbb{E}\left|\boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha} - \frac{\mathbb{E}X_1^2}{n}tr\mathbf{A}\right|^q$$
$$\leq K_q\left\{n^{-q}(v_{2q}tr(\mathbf{A}\mathbf{A}^*)^q + (v_4 tr(\mathbf{A}\mathbf{A}^*))^{q/2}) + \mathbb{P}(B_n^c(\epsilon))\|\mathbf{A}\|^q\right\}. \tag{8.3}$$

Proof. To avoid confusion, denote by K_q a positive constant large enough depending on q only. By inequality

$$\left|\frac{\|\mathbf{y}-\bar{\mathbf{y}}\|^2}{n} - 1\right| \leq \left|\frac{\sum_{i=1}^n X_i^2}{n} - 1\right| + \left(\frac{\sum_{i=1}^n X_i}{n}\right)^2, \tag{8.4}$$

we have $P\left(B_n^c(\epsilon)\right) \leq P\left(\left|\frac{\sum_{i=1}^n X_i^2}{n} - 1\right| \geq \frac{\epsilon}{2}\right) + P\left(\left|\frac{\sum_{i=1}^n X_i}{n}\right| \geq \sqrt{\frac{\epsilon}{2}}\right)$.

By Markov's inequality and Lemma 3 in the main paper, we have

$$\begin{aligned} P\left(\left|\frac{\sum_{i=1}^n X_i^2}{n} - 1\right| \geq \frac{\epsilon}{2}\right) &\leq \frac{2^q\mathbb{E}|\sum_{i=1}^n(X_i^2-1)|^q}{(n\epsilon)^q} \\ &\leq \frac{K_q(\sum_{i=1}^n \mathbb{E}|X_i^2-1|^2)^{q/2} + \mathbb{E}\sum_{i=1}^n|X_i^2-1|^q}{(n\epsilon)^q} \\ &= O(n^{-q/2}v_4^{q/2} + n^{-q+1}v_{2q}). \end{aligned} \tag{8.5}$$

Similarly, we have

$$\begin{aligned} &P\left(\left|\frac{\sum_{i=1}^n X_i}{n}\right| \geq \sqrt{\frac{\epsilon}{2}}\right) \leq \frac{K_q((\sum_{i=1}^n \mathbb{E}X_i^2)^q + \mathbb{E}\sum_{i=1}^n|X_i|^{2q})}{(n\epsilon)^{2q}} \\ &= O(n^{-q} + n^{-2q+1}v_{2q}). \end{aligned} \tag{8.6}$$

which implies $P(B_n^c(\epsilon)) = O(n^{-q/2}v_4^{q/2} + n^{-q+1}v_{2q})$.

Note that

$$\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A} = \left[\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right]I(B_n(\epsilon)) + \left[\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right]I(B_n^c(\epsilon)).$$

There exists a positive constant K_q such that

$$\mathbb{E}\left|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right|^q \leq K_q\left(\mathbb{E}\left|(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A})I(B_n(\epsilon))\right|^q + \mathbb{E}\left|(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A})I(B_n^c(\epsilon))\right|^q\right).$$

By the fact that $|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}| \leq 2\|\mathbf{A}\|$, we obtain

$$\mathbb{E}\left|(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A})I(B_n^c(\epsilon))\right|^q \tag{8.7}$$
$$\leq 2^q\|\mathbf{A}\|^q P(B_n^c(\epsilon)) = O(n^{-q/2}v_4^{q/2} + n^{-q+1}v_{2q})\|\mathbf{A}\|^q.$$

Observe that

$$\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A} = \frac{1}{\|\mathbf{y}-\bar{\mathbf{y}}\|^2}((\mathbf{y}-\bar{\mathbf{y}})^T\mathbf{A}(\mathbf{y}-\bar{\mathbf{y}}) - tr\mathbf{A}) + \left(\frac{trA}{\|\mathbf{y}-\bar{\mathbf{y}}\|^2} - \frac{1}{n}tr\mathbf{A}\right) = a_1 + a_2.$$

For $0 < \epsilon < 1/2$, there exists a positive constant K_q such that

$$\mathbb{E}\left|\left[\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right] I(B_n(\epsilon))\right|^q \leq K_q(\mathbb{E}|a_1 I(B_n(\epsilon))|^q + \mathbb{E}|a_2 I(B_n(\epsilon))|^q). \tag{8.8}$$

Consider a_2 first. It is easy to see $\frac{\|\mathbf{y}-\bar{\mathbf{y}}\|^2}{n} \geq 1-\epsilon$ on the event B_n, so that

$$\begin{aligned} \mathbb{E}|a_2 I(B_n(\epsilon))|^q &\leq K_q\left(\frac{tr\mathbf{A}}{n}\right)^q \mathbb{E}\left|1 - \frac{\|\mathbf{y}-\bar{\mathbf{y}}\|^2}{n}\right|^q \\ &\leq K_q\left(\frac{tr\mathbf{A}}{n}\right)^q \cdot \left\{\mathbb{E}\left|\frac{\sum_{i=1}^n X_i^2}{n} - 1\right|^q + \mathbb{E}\left|\left(\frac{\sum_{i=1}^n X_i}{n}\right)^2\right|^q\right\} \\ &\leq K_q\left(\frac{tr\mathbf{A}}{n}\right)^q (n^{-q/2}v_4^{q/2} + n^{-q+1}v_{2q}), \end{aligned} \tag{8.9}$$

where the second inequality follows from (8.4) and the last inequality follows from (8.5) and (8.6).

Therefore, we have

$$\mathbb{E}|a_2 I(B_n(\epsilon))|^q \leq K_q\left(\frac{tr\mathbf{A}}{n}\right)^q (n^{-q/2}v_4^{q/2} + n^{-q+1}v_{2q}). \tag{8.10}$$

Similarly, for a_1, by writing $\overline{y} = \mathbf{e}\mathbf{e}^T\mathbf{y}$, we have

$$\begin{aligned} &\mathbb{E}|a_1 I(B_n(\epsilon))|^q \leq K_q\frac{1}{n^q}\mathbb{E}|(\mathbf{y}-\bar{\mathbf{y}})^T\mathbf{A}(\mathbf{y}-\bar{\mathbf{y}}) - tr\mathbf{A}|^q \\ &\leq \frac{K_q}{n^q}\left(\mathbb{E}|\mathbf{y}^T\mathbf{A}\mathbf{y} - tr\mathbf{A}|^q + E\left|\frac{1}{n}\mathbf{y}^T\mathbf{e}\mathbf{e}^T\mathbf{A}\mathbf{y} + \frac{1}{n}\mathbf{y}^T\mathbf{A}\mathbf{e}\mathbf{e}^T\mathbf{y}\right|^q + E\left|\frac{1}{n^2}\mathbf{y}^T\mathbf{e}\mathbf{e}^T\mathbf{A}\mathbf{e}\mathbf{e}^T\mathbf{y}\right|^q\right). \end{aligned} \tag{8.11}$$

Noting that $\frac{1}{n}tr\mathbf{e}\mathbf{e}^T\mathbf{A} = \frac{1}{n}tr\mathbf{A}\mathbf{e}\mathbf{e}^T = \frac{1}{n^2}tr\mathbf{e}\mathbf{e}^T\mathbf{A}\mathbf{e}\mathbf{e}^T \leq \|\mathbf{A}\|$ and Lemma 2.2 in Bai and Silverstein (2004), we have

$$\begin{aligned} &\mathbb{E}|\mathbf{y}^T\mathbf{A}\mathbf{y} - tr\mathbf{A}|^q + \mathbb{E}|\frac{1}{n}\mathbf{y}^T\mathbf{e}\mathbf{e}^T\mathbf{A}\mathbf{y} + \frac{1}{n}\mathbf{y}^T\mathbf{A}\mathbf{e}\mathbf{e}^T\mathbf{y}|^q + \mathbb{E}|\frac{1}{n^2}\mathbf{y}^T\mathbf{e}\mathbf{e}^T\mathbf{A}\mathbf{e}\mathbf{e}^T\mathbf{y}|^q \\ &\leq K_q\ (v_{2q}tr(\mathbf{A}\mathbf{A}^*)^q + (v_4 tr(\mathbf{A}\mathbf{A}^*))^{q/2}). \end{aligned} \tag{8.12}$$

Hence, we obtain

$$\mathbb{E}|a_1 I(B_n(\epsilon))|^q \leq K_q n^{-q}(v_{2q}tr(\mathbf{A}\mathbf{A}^*)^q + (v_4 tr(\mathbf{A}\mathbf{A}^*))^{q/2}). \tag{8.13}$$

Combining (8.7),(8.10) and (8.13) together , we can conclude that

$$\begin{aligned}
&\mathbb{E}\left|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}\right|^q \\
\leq\ & K_q\ \left\{n^{-q}(v_q tr(\mathbf{A}\mathbf{A}^*)^q + (v_4 tr(\mathbf{A}\mathbf{A}^*))^{q/2}) + (n^{-q/2}v_4^{q/2} + n^{-q+1}v_{2q})\|\mathbf{A}\|^q\right\},
\end{aligned}$$

where K_q is a positive constant depending on q only.

□

8.2 Lemma 6

In addition to the condition of Lemma 5, assuming that $\mathbb{E}|X_1|^4 < \infty$, $\|\mathbf{A}\| \leq K$ and $\|\mathbf{B}\| \leq K$, then we have

$$\begin{aligned}
&\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A})(\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{B}) \qquad (8.14)\\
=\ & \sum_{i=1}^{n}\frac{1}{n^2}(\mathbb{E}|X_1|^4 - |\mathbb{E}(X_1^2)|^2 - 2)\mathbf{A}_{ii}\mathbf{B}_{ii} + \frac{|\mathbb{E}X_1^2|^2}{n^2}tr(\mathbf{A}\mathbf{B}^T) \\
&+\frac{1}{n^2}tr(\mathbf{A}\mathbf{B}) + \frac{1-\mathbb{E}|X_1|^4}{n^3}tr\mathbf{A}tr\mathbf{B} + o\left(\frac{1}{n}\right).
\end{aligned}$$

Proof. At first, we evaluate some expectations. Set $\mathbb{E}|\alpha_1|^4 = \mu_4$ and $\mathbb{E}(\bar{\alpha}_1\alpha_2)^2 = \mu_{12}$ for convenience. Note that

$$\sum_{i=1}^{n}\alpha_i = 0 \quad \sum_{i=1}^{n}(\bar{\alpha}_i\alpha_i) = 1 \quad \text{and} \quad \mathbb{E}(\bar{\alpha}_1\alpha_1) = \frac{1}{n}. \qquad (8.15)$$

It follows that for $i \neq j$

$$\begin{aligned}
\mathbb{E}(\bar{\alpha}_i\alpha_j) = \mathbb{E}(\bar{\alpha}_1\alpha_2) &= \frac{1}{n-1}\left[\mathbb{E}(\bar{\alpha}_1\sum_{i=1}^{n}\alpha_i) - \mathbb{E}(\bar{\alpha}_1\alpha_1)\right] \qquad (8.16)\\
&= -\frac{1}{n(n-1)},
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_2) &= \frac{1}{n-1}\mathbb{E}\left(\bar{\alpha}_1\alpha_1[\sum_{i=1}^{n}(\bar{\alpha}_i\alpha_i) - \bar{\alpha}_1\alpha_1]\right) \qquad (8.17)\\
&= \frac{1}{n(n-1)} - \frac{1}{n-1}\mu_4,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_1\alpha_2) &= \frac{1}{n-1}\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_1(\sum_{i=1}^{n}\alpha_i - \alpha_1)) \\
&= -\frac{1}{n-1}\mu_4, \qquad (8.18)
\end{aligned}$$

and via (8.17),

$$
\begin{aligned}
&\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_3) \qquad (8.19)\\
= &\frac{1}{n-2}\mathbb{E}[\bar{\alpha}_1\alpha_1\bar{\alpha}_2\sum_{i=1}^{n}(\bar{\alpha}_i-\alpha_1-\alpha_2)]\\
= &-\frac{1}{n-2}\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_1)-\frac{1}{n-2}\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_2)
\end{aligned}
$$

$$
\begin{aligned}
= &-\frac{1}{(n-1)(n-2)}\mathbb{E}[\bar{\alpha}_1\alpha_1\alpha_1(\sum_{i=1}^{n}\bar{\alpha}_i-\alpha_1)]-\frac{1}{n-2}\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_2)\\
= &\frac{1}{(n-1)(n-2)}\mu_4-\frac{1}{n(n-1)(n-2)}+\frac{1}{(n-1)(n-2)}\mu_4\\
= &\frac{2}{(n-1)(n-2)}\mu_4-\frac{1}{n(n-1)(n-2)}.
\end{aligned}
$$

Analogously, we can get

$$
\mathbb{E}(\bar{\alpha}_1\bar{\alpha}_1\alpha_2\alpha_3) = \frac{1}{(n-1)(n-2)}\mu_4-\frac{1}{n-2}\mu_{12}, \qquad (8.20)
$$

$$
\begin{aligned}
&\mathbb{E}(\bar{\alpha}_1\alpha_2\bar{\alpha}_3\alpha_4) \qquad (8.21)\\
= &\frac{1}{n-3}\mathbb{E}[\bar{\alpha}_1\alpha_2\bar{\alpha}_3\sum_{i=1}^{n}(\alpha_i-\alpha_1-\alpha_2-\alpha_3)]\\
= &-\frac{1}{n-3}\mathbb{E}(\bar{\alpha}_1\alpha_1\alpha_2\bar{\alpha}_3)-\frac{1}{n-3}E(\bar{\alpha}_1\alpha_2\alpha_2\bar{\alpha}_3)-\frac{1}{n-3}E(\bar{\alpha}_1\alpha_2\bar{\alpha}_3\alpha_3)\\
= &\frac{1}{(n-3)}\left(\frac{2}{n(n-1)(n-2)}-\frac{5}{(n-1)(n-2)}\mu_4+\frac{1}{n-2}\mu_{12}\right).
\end{aligned}
$$

Let's calculate μ_4. We claim that $\mu_4 = \frac{\mathbb{E}|X_1|^4}{n^2}+o(n^{-2})$. To prove it, we consider the real case only, the complex case can be proved similarly. We below use the same notation $B_n(\epsilon)$ as Lemma 5. Suppose that ϵ is a positive constant such that $\mathbb{P}(B_n^c(\epsilon)) = o(n^{-1})$ (by the convergence rate in the law of large numbers that $n\mathbb{P}\left(\left|\frac{\sum_{i=1}^n (X_i^2-1)}{n}\right| \geq \epsilon\right) \to 0$, which can be referred to Theorem 28 of Petrov (1975)). Then, we have

$$
\begin{aligned}
&\left|\mathbb{E}(\frac{|X_1-\bar{x}|^4}{\|\mathbf{y}-\bar{\mathbf{y}}\|^4}-\frac{(X_1-\bar{x})^4}{n^2})\right| \qquad (8.22)\\
\leq &\left|\mathbb{E}\frac{|X_1-\bar{x}|^4}{n^2\|\mathbf{y}-\bar{\mathbf{y}}\|^4}(n^2-\|\mathbf{y}-\bar{\mathbf{y}}\|^4)I(B_n(\epsilon))\right|\\
&+\left|\mathbb{E}\Big[\mathbb{E}\Big(\frac{|X_1-\bar{x}|^4}{\|\mathbf{y}-\bar{\mathbf{y}}\|^4}|B_n^c(\epsilon)\Big)I(B_n^c(\epsilon))\Big]\right|+\left|\mathbb{E}\frac{|X_1-\bar{x}|^4}{n^2}I(B_n^c(\epsilon))\right|\\
\leq &\left|\mathbb{E}\frac{|X_1-\bar{x}|^4}{n^2\|\mathbf{y}-\bar{\mathbf{y}}\|^4}(n-\|\mathbf{y}-\bar{\mathbf{y}}\|^2)(n+\|\mathbf{y}-\bar{\mathbf{y}}\|^2)I(B_n(\epsilon))\right|+\frac{\mathbb{P}(B_n^c(\epsilon))}{n}+o(n^{-2})\\
\leq &\frac{K\epsilon}{n^2}\mathbb{E}|X_1-\bar{x}|^4+o(n^{-2}) \leq K\epsilon\frac{1}{n^2}+o(n^{-2}),
\end{aligned}
$$

where the second part of the second inequality follows from

$$\mathbb{E}\Big[\frac{|X_1-\bar{x}|^4}{\|\mathbf{y}-\bar{\mathbf{y}}\|^4}|B_n^c(\epsilon)\Big] = \frac{1}{n}\mathbb{E}\Big[\frac{\sum_{i=1}^n |X_i-\bar{x}|^4}{\|\mathbf{y}-\bar{\mathbf{y}}\|^4}|B_n^c(\epsilon)\Big] \le \frac{1}{n},$$

in which the last inequality uses $\frac{\sum_{i=1}^n |X_i-\bar{x}|^4}{\|\mathbf{y}-\bar{\mathbf{y}}\|^4} \le 1$; the third part of the second inequality uses

$$\begin{aligned}
& \Big|\mathbb{E}|X_1-\bar{x}|^4 I(B_n^c(\varepsilon))\Big| \\
\le\ & \Big|\mathbb{E}|X_1-\bar{x}|^4 I(|X_1-\bar{x}|\le G) I(B_n^c(\varepsilon))\Big| + \Big|\mathbb{E}|X_1-\bar{x}|^4 I(|X_1-\bar{x}|\ge G) I(B_n^c(\varepsilon))\Big| \\
\le\ & G^4\cdot P(B_n^c(\varepsilon)) + o(1) = o(1),
\end{aligned}$$

in which the last inequality uses $G = \Big(P(B_n^c(\varepsilon))\Big)^{-1/5}$, $P(B_n^c(\varepsilon)) = o(1/n)$ and $\mathbb{E}|X_1|^4 < \infty$.

It means the inequality holds for any $\epsilon > 0$ and n large enough, so we have proved

$$\mathbb{E}\left(\frac{|X_1-\bar{x}|^4}{\|\mathbf{y}-\bar{\mathbf{y}}\|^4} - \frac{|X_1-\bar{x}|^4}{n^2}\right) = o(n^{-2}).$$

In a similar way, we can obtain

$$\mu_{12} = \mathbb{E}(\bar{\alpha}_1\bar{\alpha}_1\alpha_2\alpha_2) = \frac{|\mathbb{E}X_1^2|^2}{n^2} + o(n^{-2}). \tag{8.23}$$

It is easy to get

$$\begin{aligned}
& \mathbb{E}(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A})(\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{B}) \qquad (8.24)\\
=\ & \mathbb{E}(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha}\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha}) - \frac{1}{n}(tr\mathbf{A})\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{B}) \\
& -\frac{1}{n}(tr\mathbf{B})\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}) - \frac{1}{n^2}tr\mathbf{A}tr\mathbf{B}
\end{aligned}$$

and $\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{B}) = \mathbb{E}(\bar{\alpha}_1\alpha_2)\sum_{k\ne l}\mathbf{B}_{kl} = \frac{1}{n-1}\sum_{k\ne l}\mathbf{B}_{kl}E[\bar{\alpha}_1(\sum_{i=1}^n\alpha_i - \alpha_1)]$.

By (8.15), we further have

$$\begin{aligned}
\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{B}) &= -\frac{1}{n-1}\sum_{k\ne l}\mathbf{B}_{kl}\mathbb{E}[\bar{\alpha}_1\alpha_1] \qquad (8.25)\\
&= -\frac{1}{n(n-1)}\sum_{k\ne l}\mathbf{B}_{kl},
\end{aligned}$$

by which we can conclude

$$\frac{1}{n}(tr\mathbf{A})\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{B}) = -\frac{tr\mathbf{A}}{n^2(n-1)}\sum_{k\ne l}\mathbf{B}_{kl}.$$

In the same way, we can get

$$\frac{1}{n}(tr\mathbf{B})\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A}) = -\frac{tr\mathbf{B}}{n^2(n-1)}\sum_{k\ne l}\mathbf{A}_{kl}.$$

To calculate $E(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha}\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha})$, we expand the expression as

$$\begin{aligned} & \mathbb{E}(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha}\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha}) \\ = \; & \mathbb{E}(\sum_{i,j}\bar{\alpha}_i\mathbf{A}_{ij}\alpha_j\sum_{k,l}\bar{\alpha}_k\mathbf{A}_{kl}\alpha_l) = \sum_{i,j,k,l}\mathbb{E}\bar{\alpha}_i\alpha_j\bar{\alpha}_k\alpha_l\mathbf{A}_{ij}\mathbf{B}_{kl}. \end{aligned} \tag{8.26}$$

To calculate (8.26), we split it into the following cases:

1. i=j=k=l, $\sum_i(\bar{\alpha}_i\alpha_i)^2\mathbf{A}_{ii}\mathbf{B}_{ii}$;
2. i=j, k=l, i≠k, $\sum_{\substack{i,k \\ i\neq k}}(\bar{\alpha}_i\alpha_i)(\bar{\alpha}_k\alpha_k)\mathbf{A}_{ii}\mathbf{B}_{kk}$;
3. i=j, k≠ l, $\sum_{\substack{i,k,l \\ k\neq l}}(\bar{\alpha}_i\alpha_i)(\bar{\alpha}_k\alpha_l)\mathbf{A}_{ii}\mathbf{B}_{kl}$;
4. i≠j, k=l, $\sum_{\substack{i,j,k \\ i\neq j}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_k\alpha_k)\mathbf{A}_{ij}\mathbf{B}_{kk}$;
5. i≠j, k≠l, i=k,j=l $\sum_{\substack{i,j \\ i\neq j}}(\bar{\alpha}_i\bar{\alpha}_i)(\alpha_j\alpha_j)\mathbf{A}_{ij}\mathbf{B}_{ij}$;
6. i≠j, k≠l, i=l,j=k $\sum_{\substack{i,j \\ i\neq j}}(\bar{\alpha}_i\alpha_i)(\bar{\alpha}_j\alpha_j)\mathbf{A}_{ij}\mathbf{B}_{ji}$;
7. i≠j, k≠l, i=k, l≠ j, $\sum_{\substack{i,j,l \\ i\neq j\neq l}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_i\alpha_j)\mathbf{A}_{ij}\mathbf{B}_{il}$;
8. i≠j, k≠l, l=j, i≠ k, $\sum_{\substack{i,j,k \\ i\neq j\neq k}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_k\alpha_j)\mathbf{A}_{ij}\mathbf{B}_{kj}$;
9. i≠j, k≠l, k=j, i≠ l, $\sum_{\substack{i,j,l \\ i\neq j\neq l}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_j\alpha_l)\mathbf{A}_{ij}\mathbf{B}_{jl}$;
10. i≠j, k≠l, i=l, k≠ j, $\sum_{\substack{i,j,k \\ i\neq j\neq k}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_k\alpha_i)\mathbf{A}_{ij}\mathbf{B}_{ki}$;
11. i≠j, k≠l, l≠j, i≠ k, $\sum_{\substack{i,j,k,l \\ i\neq j\neq k\neq l}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_k\alpha_l)\mathbf{A}_{ij}\mathbf{B}_{kl}$.

For ease of presentation, we still keep μ_4 in the expectations although we have evaluated the value.

Case 1: $\mathbb{E}\left(\sum_i(\bar{\alpha}_i\alpha_i)^2\mathbf{A}_{ii}\mathbf{B}_{ii}\right) = \mathbb{E}(\bar{\alpha}_1\alpha_1)^2\sum_i\mathbf{A}_{ii}\mathbf{B}_{ii} = \mu_4\sum_i\mathbf{A}_{ii}\mathbf{B}_{ii}$.

Case 2: By (8.17), we have

$$\begin{aligned} \mathbb{E}\left(\sum_{\substack{i,k \\ i\neq k}}(\bar{\alpha}_i\alpha_i)(\bar{\alpha}_k\alpha_k)\mathbf{A}_{ii}\mathbf{B}_{kk}\right) &= E(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_2)\sum_{\substack{i,k \\ i\neq k}}\mathbf{A}_{ii}\mathbf{B}_{kk} \\ &= (\frac{1}{n(n-1)} - \frac{1}{n-1}\mu_4)\sum_{\substack{i,k \\ i\neq k}}\mathbf{A}_{ii}\mathbf{B}_{kk}. \end{aligned}$$

Case 3: By (8.18) and (8.19), we have

$$\begin{aligned}
& \mathbb{E}\sum_{\substack{i,k,l\\k\neq l}}(\bar{\alpha}_i\alpha_i)(\bar{\alpha}_k\alpha_l)\mathbf{A}_{ii}\mathbf{B}_{kl} \qquad (8.27)\\
= \;& \mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_3)\sum_{\substack{i,k,l\\i\neq k\neq l}}\mathbf{A}_{ii}\mathbf{B}_{kl}+\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_1\alpha_2)\sum_{\substack{i,l\\l\neq i}}\mathbf{A}_{ii}\mathbf{B}_{il}\\
& +E(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_1)\sum_{\substack{i,k\\k\neq i}}\mathbf{A}_{ii}\mathbf{B}_{ki}\\
= \;& (\frac{2}{(n-1)(n-2)}\mu_4-\frac{1}{n(n-1)(n-2)})\sum_{\substack{i,k,l\\i\neq k\neq l}}\mathbf{A}_{ii}\mathbf{B}_{kl}\\
& -\frac{1}{n-1}\mu_4\sum_{\substack{i,l\\l\neq i}}\mathbf{A}_{ii}(\mathbf{B}_{il}+\mathbf{B}_{li}).
\end{aligned}$$

Case 4: Similarly to Case 3, we obtain

$$\begin{aligned}
\mathbb{E}\sum_{\substack{i,j,k\\i\neq j}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_k\alpha_k)\mathbf{A}_{ij}\mathbf{B}_{kk} &= (\frac{2}{(n-1)(n-2)}\mu_4-\frac{1}{n(n-1)(n-2)})\sum_{\substack{i,k,l\\l\neq k\neq l}}\mathbf{B}_{ii}\mathbf{A}_{kl}\\
& -\frac{1}{n-1}\mu_4\sum_{\substack{i,l\\l\neq i}}\mathbf{B}_{ii}(\mathbf{A}_{il}+\mathbf{A}_{li}).
\end{aligned}$$

Case 5: It follows from (8.23) that

$$\begin{aligned}
\mathbb{E}\sum_{\substack{i,j\\i\neq j}}(\bar{\alpha}_i\bar{\alpha}_i\alpha_j\alpha_j)\mathbf{A}_{ij}\mathbf{B}_{ij} &= \mathbb{E}(\bar{\alpha}_1\bar{\alpha}_1\alpha_2\alpha_2)\sum_{\substack{i,j\\i\neq j}}\mathbf{A}_{ij}\mathbf{B}_{ij}=\mu_{12}\sum_{\substack{i,j\\i\neq j}}\mathbf{A}_{ij}\mathbf{B}_{ij}\\
&= (\frac{|\mathbb{E}X_1^2|^2}{n^2}+o(n^{-2}))\sum_{\substack{i,j\\i\neq j}}\mathbf{A}_{ij}\mathbf{B}_{ij}.
\end{aligned}$$

Case 6: By (8.17), we have

$$\begin{aligned}
& \mathbb{E}\left[\sum_{\substack{i,j\\i\neq j}}(\bar{\alpha}_i\alpha_i)(\bar{\alpha}_j\alpha_j)\mathbf{A}_{ij}\mathbf{B}_{ji}\right]=\mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_2)\sum_{\substack{i,j\\i\neq j}}\mathbf{A}_{ij}\mathbf{B}_{ji}\\
= \;& (\frac{1}{n(n-1)}-\frac{1}{n-1}\mu_4)\sum_{\substack{i,j\\i\neq j}}\mathbf{A}_{ij}\mathbf{B}_{ji}.
\end{aligned}$$

Case 7: In view of (8.20), we have

$$\begin{aligned}
& \mathbb{E}\left[\sum_{\substack{i,j,l\\i\neq j\neq l}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_i\alpha_l)\mathbf{A}_{ij}\mathbf{B}_{il}\right]=\mathbb{E}(\bar{\alpha}_1\alpha_2\bar{\alpha}_1\alpha_3)\sum_{\substack{i,j,l\\i\neq j\neq l}}\mathbf{A}_{ij}\mathbf{B}_{il}\\
= \;& (\frac{1}{(n-1)(n-2)}\mu_4-\frac{1}{n}\mu_{12})\sum_{\substack{i,j,l\\i\neq j\neq l}}\mathbf{A}_{ij}\mathbf{B}_{il}.
\end{aligned}$$

Case 8: Similarly to Case 7, we have

$$\mathbb{E}\left[\sum_{\substack{i,j,k\\ i\neq j\neq k}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_k\alpha_j)\mathbf{A}_{ij}\mathbf{B}_{kj}\right] = \mathbb{E}(\bar{\alpha}_1\alpha_2\bar{\alpha}_1\alpha_3)\sum_{\substack{i,j,k\\ i\neq j\neq k}}\mathbf{A}_{ij}\mathbf{B}_{kj}$$

$$= (\frac{1}{(n-1)(n-2)}\mu_4 - \frac{1}{n}\mu_{12})\sum_{\substack{i,j,k\\ i\neq j\neq k}}\mathbf{A}_{ij}\mathbf{B}_{kj}.$$

Case 9: By (8.19), we have

$$\mathbb{E}\left[\sum_{\substack{i,j,l\\ i\neq j\neq l}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_j\alpha_l)\mathbf{A}_{ij}\mathbf{B}_{jl}\right] = \mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_3)\sum_{\substack{i,j,l\\ i\neq j\neq j}}\mathbf{A}_{ij}\mathbf{B}_{kl}$$

$$= (\frac{2}{(n-1)(n-2)}\mu_4 - \frac{1}{n(n-1)(n-2)})\sum_{\substack{i,j,l\\ i\neq j\neq k}}\mathbf{A}_{ij}\mathbf{B}_{jl}.$$

Case 10: Similarly to Case 9, we have

$$\mathbb{E}\left[\sum_{\substack{i,j,k\\ i\neq j\neq k}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_k\alpha_i)\mathbf{A}_{ij}\mathbf{B}_{ki}\right] = \mathbb{E}(\bar{\alpha}_1\alpha_1\bar{\alpha}_2\alpha_3)\sum_{\substack{i,j,k\\ i\neq j\neq k}}\mathbf{A}_{ij}\mathbf{B}_{ki}$$

$$= (\frac{2}{(n-1)(n-2)}\mu_4 - \frac{1}{n(n-1)(n-2)})\sum_{\substack{i,j,k\\ i\neq j\neq l}}\mathbf{A}_{ij}\mathbf{B}_{ik}.$$

Case 11: We conclude from (8.21) that

$$\mathbb{E}\left[\sum_{\substack{i,j,k,l\\ i\neq j\neq k\neq l}}(\bar{\alpha}_i\alpha_j)(\bar{\alpha}_k\alpha_l)\mathbf{A}_{ij}\mathbf{B}_{kl}\right] = \mathbb{E}(\bar{\alpha}_1\alpha_2\bar{\alpha}_3\alpha_4)\sum_{\substack{i,j,k,l\\ i\neq j\neq k\neq l}}\mathbf{A}_{ij}\mathbf{B}_{kl}$$

$$= \frac{1}{(n-3)}\left(\frac{2}{n(n-1)(n-2)} - \frac{5}{(n-1)(n-2)}\mu_4 + \frac{1}{n-2}\mu_{12}\right)\sum_{\substack{i,j,k,l\\ i\neq j\neq k\neq l}}\mathbf{A}_{ij}\mathbf{B}_{kl}. \tag{8.28}$$

Summarizing the terms above, we conclude that

$$\mathbb{E}(\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{A})(\boldsymbol{\alpha}^*\mathbf{B}\boldsymbol{\alpha} - \frac{1}{n}tr\mathbf{B})$$
$$= \sum_{i=1}^{n}\frac{1}{n^2}(\mathbb{E}|X_1|^4 - |E(X_1^2)|^2 - 2)\mathbf{A}_{ii}\mathbf{B}_{ii} + \frac{|\mathbb{E}X_1^2|^2}{n^2}tr(\mathbf{A}\mathbf{B}^T)$$
$$+ \frac{1}{n^2}tr(\mathbf{A}\mathbf{B}) + \frac{1-\mathbb{E}|X_1|^4}{n^3}tr\mathbf{A}tr\mathbf{B} + \Omega_n,$$

where

$$\Omega_n = \frac{2\mathbb{E}|X_1|^4-2}{n^4}\sum_{\substack{i,k,l\\ i\neq k\neq l}}\mathbf{A}_{ii}\mathbf{B}_{kl} - \frac{\mathbb{E}|X_1|^4-1}{n^3}\sum_{\substack{i,l\\ l\neq i}}\mathbf{A}_{ii}(\mathbf{B}_{il}+\mathbf{B}_{li}) \tag{8.29}$$

$$+\frac{2\mathbb{E}|X_1|^4-2}{n^4}\sum_{\substack{i,k,l\\ i\neq k\neq l}}\mathbf{B}_{ii}\mathbf{A}_{kl} - \frac{\mathbb{E}|X_1|^4-1}{n^3}\sum_{\substack{i,l\\ l\neq i}}\mathbf{B}_{ii}(\mathbf{A}_{il}+\mathbf{A}_{li})$$

$$-\frac{1+|\mathbb{E}x_1^2|^2}{n(n-1)(n-2)}\sum_{\substack{i,j,l\\ i\neq j\neq l}}\mathbf{A}_{ij}(\mathbf{B}_{il}+\mathbf{B}_{li})$$

$$+\frac{2+|\mathbb{E}X_1^2|^2}{n(n-1)(n-2)(n-3)}\sum_{\substack{i,j,k,l\\ i\neq j\neq k\neq l}}\mathbf{A}_{ij}\mathbf{B}_{kl} + o\left(\frac{1}{n}\right).$$

We next prove that $\Omega_n = o(1/n)$. Actually, it is straightforward to get the following derivations:

$$\sum_{\substack{i,k,l\\ i\neq k\neq l}}\mathbf{A}_{ii}\mathbf{B}_{kl} = \sum_{i,k,l}\mathbf{A}_{ii}\mathbf{B}_{kl} - 2\sum_{i,l}\mathbf{A}_{ii}\mathbf{B}_{il} - \sum_{i,l}\mathbf{A}_{ii}\mathbf{B}_{ll} + 2\sum_{i}\mathbf{A}_{ii}B_{ii}, \tag{8.30}$$

$$\mathbb{E}|\sum_{i,k,l}\mathbf{A}_{ii}\mathbf{B}_{kl}| \leq \mathbb{E}|tr\mathbf{A}\mathbf{e}^T\mathbf{B}\mathbf{e}| = O(n^2), \tag{8.31}$$

$$\mathbb{E}|\sum_{i,l}\mathbf{A}_{ii}\mathbf{B}_{il}| \leq \mathbb{E}|\sum_{i}\mathbf{A}_{ii}\mathbf{e}_i^T\mathbf{B}\mathbf{e}| \tag{8.32}$$

$$\leq (\sum_i \mathbf{A}_{ii}\overline{\mathbf{A}}_{ii})^{1/2}\left(\mathbb{E}\left[\sum_i \mathbf{e}^T\mathbf{B}\mathbf{e}_i\overline{\mathbf{e}_i^T\mathbf{B}\mathbf{e}}\right]\right)^{1/2} = O(n),$$

$$\sum_{i,l}\mathbf{A}_{ii}\mathbf{B}_{ll} = tr\mathbf{A}tr\mathbf{B} = O(n^2) \text{ and } \sum_i \mathbf{A}_{ii}B_{ii} = O(n). \tag{8.33}$$

We conclude from (8.30)-(8.33) that $\sum_{\substack{i,k,l\\ i\neq k\neq l}}\mathbf{A}_{ii}\mathbf{B}_{kl} = O(n^2)$,

$$\sum_{\substack{i,k,l\\ i\neq k\neq l}}\mathbf{A}_{ik}\mathbf{B}_{il} = \mathbf{e}^T\mathbf{A}\mathbf{B}\mathbf{e} - tr\mathbf{A}\mathbf{B} - 2\sum_{i\neq l}\mathbf{B}_{ii}\mathbf{B}_{il} = O(n), \tag{8.34}$$

and

$$\frac{1}{n^2}\sum_{\substack{i,t,k,l\\ i\neq t\neq k\neq l}}\mathbf{A}_{it}\mathbf{A}_{kl} = \left[\frac{1}{n}(\mathbf{e}^T\mathbf{A}\mathbf{e}-tr\mathbf{A})\right]\left[\frac{1}{n}(\mathbf{e}^T\mathbf{B}\mathbf{e}-tr\mathbf{A})\right] \tag{8.35}$$

$$-\frac{2}{n^2}\sum_{\substack{i,k,l\\ i\neq k\neq l}}\mathbf{A}_{ik}\mathbf{B}_{il} = O(1).$$

We have proved the lemma. □

8.3 Lemma 7

For any $l \in \mathbf{N}^+$, $\mu_1 > (1+\sqrt{c})^2$ and $0 < \mu_2 < \mathbf{I}_{(0,1)}(c)(1-\sqrt{c})^2$, then under condition (7.6), we have

$$P(\|\mathbf{B}_n\| \geq \mu_1) = o(n^{-l}) \tag{8.36}$$

and

$$P(\lambda_{min}^{\mathbf{B}_n} \leq \mu_2) = o(n^{-l}). \tag{8.37}$$

Proof. Set $\mathbf{X}_n = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p)^*$. Denote the i-th largest eigenvalue of $\mathbf{B}_n$ by λ_i and the i-th largest eigenvalue of $C_n = \frac{1}{n}(\mathbf{X}_n - \bar{\mathbf{X}}_n)(\mathbf{X}_n - \bar{\mathbf{X}}_n)^*$ by ν_i. Noticing the trivial inequalities for any positive constant σ small enough such that $\mu_1 - \sigma > (1+\sqrt{c})^2$ and $\mu_2 + \sigma < \mathbf{I}_{(0,1)}(c)(1-\sqrt{c})^2$, we have

$$\begin{aligned} & P(\lambda_1 \geq \mu_1) \\ = \;& P(\lambda_1 \geq \mu_1, \nu_1 \geq \mu_1 - \sigma) + P(\lambda_1 \geq \mu_1, \nu_1 < \mu_1 - \sigma) \\ \leq \;& P(\nu_1 \geq \mu_1 - \sigma) + P(|\lambda_1 - \nu_1| \geq \sigma) \end{aligned} \tag{8.38}$$

and

$$\begin{aligned} & P(\lambda_p \leq \mu_2) \\ = \;& P(\lambda_p \leq \mu_2, \nu_p \leq \mu_2 - \sigma) + P(\lambda_p \leq \mu_p, \nu_p > \mu_p - \sigma) \\ \leq \;& P(\nu_p \leq \mu_2 - \sigma) + P(|\lambda_p - \nu_p| \geq \sigma). \end{aligned} \tag{8.39}$$

For the moment, we assume that

$$P(\|\mathbf{C}_n\| = \nu_1 \geq \mu_1 - \sigma) = o(n^{-l}) \tag{8.40}$$

and

$$P(\lambda_{\min}^{\mathbf{C}_n} = \nu_p \leq \mu_2 - \sigma) = o(n^{-l}). \tag{8.41}$$

It then suffices to bound $\max_{1\leq i\leq p} P(|\lambda_i - \nu_i| \geq \sigma)$. By Lemma 2, we have

$$\begin{aligned} & \max_{1\leq i\leq p} |\sqrt{\lambda_i} - \sqrt{\nu_i}| \leq \|n^{-1/2}(\mathbf{X}_n - \bar{\mathbf{X}}_n)\| \|\sqrt{n}\mathbf{D}_n - \mathbf{I}_n\| \\ & = \|n^{-1/2}(\mathbf{X}_n - \bar{\mathbf{X}}_n)\| . \max_{1\leq i\leq p} \left| \frac{n^{1/2}}{\|\mathbf{y}_i - \bar{\mathbf{y}}_i\|} - 1 \right| . \end{aligned} \tag{8.42}$$

In view of the above inequality, it is enough to show that for any fixed ϵ, we have $P\left(\max_{1\le i\le p}|\frac{n^{1/2}}{\|\mathbf{y}_i-\bar{\mathbf{y}}_i\|}-1|\ge\epsilon\right)=o(n^{-l})$, which can be guaranteed by $P\left(\max_{1\le i\le p}|\frac{\|\mathbf{y}_i-\bar{\mathbf{y}}_i\|^2}{n}-1|\ge\epsilon\right)=o(n^{-l})$.

By the inequality

$$\max_{1\le i\le p}\left|\frac{\|\mathbf{y}_i-\bar{\mathbf{y}}_i\|^2}{n}-1\right|\le\max_{1\le i\le p}\left|\frac{\sum_{j=1}^n|X_{ij}|^2}{n}-1\right|+\max_{1\le i\le p}\left|\frac{1}{n}\sum_{j=1}^n X_{ij}\right|^2,$$

it suffices to show the following two inequalities:

$$P\left(\max_{1\le i\le p}|\frac{\sum_{j=1}^n|X_{ij}|^2}{n}-1|\ge\epsilon\right)=o(n^{-l})$$

and

$$P\left(\max_{1\le i\le p}|\frac{1}{n}\sum_{j=1}^n X_{ij}|\ge\epsilon\right)=o(n^{-l}) \tag{8.43}$$

such that we can obtain (8.36) and (8.37).

To prove these two inequalities, one can refer to the proof of inequality (9) in Chen and Pan (2012) for details, we omit them here (one should note that p and n here are of the same order, which is different from Chen and Pan (2012), but the proof is almost the same).

To finish the proof, we need to show that (8.40) and (8.41). Denoting the i-th largest eigenvalue of $\mathbf{S}_n=\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*$ by τ_i, referring to Bai and Silverstein (2004), we know that

$$P(\|\mathbf{S}_n\|=\tau_1\ge\mu_1-\sigma/2)=o(n^{-l}) \tag{8.44}$$

and

$$P(\lambda_{min}^{\mathbf{S}_n}=\tau_p\le\mu_2-\sigma/2)=o(n^{-l}). \tag{8.45}$$

Similarly to (8.42), we have

$$\max_{1\le i\le p}|\sqrt{\tau_i}-\sqrt{\nu_i}|\le\|n^{-1/2}\overline{X}_n\|\le\sqrt{\frac{1}{n}\sum_{i=1}^p|\frac{1}{n}\sum_{j=1}^n X_{ij}|^2}. \tag{8.46}$$

Combining (8.43), (8.44), (8.45), (8.46) together, we have (8.40) and (8.41). □

8.4 Lemma 8

Suppose that $x_n=\mathbf{e}/\sqrt{n}=\mathbf{1}/\sqrt{n}$ is a unit vector, then for the truncated random variable satisfying (7.6), we have $\mathbb{E}|x_n^*\mathbf{D}^{-1}(z)x_n+\frac{1}{z}|^2\to 0$.

Proof. By Lemma 5, we obtain for any $2 \leq r \in \mathbb{N}^+$

$$\mathbb{E}|\mathbf{r}_j^* \mathbf{D}_j^{-1}(z) x_n x_n^T \mathbf{D}_j^{-1}(z) \mathbf{r}_j|^r = O(n^{-2}\delta_n^{2r-4}).$$

Rewrite it as a martingale

$$\begin{aligned}
& x_n^* \mathbf{D}^{-1}(z) x_n - x_n^* \mathbb{E}\mathbf{D}^{-1}(z) x_n \\
= & \sum_{j=1}^{p} x_n^* \mathbb{E}_j \mathbf{D}^{-1}(z) x_n - x_n^* \mathbb{E}_{j-1} \mathbf{D}^{-1}(z) x_n \\
= & \sum_{j=1}^{p} x_n^* \mathbb{E}_j (\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)) x_n - x_n^* \mathbb{E}_{j-1} (\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)) x_n \\
= & -\sum_{j=1}^{p} (\mathbb{E}_j - \mathbb{E}_{j-1}) \beta_j(z) \mathbf{r}_j^* \mathbf{D}_j^{-1}(z) x_n x_n^* \mathbf{D}_j^{-1}(z) \mathbf{r}_j.
\end{aligned}$$

By Burkholder's inequality and (7.61), we have

$$\begin{aligned}
& \mathbb{E}|x_n^* \mathbf{D}^{-1}(z) x_n - x_n^* \mathbb{E}\mathbf{D}^{-1}(z) x_n|^2 \qquad (8.47) \\
\leq & K \sum_{j=1}^{p} \mathbb{E}|(\mathbb{E}_j - \mathbb{E}_{j-1}) \beta_j(z) \mathbf{r}_j^* \mathbf{D}_j^{-1}(z) x_n x_n^* \mathbf{D}_j^{-1}(z) \mathbf{r}_j|^2 \\
\leq & K \sum_{j=1}^{p} (\mathbb{E}|\beta_j(z)|^4)^{1/2} (\mathbb{E}|\mathbf{r}_j^* \mathbf{D}_j^{-1}(z) x_n x_n^* \mathbf{D}_j^{-1}(z) \mathbf{r}_j|^4)^{1/2} = O(\delta_n^2).
\end{aligned}$$

Thus, we have $\mathbb{E}|x_n^* \mathbf{D}^{-1}(z) x_n - x_n^* \mathbb{E}\mathbf{D}^{-1}(z) x_n|^2 \longrightarrow 0$. If $\Im z \geq v_0 > 0$, then $|\beta_j(z)| \leq \frac{|z|}{v_0}$, so (8.47) can get a sharper bound

$$\mathbb{E}|x_n^* \mathbf{D}^{-1}(z) x_n - x_n^* \mathbb{E}\mathbf{D}^{-1}(z) x_n|^2 = O\left(\frac{1}{n}\right). \qquad (8.48)$$

From the proof above, one should note that (8.47) and (8.48) hold for $\mathbf{D}_j(z)$ and any unit vector.

Note that $\mathbf{D}(z) + (c_n z \mathbb{E} m_n(z) + z)\mathbf{I} = \sum_{j=1}^{p} r_j r_j^* + c_n z \mathbb{E} m_n(z) \mathbf{I}$.

Recalling $m_n(z) = -\frac{1}{pz} \sum_{j=1}^{p} \beta_j(z)$, $\mathbf{G}_n(z) = c_n \mathbb{E} m_n(z) \mathbf{I}_n + \mathbf{I}_n$, and using the identity

$\mathbf{r}_j^*\mathbf{D}^{-1}(z) = \beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)$, we obtain

$$\begin{aligned}
& (-z\mathbf{G}_n(z))^{-1} - \mathbb{E}\mathbf{D}^{-1}(z) \\
= & -z^{-1}\mathbf{G}_n^{-1}(z)\mathbb{E}\big[\big(\sum_{j=1}^{p}\mathbf{r}_j\mathbf{r}_j^* - (-zc_n\mathbb{E}m_n(z)\mathbf{I}_n)\big)\mathbf{D}^{-1}(z)\big] \\
= & -z^{-1}\sum_{j=1}^{p}\mathbb{E}\beta_j\big[\mathbf{G}_n^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\big] \\
& -z^{-1}\mathbb{E}\big[\mathbf{G}_n^{-1}(z)\big(-c_nz\mathbb{E}m_n(z)\big)\mathbf{I}_n\mathbf{D}^{-1}(z)\big] \\
= & -z^{-1}\sum_{n=1}^{p}\mathbb{E}\beta_j\big[\mathbf{G}_n^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\big(\underline{\mathbf{B}}_{(j)}^n - z\mathbf{I}_n\big)^{-1} - \frac{1}{n}\mathbf{G}_n^{-1}(z)\mathbb{E}\mathbf{D}^{-1}(z)\big] \\
= & -z^{-1}p\mathbb{E}\beta_1\big[\mathbf{G}_n^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z) - \frac{1}{n}\mathbf{G}_n^{-1}(z)\mathbb{E}\mathbf{D}^{-1}(z)\big].
\end{aligned}$$

Multiplying by $(-x_n^*)$ on the left and x_n on the right, we have

$$\begin{aligned}
& x_n^*\mathbf{D}^{-1}(z)x_n - x_n^*(-z\mathbf{G}_n(z))^{-1}x_n \\
= & z^{-1}p\mathbb{E}\beta_1\big[x_n^*\mathbf{G}_n^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)x_n - \frac{1}{n}x_n^*\mathbf{G}_n^{-1}(z)\mathbb{E}\mathbf{D}^{-1}(z)x_n\big] \\
\triangleq & \delta_1 + \delta_2 + \delta_3,
\end{aligned}$$

where $\delta_1 = \frac{p}{z}\mathbb{E}(\beta_1(z)\alpha_1(z))$, $\alpha_1(z) = x_n^*\mathbf{G}_n^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)x_n - \frac{1}{n}x_n^*\mathbf{G}_n^{-1}(z)\mathbf{D}_1^{-1}(z)x_n$,

$$\delta_2 = \frac{1}{z}\mathbb{E}\beta_1(z)x_n^*\mathbf{G}_n^{-1}(z)(\mathbf{D}_1^{-1}(z) - \mathbf{D}^{-1}(z))x_n,$$

and $\delta_3 = \frac{1}{z}\mathbb{E}\beta_1(z)x_n^*\mathbf{G}_n^{-1}(z)(\mathbf{D}^{-1}(z) - \mathbb{E}\mathbf{D}^{-1}(z))x_n$.

Recalling the notations defined above (7.17) and by the following equalities: $\delta_1 = \frac{p}{z}\mathbb{E}\tilde{\beta}_1(z)\alpha_1(z) - \frac{p}{z}\mathbb{E}\left[\beta_1(z)\tilde{\beta}_1(z)\varepsilon_1(z)\alpha_1(z)\right]$,

$$\tilde{\beta}_1(z) = b_n(z) - \frac{1}{n}b_n(z)\tilde{\beta}_1(z)tr(\mathbf{D}_1^{-1}(z) - \mathbb{E}\mathbf{D}_1^{-1}(z)),$$

and $\mathbb{E}\alpha_1 = -(c_n\mathbb{E}m_n(z)+1)^{-1}\frac{1}{(n-1)}[\mathbb{E}x_n^*\mathbf{D}_1^{-1}(z)x_n + o(1)]$, it is easy to see $p\mathbb{E}\beta_1(z)\alpha_1(z) = [\frac{1}{1+\mathbb{E}\underline{m}_n(z)} + o(1)]p\mathbb{E}\alpha_1(z)$.

Therefore, $\delta_1 = c_n\frac{zm_n(z)}{(c_nzm_n(z)+z)}x_n^*\mathbb{E}(\mathbf{D}_1^{-1}(z))x_n + o(1)$.

Similarly to Bai, Miao and Pan(2007), one may have $\delta_2 = o(1)$ and $\delta_3 = o(1)$. Hence, we obtain $\left(1 - \frac{c_nzm_n(z)}{c_nzm_n(z)+z}\right)x_n^*\mathbb{E}(\mathbf{D}^{-1}(z))x_n + \frac{1}{c_nzm_n(z)+z} \longrightarrow 0$, which implies $x_n^*\mathbb{E}(\mathbf{D}^{-1}(z))x_n \longrightarrow -\frac{1}{z}$.

□

Remark 7. *This is an interesting result that the limit of* $\frac{1}{n}\mathbf{e}^T\mathbb{E}(\mathbf{D}^{-1}(z))\mathbf{e}$ *is independent of the corresponding Stieltjes transform* $\underline{m}(z)$*. Meanwhile, the limit of* $x_n^*\mathbb{E}(\mathbf{D}^{-1}(z))x_n$ *depends on the limit of* x_n*, one can check this by the fact:* $\frac{1}{n}tr\mathbb{E}(\mathbf{D}^{-1}(z)) \longrightarrow \underline{m}(z)$*, which depends on the Stieltjes transform* $\underline{m}(z)$ *different from the result of Lemma 8.*

8.5 Lemma 9

For $z_1,\, z_2 \in \mathcal{C}_u$, we have

$$\frac{\partial^2}{\partial z_1 \partial z_2} J_4 \xrightarrow{i.p.} \frac{|\mathbb{E}X_{11}^2|^2 c m'(z_1) m'(z_2)}{[(1+c_1 m(z_1))(1+cm(z_2)) - c|\mathbb{E}X_{11}^2|^2 m(z_1)m(z_2)]^2}. \tag{8.49}$$

Proof. From (7.35) and bounds, we have

$$\mathbf{D}_j^{-1}(z_1) = -\mathbf{H}_n(z_1) + b_1(z_1)\mathbf{A}(z_1) + \mathbf{B}(z_1) + \mathbf{C}(z_1). \tag{8.50}$$

Therefore, recalling (7.37)–(7.39), we have

$$\begin{aligned} & \frac{1}{n} tr\big[\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_2)^T\big)\big] \\ = \; & -\mathbf{H}_n(z_1) tr \mathbb{E}_j(\mathbf{D}_j^{-1}(z_2))^T + \frac{1}{n} b_1(z_1) tr \mathbb{E}_j \mathbf{A}(z_1)(\mathbf{D}_j^{-1}(z_2))^T + o(1). \end{aligned} \tag{8.51}$$

We can write

$$tr\mathbf{E}_j\big(\mathbf{A}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2)\big)^T = B_1(z_1,z_2) + B_2(z_1,z_2) + B_3(z_1,z_2) + N(z_1,z_2),$$

where

$$\begin{aligned} B_1(z_1,z_2) &= -tr\sum_{i<j} \mathbf{H}_n(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)(\beta_{ij}(z_2)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_2))^T \\ &= -\sum_{i<j}^{p} \beta_{ij}(z_2)\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)(\mathbf{D}_{ij}^{-1}(z_2))^T\bar{\mathbf{r}}_i\mathbf{r}_i'(\mathbf{D}_{ij}^{-1}(z_2))^T\mathbf{H}_n(z_1)\mathbf{r}_i; \\ B_2(z_1,z_2) &= -tr\sum_{i<j}\mathbf{H}_n(z_1)n^{-1}\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2) - \mathbf{D}_{ij}^{-1}(z_2)\big)^T; \\ B_3(z_1,z_2) &= tr\sum_{i<j}\mathbf{H}_n(z_1)\big(\mathbf{r}_i\mathbf{r}_i^* - n^{-1}\mathbf{I}_n\big)\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)(\mathbf{D}_{ij}^{-1}(z_2))^T; \\ N(z_1,z_2) &= tr\mathbb{E}_j\sum_{i>j}\mathbf{H}_n(z_1)\big(-\frac{1}{n(n-1)}\mathbf{e}\mathbf{e}^* + \frac{1}{n(n-1)}\mathbf{I}_n\big)\mathbf{D}_{ij}^{-1}(z_1)(\mathbf{D}_j^{-1}(z_2))^T. \end{aligned}$$

It is easy to see $N(z_1,z_2) = O(1)$. We get from (7.36) and (7.38) that $|B_2(z_1,z_2)| \le \frac{1+1/v_0}{v_0^2}$. Similarly to (7.37), we have $\mathbb{E}|B_3(z_1,z_2)| \le \frac{1+1/v_0}{v_0^3} n^{1/2}$.

Using Lemma 5 and (7.31), we have, for $i<j$,

$$\begin{aligned} & \mathbb{E}\big|\beta_{ij}(z_2)\mathbf{r}_i^*\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)(\mathbf{D}_{ij}^{-1}(z_2))^T\bar{\mathbf{r}}_i\mathbf{r}_i'(\mathbf{D}_{ij}^{-1}(z_2))^T\mathbf{H}_n(z_1)\mathbf{r}_i \\ & \quad -b_1(z_2)n^{-2}|\mathbb{E}X_{11}^2|^2 tr\big(\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)(\mathbf{D}_{ij}^{-1}(z_2))^T\big)tr\big((\mathbf{D}_{ij}^{-1}(z_2))^T\mathbf{H}_n(z_1)\big)\big| \\ \le \; & Kn^{-1/2}. \end{aligned} \tag{8.52}$$

By (7.38), we have

$$\Big|tr\big(\mathbb{E}_j\big(\mathbf{D}_{ij}^{-1}(z_1)\big)(\mathbf{D}_{ij}^{-1}(z_2))^T\big)tr\big((\mathbf{D}_{ij}^{-1}(z_2))^T\mathbf{H}_n(z_1)\big) \\ -tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)(\mathbf{D}_j^{-1}(z_2))^T\big)tr\big((\mathbf{D}_j^{-1}(z_2))^T\mathbf{H}_n(z_1)\big)\Big| \le Kn. \tag{8.53}$$

It follows from (8.52) and (8.53) that

$$\mathbb{E}\Big|B_1(z_1,z_2)+\frac{j-1}{n^2}b_1(z_2)|\mathbb{E}X_{11}^2|^2 tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)(\mathbf{D}_j^{-1}(z_2))^T\big)tr\big((\mathbf{D}_j^{-1}(z_2))^T\mathbf{H}_n(z_1)\big)\Big| \\ \le Kn^{1/2}.$$

Analogously, recalling (7.48), we may obtain

$$\begin{aligned} & tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2)\big)^T\big) \\ \times\ & \big[1-\frac{j-1}{n^2}|\mathbb{E}X_{11}^2|^2 m_{c_n}(z_1)m_{c_n}(z_2)tr(\mathbf{Q}_n(z_2)\mathbf{Q}_n(z_1))\big] \\ & = \frac{1}{z_1z_2}tr\big(\mathbf{Q}_n(z_2)\mathbf{Q}_n(z_1)\big)+B_6(z_1,z_2), \end{aligned} \tag{8.54}$$

where $\mathbb{E}|B_6(z_1,z_2)| \le Kn^{1/2}$.

Rewrite (8.54) as

$$\begin{aligned} & \frac{1}{n}tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2)\big)^T\big)\big[1-\frac{j-1}{n}|\mathbb{E}X_{11}^2|^2\frac{m_{c_n}(z_1)m_n^0(z_2)}{\big(1+\frac{p-1}{n}m_{c_n}(z_2)\big)\big(1+\frac{p-1}{n}m_{c_n}(z_1)\big)}\big] \\ = \ & \frac{1}{z_1z_2}\frac{1}{\big(1+\frac{p-1}{n}m_{c_n}(z_1)\big)\big(1+\frac{p-1}{n}m_{c_n}(z_2)\big)}+\frac{1}{n}B_6(z_1,z_2). \end{aligned} \tag{8.55}$$

Therefore, we have

$$\begin{aligned} & \frac{1}{n}tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2)\big)^T\big) \\ = \ & \frac{a_n(z_1,z_2)}{z_1z_2m_{c_n}(z_1)m_{c_n}(z_2)\big[1-\frac{j-1}{p}|\mathbb{E}X_{11}^2|^2a_n(z_1,z_2)\big]}+o(1), \end{aligned} \tag{8.56}$$

where $a_n(z_1,z_2)=\frac{p}{n}\frac{m_{c_n}(z_1)m_{c_n}(z_2)}{\big(1+\frac{p-1}{n}m_{c_n}(z_1)\big(1+\frac{p-1}{n}m_{c_n}(z_2)\big)}$.

Because the limit of $a_n(z_1,z_2)$ is $a(z_1,z_2)=\frac{cm(z_1)m(z_2)}{\big(1+cm(z_1)\big)\big(1+cm(z_2)\big)}$, we have

$$\begin{aligned} & \frac{1}{n}tr\big(\mathbb{E}_j\big(\mathbf{D}_j^{-1}(z_1)\big)\big(\mathbf{D}_j^{-1}(z_2)\big)^T\big) \\ = \ & \frac{a_n(z_1,z_2)}{z_1z_2m(z_1)m(z_2)\big[1-\frac{j-1}{p}|\mathbb{E}X_{11}^2|^2a_n(z_1,z_2)\big]}+o(1). \end{aligned} \tag{8.57}$$

Therefore, J_4 can be written as

$$J_4=|\mathbb{E}X_{11}^2|^2a_n(z_1,z_2)\frac{1}{p}\sum_{j=1}^{p}\frac{1}{1-\frac{j-1}{p}|\mathbb{E}X_{11}^2|^2a_n(z_1,z_2)}+B_7(z_1,z_2),$$

where $\mathbb{E}|B_7(z_1,z_2)| \le Kn^{-1/2}$.

Thus, by (8.55), the i.p. limit of J_4 is $|\mathbb{E}X_{11}^2|^2 \int_0^{a(z_1,z_2)} \frac{1}{1-|\mathbb{E}X_{11}^2|^2 z} dz$. We can then directly get the limit of $\frac{\partial^2}{\partial z_1 \partial z_2} J_4$ and we omit the details here. □

8.6 Lemma 10

When $v_0 = \Im z > 0$ is bounded, we have

$$|\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z)| \le Kn^{-1/2}.$$

Proof. First, by (7.20), one can prove that

$$\mathbb{E}|\frac{1}{n} tr\mathbf{D}^{-1}(z) - \frac{1}{n}\mathbb{E}tr\mathbf{D}^{-1}(z)|^q \le Kn^{-q/2}, \quad \forall q \in N_+. \tag{8.58}$$

By $\mathbf{D}(z) = \sum_{i=1}^p \mathbf{r}_i\mathbf{r}_i^* - z\mathbf{I}$, we have

$$\mathbf{I} = \sum_{i=1}^p \mathbf{r}_i\mathbf{r}_i^*\mathbf{D}^{-1}(z) - z\mathbf{D}^{-1}(z) = \sum_{i=1}^p \frac{\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_i^{-1}(z)}{1+\mathbf{r}_i^*\mathbf{D}_i^{-1}(z)\mathbf{r}_i} - z\mathbf{D}^{-1}(z).$$

Taking trace and expectation on both sides, then divided by n, we have

$$1 = c_n - c_n\mathbb{E}\frac{1}{1+\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{r}_1} - z\mathbb{E}\underline{m}_n(z). \tag{8.59}$$

Denote

$$\rho_n(z) = c_n(\mathbb{E}\frac{1}{1+\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{r}_1} - \frac{1}{1+\frac{1}{n}\mathbb{E}tr\mathbf{D}_1^{-1}(z)})/\mathbb{E}\underline{m}_n(z). \tag{8.60}$$

Combining (8.59) and (8.60) together, we obtain

$$\mathbb{E}\underline{m}_n(z) = \frac{1}{c_n\frac{1}{1+\mathbb{E}\underline{m}_n(z)} - z + \rho_n(z)}.$$

As we know that $\underline{m}_{c_n}(z)$ satisfies the following equation

$$\underline{m}_{c_n}(z) = \frac{1}{c_n\frac{1}{1+\underline{m}_{c_n}(z)} - z}.$$

Then we have

$$\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z) = \frac{c_n(\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z))\frac{1}{(1+\mathbb{E}\underline{m}_n(z))(1+\underline{m}_{c_n}(z))}}{(c_n\frac{1}{1+\mathbb{E}\underline{m}_n(z)} - z + \rho_n(z))(c_n\frac{1}{1+\underline{m}_{c_n}(z)} - z)}$$
$$+ (\mathbb{E}\underline{m}_n(z))\underline{m}_{c_n}(z)\rho_n(z).$$

Rewrite it as

$$(\mathbb{E}\underline{m}_n(z) - \underline{m}_{c_n}(z))(1 - c_n\frac{\mathbb{E}\underline{m}_n(z)\underline{m}_{c_n}(z))}{(1+\mathbb{E}\underline{m}_n(z))(1+\underline{m}_{c_n}(z))}) = (\mathbb{E}\underline{m}_n(z))\underline{m}_{c_n}(z)\rho_n(z).$$

Because $\Im z$ is bounded, it is straightforward to obtain $\underline{m}_{c_n}(z) = O(1)$. By the definition of $\rho_n(z)$, there exists a constant C such that

$$
\begin{aligned}
&|\mathbb{E}\underline{m}_n(z)\rho_n(z)| \le C\mathbb{E}|\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{r}_1 - \frac{1}{n}\mathbb{E}tr\mathbf{D}_1^{-1}(z)| \\
&\le C(\mathbb{E}|\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{r}_1 - \frac{1}{n}tr\mathbf{D}_1^{-1}(z)| + \mathbb{E}|\mathbb{E}\frac{1}{n}tr\mathbf{D}^{-1}(z) - \frac{1}{n}tr\mathbf{D}^{-1}(z)| \\
&+\mathbb{E}|\frac{1}{n}tr\mathbf{D}^{-1}(z) - \frac{1}{n}tr\mathbf{D}_1^{-1}(z)|) \le Cn^{-1/2},
\end{aligned}
\tag{8.61}
$$

where the last inequality follows from Lemma 5 and (8.58). Similar to (2.19) of Bai and Silverstein (2004), combining with $|\mathbb{E}\underline{m}_n(z)\rho_n(z)| \le Cn^{-1/2}$ and $\underline{m}_{c_n}(z) = O(1)$, we have

$$|c_n \frac{\mathbb{E}\underline{m}_n(z)\underline{m}_{c_n}(z))}{(1+\mathbb{E}\underline{m}_n(z))(1+\underline{m}_{c_n}(z))}| < 1.$$

Thus, we have proved it, and therefore completed all the proofs in this supplementary document. □

References

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

Bai, Z. D., Miao, B. Q. and Pan, G. M. (2007). On the asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35** 1532–1572.

Bao, Z. G., Pan, G. M. and Zhou, W. (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Elctron. J. Probab.* **17** 1-32.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Billingsley, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 1007–1026.

Chen, B. B. and Pan, G. M. (2012). Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero. *Bernoulli* **18** 1405–1420.

Cai, T. T. and Jiang, T. F. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525.

Cai, T. T., Zhang, C. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144.

Horn, R. N. and Johnson, C. R. (1999). *Matrix Analysis.* Cambridge University Press.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327.

Jiang, T. F. (2004). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66** 35–48.

Morrison, D. F. (2005). *Multivariate Statistical Methods* 4th ed. CA: Brooks/Cole.

Pan, G. M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270.

Petrov, V. V. (1975). *Sums of Independent Random Variables* 1st ed. Springer.

Pillai, N. S. and Yin, J. (2012). Edge university of correlation matrices. *Ann. Statist.* **40** 1737–1763.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55** 331–339.

Xiao, H. and Zhou, W. (2010). Almost sure limit of the smallest eigenvalue of some sample correlation matrices. *J. Theor. Probab.* **23** 1–20.

Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988). On limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Relat. Field* **78** 509–521.