Bernoulli **21**(2), 2015, 1089–1133
DOI: 10.3150/14-BEJ599

CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size

BINBIN CHEN* and GUANGMING PAN**

*Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. E-mail: *CHEN0635@e.ntu.edu.sg; **gmpan@ntu.edu.sg*

Let $\mathbf{A} = \frac{1}{\sqrt{np}}(\mathbf{X}^T\mathbf{X} - p\mathbf{I}_n)$ where $\mathbf{X}$ is a $p \times n$ matrix, consisting of independent and identically distributed (i.i.d.) real random variables X_{ij} with mean zero and variance one. When $p/n \to \infty$, under fourth moment conditions a central limit theorem (CLT) for linear spectral statistics (LSS) of $\mathbf{A}$ defined by the eigenvalues is established. We also explore its applications in testing whether a population covariance matrix is an identity matrix.

Keywords: central limit theorem; empirical spectral distribution; hypothesis test; linear spectral statistics; sample covariance matrix

1. Introduction

The last few decades have seen explosive growth in data analysis, due to the rapid development of modern information technology. We are now in a setting where many very important data analysis problems are high-dimensional. In many scientific areas, the data dimension can even be much larger than the sample size. For example, in micro-array expression, the number of genes can be tens of thousands or hundreds of thousands while there are only hundreds of samples. Such kind of data also arises in genetic, proteomic, functional magnetic resonance imaging studies and so on (see Chen *et al.* [11], Donoho [13], Fan and Fan [14]).

The main purpose of this paper is to establish a central limit theorem (CLT) of linear functionals of eigenvalues of the sample covariance matrix when the dimension p is much larger than the sample size n. Consider the sample covariance matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$, where $\mathbf{X} = (X_{ij})_{p\times n}$ and $X_{ij}, i = 1, \dots, p, j = 1, \dots, n$ are i.i.d. real random variables with mean zero and variance one. As we know, linear functionals of eigenvalues of $\mathbf{S}$ are closely related to its empirical spectral distribution (ESD) function $F^{\mathbf{S}}(x)$. Here for any $n \times n$ Hermitian matrix $\mathbf{M}$ with real eigenvalues $\lambda_1, \dots, \lambda_n$, the empirical spectral distribution of $\mathbf{M}$ is defined by

$$F^{\mathbf{M}} = \frac{1}{n}\sum_{j=1}^{n} I(\lambda_j \le x),$$

where $I(\cdot)$ is the indicator function. However, it is inappropriate to investigate $F^{\mathbf{S}}(x)$ when $p/n \to \infty$ since $\mathbf{S}$ has $(p-n)$ zero eigenvalues and hence $F^{\mathbf{S}}(x)$ converges to a degenerate distribution with probability one. Note that the eigenvalues of $\mathbf{S}$ are the same as those of $\frac{1}{n}\mathbf{X}^T\mathbf{X}$ except $(p-n)$ zero eigenvalues. Thus, instead, we turn to the eigenvalues of $\frac{1}{p}\mathbf{X}^T\mathbf{X}$ and renormalize it as

$$\mathbf{A} = \sqrt{\frac{p}{n}}\left(\frac{1}{p}\mathbf{X}^T\mathbf{X} - \mathbf{I}_n\right), \tag{1.1}$$

where $\mathbf{I}_n$ is the identity matrix of order n.

The first breakthrough regarding the ESD of $\mathbf{A}$ was made in Bai *et al.* [7]. They proved that with probability one

$$F^{\mathbf{A}}(x) \to F(x),$$

which is the so-called semicircle law with the density

$$F'(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & \text{if } |x| \le 2, \\ 0, & \text{if } |x| > 2. \end{cases} \tag{1.2}$$

In random matrix theory, $F(x)$ is referred to as the limiting spectral distribution (LSD) of $\mathbf{A}$. For such matrices, Chen and Pan [10] proved that the largest eigenvalue converges to the right endpoint of the support of $F(x)$ with probability one. When $X_{11} \sim N(0,1)$, Karoui [20] reported that the largest eigenvalue of $\mathbf{X}\mathbf{X}^T$ after properly centering and scaling converges in distribution to the Tracy–Widom law, and Birke and Dette [9] established central limit theorems for the quadratic function of the eigenvalues of $\mathbf{A}$. Recently, Pan and Gao [24] further derived the LSD of a general form of (1.1), which is determined by its Stieltjes transform. Here, the Stieltjes transform for any distribution function G is given by

$$m_G(z) = \int \frac{1}{x-z}\,\mathrm{d}G(x), \qquad \Im(z) > 0,$$

where $\Im(z)$ represents the imaginary part of z.

Gaussian fluctuations in random matrices are investigated by different authors, starting with Costin and Lebowitz [12]. Johansson [18] considered an extended random ensembles whose entries follow a specific class of densities and established a CLT of the linear spectral statistics (LSS). Recently, a CLT for LSS of sample covariance matrices is studied by Bai and Silverstein [5] and of Wigner matrices is studied by Bai and Yao [6].

We introduce some notation before stating our results. Denote the Stieltjes transform of the semicircle law F by $m(z)$. $\Im(z)$ is used to denote the imaginary part of a complex number z. For any given square matrix $\mathbf{B}$, let $\operatorname{tr}\mathbf{B}$ and $\overline{\mathbf{B}}$ denote the trace and the complex conjugate matrix of $\mathbf{B}$, respectively. The norm $\|\mathbf{B}\|$ represents the spectral norm of $\mathbf{B}$, that is, $\|\mathbf{B}\| = \sqrt{\lambda_1(\mathbf{B}\overline{\mathbf{B}})}$ where $\lambda_1(\mathbf{B}\overline{\mathbf{B}})$ means the maximum eigenvalue of $\mathbf{B}\overline{\mathbf{B}}$. The notation $\xrightarrow{d}$ means "convergence in distribution to". Let $\mathcal{S}$ denote any open region on the real plane including $[-2,2]$, which is the support of $F(x)$, and $\mathcal{M}$ be the set of functions which are analytic on $\mathcal{S}$. For any $f \in \mathcal{M}$, define

$$G_n(f) \triangleq n \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}\bigl(F^{\mathbf{A}}(x) - F(x)\bigr) - \frac{n}{2\pi \mathrm{i}} \oint_{|m|=\rho} f\bigl(-m - m^{-1}\bigr)\mathcal{X}_n(m)\frac{1-m^2}{m^2}\,\mathrm{d}m, \tag{1.3}$$

where

$$\begin{aligned} \mathcal{X}_n(m) &\triangleq \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}, \qquad \mathcal{A} = m - \sqrt{\frac{n}{p}}\bigl(1+m^2\bigr), \\ \mathcal{B} &= m^2 - 1 - \sqrt{\frac{n}{p}}m\bigl(1+2m^2\bigr), \qquad \mathcal{C} = \frac{m^3}{n}\biggl(\frac{m^2}{1-m^2} + \nu_4 - 2\biggr) - \sqrt{\frac{n}{p}}m^4, \end{aligned} \tag{1.4}$$

$\nu_4 = EX_{11}^4$ and $\sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}$ is a complex number whose imaginary part has the same sign as that of $\mathcal{B}$. The integral's contour is taken as $|m| = \rho$ with $\rho < 1$.

Let $\{T_k\}$ be the family of Chebyshev polynomials, which is defined as $T_0(x) = 1$, $T_1(x) = x$ and $T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$. To give an alternative way of calculating the asymptotic covariance of $X(f)$ in Theorem 1.1 below, for any $f \in \mathcal{M}$ and any integer $k > 0$, we define

$$\begin{aligned} \Psi_k(f) &\triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos\theta)\mathrm{e}^{\mathrm{i}k\theta}\,\mathrm{d}\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos\theta)\cos k\theta\,\mathrm{d}\theta = \frac{1}{\pi}\int_{-1}^{1} f(2x)T_k(x)\frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x. \end{aligned}$$

The main result is formulated in the following.

Theorem 1.1. *Suppose that*

(a) $\mathbf{X} = (X_{ij})_{p\times n}$ *where* $\{X_{ij} : i = 1, 2, \dots, p;\ j = 1, 2, \dots, n\}$ *are i.i.d. real random variables with* $EX_{11} = 0$, $EX_{11}^2 = 1$ *and* $\nu_4 = EX_{11}^4 < \infty$.
(b1) $n/p \to 0$ *as* $n \to \infty$.

Then, for any $f_1, \dots, f_k \in \mathcal{M}$, *the finite-dimensional random vector* $(G_n(f_1), \dots, G_n(f_k))$ *converges weakly to a Gaussian vector* $(Y(f_1), \dots, Y(f_k))$ *with mean function* $EY(f) = 0$ *and covariance function*

$$\operatorname{cov}\bigl(Y(f_1), Y(f_2)\bigr) = (\nu_4 - 3)\Psi_1(f_1)\Psi_1(f_2) + 2\sum_{k=1}^{\infty} k\Psi_k(f_1)\Psi_k(f_2) \tag{1.5}$$

$$= \frac{1}{4\pi^2}\int_{-2}^{2}\int_{-2}^{2} f_1'(x)f_2'(y)H(x,y)\,\mathrm{d}x\,\mathrm{d}y, \tag{1.6}$$

where

$$H(x,y) = (\nu_4 - 3)\sqrt{4-x^2}\sqrt{4-y^2} + 2\log\left(\frac{4 - xy + \sqrt{(4-x^2)(4-y^2)}}{4 - xy - \sqrt{(4-x^2)(4-y^2)}}\right).$$

Remark 1.1. Note that $\mathcal{X}_n(m)$ in (1.3) and $\underline{\mathcal{X}}_n(m) \triangleq \frac{-\mathcal{B}-\sqrt{\mathcal{B}^2-4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}$ are the two roots of the equation $\mathcal{A}x^2+\mathcal{B}x+\mathcal{C}=0$. Since $n/p \to 0$, an easy calculation shows $\mathcal{X}_n(m)=\mathrm{o}(1)$ and $\underline{\mathcal{X}}_n(m)=\frac{1-m^2}{m}+\mathrm{o}(1)$. Hence in practice, one may implement the mean correction in (1.3) by taking

$$\mathcal{X}_n(m)=\min\left\{\left|\frac{-\mathcal{B}+\sqrt{\mathcal{B}^2-4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}\right|,\left|\frac{-\mathcal{B}-\sqrt{\mathcal{B}^2-4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}\right|\right\},$$

and $m=\rho\mathrm{e}^{\mathrm{i}\theta}$ with $\theta\in[-2\pi,2\pi]$ and $0<\rho<1$.

The mean correction term, the last term in (1.3), can be simplified when $n^3/p=\mathrm{O}(1)$. Indeed, if $n^3/p=\mathrm{O}(1)$, we have $4\mathcal{A}\mathcal{C}=\mathrm{o}(1)$, $\mathcal{B}=m^2-1$. By (1.4),

$$\begin{aligned}
n\mathcal{X}_n(m)&=n\cdot\frac{-\mathcal{B}+\sqrt{\mathcal{B}^2-4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}=\frac{-2nC}{\mathcal{B}+\sqrt{\mathcal{B}^2-4\mathcal{A}\mathcal{C}}}\\
&=\frac{m^3}{1-m^2}\left(\frac{m^2}{1-m^2}-\nu_4-2\right)+\sqrt{\frac{n^3}{p}}\frac{m^4}{1-m^2}+\mathrm{o}(1).
\end{aligned}$$

Hence, by using the same calculation as that in Section 5.1 of Bai and Yao [6], we have

$$\begin{aligned}
&-\frac{n}{2\pi\mathrm{i}}\oint_{|m|=\rho} f\left(-m-m^{-1}\right)\mathcal{X}_n(m)\frac{1-m^2}{m^2}\,\mathrm{d}m\\
&=-\frac{1}{2\pi\mathrm{i}}\oint_{|m|=\rho} f\left(-m-m^{-1}\right)m\left[\frac{m^2}{1-m^2}-\nu_4-2+\sqrt{\frac{n^3}{p}}m\right]\mathrm{d}m+\mathrm{o}(1)\\
&=-\frac{1}{4}\big(f(2)+f(-2)\big)-\frac{1}{\pi}\int_{-1}^{1} f(2x)\left[2(\nu_4-3)x^2-\left(\nu_4-\frac{5}{2}\right)\right]\frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x \qquad (1.7)\\
&\quad-\frac{1}{\pi}\sqrt{\frac{n^3}{p}}\int_{-1}^{1} f(2x)\frac{4x^3-3x}{\sqrt{1-x^2}}\,\mathrm{d}x\\
&=-\left[\frac{1}{4}\big(f(2)+f(-2)\big)-\frac{1}{2}\Psi_0(f)+(\nu_4-3)\Psi_2(f)\right]-\sqrt{\frac{n^3}{p}}\Psi_3(f)+\mathrm{o}(1).
\end{aligned}$$

Define

$$Q_n(f)\triangleq n\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}\big(F^{\mathbf{A}}(x)-F(x)\big)-\sqrt{\frac{n^3}{p}}\Psi_3(f). \qquad (1.8)$$

Under the condition $n^3/p=\mathrm{O}(1)$, we then give a simple and explicit expression of the mean correction term of (1.3) in the following corollary.

Corollary 1.1. *Suppose that*

(a) $\mathbf{X} = (X_{ij})_{p\times n}$ *where* $\{X_{ij} : i = 1, 2, \ldots, p; j = 1, 2, \ldots, n\}$ *are i.i.d. real random variables with* $EX_{11} = 0$, $EX_{11}^2 = 1$ *and* $\nu_4 = EX_{11}^4 < \infty$.

(b2) $n^3/p = \mathrm{O}(1)$ *as* $n \to \infty$.

Then, for any $f_1, \ldots, f_k \in \mathcal{M}$, *the finite-dimensional random vector* $(Q_n(f_1), \ldots, Q_n(f_k))$ *converges weakly to a Gaussian vector* $(X(f_1), \ldots, X(f_k))$ *with mean function*

$$EX(f) = \tfrac{1}{4}\bigl(f(2) + f(-2)\bigr) - \tfrac{1}{2}\Psi_0(f) + (\nu_4 - 3)\Psi_2(f) \tag{1.9}$$

and covariance function $\operatorname{cov}(X(f), X(g))$ *being the same as that given in* (1.5) *and* (1.6).

Remark 1.2. The result of Bai *et al.* [2] suggests that, for large p and n with $p/n \to \infty$, the matrix $\sqrt{n}\mathbf{A}$ is close to a $n \times n$ Wigner matrix although its entries are not independent but weakly dependent. It is then reasonable to conjecture that the CLT for the LSS of $\mathbf{A}$ resembles that of a Wigner matrix described in Bai and Yao [6]. More precisely, by writing $\mathbf{A} = \frac{1}{\sqrt{n}}(w_{ij})$, where $w_{ii} = (\mathbf{s}_i^T\mathbf{s}_i - p)/\sqrt{p}$, $w_{ij} = \mathbf{s}_i^T\mathbf{s}_j/\sqrt{p}$ for $i \neq j$ and $\mathbf{s}_j$ is the jth column of $\mathbf{X}$, we have

$$\operatorname{Var}(w_{11}) = \nu_4 - 1, \qquad \operatorname{Var}(w_{12}) = 1, \qquad E\bigl(w_{12}^2 - 1\bigr)^2 = \frac{1}{p}\bigl(\nu_4^2 - 3\bigr).$$

Then, (1.9), (1.5) and (1.6) are consistent with (1.4), (1.5) and (1.6) of Bai and Yao [6], respectively, by taking their parameters as $\sigma^2 = \nu_4 - 1$, $\kappa = 2$ (the real variable case) and $\beta = 0$.

However, we remark that the mean correction term of $Q_n(f)$, the last term of (1.8), cannot be speculated from the result of Bai and Yao [6]. Note that this correction term will vanish in the case of the function f to be even or $n^3/p \to 0$. By the definition of $\Psi_k(f)$, one may verify that

$$\Psi_3(f) = \frac{1}{\pi}\sqrt{\frac{n^3}{p}}\int_{-1}^{1} f(2x)\frac{4x^3 - 3x}{\sqrt{1-x^2}}\,\mathrm{d}x,$$

$$-\frac{1}{2}\Psi_0(f) + (\nu_4 - 3)\Psi_2(f) = \frac{1}{\pi}\int_{-1}^{1} f(2x)\left[2(\nu_4 - 3)x^2 - \left(\nu_4 - \frac{5}{2}\right)\right]\frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x.$$

Remark 1.3. If we interchange the roles of p and n, Birke and Dette [9] established the CLT for $Q_n(f)$ in their Theorem 3.4 when $f = x^2$ and $X_{ij} \sim N(0, 1)$. We below show that our Corollary 1.1 can recover their result. First, since $f = x^2$ is an even function, it implies that the last term of (1.8) is exactly zero. Therefore, the mean in Theorem 3.4 of Birke and Dette [9] is the same as (1.9), which equals one. Second, the variance in Theorem 3.4 of Birke and Dette [9] is also consistent with (1.5). In fact, the variance of Birke and Dette [9] equals 4 when taking their parameter $y = 0$. On the other hand, since $X_{ij} \sim N(0, 1)$, we have $\nu_4 = 3$ and the first term of (1.5) is zero. Furthermore, by a direct evaluation, we obtain that

$$\Psi_1(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} 4\cos^3\theta\,\mathrm{d}\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi}(\cos 3\theta + 3\cos\theta)\,\mathrm{d}\theta = 0,$$

$$\Psi_2(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} 4\cos^2\theta\cos 2\theta\,\mathrm{d}\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi}(\cos 4\theta + 1 + 2\cos 2\theta)\,\mathrm{d}\theta = 1,$$

$$\Psi_k(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} 4\cos^2\theta \cos k\theta \,\mathrm{d}\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} 2(\cos 2\theta + 1)\cos k\theta \,\mathrm{d}\theta$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \big(\cos(k-2)\theta + \cos(k+2)\theta + 2\cos k\theta\big)\,\mathrm{d}\theta = 0, \qquad \text{for } k \geq 3.$$

It implies that $\operatorname{cov}(X(x^2), X(x^2)) = 4$, which equals the variance of Birke and Dette [9].

The main contribution of this paper is summarized as follows. We have established the central limit theorems of linear spectral statistics of the eigenvalues of the normalized sample covariance matrices when both the dimension and the sample size go to infinity with the dimension dominating the sample size (for the case $p/n \to \infty$). Theorem 1.1 and Corollary 1.1 are both applicable to the data with the dimension dominating the sample size while Corollary 1.1 provides a simplified correction term (hence, CLT) in the ultrahigh dimension cases ($n^3/p = \mathrm{O}(1)$). Such an asymptotic theory complements the results of Bai and Silverstein [5] and Pan [23] for the case $p/n \to c \in (0, \infty)$ and Bai and Yao [6] for Wigner matrix.

This paper is organized as follows. Section 2 provides a calibration of the mean correction term in (1.3), runs simulations to check the accuracy of the calibrated CLTs in Theorem 1.1, and considers a statistical application of Theorem 1.1 and a real data analysis. Section 3 gives the strategy of proving Theorem 1.1 and two intermediate results, Propositions 3.1 and 3.2, and truncation steps of the underlying random variables are given as well. Some preliminary results are given in Section 4. Sections 5 and 6 are devoted to the proof of Proposition 3.1. We present the proof of Proposition 3.2 in Section 7. Section 8 derives mean and covariance in Theorem 1.1.

2. Calibration, application and empirical studies

Section 2.1 considers a calibration to the mean correction term of (1.3). A statistical application is performed in Section 2.2 and the empirical studies are carried out in Section 2.3.

2.1. Calibration of the mean correction term in (1.3)

Theorem 1.1 provides a CLT for $G_n(f)$ under the general framework $p/n \to \infty$, which only requires zero mean, unit variance and the bounded fourth moment. However, the simulation results show that the asymptotic distributions of $G_n(f)$, especially the asymptotic means, are sensitive to the skewness and the kurtosis of the random variables for some particular functions f, for example, $f(x) = \frac{1}{2}x(x^2 - 3)$. This phenomenon is caused by the slow convergence rate of $EG_n(f)$ to zero, which is illustrated as follows. Suppose that $EX_{11}^8 < \infty$. We then have $|EG_n(f)| = \mathrm{O}(\sqrt{n/p}) + \mathrm{O}(1/\sqrt{n})$ by the arguments in Section 6. Also, the remaining terms (see (2.1) below) have a coefficient $(\nu_4 - 1)\sqrt{n/p}$ which converges to zero theoretically since $p/n \to \infty$. However, if $n = 100$, $p = n^2$ and the variables X_{ij} are from central exp(1) then $(\nu_4 - 1)\sqrt{n/p}$ could be as big as 0.8.

In view of this, we will regain such terms and give a calibration for the mean correction term in (1.3). From Section 6, we observe that the convergence rate of $|EG_n(f)|$ relies on the rate of

$|nE\omega_n - m^3(z)(m'(z) + \nu_4 - 2)|$ in Lemma 6.1. By the arguments in Section 6, only S_{22} below (6.13) has the coefficient $(\nu_4 - 1)\sqrt{n/p}$. A simply calculation implies that

$$S_{22} = -2(\nu_4 - 1)\sqrt{n/p}m(z) + \mathrm{o}(1). \tag{2.1}$$

Hence, the limit of $nE\omega_n$ is calibrated as

$$nE\omega_n = m^3(z)\big[\nu_4 - 2 + m'(z) - 2(\nu_4 - 1)\sqrt{n/p}m(z)\big] + \mathrm{o}(1). \tag{2.2}$$

We then calibrate $G_n(f)$ as

$$\begin{aligned} G_n^{\mathrm{Calib}}(f) \triangleq{}& n\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}\big(F^{\mathbf{A}}(x) - F(x)\big) \\ &- \frac{n}{2\pi\mathrm{i}}\oint_{|m|=\rho} f\big(-m - m^{-1}\big)\mathcal{X}_n^{\mathrm{Calib}}(m)\frac{1 - m^2}{m^2}\,\mathrm{d}m, \end{aligned} \tag{2.3}$$

where, via (2.2),

$$\begin{aligned} \mathcal{X}_n^{\mathrm{Calib}}(m) &\triangleq \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}^{\mathrm{Calib}}}}{2\mathcal{A}}, \\ \mathcal{C}^{\mathrm{Calib}} &= \frac{m^3}{n}\left[\nu_4 - 2 + \frac{m^2}{1 - m^2} - 2(\nu_4 - 1)m\sqrt{n/p}\right] - \sqrt{\frac{n}{p}}m^4, \end{aligned} \tag{2.4}$$

$\mathcal{A}, \mathcal{B}$ are defined in (1.4) and $\sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}^{\mathrm{Calib}}}$ is a complex number whose imaginary part has the same sign as that of $\mathcal{B}$. Theorem 1.1 still holds if we replace $G_n(f)$ with $G_n^{\mathrm{Calib}}(f)$.

We next perform a simulation study to check the accuracy of the CLT in Theorem 1.1 with $G_n(f)$ replaced by the calibrated expression $G_n^{\mathrm{Calib}}(f)$ in (2.3). Two combinations of (p, n), $p = n^2, n^{2.5}$, and the test function $f(x) = \frac{1}{2}x(x^2 - 3)$ are considered in the simulations, as suggested by one of the referees. To inspect the impact of the skewness and the kurtosis of the variables, we use three types of random variables, $N(0, 1)$, central exp(1) and central $t(6)$. The skewnesses of these variables are 0, 2 and 0 while the fourth moments of these variables are 3, 9 and 6, respectively. The empirical means and empirical standard deviations of $G_n^{\mathrm{Calib}}(f)/(\mathrm{Var}(Y(f)))^{1/2}$ from 1000 independent replications are shown in Table 1.

It is observed from Table 1 that both the empirical means and standard deviations for $N(0, 1)$ random variables are very accurate. The empirical means for central exp(1) and central $t(6)$ also show their good accuracy. We note that the standard deviations for central exp(1) and central $t(6)$ random variables are not good when n is small (e.g., $n = 50$). But it gradually tends to 1 as the sample size n increases.

Q–Q plots are employed to illustrate the accuracy of the normal approximation in Figures 1 and 2 corresponding to the scenarios $p = n^2$ and $p = n^{2.5}$, respectively. In each figure, Q–Q plots from left to right correspond to $n = 50, 100, 150, 200$, respectively with random variables generated from $N(0, 1)$ ($\triangledown$), central exp(1) ($\triangle$) and central $t(6)$ (+). We observe the same phenomenon that the normal approximation is very accurate for normal variables while the approximation is gradually better when n increases for central exp(1) and $t(6)$ variables.

Table 1. Empirical means of $G_n^{\text{Calib}}(f)/(\text{Var}(Y(f)))^{1/2}$ (cf. (2.3)) for the function $f(x) = \frac{1}{2}x(x^2 - 3)$ with the corresponding standard deviations in the parentheses

n	50	100	150	200
	$p = n^2$			
$N(0, 1)$	−0.314 (1.227)	−0.221 (1.038)	−0.188 (1.051)	−0.093 (0.940)
exp(1)	−0.088 (2.476)	−0.079 (1.447)	−0.140 (1.400)	−0.161 (1.154)
$t(6)$	−0.084 (2.813)	−0.077 (1.541)	−0.095 (1.246)	−0.0897 (1.104)
	$p = n^{2.5}$			
$N(0, 1)$	−0.068 (1.049)	−0.053 (1.077)	−0.0476 (0.944)	−0.016 (1.045)
exp(1)	−0.049 (1.879)	−0.029 (1.390)	−0.046 (1.162)	−0.045 (1.156)
$t(6)$	−0.075 (1.693)	0.050 (1.252)	−0.044 (1.145)	−0.027 (1.044)

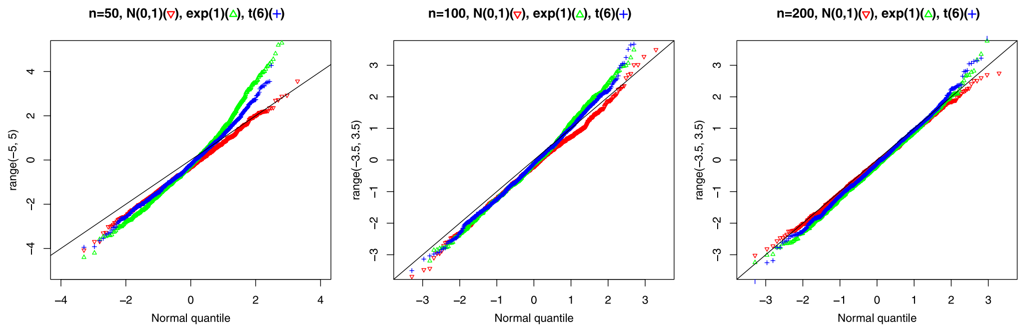

Figure 1. The Q–Q plots of the standard Gaussian distribution versus $G_n^{\text{Calib}}(f)/(\text{Var}(Y(f)))^{1/2}$ based on the sample generating from $N(0, 1)$ ($\triangledown$), standardized exp(1) ($\triangle$) and standardized $t(6)$ (+) with the sample sizes $n = 50, 100, 200$ from left to right and the dimension $p = n^2$.

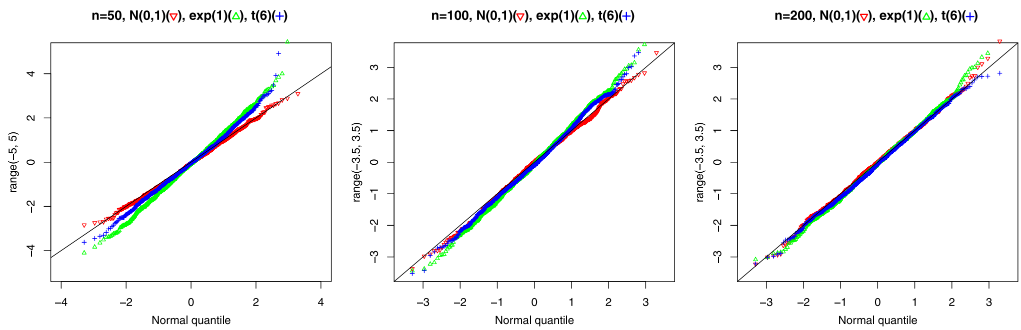

Figure 2. The Q–Q plots of the standard Gaussian distribution versus $G_n^{\text{Calib}}(f)/(\text{Var}(Y(f)))^{1/2}$ based on the sample generating from $N(0, 1)$ ($\triangledown$), standardized exp(1) ($\triangle$) and standardized $t(6)$ (+) with the sample sizes $n = 50, 100, 200$ from left to right and the dimension $p = n^{2.5}$.

2.2. Application of CLTs to hypothesis test

This subsection is to consider an application of Theorem 1.1 which is about hypothesis testing for the covariance matrix. Suppose that $\mathbf{y} = \Gamma\mathbf{s}$ is a p-dimensional vector where Γ is a $p \times p$ matrix with positive eigenvalues and the entries of $\mathbf{s}$ are i.i.d. random variables with mean zero and variance one. Hence, the covariance matrix of $\mathbf{y}$ is $\Sigma = \Gamma\Gamma^T$. Suppose that one wishes to test the hypothesis

$$H_0 : \Sigma = \mathbf{I}_p, \qquad H_1 : \Sigma \neq \mathbf{I}_p. \tag{2.5}$$

Based on the i.i.d. samples $\mathbf{y}_1, \dots, \mathbf{y}_n$ (from $\mathbf{y}$), many authors have considered (2.5) in terms of the relationship of p and n. For example, John [19] and Nagao [22] considered the fixed-dimensional case; Ledoit and Wolf [21], Fisher *et al.* [16] and Bai *et al.* [2] studied the case of $\frac{p}{n} \to c \in (0, \infty)$; Srivastava [26], Srivastava, Kollo and von Rosen [27], Fisher [15] and Chen *et al.* [11] proposed the testing statistics which can accommodate large p and small n.

We are interesting in testing (2.5) in the setting of $\frac{p}{n} \to \infty$. As in Ledoit and Wolf [21] and Birke and Dette [9], we set $f = x^2$. We then propose the following test statistic for the hypothesis of (2.5):

$$\begin{aligned} L_n = \frac{1}{2}\biggl[& n\biggl(\int x^2 \, \mathrm{d}F^{\mathbf{B}}(x) - \int x^2 \, \mathrm{d}F(x)\biggr) \\ & - \biggl(\frac{n}{2\pi \mathrm{i}} \oint_{|m|=\rho} \bigl(m + m^{-1}\bigr)^2 \mathcal{X}_n^{\mathrm{Calib}}(m) \frac{1-m^2}{m^2} \, \mathrm{d}m\biggr)\biggr], \end{aligned} \tag{2.6}$$

where $\mathcal{X}_n^{\mathrm{Calib}}(m)$ is given in (2.4) and $\mathbf{B} = \sqrt{\frac{p}{n}}(\frac{1}{p}\mathbf{Y}^T\mathbf{Y} - \mathbf{I}_n)$ is the normalized sample covariance matrix with $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$. The asymptotic mean and variance of L_n are 0 and 1, respectively, see Theorem 1.1 or Remark 1.3 for details. Since there is no close form for the mean correction term in (2.6), we use *Matlab* to calculate this correction term. It shows that as $n/p \to 0$,

$$\frac{n}{2\pi \mathrm{i}} \oint_{|m|=\rho} \bigl(m + m^{-1}\bigr)^2 \mathcal{X}_n^{\mathrm{Calib}}(m) \frac{1-m^2}{m^2} \, \mathrm{d}m = \nu_4 - 2.$$

We also note the fact that

$$E\biggl[n \int x^2 \, \mathrm{d}\bigl(F^{\mathbf{B}}(x) - F(x)\bigr)\biggr] = E\bigl[\operatorname{tr} \mathbf{B}\mathbf{B}^T - n\bigr] = \nu_4 - 2.$$

Thus, we use the following test statistic in the simulations:

$$L_n = \frac{1}{2}\biggl[n\biggl(\int x^2 \, \mathrm{d}F^{\mathbf{B}}(x) - \int x^2 \, \mathrm{d}F(x)\biggr) - (\nu_4 - 2)\biggr] = \frac{1}{2}\bigl(\operatorname{tr} \mathbf{B}\mathbf{B}^T - n - (\nu_4 - 2)\bigr). \tag{2.7}$$

Since $\Gamma^T\Gamma = \mathbf{I}_p$ is equivalent to $\Gamma\Gamma^T = \mathbf{I}_p$, under the null hypothesis H_0 in (2.5), we have

$$L_n \xrightarrow{d} N(0, 1). \tag{2.8}$$

By the law of large numbers, a consistent estimator of ν_4 is $\widehat{\nu}_4 = \frac{1}{np}\sum_{i,j} Y_{ij}^4$ under the null hypothesis H_0. By Slutsky's theorem, (2.8) also holds if we replace ν_4 of (2.7) with $\widehat{\nu}_4$.

The numerical performance of the proposed statistic L_n is carried out by Monte Carlo simulations. Let $Z_{\alpha/2}$ and $Z_{1-\alpha/2}$, respectively, be the $100\alpha/2\%$ and $100(1-\alpha/2)\%$ quantiles of the asymptotic null distribution of the test statistic L_n. With T replications of the data set simulated under the null hypothesis, we calculate the empirical size as

$$\hat{\alpha} = \frac{\{\#L_n^{\text{null}} \le Z_{\alpha/2}\} + \{\#L_n^{\text{null}} > Z_{1-\alpha/2}\}}{T},$$

where # denotes the number and L_n^{null} represents the values of the test statistic L_n based on the data set simulated under the null hypothesis. The empirical power is calculated by

$$\hat{\beta} = \frac{\{\#L_n^{\text{alter}} \le Z_{\alpha/2}\} + \{\#L_n^{\text{alter}} > Z_{1-\alpha/2}\}}{T},$$

where L_n^{alter} represents the values of the test statistic L_n based on the data set simulated under the alternative hypothesis. In our simulations, we fix $T = 1000$ as the number of replications and set the nominal significance level $\alpha = 5\%$. By asymptotic normality, we have $Z_{\alpha/2} = -1.96$ and $Z_{1-\alpha/2} = 1.96$.

Our proposed test is intended for the situation "large p, small n". To inspect the impact caused by the sample size and/or the dimension, we set

$$n = 20, 40, 60, 80,$$
$$p = 600, 1500, 3000, 5500, 8000, 10\,000.$$

The entries of **s** are generated from three types of distributions, Gaussian distribution, standardized Gamma(4, 0.5) and Bernoulli distribution with $P(X_{ij} = \pm 1) = 0.5$.

The following two types of covariance matrices are considered in the simulations to investigate the empirical power of the test.

1. (Diagonal covariance.) $\Sigma = \operatorname{diag}(\sqrt{2}\mathbf{1}_{[\nu p]}, \mathbf{1}_{1-[\nu p]})$, where $\nu = 0.08$ or $\nu = 0.25$, $[a]$ denotes the largest integer that is not greater than a.
2. (Banded covariance.) $\Sigma = \operatorname{diag}(A_1, \operatorname{diag}(\mathbf{1}_{p-[v_2 p]}))$, where A_1 is a $[v_2 p] \times [v_2 p]$ tridiagonal symmetric matrix with the diagonal elements being equal to 1 and elements below and above the diagonal all being equal to v_1.

Since the test in Chen *et al.* [11] accommodates a wider class of variates and has less restrictions on the ratio p/n, we below compare performance of our test with that of Chen *et al.* [11]. To simplify the notation, denote their test by the CZZ test. Table 2 reports empirical sizes of the proposed test and of the CZZ test for the preceding three distributions. We observe from Table 2 that the sizes of both tests are roughly the same, when the underlying variables are normally or Bernoulli distributed. It seems that the CZZ test looks better for skewed data, for example, gamma distribution. We believe additional corrections such as the Edgeworth expansion will be helpful, which is beyond the scope of this paper. However, our test still performs well for skewed data if $p \gg n$.

Table 2. Empirical sizes of CZZ test and L_n at the significant level $\alpha = 5\%$ for normal, gamma, Bernoulli random vectors

	CZZ test n				L_n n			
p	20	40	60	80	20	40	60	80
	Normal random vectors							
600	0.069	0.071	0.052	0.052	0.063	0.077	0.066	0.082
1500	0.057	0.059	0.061	0.059	0.055	0.058	0.058	0.062
3000	0.067	0.068	0.057	0.053	0.048	0.067	0.056	0.052
5500	0.064	0.06	0.067	0.058	0.054	0.055	0.071	0.068
8000	0.071	0.062	0.062	0.054	0.055	0.049	0.06	0.059
10 000	0.055	0.059	0.063	0.06	0.037	0.058	0.057	0.054
	Gamma random vectors							
600	0.055	0.073	0.056	0.062	0.103	0.119	0.125	0.123
1500	0.064	0.047	0.059	0.059	0.094	0.072	0.072	0.088
3000	0.069	0.071	0.059	0.052	0.066	0.074	0.071	0.061
5500	0.065	0.069	0.048	0.041	0.077	0.073	0.047	0.045
8000	0.069	0.065	0.07	0.053	0.078	0.075	0.063	0.059
10 000	0.072	0.06	0.06	0.057	0.078	0.082	0.065	0.06
	Bernoulli random vectors							
600	0.078	0.079	0.056	0.037	0.048	0.064	0.046	0.037
1500	0.065	0.050	0.051	0.053	0.039	0.040	0.049	0.050
3000	0.048	0.053	0.058	0.060	0.040	0.052	0.052	0.056
5500	0.059	0.061	0.059	0.042	0.040	0.052	0.060	0.040
8000	0.065	0.074	0.065	0.059	0.046	0.052	0.05	0.051
10 000	0.07	0.057	0.047	0.048	0.044	0.037	0.038	0.047

Table 3 to Table 5 summarize the empirical powers of the proposed tests as well as those of the CZZ test for both the diagonal and the banded covariance matrix. Table 3 assumes the underlying variables are normally distributed while Tables 4 and 5 assume the central gamma and the central bernoulli random variables, respectively. For the diagonal covariance matrix, we observe that the proposed test consistently outperforms the CZZ test for all types of distributions, especially for "small" n. For example, when $n = 20$, even $n = 40, 60, 80$ for $\nu = 0.08$, the CZZ test results in power ranging from 0.2–0.8, while our test still gains very satisfying power exceeding 0.932.

For the banded covariance matrix, we observe an interesting phenomenon. Our test seems to be more sensitive to the dimension p. When $p = 600, 1500, 3000$, the power of our test is not that good for small υ_2 $(= 0.4)$. Fortunately, when $p = 5500, 8000, 10\,000$, the performance is much better, where the power is one or close to one. Similar results are also observed for $\upsilon_2 = 0.8$. We also note that large υ_2 outperforms smaller υ_2 because when υ_2 becomes larger, the corresponding covariance matrix becomes more "different" from the identity matrix. As for the CZZ test, its power is mainly affected by n. But generally speaking, our test gains better power than the CZZ test for extremely larger p and small n.

Table 3. Empirical powers of CZZ test and L_n at the significant level $\alpha = 5\%$ for normal random vectors. Two types of population covariance matrices are considered. In the first case, $\Sigma_1 = \mathrm{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$, respectively. In the second case, $\Sigma_2 = \mathrm{diag}(A_1, \mathrm{diag}(\mathbf{1}_{p-[v_2 p]}))$, where A_1 is a $[v_2 p] \times [v_2 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5$, $v_2 = 0.8$ and $v_1 = 0.5$, $v_2 = 0.4$, respectively

	CZZ test n				L_n n			
p	20	40	60	80	20	40	60	80
	Normal random vectors ($\nu = 0.08$)							
600	0.186	0.392	0.648	0.826	0.932	1	1	1
1500	0.179	0.397	0.642	0.822	0.999	1	1	1
3000	0.197	0.374	0.615	0.867	1.000	1	1	1
5500	0.225	0.382	0.615	0.85	1	1	1	1
8000	0.203	0.391	0.638	0.843	1	1	1	1
10 000	0.204	0.381	0.639	0.835	1	1	1	1
	Normal random vectors ($\nu = 0.25$)							
600	0.571	0.952	0.997	1	1	1	1	1
1500	0.585	0.959	1.000	1	1	1	1	1
3000	0.594	0.961	1.000	1	1	1	1	1
5500	0.617	0.954	1	1	1	1	1	1
8000	0.607	0.957	0.999	1	1	1	1	1
10 000	0.595	0.949	1	1	1	1	1	1
	Normal random vectors ($v_1 = 0.5$, $v_2 = 0.8$)							
600	0.333	0.874	0.997	1	0.443	0.493	0.492	0.488
1500	0.310	0.901	0.999	1	0.987	0.997	0.997	0.998
3000	0.348	0.889	0.998	1	1.000	1.000	1.000	1.000
5500	0.382	0.871	0.998	1	1	1	1	1
8000	0.33	0.867	0.998	1	1	1	1	1
10 000	0.359	0.868	0.998	1	1	1	1	1
	Normal random vectors ($v_1 = 0.5$, $v_2 = 0.4$)							
600	0.142	0.364	0.668	0.896	0.078	0.089	0.069	0.102
1500	0.131	0.354	0.653	0.890	0.220	0.235	0.230	0.226
3000	0.139	0.361	0.662	0.899	0.635	0.660	0.647	0.684
5500	0.148	0.352	0.645	0.898	0.97	0.979	0.989	0.989
8000	0.152	0.36	0.688	0.905	0.981	0.978	0.986	0.989
10 000	0.137	0.328	0.674	0.886	1	1	1	1

2.3. Empirical studies

As empirical applications, we consider two classic datasets: the colon data of Alon *et al.* [1] and the leukemia data of Golub *et al.* [17]. Both datasets are publicly available on the web site of Tatsuya Kubokawa: http://www.tatsuya.e.u-tokyo.ac.jp/. Such data were used in Fisher [15] as well. The sample sizes and dimensions (n, p) of the colon data and the leukemia data

Table 4. Empirical powers of CZZ test and L_n at the significant level $\alpha = 5\%$ for standardized gamma random vectors. Two types of population covariance matrices are considered. In the first case, $\Sigma_1 = \operatorname{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$, respectively. In the second case, $\Sigma_2 = \operatorname{diag}(A_1, \operatorname{diag}(\mathbf{1}_{p-[v_2 p]}))$, where A_1 is a $[v_2 p] \times [v_2 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5$, $v_2 = 0.8$ and $v_1 = 0.5$, $v_2 = 0.4$, respectively

	CZZ test n				L_n n			
p	20	40	60	80	20	40	60	80
	Gamma random vectors ($\nu = 0.08$)							
600	0.331	0.638	0.891	0.982	0.999	1	1	1
1500	0.356	0.636	0.901	0.979	1	1	1	1
3000	0.197	0.383	0.638	0.823	1	1	1	1
5500	0.178	0.361	0.658	0.845	1	1	1	1
8000	0.199	0.399	0.642	0.85	1	1	1	1
10 000	0.216	0.353	0.636	0.843	1	1	1	1
	Gamma random vectors ($\nu = 0.25$)							
600	0.621	0.943	1.000	1	1	1	1	1
1500	0.610	0.946	0.999	1	1	1	1	1
3000	0.579	0.946	0.997	1	1	1	1	1
5500	0.596	0.957	0.999	1	1	1	1	1
8000	0.616	0.962	0.999	1	1	1	1	1
10 000	0.614	0.955	0.999	1	1	1	1	1
	Gamma random vectors ($v_1 = 0.5$, $v_2 = 0.8$)							
600	0.192	0.871	0.998	0.972	0.122	0.413	0.423	0.133
1500	0.198	0.883	0.995	0.980	0.440	0.992	0.993	0.433
3000	0.343	0.885	0.995	1	1	1	1	1
5500	0.342	0.88	0.996	1	1	1	1	1
8000	0.349	0.877	0.998	1	1	1	1	1
10 000	0.337	0.879	0.998	1	1	1	1	1
	Gamma random vectors ($v_1 = 0.5$, $v_2 = 0.4$)							
600	0.117	0.353	0.650	0.780	0.087	0.111	0.114	0.120
1500	0.138	0.365	0.661	0.799	0.183	0.215	0.226	0.157
3000	0.129	0.349	0.646	0.89	0.593	0.621	0.627	0.61
5500	0.124	0.335	0.678	0.889	0.945	0.972	0.981	0.986
8000	0.142	0.369	0.668	0.901	0.999	1	1	1
10 000	0.142	0.336	0.668	1	1	1	1	1

are $(62, 2000)$ and $(72, 3571)$, respectively. Simulations show that these two datasets have zero mean (10^{-8} to 10^{-11}) and unit variance. Therefore, we consider the hypothesis test in (2.5) by using the test statistic L_n in (2.7). The computed values are $L_n = 33\,933.7$ for the colon data and $L_n = 60\,956$ for the leukemia data. It is also interesting to note that the statistic values of Fisher [15] are 6062.642 for the colon data and 6955.651 for the leukemia data when testing

Table 5. Empirical powers of CZZ test and L_n at the significant level $\alpha = 5\%$ for standardized Bernoulli random vectors. Two types of population covariance matrices are considered. In the first case, $\Sigma_1 = \mathrm{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$, respectively. In the second case, $\Sigma_2 = \mathrm{diag}(A_1, \mathrm{diag}(\mathbf{1}_{p-[v_2 p]}))$, where A_1 is a $[v_2 p] \times [v_2 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5$, $v_2 = 0.8$ and $v_1 = 0.5$, $v_2 = 0.4$, respectively

	CZZ test n				L_n n			
p	20	40	60	80	20	40	60	80
	Bernoulli random vectors ($\nu = 0.08$)							
600	0.216	0.381	0.622	0.849	0.972	1	1	1
1500	0.198	0.401	0.632	0.837	1	1	1	1
3000	0.203	0.362	0.622	0.823	1	1	1	1
5500	0.196	0.354	0.627	0.829	1	1	1	1
8000	0.203	0.373	0.638	0.834	1	1	1	1
10 000	0.213	0.397	0.637	0.822	1	1	1	1
	Bernoulli random vectors ($\nu = 0.25$)							
600	0.594	0.952	0.998	1	1	1	1	1
1500	0.619	0.960	1.000	1	1	1	1	1
3000	0.594	0.964	0.999	1	1	1	1	1
5500	0.609	0.948	1.000	1	1	1	1	1
8000	0.589	0.952	1	1	1	1	1	1
10 000	0.603	0.957	0.999	1	1	1	1	1
	Bernoulli random vectors ($v_1 = 0.5$, $v_2 = 0.8$)							
600	0.356	0.870	0.996	1	0.507	0.512	0.526	0.558
1500	0.359	0.892	0.995	1	0.999	1	1	0.999
3000	0.343	0.877	0.998	1	1.000	1	1	1.000
5500	0.355	0.868	0.997	1	1.000	1.000	1.000	1.000
8000	0.332	0.873	0.997	1	1	1	1	1
10 000	0.353	0.872	1	1	1	1	1	1
	Bernoulli random vectors ($v_1 = 0.5$, $v_2 = 0.4$)							
600	0.153	0.348	0.643	0.901	0.092	0.086	0.079	0.085
1500	0.154	0.372	0.643	0.878	0.239	0.255	0.235	0.241
3000	0.141	0.339	0.649	0.882	0.682	0.680	0.680	0.674
5500	0.156	0.343	0.656	0.893	0.997	0.994	0.994	0.994
8000	0.144	0.353	0.664	0.904	1	1	1	1
10 000	0.139	0.356	0.685	0.889	1	1	1	1

the identity hypothesis. Also, the statistics of Fisher [15] and L_n in (2.7) are both asymptotic normality (standard normal). As in Fisher [15], we conclude that p-values of the test statistics are zero which shows evidence to reject the null hypothesis. This is consistent with Fisher's [15] conclusion for these two datasets.

3. Truncation and strategy for the proof of Theorem 1.1

In the rest of the paper, we use K to denote a constant which may take different values at different places. The notation $o_{L_p}(1)$ stands for a term converging to zero in L_p norm; $\xrightarrow{\text{a.s.}}$ means "convergence almost surely to"; $\xrightarrow{\text{i.p.}}$ means "convergence in probability to".

3.1. Truncation

In this section, we truncate the underlying random variables as in Pan and Gao [24]. Choose δ_n satisfying

$$\lim_{n\to\infty} \delta_n^{-4} E|X_{11}|^4 I\bigl(|X_{11}| > \delta_n \sqrt[4]{np}\bigr) = 0, \qquad \delta_n \downarrow 0, \delta_n \sqrt[4]{np} \uparrow \infty. \tag{3.1}$$

In what follows, we will use δ to represent δ_n for convenience. We first truncate the variables $\hat{X}_{ij} = X_{ij} I(|X_{ij}| < \delta \sqrt[4]{np})$ and then normalize it as $\tilde{X}_{ij} = (\hat{X}_{ij} - E\hat{X}_{ij})/\sigma$, where σ is the standard deviation of $\hat{X}_{ij}$. Let $\hat{\mathbf{X}} = (\hat{X}_{ij})$ and $\tilde{\mathbf{X}} = (\tilde{X}_{ij})$. Define $\hat{\mathbf{A}}, \tilde{\mathbf{A}}$ and $\hat{G}_n(f), \tilde{G}_n(f)$ similarly by means of (1.1) and (1.3), respectively. We then have

$$P(\mathbf{A} \neq \hat{\mathbf{A}}) \le npP\bigl(|X_{11}| \ge \delta \sqrt[4]{np}\bigr) \le K\delta^{-4} E|X_{11}|^4 I\bigl(|X_{11}| > \delta \sqrt[4]{np}\bigr) = o(1).$$

It follows from (3.1) that

$$\begin{aligned}
\bigl|1 - \sigma^2\bigr| &\le 2\bigl|EX_{11}^2 I\bigl(|X_{11} > \delta \sqrt[4]{np}|\bigr)\bigr| \\
&\le 2(np)^{-1/2} \delta^{-2} E|X_{11}|^4 I\bigl(|X_{11}| > \delta \sqrt[4]{np}\bigr) = o\bigl((np)^{-1/2}\bigr)
\end{aligned}$$

and

$$|E\hat{X}_{11}| \le \delta^{-3} (np)^{-3/4} E|X_{11}|^4 I\bigl(|X_{11}| > \delta \sqrt[4]{np}\bigr) = o\bigl((np)^{-3/4}\bigr).$$

Therefore

$$\begin{aligned}
E \operatorname{tr}(\tilde{\mathbf{X}} - \hat{\mathbf{X}})^T (\tilde{\mathbf{X}} - \hat{\mathbf{X}}) &\le \sum_{i,j} E|\hat{X}_{ij} - \tilde{X}_{ij}|^2 \\
&\le Kpn\biggl(\frac{(1-\sigma)^2}{\sigma^2} E|\hat{X}_{11}|^2 + \frac{1}{\sigma^2}|E\hat{X}_{ij}|^2\biggr) = o(1)
\end{aligned}$$

and

$$E \operatorname{tr} \hat{\mathbf{X}}^T \hat{\mathbf{X}} \le \sum_{i,j} E|\hat{X}_{ij}|^2 \le Knp, \qquad E \operatorname{tr} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \le \sum_{i,j} E|\tilde{X}_{ij}|^2 \le Knp.$$

Recalling that the notation $\lambda_j(\cdot)$ represents the jth largest eigenvalue, we then have $\lambda_j(\mathbf{X}^T\mathbf{X}) = \sqrt{np}\lambda_j(\mathbf{A}) + p$. Similar equalities also hold if $\mathbf{X}, \mathbf{A}$ are replaced by $\hat{\mathbf{X}}, \hat{\mathbf{A}}$ or $\tilde{\mathbf{X}}, \tilde{\mathbf{A}}$.

Consequently, applying the argument used in Theorem 11.36 in Bai and Silverstein [3] and Cauchy–Schwarz's inequality, we have

$$\begin{aligned}
E\big|\tilde{G}_n(f)-\hat{G}_n(f)\big| &\le \sum_{j=1}^{n} E\big|f\big(\lambda_j(\hat{\mathbf{A}})\big)-f\big(\lambda_j(\tilde{\mathbf{A}})\big)\big| \\
&\le K_f\sum_{j=1}^{n} E\big|\lambda_j(\hat{\mathbf{A}})-\lambda_j(\tilde{\mathbf{A}})\big| \le \frac{K_f}{\sqrt{np}}\sum_{j=1}^{n} E\big|\lambda_j\big(\hat{\mathbf{X}}^T\hat{\mathbf{X}}\big)-\lambda_j\big(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\big)\big| \\
&\le \frac{K_f}{\sqrt{np}} E\big[\operatorname{tr}(\tilde{\mathbf{X}}-\hat{\mathbf{X}})^T(\tilde{\mathbf{X}}-\hat{\mathbf{X}})\cdot 2\big(\operatorname{tr}\hat{\mathbf{X}}^T\hat{\mathbf{X}}+\operatorname{tr}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\big)\big]^{1/2} \\
&\le \frac{2K_f}{\sqrt{np}}\big[E\operatorname{tr}(\tilde{\mathbf{X}}-\hat{\mathbf{X}})^T(\tilde{\mathbf{X}}-\hat{\mathbf{X}})\cdot\big(E\operatorname{tr}\hat{\mathbf{X}}^T\hat{\mathbf{X}}+E\operatorname{tr}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\big)\big]^{1/2} \\
&= \mathrm{o}(1),
\end{aligned}$$

where K_f is a bound on $|f'(x)|$. Thus, the weak convergence of $G_n(f)$ is not affected if we replace the original variables X_{ij} by the truncated and normalized variables $\tilde{X}_{ij}$. For convenience, we still use X_{ij} to denote $\tilde{X}_{ij}$, which satisfies the following additional assumption (c):

(c) The underlying variables satisfy

$$|X_{ij}|\le\delta\sqrt[4]{np},\qquad EX_{ij}=0,\qquad EX_{ij}^2=1,\qquad EX_{ij}^4=\nu_4+\mathrm{o}(1),$$

where $\delta=\delta_n$ satisfies $\lim_{n\to\infty}\delta_n^{-4}E|X_{11}|^4I(|X_{11}|>\delta_n\sqrt[4]{np})=0$, $\delta_n\downarrow 0$, and $\delta_n\sqrt[4]{np}\uparrow\infty$.

For any $\varepsilon>0$, define the event $F_n(\varepsilon)=\{\max_{j\le n}|\lambda_j(\mathbf{A})|\ge 2+\varepsilon\}$ where $\mathbf{A}$ is defined by the truncated and normalized variables satisfying assumption (c). By Theorem 2 in Chen and Pan [10], for any $\ell>0$

$$P\big(F_n(\varepsilon)\big)=\mathrm{o}\big(n^{-\ell}\big). \tag{3.2}$$

Here we would point out that the result regarding the minimum eigenvalue of $\mathbf{A}$ can be obtained similarly by investigating the maximum eigenvalue of $-\mathbf{A}$.

3.2. Strategy of the proof

We shall follow the strategy of Bai and Yao [6]. Specifically speaking, assume that u_0, v are fixed and sufficiently small so that $\varsigma\subset\mathcal{S}$ (see the definition in the introduction), where ς is the contour formed by the boundary of the rectangle with $(\pm u_0,\pm \mathrm{i}v)$ where $u_0>2, 0<v\le 1$. By Cauchy's integral formula, with probability one,

$$G_n(f)=-\frac{1}{2\pi\mathrm{i}}\oint_{\varsigma} f(z)n\big[m_n(z)-m(z)-\mathcal{X}_n\big(m(z)\big)\big]\,\mathrm{d}z,$$

where $m_n(z), m(z)$ denote the Stieltjes transform of $F^{\mathbf{A}}(x)$ and $F(x)$, respectively.

Let

$$M_n(z) = n\big[m_n(z) - m(z) - \mathcal{X}_n\big(m(z)\big)\big], \qquad z \in \varsigma.$$

For $z \in \varsigma$, write $M_n(z) = M_n^{(1)}(z) + M_n^{(2)}(z)$ where

$$M_n^{(1)}(z) = n\big[m_n(z) - Em_n(z)\big], \qquad M_n^{(2)}(z) = n\big[Em_n(z) - m(z) - \mathcal{X}_n\big(m(z)\big)\big].$$

Split the contour ς as the union of $\varsigma_u, \varsigma_l, \varsigma_r, \varsigma_0$ where $\varsigma_l = \{z = -u_0 + \mathrm{i}v, \xi_n n^{-1} < |v| < v_1\}$, $\varsigma_r = \{z = u_0 + \mathrm{i}v, \xi_n n^{-1} < |v| < v_1\}$, $\varsigma_0 = \{z = \pm u_0 + \mathrm{i}v, |v| \le \xi_n n^{-1}\}$ and $\varsigma_u = \{z = u \pm \mathrm{i}v_1, |u| \le u_0\}$ and where ξ_n is a slowly varying sequence of positive constants and v_1 is a positive constant which is independent of n. Throughout this paper, let $\mathbb{C}_1 = \{z : z = u + \mathrm{i}v, u \in [-u_0, u_0], |v| \ge v_1\}$.

Proposition 3.1. *Under assumptions* (b1), (c), *the empirical process* $\{M_n(z), z \in \mathbb{C}_1\}$ *converges weakly to a Gaussian process* $\{M(z), z \in \mathbb{C}_1\}$ *with the mean function*

$$\Delta(z) = 0 \tag{3.3}$$

and the covariance function

$$\Lambda(z_1, z_2) = m'(z_1)m'(z_2)\big[\nu_4 - 3 + 2\big(1 - m(z_1)m(z_2)\big)^{-2}\big]. \tag{3.4}$$

As in Bai and Yao [6], the process of $\{M(z), z \in \mathbb{C}_1\}$ can be extended to $\{M(z), \Re(z) \notin [-2, 2]\}$ due to the facts that (i) $M(z)$ is symmetric, for example, $M(\bar{z}) = \overline{M(z)}$; (ii) the mean and the covariance function of $M(z)$ are independent of v_1 and they are continuous except for $\Re(z) \notin [-2, 2]$. By Proposition 3.1 and the continuous mapping theorem,

$$-\frac{1}{2\pi \mathrm{i}} \int_{\varsigma_u} f(z) M_n(z)\,\mathrm{d}z \xrightarrow{d} -\frac{1}{2\pi \mathrm{i}} \int_{\varsigma_u} f(z) M(z)\,\mathrm{d}z.$$

Thus, to prove Theorem 1.1, it is also necessary to prove the following proposition.

Proposition 3.2. *Let* $z \in \mathbb{C}_1$. *Under assumptions* (b1), (c), *there exists some event* U_n *with* $P(U_n) \to 0$, *as* $n \to \infty$, *such that*

$$\lim_{v_1 \downarrow 0} \limsup_{n \to \infty} E\left|\int_{\bigcup_{i=l,r,0} \varsigma_i} M_n^{(1)}(z) I\big(U_n^c\big)\,\mathrm{d}z\right|^2 = 0, \tag{3.5}$$

$$\lim_{v_1 \downarrow 0} \limsup_{n \to \infty} \left|\int_{\bigcup_{i=l,r,0} \varsigma_i} E M_n(z) I\big(U_n^c\big)\,\mathrm{d}z\right| = 0 \tag{3.6}$$

and

$$\lim_{v_1 \downarrow 0} E\left|\int_{\varsigma_i} M^{(1)}(z)\,\mathrm{d}z\right|^2 = 0, \qquad \lim_{v_1 \downarrow 0} E\left|\int_{\varsigma_i} M(z)\,\mathrm{d}z\right|^2 = 0. \tag{3.7}$$

Since $E|M^{(1)}(z)|^2 = \Lambda(z, \bar{z})$ and $E|M(z)|^2 = \Lambda(z, \bar{z}) + |EM(z)|^2$, (3.7) can be easily obtained from Proposition 3.1. For $i = 0$, if we choose $U_n = F_n(\varepsilon)$ with the $\varepsilon = (u_0 - 2)/2$, then when U_n^c happens, $\forall z \in \varsigma_0$, we have $|m_n(z)| \leq 2/(u_0 - 2)$ and $|m(z)| \leq 1/(u_0 - 2)$. Thus

$$\left|\int_{\varsigma_0} M_n^{(1)}(z) I\left(U_n^c\right) \mathrm{d}z\right| \leq n\left(\frac{4}{u_0 - 2}\right)^2 \|\varsigma_0\| \leq \frac{4\xi_n}{(u_0 - 2)^2},$$

where $\|\varsigma_0\|$ represents the length of ς_0. Furthermore,

$$\left|\int_{\varsigma_0} M_n(z) I\left(U_n^c\right) \mathrm{d}z\right| \leq n\left(\frac{2}{u_0 - 2} + \frac{1}{u_0 - 2} + K\frac{n}{p}\right)^2 \|\varsigma_0\|.$$

These imply that (3.6) and (3.5) are true for $z \in \varsigma_0$ by noting that $\xi_n \to 0$ as $p \to \infty$.

Sections 5 and 6 are devoted to the proof of Proposition 3.1. The main steps are summarized in the following:

- According to Theorem 8.1 in Billingsley [8], to establish the convergence of the process $\{M_n(z), z \in \mathbb{C}_1\}$, it suffices to prove the finite-dimensional convergence of the random part $M_n^{(1)}(z)$ and its tightness, and the convergence of the non-random part $M_n^{(2)}(z)$.
- For the random part $M_n^{(1)}(z)$, we rewrite it in terms of a martingale expression so that we may apply the central limit theorem of martingales to find its asymptotic mean and covariance.
- For the non-random part $M_n^{(2)}(z)$, by the formula of the inverse of a matrix and the equation satisfied by $m(z)$ we develop an equation for $(Em_n(z) - m(z))$. Based on it, we then find its limit under assumptions $n/p \to 0$ and $n^3/p = \mathrm{O}(1)$ for Theorem 1.1 and Corollary 1.1, respectively.

Section 7 uses Lemma 4.4 below to finish the proofs of (3.5) for $i = l, r$ so that the proof of Proposition 3.2 is completed. Section 8 uses Bai and Yao's [6] asymptotic mean and covariance function to conclude the proof of Theorem 1.1.

4. Preliminary results

This section is to provide simplification of $M_n^{(1)}(z)$ and some useful lemmas needed to prove Proposition 3.1.

4.1. Simplification of $M_n^{(1)}(z)$

The aim of this subsection is to simplify $M_n^{(1)}(z)$ so that $M_n^{(1)}(z)$ can be written in the form of martingales. Some moment bounds are also proved.

Define $\mathbf{D} = \mathbf{A} - z\mathbf{I}_n$. Let $\mathbf{s}_k$ be the kth column of $\mathbf{X}$ and $\mathbf{X}_k$ be a $p \times (n-1)$ matrix constructed from $\mathbf{X}$ by deleting the kth column. We then similarly define $\mathbf{A}_k = \frac{1}{\sqrt{np}}(\mathbf{X}_k^T\mathbf{X}_k - p\mathbf{I}_{n-1})$ and $\mathbf{D}_k = \mathbf{A}_k - z\mathbf{I}_{n-1}$. The kth diagonal element of $\mathbf{D}$ is $a_{kk}^{\mathrm{diag}} = \frac{1}{\sqrt{np}}(\mathbf{s}_k^T\mathbf{s}_k - p) - z$ and the kth row of $\mathbf{D}$ with the kth element deleted is $\mathbf{q}_k^T = \frac{1}{\sqrt{np}}\mathbf{s}_k^T\mathbf{X}_k$. The Stieltjes transform of $F^{\mathbf{A}}$ has the form $m_n(z) = \frac{1}{n}\operatorname{tr}\mathbf{D}^{-1}$. The limiting Stieltjes transform $m(z)$ satisfies

$$m(z) = -\frac{1}{z + m(z)}, \qquad \bigl|m(z)\bigr| \le 1 \tag{4.1}$$

(one may see Bai and Yao [6]).

Define the σ-field $\mathcal{F}_k = \sigma(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k)$ and the conditional expectation $E_k(\cdot) = E(\cdot|\mathcal{F}_k)$. By the matrix inversion formula, we have (see (3.9) of Bai [4])

$$\operatorname{tr}\bigl(\mathbf{D}^{-1} - \mathbf{D}_k^{-1}\bigr) = -\frac{(1 + \mathbf{q}_k^T\mathbf{D}_k^{-2}\mathbf{q}_k)}{-a_{kk}^{\mathrm{diag}} + \mathbf{q}_k^T\mathbf{D}_k^{-1}\mathbf{q}_k}. \tag{4.2}$$

We then obtain

$$M_n^{(1)}(z) = \operatorname{tr}\mathbf{D}^{-1} - E\operatorname{tr}\mathbf{D}^{-1} = \sum_{k=1}^{n}(E_k - E_{k-1})\operatorname{tr}\bigl(\mathbf{D}^{-1} - \mathbf{D}_k^{-1}\bigr) = \sum_{k=1}^{n}\varrho_k \tag{4.3}$$

$$= (E_k - E_{k-1})\iota_k - E_k\kappa_k, \tag{4.4}$$

where

$$\begin{aligned}
\varrho_k &= -(E_k - E_{k-1})\beta_k\bigl(1 + \mathbf{q}_k^T\mathbf{D}_k^{-2}\mathbf{q}_k\bigr),\\
\iota_k &= -\beta_k^{\mathrm{tr}}\beta_k\eta_k\bigl(1 + \mathbf{q}_k^T\mathbf{D}_k^{-2}\mathbf{q}_k\bigr),\\
\eta_k &= \frac{1}{\sqrt{np}}\bigl(\mathbf{s}_k^T\mathbf{s}_k - p\bigr) - \gamma_{k1}, \qquad \beta_k = \frac{1}{-a_{kk}^{\mathrm{diag}} + \mathbf{q}_k^T\mathbf{D}_k^{-1}\mathbf{q}_k},\\
\beta_k^{\mathrm{tr}} &= \frac{1}{z + (1/(np))\operatorname{tr}\mathbf{M}_k^{(1)}}, \qquad \mathbf{M}_k^{(s)} = \mathbf{X}_k\mathbf{D}_k^{-s}\mathbf{X}_k^T, s = 1, 2,\\
\gamma_{ks} &= \mathbf{q}_k^T\mathbf{D}_k^{-s}\mathbf{q}_k - (np)^{-1}\operatorname{tr}\mathbf{M}_k^{(s)}, \qquad \kappa_k = \beta_k^{\mathrm{tr}}\gamma_{k2}.
\end{aligned}$$

In the above equality, ϱ_k is obtained by (4.2) and the last equality uses the facts that

$$\beta_k = \beta_k^{\mathrm{tr}} + \beta_k\beta_k^{\mathrm{tr}}\eta_k \tag{4.5}$$

and

$$(E_k - E_{k-1})\left[\beta_k^{\mathrm{tr}}\left(1 + \frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(2)}\right)\right] = 0, \qquad E_{k-1}\kappa_k = 0.$$

We remind the readers that the variable z has been dropped from the expressions such as $\mathbf{D}^{-1}, \mathbf{D}_k^{-1}, \beta_k, \gamma_{ks}$ and so on. When necessary, we will also indicate them as $\mathbf{D}^{-1}(z), \mathbf{D}_k^{-1}(z)$, $\beta_k(z)$, $\gamma_{ks}(z)$, etc.

We next provide some useful bounds. It follows from the definitions of $\mathbf{D}$ and $\mathbf{D}_k$ that

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{X}^T\mathbf{X} &= p\mathbf{D}^{-1} + \sqrt{np}\big(\mathbf{I}_n + z\mathbf{D}^{-1}\big),\\ \mathbf{D}_k^{-1}\mathbf{X}_k^T\mathbf{X}_k &= p\mathbf{D}_k^{-1} + \sqrt{np}\big(\mathbf{I}_{n-1} + z\mathbf{D}_k^{-1}\big).\end{aligned} \tag{4.6}$$

Since the eigenvalues of $\mathbf{D}^{-1}$ have the form $1/(\lambda_j(\mathbf{A})-z)$, $\|\mathbf{D}^{-1}\| \le 1/v_1$ and similarly $\|\mathbf{D}_k^{-1}\| \le 1/v_1$. From Theorem 11.4 in Bai and Silverstein [3], we note that $-\beta_k(z)$ is the kth diagonal element of $\mathbf{D}^{-1}$ so that $|\beta_k| \le 1/v_1$. Moreover, considering the imaginary parts of $1/\beta_k^{\mathrm{tr}}$ and $1/\beta_k$ and by (4.6) we have

$$\big|\beta_k^{\mathrm{tr}}\big| \le 1/v_1, \qquad \left|1 + \frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(s)}\right| \le \big(1 + 1/v_1^{2s}\big), \qquad s = 1, 2 \tag{4.7}$$

and

$$\big|\big(1 + \mathbf{q}_k^T\mathbf{D}_k^{-2}\mathbf{q}_k\big)\beta_k\big| \le \frac{1 + \mathbf{q}_k^T\mathbf{D}_k^{-1}\overline{\mathbf{D}}_k^{-1}\mathbf{q}_k}{v_1(1 + \mathbf{q}_k^T\mathbf{D}_k^{-1}\overline{\mathbf{D}}_k^{-1}\mathbf{q}_k)} = 1/v_1. \tag{4.8}$$

Applying (4.5), we split ι_k as

$$\begin{aligned}\iota_k &= -\left(1 + \frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(2)}\right)\big(\beta_k^{\mathrm{tr}}\big)^2\eta_k - \gamma_{k1}\big(\beta_k^{\mathrm{tr}}\big)^2\eta_k - \left(1 + \frac{1}{np}\mathbf{q}_k^T\mathbf{D}_k^{-2}\mathbf{q}_k\right)\big(\beta_k^{\mathrm{tr}}\big)^2\beta_k\eta_k^2\\ &= \iota_{k1} + \iota_{k2} + \iota_{k3}.\end{aligned}$$

As will be seen, ι_{k1}, ι_{k2} could be negligible by Lemma 4.1 below.

By Lemma 4.1, (4.7) and (4.8), we have

$$E\left|\sum_{k=1}^n (E_k - E_{k-1})\iota_{k3}\right|^2 \le \sum_{k=1}^n E\left|\left(1 + \frac{1}{np}\mathbf{s}_k^T\mathbf{M}_k^{(2)}\mathbf{s}_k\right)\big(\beta_k^{\mathrm{tr}}\big)^2\beta_k\eta_k^2\right|^2 \le K\delta^4,$$

and that

$$E\left|\sum_{k=1}^n (E_k - E_{k-1})\iota_{k2}\right|^2 \le \sum_{k=1}^n E\big|\gamma_{k1}\big(\beta_k^{\mathrm{tr}}\big)^2\eta_k\big|^2 \le K\sum_{k=1}^n \big(E|\gamma_{k1}|^4 E|\eta_k|^4\big)^{1/2} \le \frac{Kn}{p} + K\delta^2.$$

Therefore, $M_n^{(1)}(z)$ is simplified as

$$\begin{aligned}M_n^{(1)}(z) &= \sum_{k=1}^n E_k\left[-\left(1 + \frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(2)}\right)\big(\beta_k^{\mathrm{tr}}\big)^2\eta_k - \kappa_k\right] + \mathrm{o}_{L_2}(1)\\ &= \sum_{k=1}^n E_k\big(\alpha_k(z)\big) + \mathrm{o}_{L_2}(1),\end{aligned} \tag{4.9}$$

where $\alpha_k(z)$ represents the term in the square bracket. Thus, to prove finite-dimensional convergence of $M_n^{(1)}(z)$, $z \in \mathbb{C}_1$ we need only consider the sum

$$\sum_{j=1}^{l} a_j \sum_{k=1}^{n} E_k\big(\alpha_k(z_j)\big) = \sum_{k=1}^{n}\sum_{j=1}^{l} a_j E_k\big(\alpha_k(z_j)\big), \tag{4.10}$$

where $a_1, \dots, a_l$ are complex numbers and l is any positive integer.

4.2. Useful lemmas

The aim of this subsection is to provide some useful lemmas.

Lemma 4.1. *Let $z \in \mathbb{C}_1$. Under assumptions* (b1), (c), *we have*

$$E|\gamma_{ks}|^2 \le Kn^{-1}, \qquad E|\gamma_{ks}|^4 \le K\left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np}\right), \tag{4.11}$$

$$E|\eta_k|^2 \le Kn^{-1}, \qquad E|\eta_k|^4 \le K\frac{\delta^4}{n} + K\left(\frac{1}{n^2} + \frac{p}{n^2} + \frac{1}{np}\right). \tag{4.12}$$

Proof. From Lemma 5 in Pan and Zhou [25], we obtain

$$E\left|\mathbf{s}_k^T \mathbf{H}\mathbf{s}_k - \operatorname{tr}\mathbf{H}\right|^4 \le K\left(EX_{11}^4\right)^2 E(\operatorname{tr}\mathbf{H}\mathbf{H})^2 \le KE\big(\operatorname{tr}\mathbf{M}_k^{(s)}\overline{\mathbf{M}}_k^{(s)}\big)^2 \le Kn^2p^4, \tag{4.13}$$

where $\mathbf{H} = \mathbf{M}_k^{(s)} - \operatorname{diag}(a_{11}^{(s)}, \dots, a_{nn}^{(s)})$ and $a_{jj}^{(s)}$ is the jth diagonal element of the matrix $\mathbf{M}_k^{(s)}$. To get the third inequality in (4.13), by (4.6) and the uniform bound for $\|\mathbf{D}_k^{-1}\|$, we obtain

$$\begin{aligned} \big|\operatorname{tr}\mathbf{M}_k^{(s)}\overline{\mathbf{M}}_k^{(s)}\big| &= \big|\operatorname{tr}\mathbf{D}_k^{-s}\mathbf{X}_k^T\mathbf{X}_k\overline{\mathbf{D}}_k^{-s}\mathbf{X}_k^T\mathbf{X}_k\big| \le \frac{n}{v_1^{2(s-1)}}\big\|\mathbf{D}_k^{-1}\mathbf{X}_k^T\mathbf{X}_k\big\|^2 \\ &\le \frac{n}{v_1^{2(s-1)}}\big\|p\mathbf{D}_k^{-1} + \sqrt{np}\big(\mathbf{I}_{n-1} + z\mathbf{D}_k^{-1}\big)\big\|^2 \le \frac{Kn^2p^4}{v_1^{2s}}. \end{aligned} \tag{4.14}$$

Let $\mathbb{E}_j(\cdot) = E(\cdot|X_{1k}, X_{2k}, \dots, X_{jk})$, $j = 1, \dots, p$. Since $\{X_{jk}\}_{j=1}^k$ are independent of $a_{jj}^{(s)}$, $(X_{jk}^2 - 1)a_{jj}^{(s)} = (\mathbb{E}_j - \mathbb{E}_{j-1})(X_{jk}^2 - 1)a_{jj}^{(s)}$. By Burkholder's inequality and assumption (c)

$$\begin{aligned} E\left|\sum_{j=1}^{p}\big(X_{jk}^2 - 1\big)a_{jj}^{(s)}\right|^4 &= E\left|\sum_{j=1}^{p}(\mathbb{E}_j - \mathbb{E}_{j-1})\big(X_{jk}^2 - 1\big)a_{jj}^{(s)}\right|^4 \\ &\le KE\left(\sum_{j=1}^{n} E|X_{11}|^4\big|a_{jj}^{(s)}\big|^2\right)^2 + K\sum_{j=1}^{p} E|X_{11}|^8 E|a_{jj}|^4 \\ &\le Kn^5p^2 + n^3p^3, \end{aligned} \tag{4.15}$$

where we use the fact that, with $\mathbf{w}_j^T$ being the jth row of $\mathbf{X}_k$,

$$\begin{aligned} E\left|a_{jj}^{(s)}\right|^4 &= E\left|\ddot{\mathbf{e}}_j^T\mathbf{X}_k\mathbf{D}_k^{-s}\mathbf{X}_k^T\ddot{\mathbf{e}}_j\right|^4 \\ &= E\left|\mathbf{w}_j^T\mathbf{D}_k^{-s}\mathbf{w}_j\right|^4 \le v_1^{-4s}E\left\|\mathbf{w}_j^T\right\|^8 \le Kn^4 + Kn^2p. \end{aligned} \tag{4.16}$$

Here for $j = 1, \dots, p$, $\ddot{\mathbf{e}}_j$ denotes the p-dimensional unit vector with the jth element being 1 and all the remaining being zero. It follows from (4.13) and (4.15) that

$$\begin{aligned} E|\gamma_{ks}|^4 &\le \frac{K}{n^4p^4}E\left|\sum_{j=1}^{p}\left(X_{jk}^2 - 1\right)a_{jj}^{(s)}\right|^4 + \frac{K}{n^4p^4}E\left|\mathbf{s}_k^T\mathbf{H}\mathbf{s}_k - \operatorname{tr}\mathbf{H}\right|^4 \\ &\le K\left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np}\right). \end{aligned}$$

Moreover, applying Lemma 8.10 in Bai and Silverstein [3], we have

$$E|\eta_k|^4 \le \frac{K}{n^2p^2}E\left|\mathbf{s}_k^T\mathbf{s}_k - n\right|^4 + KE\left|\gamma_{k1}(z)\right|^4 \le K\frac{\delta^4}{p} + K\left(\frac{1}{n^2} + \frac{p}{n^2} + \frac{1}{np}\right).$$

The bounds of the absolute second moments for γ_{ks}, η_k follow from a direct application of Lemma 8.10 in Bai and Silverstein [3], (4.6) and the uniform bound for $\|\mathbf{D}_k^{-1}\|$. □

When $z \in \varsigma_l \cup \varsigma_r$, the spectral norm of $\mathbf{D}^{-1}(z)$ as well as the quantities in (4.7) or Lemma 4.1 are unbounded. In order to prove Lemma 6.1, we will establish the bounds similar to those in (4.7) and in Lemma 4.1 for $z \in \varsigma_l \cup \varsigma_r$ below.

Let the event $U_n = \{\max_{j\le n}|\lambda_j(\mathbf{A})| \ge u_0/2 + 1\}$ and $U_{nk} = \{\max_{j\le n}|\lambda_j(\mathbf{A}_k)| \ge 1 + u_0/2\}$. The Cauchy interlacing theorem ensures that

$$\lambda_1(\mathbf{A}) \ge \lambda_1(\mathbf{A}_k) \ge \lambda_2(\mathbf{A}) \ge \lambda_2(\mathbf{A}_k) \ge \dots \ge \lambda_{n-1}(\mathbf{A}_k) \ge \lambda_n(\mathbf{A}). \tag{4.17}$$

Thus, $U_{nk} \subset U_n$. By (3.2) for any $\ell > 0$

$$P(U_{nk}) \le P(U_n) = \mathrm{o}\left(n^{-\ell}\right). \tag{4.18}$$

We claim that

$$\max\left\{\left\|\mathbf{D}^{-1}(z)\right\|, \left\|\mathbf{D}^{-1}(z)\right\|, |\beta_k|\right\} \le \xi_n^{-1}n; \tag{4.19}$$

$$\begin{aligned} &\frac{I(U_n^c)}{|\lambda_j(\mathbf{A}) - z|} \le K, \qquad j = 1, 2, \dots, n, \\ &\frac{I(U_{nk}^c)}{|\lambda_j(\mathbf{A}_k) - z|} \le K, \qquad i = 1, 2, \dots, (n-1); \end{aligned} \tag{4.20}$$

$$\left\|\mathbf{D}^{-1}(z)\right\|I\left(U_n^c\right) \le 2/(u_0 - 2), \qquad \left\|\mathbf{D}_k^{-1}(z)\right\|I\left(U_{nk}^c\right) \le 2/(u_0 - 2). \tag{4.21}$$

Indeed, the quantities in (4.19) are bounded due to $|1/\Im(z)| \le \xi_n^{-1} n$ while (4.20) holds because $I(U_n^c)/|\lambda_j(\mathbf{A}) - z|$ (or $I(U_n^c)/|\lambda_j(\mathbf{A}_k) - z|$) is bounded by v_1^{-1} when $z \in \varsigma_u$ and bounded by $2/(u_0-2)$ when $z \in \varsigma_l \cup \varsigma_r$. The estimates in (4.21) hold because of the eigenvalues of $\mathbf{D}^{-1} I(U_n^c)$ (or $\mathbf{D}^{-1} I(U_n^c)$) having the form $I(U_n^c)/(\lambda_j(\mathbf{A}) - z)$ (or $I(U_n^c)/(\lambda_j(\mathbf{A}_k) - z)$).

Lemma 4.2. *Let $z \in \varsigma_n$. The following bound*

$$|\beta_k| I\big(U_n^c\big) \le K, \tag{4.22}$$

holds.

Proof. In view of (4.2), to prove (4.22), we need to find an upper bound for $|\operatorname{tr}\mathbf{D}^{-1} - \operatorname{tr}\mathbf{D}_k^{-1}| I(U_n^c)$ and a lower bound for $|1 + \mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k| I(U_n^c)$. It follows from (4.20) and (4.17) that

$$\begin{aligned}
\big|\operatorname{tr}\mathbf{D}^{-1} - \operatorname{tr}\mathbf{D}_k^{-1}\big| I\big(U_n^c\big) &\le \left| \sum_{j=1}^{n} \frac{1}{\lambda_j(\mathbf{A}) - z} - \sum_{j=1}^{n-1} \frac{1}{\lambda_j(\mathbf{A}_k) - z} \right| I\big(U_n^c\big) \\
&\le \left(\sum_{j=1}^{n-1} \frac{\lambda_j(\mathbf{A}) - \lambda_j(\mathbf{A}_k)}{|\lambda_j(\mathbf{A}) - z||\lambda_j(\mathbf{A}_k) - z|} + \frac{1}{|\lambda_n(\mathbf{A}) - z|} \right) I\big(U_n^c\big) \\
&\le K \left(\sum_{j=1}^{n-1} \big(\lambda_j(\mathbf{A}) - \lambda_j(\mathbf{A}_k)\big) + 1 \right) I\big(U_n^c\big) \\
&\le K\big(\lambda_1(\mathbf{A}) - \lambda_n(\mathbf{A}) + 1\big) I\big(U_n^c\big) \le K(u_0 + 3).
\end{aligned} \tag{4.23}$$

Let $\mathbf{u}_j(\mathbf{A}_k)$, $j = 1, \dots, n-1$ be the eigenvectors corresponding to the eigenvalues $\lambda_j(\mathbf{A}_k)$, $j = 1, \dots, n-1$. Then $\sum_{j=1}^{n-1} \frac{\mathbf{u}_j(\mathbf{A}_k)\mathbf{u}_j^T(\mathbf{A}_k)}{(\lambda_j(\mathbf{A}_k)-z)^2}$ is the spectral decomposition of $\mathbf{D}_k^{-2}$. We distinguish two cases:

(i) When $z \in V_1 = \varsigma_u \cup \{z : |\Im(z)| > (u_0 - 2)/4\}$, via (4.7), we then obtain

$$|\beta_k| I\big(U_n^c\big) \le 1/\big|\Im(z)\big| \le \max\big\{v_1^{-1}, 4/(u_0 - 2)\big\} \le K.$$

Thus, (4.22) is true for $z \in V_1$.

(ii) When $z \in V_2 = (\varsigma_l \cup \varsigma_r) \cap \{z : |\Im(z)| < (u_0 - 2)/4\}$, if U_n^c happens, we have $|\lambda_j(\mathbf{A}_k) - \Re(z)| \ge \frac{u_0-2}{2}$ since $\Re(z) = \pm u_0$ for $z \in V_2$. A direct calculation shows

$$\Re\big(\big(1+\mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k\big) I\big(U_n^c\big)\big) = 1 + \sum_{j=1}^{n-1} \frac{(\lambda_j(\mathbf{A}_k) - \Re(z))^2 - |\Im(z)|^2}{|\lambda_j(\mathbf{A}_k) - z|^4} \big(\mathbf{q}_k^T \mathbf{u}_j(\mathbf{A}_k)\big)^2 I\big(U_n^c\big) > 1.$$

Therefore, $|1 + \mathbf{q}^T \mathbf{D}_k^{-2} \mathbf{q}| I(U_n^c)$ has a lower bound which, together with (4.23), implies (4.22) is true for $z \in V_2$.

Since $\varsigma_n = V_1 \cup V_2$, we finish the proof of Lemma 4.2. □

Lemma 4.3. *Let* $z \in \varsigma_n$ *and* $\bar{\mu}_k = \frac{1}{\sqrt{np}}(\mathbf{s}_k^T\mathbf{s}_k - p) - \mathbf{q}_k^T\mathbf{D}_k^{-1}(z)\mathbf{q}_k + E\frac{1}{np}\operatorname{tr}\mathbf{X}\mathbf{D}^{-1}(z)\mathbf{X}^T$. *The following bounds hold*

$$E|\bar{\mu}_k|^4 \le K\frac{\delta^4}{n} + K\left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np}\right) \tag{4.24}$$

and

$$\left|E\bar{\mu}_k^3\right| = \mathrm{o}\left(n^{-1}\right). \tag{4.25}$$

Proof. Write

$$\begin{aligned}
\bar{\mu}_k &= \frac{1}{\sqrt{np}}\left(\mathbf{s}_k^T\mathbf{s}_k - p\right) - \gamma_{k1} + \left(1 + z\sqrt{\frac{p}{n}}\right)\left(\frac{1}{n}\operatorname{tr}\mathbf{D}^{-1}(z) - \frac{1}{n}\operatorname{tr}\mathbf{D}_k^{-1}(z)\right) \\
&\quad - \left(1 + z\sqrt{\frac{p}{n}}\right)\left(\frac{1}{n}\operatorname{tr}\mathbf{D}^{-1}(z) - E\frac{1}{n}\operatorname{tr}\mathbf{D}^{-1}(z)\right) + \frac{1}{\sqrt{np}} \\
&= L_1 - \gamma_{k1} + L_3 + L_4 + L_5.
\end{aligned}$$

When the event U_n^c happens, reviewing the proof of the second result of (4.11) and via (4.21), we also have

$$E|\gamma_{ks}|^4 I\left(U_n^c\right) \le K\left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np}\right), \qquad m = 1, 2.$$

Moreover, by (4.18) and (4.19)

$$E|\gamma_{ks}|^4 I(U_n) = \mathrm{o}\left(n^{-\ell}\right).$$

It follows that

$$E|\gamma_{ks}|^4 \le K\left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np}\right), \qquad m = 1, 2. \tag{4.26}$$

Using Lemma 8.10 in Bai and Silverstein [3], (4.18), (4.19) and (4.23) we then have

$$E|L_1|^4 \le K\delta^4 n^{-1}, \qquad E|L_3|^4 \le Kn^{-4}, \qquad E|L_5|^4 \le Kn^{-2}p^{-2}. \tag{4.27}$$

As for L_4, by Burkholder's inequality, (4.3) and (4.23), we have

$$\begin{aligned}
E|L_4|^4 &\le Kn^{-4}E\left|\sum_{k=1}^n (E_k - E_{k-1})\left(\operatorname{tr}\mathbf{D}^{-1} - \operatorname{tr}\mathbf{D}_k^{-1}\right)\right|^4 \\
&\le Kn^{-4}\sum_{k=1}^n E\left|\operatorname{tr}\mathbf{D}^{-1}(z) - \operatorname{tr}\mathbf{D}_k^{-1}(z)\right|^4 + Kn^{-1}E\left(\sum_{k=1}^n E_k\left|\operatorname{tr}\mathbf{D}^{-1}(z) - \operatorname{tr}\mathbf{D}_k^{-1}(z)\right|^2\right)^2
\end{aligned}$$

$$\leq Kn^{-4}\sum_{k=1}^{n} E\left|\operatorname{tr}\mathbf{D}^{-1}(z)-\operatorname{tr}\mathbf{D}_k^{-1}(z)\right|^4 I\left(U_n^c\right) \tag{4.28}$$

$$+ Kn^{-4}E\left(\sum_{k=1}^{n} E_k\left|\operatorname{tr}\mathbf{D}^{-1}(z)-\operatorname{tr}\mathbf{D}_k^{-1}(z)\right|^2\right)^2 I\left(U_n^c\right)+\mathrm{o}\left(n^{-\ell}\right)$$

$$\leq Kn^{-2}.$$

Therefore, the proof of (4.24) is completed. Also, the analysis above yields

$$E|L_1-\gamma_{k1}|^4 \leq K\left(\frac{\delta^2}{n}+\frac{1}{n^2}+\frac{n}{p^2}+\frac{1}{np}\right) \leq K\delta^2 n^{-1}, \qquad E|L_3+L_4+L_5|^4 \leq Kn^{-2}. \tag{4.29}$$

It is also easy to verify that, for $z \in \varsigma_n$,

$$E\left|\frac{1}{\sqrt{np}}\left(\mathbf{s}_k^T\mathbf{s}_k-p\right)\right|^2 \leq Kn^{-1}, \qquad E|\gamma_{km}|^2 \leq Kn^{-1}. \tag{4.30}$$

We proceed to prove (4.25). First of all

$$\left|EL_1^3\right| = \frac{1}{(np)^{3/2}}\left|E\left(\sum_{j=1}^{p}\left(X_{jk}^2-1\right)\right)^3\right| = \frac{1}{(np)^{3/2}}\sum_{j=1}^{p} E\left(X_{jk}^2-1\right)^3 \leq K\delta^2/n. \tag{4.31}$$

For $s=1,2$, denoting $\mathbf{M}_k^{(s)}=(a_{ij}^{(s)})_{p\times p}$, we then have

$$E\gamma_{ks}^3 = \frac{1}{n^3p^3}E\left(\sum_{i\neq j} X_{ik}X_{jk}a_{ij}^{(s)}+\sum_{i=1}^{n}\left(X_{ik}^2-1\right)a_{ii}^{(s)}\right)^3$$
$$= J_1+J_2+J_3+J_4,$$

where

$$J_1=\frac{1}{n^3p^3}E\left(\sum_{i\neq j, j\neq t, t\neq i} X_{ik}^2X_{jk}^2X_{tk}^2a_{ij}^{(s)}a_{jt}^{(s)}a_{ti}^{(s)}\right)+\frac{4}{n^3p^3}E\left(\sum_{i\neq j} X_{ik}^3X_{jk}^3\left(a_{ij}^{(s)}\right)^3\right) \triangleq J_{11}+J_{12},$$

$$J_2=\frac{1}{n^3p^3}E\left(\sum_{i=1}^{p}\left(X_{ik}^2-1\right)^3\left(a_{ii}^{(s)}\right)^3\right),$$

$$J_3=3\frac{1}{n^3p^3}E\left(\sum_{i\neq j} X_{ik}\left(X_{ik}^2-1\right)X_{jk}\left(X_{jk}^2-1\right)a_{ij}^{(s)}a_{ii}^{(s)}a_{jj}^{(s)}\right),$$

$$J_4=3\frac{2}{n^3p^3}E\left(\sum_{i\neq j} X_{ik}^2\left(X_{ik}^2-1\right)X_{jk}^2a_{ij}^{(s)}a_{ii}^{(s)}a_{ji}^{(s)}\right).$$

The inequality (4.16) can be extended to the range $z \in \varsigma_n$ by a similar method as that in (4.26). Therefore,

$$|J_2| \le K \frac{1}{n^3 p^3} p\delta^2 \sqrt{np}\left(n^4 + n^2 p\right)^{3/4} \le K\delta^2 n^{-1},$$
$$|J_3| \le K \frac{1}{n^3 p^3} p^2 E\|\mathbf{w}_i\|^3 E\|\mathbf{w}_j\|^3 + o\left(n^{-\ell}\right) \le Kp^{-1} + o\left(n^{-\ell}\right), \qquad J_4 \le Kp^{-1} + o\left(n^{-\ell}\right),$$

where $\mathbf{w}_j^T$ is the jth row of $\mathbf{X}_k$.

Consider J_1 now. We first note that $J_{12} = O(p^{-1})$. Split J_{12} as

$$\begin{aligned}
J_{12} &= \frac{1}{n^3 p^3} E \operatorname{tr}\left(\mathbf{X}_k \mathbf{D}_k^{-s} \mathbf{X}_k^T\right)^3 - \frac{1}{n^3 p^3} E \sum_{i \ne t} a_{ii}^{(s)} a_{it}^{(s)} a_{ti}^{(s)} \\
&\quad + \frac{1}{n^3 p^3} E \sum_{i \ne j} a_{ij}^{(s)} a_{jj}^{(s)} a_{ji}^{(s)} + \frac{1}{n^3 p^3} E \sum_{i \ne j} a_{ij}^{(s)} a_{ji}^{(s)} a_{ii}^{(s)} + \frac{1}{n^3 p^3} E \sum_{i=1}^{p} \left(a_{ii}^{(s)}\right)^3 \\
&\le Kn^{-2} + Kp^{-1}.
\end{aligned}$$

Thus, we obtain

$$\left|E\gamma_{ks}^3\right| \le K\left(\delta^2 n^{-1} + p^{-1}\right). \tag{4.32}$$

It follows from (4.29), (4.30) and (4.32) that

$$\begin{aligned}
\left|E\bar{\mu}_k^3\right| &\le \left|E(L_1 - \gamma_{k1})^3\right| + \left|E(L_3 + L_4 + L_5)^3\right| + 3\left|E(L_1 - \gamma_{k1})(L_3 + L_4 + L_5)^2\right| \\
&\quad + 3\left|E(L_1 - \gamma_{k1})^2 (L_3 + L_4 + L_5)\right| \\
&\le \left|EL_1^3\right| + \left|E\gamma_{k3}^3\right| + 3E^{1/2} EL_1^4 \cdot E^{1/2}\gamma_{ks}^2 + 3E^{1/2} L_1^2 \cdot E^{1/2}\gamma_{ks}^4 + Kn^{-3/2} + K\delta n^{-1} \\
&= o\left(n^{-1}\right).
\end{aligned}$$

The proof of Lemma 4.3 is completed. □

The following lemma will be used to prove the first result of (3.5) and (6.15) below.

Lemma 4.4. *For $z \in \varsigma_n$ we have*

$$E\left|M_n^{(1)}(z)\right| \le K,$$

where $M_n^{(1)}(z) = n(m_n(z) - Em_n(z))$.

Proof. Note that the expression $M_n^{(1)}(z)$ in (4.3) may not be suitable for $z \in \varsigma_n$, since β_k^{tr} or even $\beta_k^{\text{tr}} I(U_n^c)$ may be not bounded. For this reason, we introduce the following notations with the purpose to obtain a similar expression to (4.3). Let

$$\acute{\varepsilon}_k = \frac{1}{z + (1/(np))E\operatorname{tr}\mathbf{M}_k^{(1)}}, \qquad \acute{\mu}_k = \frac{1}{\sqrt{np}}\big(\mathbf{s}_k^T\mathbf{s}_k - p\big) - \gamma_{k1} - \left(\frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(1)} - \frac{1}{np}E\operatorname{tr}\mathbf{M}_k^{(1)}\right).$$

Hence

$$\beta_k = \acute{\varepsilon}_k + \beta_k\acute{\varepsilon}_k\acute{\mu}_k. \tag{4.33}$$

As in (4.3) and a few lines below it, by (4.33), we write

$$M^{(1)}(z) = \sum_{k=1}^{n}(E_k - E_{k-1})(\acute{\iota}_{k1} + \acute{\iota}_{k2} + \acute{\iota}_{k3} + \acute{\kappa}_k),$$

where

$$\acute{\iota}_{k1}(z) = -\left(1 + \frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(2)}\right)(\acute{\varepsilon}_k)^2\acute{\mu}_k, \qquad \acute{\iota}_{k2}(z) = -\gamma_{k1}(\acute{\varepsilon}_k)^2\acute{\mu}_k,$$

$$\acute{\iota}_{k3}(z) = -\left(1 + \frac{1}{np}\mathbf{q}_k^T\mathbf{D}_k^{-2}(z)\mathbf{q}_k\right)\beta_k(\acute{\varepsilon}_k)^2\acute{\mu}_k^2, \qquad \acute{\kappa}_k = \acute{\varepsilon}_k\gamma_{k2}(z).$$

We next derive the bounds for $\acute{\varepsilon}_k$ and the forth moment of $\acute{\mu}_k$. Since $F_n \xrightarrow{\text{a.s.}} F$ as $n \to \infty$, we conclude from (4.18), (4.19), (4.21) and the dominated convergence theorem that, for any fixed positive integer t

$$E\big|m_n(z) - m(z)\big|^t \to 0. \tag{4.34}$$

By (4.6), (4.23) and (4.34), we then have

$$E\frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(1)} = E\left[\left(1 + z\sqrt{\frac{n}{p}}\right)m_n(z) - \left(1 + z\sqrt{\frac{n}{p}}\right)\frac{1}{n}\big(\operatorname{tr}\mathbf{D}^{-1} - \operatorname{tr}\mathbf{D}_k^{-1}\big) + \frac{n-1}{\sqrt{np}}\right] \to m(z).$$

Hence,

$$|\acute{\varepsilon}_k| = \left|\frac{1}{z + m(z) + \mathrm{o}(1)}\right| \le \left|\frac{2}{z + m(z)}\right| \le 2. \tag{4.35}$$

On the other hand, via (4.6), (4.23) and (4.28)

$$E\left|\frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(1)} - E\frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(1)}\right|^4 \le \left(1 + z\sqrt{\frac{n}{p}}\right)^4 n^{-4}E\big|\operatorname{tr}\mathbf{D}^{-1} - E\operatorname{tr}\mathbf{D}^{-1}\big|^4 \le Kn^{-2},$$

and this, together with (4.26), implies

$$E|\acute{\mu}_k|^4 \le K\frac{\delta^4}{n} + K\left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np}\right). \tag{4.36}$$

Combining (4.35), (4.36), Lemma 4.2, (4.18), (4.19), (4.21) with Burkholder's inequality, we obtain

$$E\left|M_n^{(1)}(z)\right|^2 \le K.$$

The proof of the lemma is completed. □

5. Convergence of $M_n^{(1)}(z)$

To prove Proposition 3.1, we need to establish (i) the finite-dimensional convergence and the tightness of $M_n^{(1)}(z)$; (ii) the convergence of the mean function $EM(z)$. This section is devoted to the first target. Throughout this section, we assume that $z \in \mathbb{C}_1$ and K denotes a constant which may change from line to line and may depend on v_1 but is independent of n.

5.1. Application of central limit theorem for martingales

In order to establish the central limit theorem for the martingale (4.10), we have to check the following two conditions:

Condition 5.1 (Lyapunov condition). *For some $a > 2$,*

$$\sum_{k=1}^{n} E_{k-1}\left[\left|E_k\left(\sum_{j=1}^{l} a_j E_k\big(\alpha_k(z_j)\big)\right)\right|^a\right] \xrightarrow{i.p.} 0.$$

Condition 5.2. *The covariance*

$$\Lambda_n(z_1, z_2) \triangleq \sum_{k=1}^{n} E_{k-1}\big[E_k\alpha_k(z_1)\cdot E_k\alpha_k(z_2)\big] \tag{5.1}$$

converges in probability to $\Lambda(z_1, z_2)$ whose explicit form will be given in (5.29).

Condition 5.1 is satisfied by choosing $a = 4$, using Lemma 4.1, and the fact that via (4.7)

$$\left|\alpha_k(z)\right| = \left|\left(1 + \frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(2)}\right)\left(\beta_k^{\mathrm{tr}}\right)^2\eta_k + \beta_k^{\mathrm{tr}}\gamma_k\right| \le \frac{1 + v_1^{-2}}{v_1^2}|\eta_k| + \frac{1}{v_1}|\gamma_k|.$$

Consider Condition 5.2 now. Note that

$$\alpha_k(z) = -\left(1 + \frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(2)}\right)\left(\beta_k^{\mathrm{tr}}\right)^2\eta_k - \gamma_k\beta_k^{\mathrm{tr}} = \frac{\partial}{\partial z}\left(\beta_k^{\mathrm{tr}}\eta_k\right).$$

By the dominated convergence theorem, we have

$$\Lambda_n(z_1, z_2) = \frac{\partial^2}{\partial z_2\,\partial z_1}\sum_{k=1}^{n} E_{k-1}\Big[E_k\big(\beta_k^{\mathrm{tr}}(z_1)\eta_k(z_1)\big)\cdot E_k\big(\beta_k^{\mathrm{tr}}(z_2)\eta_k(z_2)\big)\Big]. \tag{5.2}$$

By (4.6), (4.2), (4.8), (4.1) and the fact $m_n(z) \xrightarrow{\text{a.s.}} m(z)$, and the dominated convergence theorem again, for any fixed t,

$$E\left|\frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(1)} - m(z)\right|^t \to 0, \qquad E\left|\beta_k^{\mathrm{tr}}(z) + m(z)\right|^t \to 0, \qquad \text{as } n \to \infty. \tag{5.3}$$

Substituting (5.3) into (5.2) yields

$$\begin{aligned}\Lambda_n(z_1, z_2) &= \frac{\partial^2}{\partial z_2\,\partial z_1}\left[m(z_1)m(z_2)\sum_{k=1}^{n} E_{k-1}\big(E_k\eta_k(z_1)\cdot E_k\eta_k(z_2)\big) + \mathrm{o}_{\mathrm{i.p.}}(1)\right] \\ &= \frac{\partial^2}{\partial z_2\,\partial z_1}\big[m(z_1)m(z_2)\tilde{\Lambda}_n(z_1, z_2) + \mathrm{o}_{\mathrm{i.p.}}(1)\big].\end{aligned} \tag{5.4}$$

By Vitali's theorem (see Titchmarsh [28], page 168), it is enough to find the limit of $\tilde{\Lambda}_n(z_1, z_2)$. To this end, with notation $E_k(\mathbf{M}_k^{(1)}(z)) = (a_{ij}(z))_{n\times n}$, write

$$E_k\eta_k(z) = \frac{1}{\sqrt{np}}\sum_{j=1}^{p}\big(X_{jk}^2 - 1\big) - \frac{1}{np}\left(\sum_{i\neq j} X_{ik}X_{jk}a_{ij}(z) + \sum_{i=1}^{p}\big(X_{ik}^2 - 1\big)a_{ii}(z)\right).$$

By the above formula and independence between $\{X_{ik}\}_{i=1}^p$ and $E_k(\mathbf{M}_k^{(1)})$, a straightforward calculation yields

$$E_{k-1}\big[E_k\eta_k(z_1)\cdot E_k\eta_k(z_2)\big] = \frac{1}{n}E\big(X_{11}^2 - 1\big)^2 + A_1 + A_2 + A_3 + A_4, \tag{5.5}$$

where

$$A_1 = -\frac{1}{np\sqrt{np}}E\big(X_{11}^2 - 1\big)^2\sum_{i=1}^{p} a_{ii}(z_1), \qquad A_2 = -\frac{1}{np\sqrt{np}}E\big(X_{11}^2 - 1\big)^2\sum_{i=1}^{p} a_{ii}(z_2),$$

$$A_3 = \frac{2}{n^2p^2}\sum_{i\neq j}^{p} a_{ij}(z_1)a_{ij}(z_2), \qquad A_4 = \frac{1}{n^2p^2}E\big(X_{11}^2 - 1\big)^2\sum_{i=1}^{p} a_{ii}(z_1)a_{ii}(z_2).$$

Note that $a_{ii}(z)$ is precisely $E_k a_{ii}^{(1)}$ in (4.16). From (4.16), we then obtain for $j = 1, 2, 4$

$$E\left|\sum_{k=1}^{n} A_j\right| \to 0.$$

Also, we conclude from (4.16) that

$$\sum_{k=1}^{n} A_3 = \frac{2}{n}\sum_{k=1}^{n}\mathbb{Z}_k - \frac{2}{n^2p^2}\sum_{k=1}^{n}\sum_{i=1}^{p} a_{ii}(z_1)a_{ii}(z_2) = \frac{2}{n}\sum_{k=1}^{n}\mathbb{Z}_k + \mathrm{o}_{L_1}(1),$$

where

$$\mathbb{Z}_k = \frac{1}{np^2} \operatorname{tr} E_k \mathbf{M}_k^{(1)}(z_1) \cdot E_k \mathbf{M}_k^{(1)}(z_2).$$

Summarizing the above we see that

$$\tilde{\Lambda}_n(z_1, z_2) = \frac{2}{n} \sum_{k=1}^{n} \mathbb{Z}_k + \nu_4 - 1 + \mathrm{o}_{L_1}(1). \tag{5.6}$$

5.2. The asymptotic expression of $\mathbb{Z}_k$

The goal is to derive an asymptotic expression of $\mathbb{Z}_k$ with the purpose of obtaining the limit of $\tilde{\Lambda}_n(z_1, z_2)$.

5.2.1. *Decomposition of* $\mathbb{Z}_k$

To evaluate $\mathbb{Z}_k$, we need two different decompositions of $E_k \mathbf{M}_k^{(1)}(z)$. With slight abuse of notation, let $\{\mathbf{e}_i, i = 1, \ldots, k-1, k+1, \ldots, n\}$ be the $(n-1)$-dimensional unit vectors with the ith (or $(i-1)$th) element equal to 1 and the remaining equal to 0 according as $i < k$ (or $i > k$). Write $\mathbf{X}_k = \mathbf{X}_{ki} + \mathbf{s}_i \mathbf{e}_i^T$. Define

$$\begin{aligned}
\mathbf{D}_{ki,r} &= \mathbf{D}_k - \mathbf{e}_i \mathbf{h}_i^T = \frac{1}{\sqrt{np}} \big(\mathbf{X}_{ki}^T \mathbf{X}_k - p \mathbf{I}_{(i)} \big) - z \mathbf{I}_{n-1}, \\
\mathbf{D}_{ki} &= \mathbf{D}_k - \mathbf{e}_i \mathbf{h}_i^T - \mathbf{r}_i \mathbf{e}_i^T = \frac{1}{\sqrt{np}} \big(\mathbf{X}_{ki}^T \mathbf{X}_{ki} - p \mathbf{I}_{(i)} \big) - z \mathbf{I}_{n-1}, \\
\mathbf{h}_i^T &= \frac{1}{\sqrt{np}} \mathbf{s}_i^T \mathbf{X}_{ki} + \frac{1}{\sqrt{np}} \big(\mathbf{s}_i^T \mathbf{s}_i - p \big) \mathbf{e}_i^T, \qquad \mathbf{r}_i = \frac{1}{\sqrt{np}} \mathbf{X}_{ki}^T \mathbf{s}_i, \\
\zeta_i &= \frac{1}{1 + \vartheta_i}, \qquad \vartheta_i = \mathbf{h}_i^T \mathbf{D}_{ki,r}^{-1}(z) \mathbf{e}_i, \qquad \mathbf{M}_{ki} = \mathbf{X}_{ki} \mathbf{D}_{ki}^{-1}(z) \mathbf{X}_{ki}^T.
\end{aligned} \tag{5.7}$$

Here $\mathbf{I}_{(i)}$ is obtained from $\mathbf{I}_{n-1}$ with the ith (or $(i-1)$th) diagonal element replaced by zero if $i < k$ (or $i > k$). With respect to the above notations we would point out that, for $i < k$ (or $i > k$), the matrix $\mathbf{X}_{ki}$ is obtained from $\mathbf{X}_k$ with the entries on the ith (or $(i-1)$th) column replaced by zero; $\mathbf{h}_i^T$ is the ith (or $(i-1)$th) row of $\mathbf{A}_k$ and $\mathbf{r}_i$ is the ith (or $(i-1)$th) column of $\mathbf{A}_k$ with the ith (or $(i-1)$th) element replaced by zero. $(\mathbf{X}_{ki}^T \mathbf{X}_k - p \mathbf{I}_{(i)})$ is obtained from $(\mathbf{X}_k^T \mathbf{X}_k - p \mathbf{I}_{n-1})$ with the entries on the ith (or $(i-1)$th) row and ith (or $(i-1)$th) column replaced by zero.

The notation defined above may depend on k. When we obtain bounds or limits for them such as $\frac{1}{n} \operatorname{tr} \mathbf{D}_{ki}^{-1}$ the results hold uniformly in k.

Observing the structure of the matrices $\mathbf{X}_{ki}$ and $\mathbf{D}_{ki}^{-1}$, we have some crucial identities,

$$\mathbf{X}_{ki} \mathbf{e}_i = \mathbf{0}, \qquad \mathbf{e}_i^T \mathbf{D}_{ki,r}^{-1} = \mathbf{e}_i^T \mathbf{D}_{ki}^{-1} = -z^{-1} \mathbf{e}_i, \tag{5.8}$$

where $\mathbf{0}$ is a p-dimensional vector with all the elements equal to 0. By (5.8) and the frequently used formulas

$$\begin{aligned}
\mathbf{Y}^{-1}-\mathbf{W}^{-1}&=-\mathbf{W}^{-1}(\mathbf{Y}-\mathbf{W})\mathbf{Y}^{-1},\\
\left(\mathbf{Y}+\mathbf{a}\mathbf{b}^T\right)^{-1}\mathbf{a}&=\frac{\mathbf{Y}^{-1}\mathbf{a}}{1+\mathbf{b}^T\mathbf{Y}^{-1}\mathbf{a}},\\
\mathbf{b}^T\left(\mathbf{Y}+\mathbf{a}\mathbf{b}^T\right)^{-1}&=\frac{\mathbf{b}^T\mathbf{Y}^{-1}}{1+\mathbf{b}^T\mathbf{Y}^{-1}\mathbf{a}},
\end{aligned}\tag{5.9}$$

we have

$$\begin{aligned}
\mathbf{D}_k^{-1}-\mathbf{D}_{ki,r}^{-1}&=-\zeta_i\mathbf{D}_{ki,r}^{-1}\mathbf{e}_i\mathbf{h}_i^T\mathbf{D}_{ki,r}^{-1},\\
\mathbf{D}_{ki,r}^{-1}-\mathbf{D}_{ki}^{-1}&=\frac{1}{z\sqrt{np}}\mathbf{D}_{ki}^{-1}\mathbf{X}_{ki}^T\mathbf{s}_i\mathbf{e}_i^T.
\end{aligned}\tag{5.10}$$

We first claim the following decomposition of $E_k\mathbf{M}_k^{(1)}(z)$, for $i<k$,

$$\begin{aligned}
E_k\mathbf{M}_k^{(1)}(z)&=E_k\mathbf{M}_{ki}-E_k\left(\frac{\zeta_i}{znp}\mathbf{M}_{ki}\mathbf{s}_i\mathbf{s}_i^T\mathbf{M}_{ki}\right)+E_k\left(\frac{\zeta_i}{z\sqrt{np}}\mathbf{M}_{ki}\right)\mathbf{s}_i\mathbf{s}_i^T\\
&\quad+\mathbf{s}_i\mathbf{s}_i^T E_k\left(\frac{\zeta_i}{z\sqrt{np}}\mathbf{M}_{ki}\right)-E_k\left(\frac{\zeta_i}{z}\right)\mathbf{s}_i\mathbf{s}_i^T\\
&=B_1(z)+B_2(z)+B_3(z)+B_4(z)+B_5(z).
\end{aligned}\tag{5.11}$$

Indeed, by the decomposition of $\mathbf{X}_k$, write

$$\mathbf{M}_k^{(1)}=\mathbf{X}_{ki}\mathbf{D}_k^{-1}\mathbf{X}_{ki}^T+\mathbf{X}_{ki}\mathbf{D}_k^{-1}\mathbf{e}_i\mathbf{s}_i^T+\mathbf{s}_i\mathbf{e}_i^T\mathbf{D}_k^{-1}\mathbf{X}_{ki}^T+\mathbf{s}_i\mathbf{e}_i^T\mathbf{D}_k^{-1}\mathbf{e}_i\mathbf{s}_i^T.$$

Applying (5.7), (5.8) and (5.10), we obtain

$$\begin{aligned}
\mathbf{X}_{ki}\mathbf{D}_k^{-1}\mathbf{X}_{ki}^T&=\mathbf{X}_{ki}\mathbf{D}_{ki,r}^{-1}\mathbf{X}_{ki}^T-\zeta_i\mathbf{X}_{ki}^T\mathbf{D}_{ki,r}^{-1}\mathbf{e}_i\mathbf{h}_i^T\mathbf{D}_{ki,r}^{-1}\mathbf{X}_{ki}^T\\
&=\mathbf{M}_{ki}-\frac{\zeta_i}{z\sqrt{np}}\mathbf{M}_{ki}\mathbf{s}_i\cdot\frac{1}{\sqrt{np}}\mathbf{s}_i^T\mathbf{X}_{ki}\mathbf{D}_{ki,r}^{-1}\mathbf{X}_{ki}^T\\
&=\mathbf{M}_{ki}-\frac{\zeta_i}{znp}\mathbf{M}_{ki}\mathbf{s}_i\mathbf{s}_i^T\mathbf{M}_{ki}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbf{X}_{ki}\mathbf{D}_k^{-1}\mathbf{e}_i\mathbf{s}_i^T=\frac{\zeta_i}{z\sqrt{np}}\mathbf{M}_{ki}\mathbf{s}_i\mathbf{s}_i^T,\qquad \mathbf{s}_i\mathbf{e}_i^T\mathbf{D}_k^{-1}\mathbf{X}_{ki}^T=\frac{\zeta_i}{z\sqrt{np}}\mathbf{s}_i\mathbf{s}_i^T\mathbf{M}_{ki},\\
&\mathbf{s}_i\mathbf{e}_i^T\mathbf{D}_k^{-1}\mathbf{e}_i\mathbf{s}_i^T=\zeta_i\mathbf{s}_i\mathbf{e}_i^T\mathbf{D}_{ki,r}^{-1}\mathbf{e}_i\mathbf{s}_i^T=-\frac{\zeta_i}{z}\mathbf{s}_i\mathbf{s}_i^T.
\end{aligned}$$

Summarizing the above and noting $E_k(\mathbf{s}_i)=\mathbf{s}_i$ for $i<k$ yield (5.11), as claimed.

On the other hand, write

$$\mathbf{D}_k=\sum_{i=1(\neq k)}^{n}\mathbf{e}_i\mathbf{h}_i^T-z\mathbf{I}_{n-1}.$$

Multiplying by $\mathbf{D}_k^{-1}$ on both sides, we have

$$z\mathbf{D}_k^{-1}=-\mathbf{I}_{n-1}+\sum_{i=1(\neq k)}^{n}\mathbf{e}_i\mathbf{h}_i^T\mathbf{D}_k^{-1}. \tag{5.12}$$

Therefore, by (5.8), (5.10) and the fact that $\mathbf{X}_k\mathbf{X}_k^T=\sum_{i\neq k}\mathbf{s}_i\mathbf{s}_i^T$, we have

$$\begin{aligned}
zE_k\big(\mathbf{M}_k^{(1)}(z)\big)&=-E_k\big(\mathbf{X}_k\mathbf{X}_k^T\big)+\sum_{i=1(\neq k)}^{n}E_{k-1}\big(\mathbf{X}_k\mathbf{e}_i\mathbf{h}_i^T\mathbf{D}_k^{-1}\mathbf{X}_k^T\big)\\
&=-E_k\left(\sum_{i=1(\neq k)}^{n}\mathbf{s}_i\mathbf{s}_i^T\right)+\sum_{i=1(\neq k)}^{n}E_k\big(\zeta_i\mathbf{s}_i\mathbf{h}_i^T\mathbf{D}_{ki,r}^{-1}\big(\mathbf{X}_{ki}^T+\mathbf{e}_i\mathbf{s}_i^T\big)\big)\\
&=-(n-k)\mathbf{I}_{n-1}-\sum_{i<k}\mathbf{s}_i\mathbf{s}_i^T+\sum_{i=1(\neq k)}^{n}E_k\left(\frac{\zeta_i}{\sqrt{np}}\mathbf{s}_i\mathbf{s}_i^T\mathbf{M}_{ki}\right)\\
&\quad+\sum_{i=1(\neq k)}^{n}E_k\big(\zeta_i\vartheta_i\mathbf{s}_i\mathbf{s}_i^T\big).
\end{aligned} \tag{5.13}$$

Consequently, by splitting $E_k(\mathbf{M}_k^{(1)}(z_2))$ as in (5.11) for $i<k$ and $z_1E_k(\mathbf{M}_k^{(1)}(z_1))$ as in (5.13), we obtain

$$\begin{aligned}
z_1\mathbb{Z}_k&=\frac{z_1}{np^2}\operatorname{tr}E_k\mathbf{M}_k^{(1)}(z_1)\cdot E_k\mathbf{M}_k^{(1)}(z_2)\\
&=C_1(z_1,z_2)+C_2(z_1,z_2)+C_3(z_1,z_2)+C_4(z_1,z_2),
\end{aligned} \tag{5.14}$$

where

$$C_1(z_1,z_2)=-\frac{1}{np^2}(n-k)\operatorname{tr}E_k\mathbf{M}_k^{(1)}(z_2),$$

$$C_2(z_1,z_2)=-\frac{1}{np^2}\sum_{i<k}\mathbf{s}_i^T\left(\sum_{j=1}^{5}B_j(z_2)\right)\mathbf{s}_i=\sum_{j=1}^{5}C_{2j},$$

$$C_3(z_1,z_2) = \frac{1}{np^2}\sum_{i<k} E_k\left[\frac{\zeta_i(z_1)}{\sqrt{np}}\mathbf{s}_i^T\mathbf{M}_{ki}(z_1)\left(\sum_{j=1}^5 B_j(z_2)\right)\mathbf{s}_i\right]$$

$$+\frac{1}{np^2}\sum_{i>k} E_k\left[\frac{\zeta_i(z_1)}{\sqrt{np}}\mathbf{s}_i^T\mathbf{M}_{ki}(z_1)E_k\mathbf{M}_k^{(1)}(z_2)\mathbf{s}_i\right] = \sum_{j=1}^6 C_{3j},$$

$$C_4(z_1,z_2) = \frac{1}{np^2}\sum_{i<k} E_k\left[\zeta_i(z_1)\vartheta_i(z_1)\mathbf{s}_i^T\left(\sum_{j=1}^5 B_j(z_2)\right)\mathbf{s}_i\right]$$

$$+\frac{1}{np^2}\sum_{i>k} E_k\left[\zeta_i(z_1)\vartheta_i(z_1)\mathbf{s}_i^T E_k\mathbf{M}_k^{(1)}(z_2)\mathbf{s}_i\right] = \sum_{j=1}^6 C_{4j},$$

where C_{2j} corresponds to B_j, $j=1,\ldots,5$, for example, $C_{21} = -\frac{1}{np^2}\sum_{i<k}\mathbf{s}_i^T(B_1(z_2))\mathbf{s}_i$, and C_{3j} and C_{4j} are similarly defined. Here both $C_3(z_1,z_2)$ and $C_4(z_1,z_2)$ are broken up into two parts in terms of $i>k$ or $i<k$. As will be seen, the terms in (5.14) tend to 0 in L_1, except C_{25}, C_{34}, C_{45}. Next let us demonstrate the details.

5.2.2. *Conclusion of the asymptotic expansion of* $\mathbb{Z}_k$

The purpose is to analyze each term in $C_j(z_1,z_2)$, $j=1,2,3,4$. We first claim the limits of ζ_i, ϑ_i which appear in $C_j(z_1,z_2)$ for $j=2,3,4$:

$$\vartheta_i \xrightarrow{L_4} m(z)/z, \qquad \zeta_i(z) \xrightarrow{L_4} -zm(z), \qquad \text{as } n\to\infty. \tag{5.15}$$

Indeed, by (5.8) and (5.10), we have

$$\vartheta_i = \frac{1}{znp}\mathbf{s}_i^T\mathbf{M}_{ki}\mathbf{s}_i - \frac{1}{z\sqrt{np}}\big(\mathbf{s}_i^T\mathbf{s}_i - p\big). \tag{5.16}$$

Replacing $\mathbf{M}_k^{(m)}$ in $\gamma_{km}(z)$ by $\mathbf{M}_{ki}$, by a proof similar to that of (4.11), we have

$$E\left|\frac{1}{np}\mathbf{s}_i^T\mathbf{M}_{ki}\mathbf{s}_i - \frac{1}{np}\operatorname{tr}\mathbf{M}_{ki}\right|^4 \le K\left(\frac{1}{n^2}+\frac{1}{np}\right). \tag{5.17}$$

By (4.6), we then have $\vartheta_i - \frac{1}{zn}\operatorname{tr}\mathbf{D}_{ki}^{-1} \xrightarrow{L_4} 0$. To investigate the distance between $\operatorname{tr}\mathbf{D}_{ki}^{-1}$ and $\operatorname{tr}\mathbf{D}_k^{-1}$, let $\dot{\mathbf{A}}_{ki}$ be the matrix constructed from $\mathbf{A}_k$ by deleting its ith (or $(i-1)$th) row and ith (or $(i-1)$th) column and write $\dot{\mathbf{D}}_{ki} \triangleq \dot{\mathbf{D}}_{ki}(z) = \dot{\mathbf{A}}_{ki} - z\mathbf{I}_{n-2}$ if $i<k$ (or $i>k$). We observe that $\dot{\mathbf{D}}_{ki}^{-1}$ can be obtained from $\mathbf{D}_{ki}^{-1}$ by deleting the ith (or $(i-1)$th) row and ith (or $(i-1)$th) column if $i<k$ (or $i>k$). Then $\operatorname{tr}\mathbf{D}_{ki}^{-1} - \operatorname{tr}\dot{\mathbf{D}}_{ki}^{-1} = -\frac{1}{z}$. By an identity similar to (4.2) and an inequality similar to the bound (4.8), we also have $|\operatorname{tr}\mathbf{D}_k^{-1} - \operatorname{tr}\dot{\mathbf{D}}_{ki}| \le 1/v_1$. Hence $|\operatorname{tr}\mathbf{D}_k^{-1} - \operatorname{tr}\mathbf{D}_{ki}^{-1}| \le (1/v_1 + 1/|z|)$. From (4.2), we have $|\operatorname{tr}\mathbf{D}_k^{-1} - \operatorname{tr}\dot{\mathbf{D}}| \le 1/v_1$ as well. As $\frac{1}{n}\operatorname{tr}\mathbf{D}^{-1} \xrightarrow{L_t} m(z)$ for any fixed t by the Helly–Bray theorem and the dominated convergence theorem, we obtain the first conclusion of (5.15).

Since the imaginary part of $(z\zeta_i)^{-1}$ is $(\Im(z) + \frac{1}{np}\Im(\mathbf{s}_i^T \mathbf{M}_{ki}\mathbf{s}_i))$ whose absolute value is greater than v_1, we have $|\zeta_i| \le |z|/v_1$. Consequently, via (4.1), we complete the proof of the second consequence of (5.15), as claimed.

Consider $C_1(z_1, z_2)$ first. By (4.6),

$$\begin{aligned} E\big|C_1(z_1,z_2)\big| &= E\left|-\frac{1}{n^2 p}(n-k)\operatorname{tr} E_k \mathbf{M}_k^{(1)}(z_2)\right| \\ &\le \frac{K}{np^2} n^2 p = K\frac{n}{p} \to 0. \end{aligned} \tag{5.18}$$

Before proceeding, we introduce the inequalities for further simplification in the following. By Lemma 8.10 in Bai and Silverstein [3] and (4.6), for any matrix $\mathbf{B}$ independent of $\mathbf{s}_i$,

$$E\big|\mathbf{s}_i^T \mathbf{M}_{ki}\mathbf{B}\mathbf{s}_i\big|^2 \le K\big(E\big|\mathbf{s}_i^T \mathbf{M}_{ki}\mathbf{B}\mathbf{s}_i - \operatorname{tr}\mathbf{M}_{ki}\mathbf{B}\big|^2 + KE|\operatorname{tr}\mathbf{M}_{ki}\mathbf{B}|^2\big) \le Kp^2n^2E\|\mathbf{B}\|^2, \tag{5.19}$$

where we also use the fact that, via (4.6),

$$\begin{aligned} |\operatorname{tr}\mathbf{M}_{ki}\mathbf{B}\overline{\mathbf{B}}\overline{\mathbf{M}}_{ki}| &= \big|\operatorname{tr}\mathbf{D}_{ki}^{-1/2}\mathbf{X}_{ki}^T\mathbf{B}\overline{\mathbf{B}}\mathbf{X}_{ki}\overline{\mathbf{D}}_{ki}^{-1}\mathbf{X}_{ki}^T\mathbf{X}_{ki}\mathbf{D}_{ki}^{-1/2}\big| \\ &\le n\big\|\mathbf{D}_{ki}^{-1/2}\mathbf{X}_{ki}^T\big\|^2 \cdot \|\mathbf{B}\|^2 \cdot \big\|\mathbf{X}_{ki}\overline{\mathbf{D}}_{ki}^{-1}\mathbf{X}_{ki}^T\big\| \\ &= n\cdot\|\mathbf{B}\|^2\cdot\big\|\mathbf{D}_{ki}^{-1}\mathbf{X}_{ki}^T\mathbf{X}_{ki}\big\|^2 \\ &= n\cdot\|\mathbf{B}\|^2\cdot\big\|p\mathbf{D}_{ki}^{-1} + \sqrt{np}\big(I_{n-1} + z\mathbf{D}_{ki}^{-1}\big)\big\|^2 \\ &\le Knp^2\|\mathbf{B}\|^2. \end{aligned}$$

For $i > k$, since $E_k\mathbf{M}_k$ is independent of $\mathbf{s}_i$, we similarly have

$$E\big|\mathbf{s}_i^T E_k\mathbf{M}_k\mathbf{B}\mathbf{s}_i\big|^2 \le Kn^2p^2. \tag{5.20}$$

Applying Cauchy–Schwarz's inequality, (5.19) with $\mathbf{B} = \mathbf{I}_{n-1}$ and the fact that $|\zeta_i|$ is bounded by $|z|/v_1$ we have

$$E|C_{2j}| \le K\sqrt{\frac{n}{p}}, \qquad j = 1,2,3,4. \tag{5.21}$$

Using (5.19) with $\mathbf{B} = E_k\mathbf{M}_{ki}(z_2)$ or $\mathbf{B} = E_k\mathbf{M}_k$ in (5.19), we also have

$$E|C_{3j}| \le K\sqrt{\frac{n}{p}}, \qquad j = 1,2,3,4. \tag{5.22}$$

By (5.20), (5.15) and (5.19) with $\mathbf{B} = \mathbf{I}_{n-1}$, we obtain

$$E|C_{4j}| \le K\frac{n}{p}, \qquad j = 1,2,3,4,6. \tag{5.23}$$

Consider C_{32} now. Define $\breve{\zeta}_i$ and $\breve{\mathbf{M}}_{ki}$, the analogues of $\zeta_i(z)$ and $\mathbf{M}_{ki}(z)$ respectively, by $(\mathbf{s}_1, \ldots, \mathbf{s}_k, \breve{\mathbf{s}}_{k+1}, \ldots, \breve{\mathbf{s}}_n)^T$, where $\breve{\mathbf{s}}_{k+1}, \ldots, \breve{\mathbf{s}}_n$ are i.i.d. copies of $\mathbf{s}_{k+1}, \ldots, \mathbf{s}_n$ and independent of $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Then $\breve{\zeta}_i, \breve{\mathbf{M}}_{ki}$ have the same properties as $\zeta_i(z), \mathbf{M}_{ki}(z)$, respectively. Therefore, $|\breve{\zeta}_i| \le |z|/v_1$ and $\|\breve{\mathbf{M}}_{ki}\| \le Kp$. Applying (5.19) with $\mathbf{B} = \breve{\mathbf{M}}_{ki}(z_1)$, we have

$$\begin{aligned}
E|C_{32}| &= E\left|\frac{1}{np^2}\sum_{i<k} E_k E_k\left(\frac{\zeta_i(z_1)}{\sqrt{np}}\mathbf{s}_i^T\mathbf{M}_{ki}(z_1)\frac{\breve{\zeta}_i(z_2)}{z_2 np}\breve{\mathbf{M}}_{ki}(z_2)\mathbf{s}_i\mathbf{s}_i^T\breve{\mathbf{M}}_{ki}(z_2)\mathbf{s}_i\right)\right| \\
&\le \frac{K}{n^2p^3\sqrt{np}}\sum_{i<k} E^{1/2}\left|\mathbf{s}_i^T\mathbf{M}_{ki}(z_1)\breve{\mathbf{M}}_{ki}(z_2)\mathbf{s}_i\right|^2 \cdot E^{1/2}\left|\mathbf{s}_i^T\breve{\mathbf{M}}_{ki}(z_2)\mathbf{s}_i\right|^2 \qquad (5.24)\\
&\le K\sqrt{\frac{n}{p}}.
\end{aligned}$$

Third, consider C_{25}. In view of (5.15), it is straightforward to check that

$$C_{25} = -\frac{k}{n}m(z_2) + o_{L_1}(1). \qquad (5.25)$$

Further, consider C_{34}. By (5.15) and (5.19), we have

$$\begin{aligned}
C_{34} &= \frac{1}{np^2}\sum_{i<k} E_k\left[\frac{\zeta_i(z_1)}{\sqrt{np}}\mathbf{s}_i^T\mathbf{M}_{ki}(z_1)B_4(z_2)\mathbf{s}_i\right] \\
&= \frac{1}{np^2}\sum_{i<k} E_k\left[\frac{\zeta_i(z_1)}{\sqrt{np}}\mathbf{s}_i^T\mathbf{M}_{ki}(z_1)E_k\left(\frac{\zeta_i(z_2)}{z_2\sqrt{np}}\mathbf{M}_{ki}(z_2)\right)\mathbf{s}_i\mathbf{s}_i^T\mathbf{s}_i\right] \\
&= z_1 m(z_1)m(z_2)\frac{1}{n^2p^2}\sum_{i<k}\mathbf{s}_i^T E_k\mathbf{M}_{ki}(z_1)\cdot E_k\mathbf{M}_{ki}(z_2)\mathbf{s}_i + o_{L_1}(1) \qquad (5.26)\\
&= z_1 m(z_1)m(z_2)\frac{1}{n^2p^2}\sum_{i<k}\operatorname{tr}\big(E_k\mathbf{M}_{ki}(z_1)\cdot E_k\mathbf{M}_{ki}(z_2)\big) + o_{L_1}(1) \\
&= z_1 m(z_1)m(z_2)\frac{k}{n}\mathbb{Z}_k + o_{L_1}(1),
\end{aligned}$$

where the last step uses the fact that via (5.11), (5.19), (5.8) and a tedious but elementary calculation

$$\frac{1}{np^2}\left|\operatorname{tr}\big(E_k\mathbf{M}_{ki}(z_1)\cdot E_k\mathbf{M}_{ki}(z_2)\big) - \operatorname{tr} E_k\big(\mathbf{X}_k\mathbf{D}_k^{-1}(z_1)\mathbf{X}_k^T\big)\cdot E_k\big(\mathbf{X}_k\mathbf{D}_k^{-1}(z_2)\mathbf{X}_k^T\big)\right| \le \frac{K}{n}.$$

Consider C_{45} finally. By (5.15), we have

$$C_{45} = -m^2(z_1)m(z_2)\frac{k}{n} + o_{L_1}(1). \qquad (5.27)$$

We conclude from (5.14), (5.18), (5.21)–(5.27) and the fact $m^2(z)+zm(z)+1=0$ that

$$\begin{aligned}z_1\mathbb{Z}_k &= -\frac{k}{n}m(z_2)-\frac{k}{n}m^2(z_1)m(z_2)+\frac{k}{n}z_1m(z_1)m(z_2)\mathbb{Z}_k+\mathrm{o}_{L_1}(1)\\ &= \frac{k}{n}z_1m(z_1)m(z_2)+\frac{k}{n}z_1m(z_1)m(z_2)\mathbb{Z}_k+\mathrm{o}_{L_1}(1),\end{aligned}$$

which is equivalent to

$$\mathbb{Z}_k=\frac{(k/n)m(z_1)m(z_2)}{1-(k/n)m(z_1)m(z_2)}+\mathrm{o}_{L_1}(1). \tag{5.28}$$

5.3. Proof of Condition 5.1

The equality (5.28) ensures that

$$\begin{aligned}&\frac{1}{n^2p^2}\sum_{k=1}^{n}\operatorname{tr}E_k\mathbf{M}_k^{(1)}(z_1)\cdot E_k\mathbf{M}_k^{(1)}(z_2)=\frac{1}{n}\sum_{k=1}^{n}\mathbb{Z}_k\\ &\to\quad\int_0^1\frac{tm(z_1)m(z_2)}{1-tm(z_1)m(z_2)}\,\mathrm{d}t=-1-\big(m(z_1)m(z_2)\big)^{-1}\log\big(1-m(z_1)m(z_2)\big).\end{aligned}$$

Thus, via (5.6), we obtain

$$\tilde{\Lambda}_n(z_1,z_2)\xrightarrow{\text{i.p.}}\nu_4-3-2\big(m(z_1)m(z_2)\big)^{-1}\log\big(1-m(z_1)m(z_2)\big).$$

Consequently, by (5.4)

$$\begin{aligned}\Lambda(z_1,z_2)&=\frac{\partial^2}{\partial z_1\,\partial z_2}\big[(\nu_4-3)m(z_1)m(z_2)-2\log\big(1-m(z_1)m(z_2)\big)\big]\\ &=m'(z_1)m'(z_2)\big[\nu_4-3+2\big(1-m(z_1)m(z_2)\big)^{-2}\big].\end{aligned} \tag{5.29}$$

5.4. Tightness of $M_n^{(1)}(z)$

This section is to prove the tightness of $M_n^{(1)}(z)$ for $z\in\mathbb{C}_1$. By (4.7) and Lemma 4.1,

$$E\left|\sum_{k=1}^{n}\sum_{j=1}^{l}a_jE_{k-1}\big(\alpha_k(z_j)\big)\right|^2\le K\sum_{k=1}^{n}\sum_{j=1}^{l}|a_j|^2E\big|\alpha_k(z_j)\big|^2\le K,$$

which ensures condition (i) of Theorem 12.3 of Billingsley [8]. Condition (ii) of Theorem 12.3 of Billingsley [8] will be verified by proving

$$\frac{E|M_n^{(1)}(z_1)-M_n^{(1)}(z_2)|^2}{|z_1-z_2|^2}\le K,\qquad z_1,z_2\in\mathbb{C}_1. \tag{5.30}$$

We employ the same notations as those in Section 4.1. Let

$$\begin{aligned}\Upsilon_{k1} &= \frac{1}{np}\mathbf{s}_k^T\mathbf{X}_k\mathbf{D}_k^{-1}(z_1)\big(\mathbf{D}_k^{-1}(z_1)+\mathbf{D}_k^{-1}(z_2)\big)\mathbf{D}_k^{-1}(z_2)\mathbf{X}_k^T\mathbf{s}_k \\ &\quad - \frac{1}{np}\operatorname{tr}\mathbf{X}_k\mathbf{D}_k^{-1}(z_1)\big(\mathbf{D}_k^{-1}(z_1)+\mathbf{D}_k^{-1}(z_2)\big)\mathbf{D}_k^{-1}(z_2)\mathbf{X}_k^T, \\ \Upsilon_{k2} &= \frac{1}{np}\big(\mathbf{s}_k^T\mathbf{X}_k\mathbf{D}_k^{-1}(z_2)\mathbf{D}_k^{-1}(z_1)\mathbf{X}_k^T\mathbf{s}_k - \operatorname{tr}\mathbf{X}_k\mathbf{D}_k^{-1}(z_2)\mathbf{D}_k^{-1}(z_1)\mathbf{X}_k^T\big), \\ d_{k1}(z) &= \beta_k(z)\left(1+\frac{1}{np}\mathbf{s}_k^T\mathbf{M}_k^{(2)}(z)\mathbf{s}_k\right), \\ d_{k2}(z) &= 1+\frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(2)}(z), \\ d_{k3} &= 1+\frac{1}{np}\operatorname{tr}\mathbf{X}_k\mathbf{D}_k^{-1}(z_2)\mathbf{D}_k^{-1}(z_1)\mathbf{X}_k^T, \\ d_{k4} &= \frac{1}{np}\operatorname{tr}\mathbf{X}_k\mathbf{D}_k^{-1}(z_1)\big(\mathbf{D}_k^{-1}(z_1)+\mathbf{D}_k^{-1}(z_2)\big)\mathbf{D}_k^{-1}(z_2)\mathbf{X}_k^T.\end{aligned}$$

As in (4.3), we write

$$\begin{aligned}&M_n^{(1)}(z_1) - M_n^{(1)}(z_2) \\ &\quad = -\sum_{k=1}^n (E_k - E_{k-1})\big(d_{k1}(z_1) - d_{k1}(z_2)\big) \\ &\quad = -(z_1 - z_2)\sum_{k=1}^n (E_k - E_{k-1})\big[\beta_k(z_1)(\Upsilon_{k1}+d_{k4}) - \beta_k(z_1)d_{k1}(z_2)(\Upsilon_{k2}+d_{k3})\big] \\ &\quad = -(z_1 - z_2)\sum_{k=1}^n (E_k - E_{k-1}) \\ &\qquad\qquad \times \big[(l_1+l_2)+l_3-\beta_k(z_1)\beta_k(z_2)d_{k2}d_{k3}-\beta_k(z_1)\beta_k(z_2)d_{k3}\gamma_k(z_2)\big] \\ &\quad = -(z_1 - z_2)\sum_{k=1}^n (E_k - E_{k-1})(l_1+l_2+l_3+l_4+l_5+l_6),\end{aligned}$$

where

$$\begin{aligned}&l_1 = \beta_k(z_1)\Upsilon_{k1}, \qquad l_2 = \beta_k(z_1)\beta_k^{\mathrm{tr}}(z_1)\eta_k(z_1)d_{k4}, \\ &l_3 = -\beta_k(z_1)\Upsilon_{k2}d_{k1}(z_1), \qquad l_4 = -\beta_k(z_1)\beta_k^{\mathrm{tr}}(z_1)\eta_k(z_1)\beta_k(z_2)d_{k2}(z_2)d_{k3}, \\ &l_5 = -\beta_k^{\mathrm{tr}}(z_1)\beta_k(z_2)\beta_k^{\mathrm{tr}}(z_2)\eta_k(z_2)d_{k2}(z_2)d_{k3}, \qquad l_6 = -\beta_k(z_1)\beta_k(z_2)d_{k3}\gamma_k(z_2).\end{aligned}$$

Here the last step uses (4.5) for $\beta_k(z_1)$ and the facts that

$$\mathbf{D}_k^{-2}(z_1) - \mathbf{D}_k^{-2}(z_2) = (z_1 - z_2)\mathbf{D}_k^{-1}(z_1)\big(\mathbf{D}_k^{-1}(z_1) + \mathbf{D}_k^{-1}(z_2)\big)\mathbf{D}_k^{-1}(z_2),$$

$$\beta_k(z_1) - \beta_k(z_2) = (z_2 - z_1)\beta_k(z_1)\beta_k(z_2)\Upsilon_{k2} + (z_2 - z_1)\beta_k(z_1)\beta_k(z_2)d_{k3},$$

$$(E_k - E_{k-1})\beta_k^{\mathrm{tr}}(z_1)d_{k4} = 0, \qquad (E_k - E_{k-1})\beta_k^{\mathrm{tr}}(z_1)\beta_k^{\mathrm{tr}}(z_2)d_{k2}(z_2)d_{k3} = 0.$$

By (4.6) and Lemma 8.10 in Bai and Silverstein [3], without any tedious calculations, one may verify that

$$\big|d_{kj}(z)\big| \le K, \qquad j = 1, 2, 3, 4, \quad \text{and} \quad E|\Upsilon_{kj}|^2 \le Kp^{-1}, \qquad j = 1, 2.$$

The above inequalities, together with Burkholder's inequality, imply (5.30).

6. Uniform convergence of $EM_n(z)$

To finish the proof of Proposition 3.1, it remains to derive an asymptotic expansion of $n(Em_n(z) - m(z))$ for $z \in \mathbb{C}_1$ (defined in Section 3.2). In order to unify the proof of Theorem 1.1 and Corollary 1.1, we derive the asymptotic expansion of $n(Em_n(z) - m(z))$ under both assumptions $n/p \to 0$ and $n^3/p = \mathrm{O}(1)$ in Proposition 6.1. For the purpose of proving (3.6), we will prove a stronger result in Proposition 6.1, namely uniform convergence of $n(Em_n(z) - m(z))$ for $z \in \varsigma_n = \bigcup_{i=l,r,u} \varsigma_i$. For z located in the wider range ς_n, the bounds or limits in Section 3 (e.g., Lemma 4.1, (5.3), (5.15)), cannot be applied directly. Hence in Section 4, we re-establish these and other useful results. Throughout this section, we assume $z \in \varsigma_n$ and use the same notations as those in Section 3.

Proposition 6.1. *Suppose that assumption* (c) *is satisfied.*

(i) *Under assumption* (b1): $n/p \to 0$, *we have the following asymptotic expansion*

$$n\big[Em_n(z) - m(z) - \mathcal{X}_n\big(m(z)\big)\big] = \mathrm{o}(1), \tag{6.1}$$

uniformly for $z \in \varsigma_n = \bigcup_{i=l,r,u} \varsigma_i$, *where* $\mathcal{X}_n(m)$ *is defined in* (1.4).

(ii) *Under assumption* (b2): $n^3/p = \mathrm{O}(1)$, *we have the following asymptotic expansion*

$$\begin{aligned} &n\left[Em_n(z) - m(z) + \sqrt{\frac{n}{p}}m^4(z)\big(1 + m'(z)\big)\right] \\ &\quad = m^3(z)\big(m'(z) + \nu_4 - 2\big)\big(1 + m'(z)\big) + \mathrm{o}(1), \end{aligned} \tag{6.2}$$

uniformly for $z \in \varsigma_n = \bigcup_{i=l,r,u} \varsigma_i$.

This, together with (5.29) and the tightness of $M_n^{(1)}(z)$ in Section 5.4, implies Proposition 3.1. It remains to prove Proposition 6.1. To facilitate statements, let

$$\omega_n = \frac{1}{n}\sum_{k=1}^{n} m(z)\beta_k\bar{\mu}_k, \qquad \bar{\varepsilon}_n = \frac{1}{z + E(1/(np))\operatorname{tr}\mathbf{X}\mathbf{D}^{-1}(z)\mathbf{X}^T}.$$

Here, $\omega_n, \bar{\varepsilon}_n$ all depend on z and n, and $\bar{\varepsilon}_n$ are non-random.

Lemma 6.1. *Let $z \in \varsigma_n$. We have*

$$nE\omega_n = m^3(z)\big(m'(z) + \nu_4 - 2\big) + \mathrm{o}(1).$$

Assuming that Lemma 6.1 is true for the moment, whose proof is given in Section 6.1 below, let us demonstrate how to get Proposition 6.1. By (3.8) in Bai [4], we obtain

$$m_n(z) = \frac{1}{n}\operatorname{tr}\mathbf{D}^{-1}(z) = -\frac{1}{n}\sum_{k=1}^{n}\beta_k. \tag{6.3}$$

Applying (4.1), (6.3), (4.6) and taking the difference between β_k and $\frac{1}{z+m(z)}$, we have

$$\begin{aligned} Em_n(z) - m(z) &= -\frac{1}{n}\sum_{k=1}^{n}E\beta_k + \frac{1}{z+m(z)} \\ &= E\frac{1}{n}\sum_{k=1}^{n}\beta_k m(z)\left[\bar{\mu}_k - \big(Em_n(z) - m(z)\big) - \sqrt{\frac{n}{p}}\big(1 + zEm_n(z)\big)\right] \qquad (6.4) \\ &= E\omega_n + m(z)Em_n(z)\big(Em_n(z) - m(z)\big) + \sqrt{\frac{n}{p}}m(z)Em_n(z)\big(1 + zEm_n(z)\big). \end{aligned}$$

Under assumption $n/p \to 0$: Let Em_n, m respectively, denote $Em_n(z), m(z)$ to simplify the notations below. By (4.1) and (6.4), we have

$$\begin{aligned} Em_n - m &= E\omega_n + m^2(Em_n - m) + m(Em_n - m)^2 + \sqrt{\frac{n}{p}}m(Em_n - m)(1 + zm) \\ &\quad + \sqrt{\frac{n}{p}}m^2(1 + zm) + \sqrt{\frac{n}{p}}zm(Em_n - m)^2 + \sqrt{\frac{n}{p}}zm^2(Em_n - m) \\ &= \mathcal{A}(Em_n - m)^2 + (\mathcal{B} + 1)(Em_n - m) + \mathcal{C}_n, \end{aligned}$$

where $\mathcal{A}$, $\mathcal{B}$ are defined in (1.4) and

$$\mathcal{C}_n = E\omega_n - \sqrt{\frac{n}{p}}m^4.$$

Rearranging the above equation, we observe that $(Em_n - m)$ satisfies the equation $\mathcal{A}x^2 + \mathcal{B}x + \mathcal{C}_n = 0$. Solving the equation, we obtain

$$x_{(1)} = \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}_n}}{2\mathcal{A}}, \qquad x_{(2)} = \frac{-\mathcal{B} - \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}_n}}{2\mathcal{A}},$$

where $\sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}_n}$ is a complex number whose imaginary part has the same sign as that of $\mathcal{B}$. By the assumption $n/p \to 0$ and Lemma 6.1, we have $4\mathcal{A}\mathcal{C}_n \to 0$. Then $x_{(1)} = \mathrm{o}(1)$ and $x_{(2)} = \frac{1-m^2}{m} + \mathrm{o}(1)$. Since $Em_n - m = \mathrm{o}(1)$ by (6.5), we choose $Em_n - m = x_{(1)}$. Applying Lemma 6.1 and the definition of $\mathcal{X}_n(m)$ in (1.4), we have

$$\begin{aligned} n\big[Em_n(z) - m(z) - \mathcal{X}_n\big(m(z)\big)\big] &= \frac{-4\mathcal{A}[nE\omega_n - m^3(z)(m'(z) + \nu_4 - 2)]}{2\mathcal{A}(\sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}_n} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}})} \\ &\to 0. \end{aligned}$$

Hence Proposition 6.1(i) is proved.

Under assumption $n^3/p = \mathrm{O}(1)$: subtracting $m(z)Em_n(z)(Em_n(z) - m(z))$ on the both sides of (6.4) and then dividing $\frac{1}{n}(1 - m(z)Em_n(z))$, we have

$$\begin{aligned} n\big(Em_n(z) - m(z)\big) &= \frac{nE\omega_n}{1 - m(z)Em_n(z)} + \sqrt{\frac{n^3}{p}} \frac{m(z)Em_n(z)(1 + zEm_n(z))}{1 - m(z)Em_n(z)} \\ &= \frac{m^3(z)}{1 - m^2(z)}\big(m'(z) + \nu_4 - 2\big) - \sqrt{\frac{n^3}{p}} \frac{m^4(z)}{1 - m^2(z)} + \mathrm{o}\bigg(\sqrt{\frac{n^3}{p}}\bigg), \end{aligned}$$

where we use (4.34), Lemma 6.1, (4.1) and the fact that $m'(z) = \frac{m^2(z)}{1-m^2(z)}$. Proposition 6.1(ii) is proved. Hence, the proof of Proposition 6.1 is completed. Now it remains to prove Lemma 6.1.

6.1. Proof of Lemma 6.1

From the definitions of $\beta_k, \bar{\varepsilon}_n$ and $\bar{\mu}_k$ (see Lemma 4.3), we obtain

$$\beta_k = \bar{\varepsilon}_n + \beta_k \bar{\varepsilon}_n \bar{\mu}_k. \tag{6.5}$$

By (6.5), we further write β_k as $\beta_k = \bar{\varepsilon}_n + \bar{\varepsilon}_n^2 \bar{\mu}_k + \bar{\varepsilon}_n^3 \bar{\mu}_k^2 + \beta_k \bar{\varepsilon}_n^3 \bar{\mu}_k^3$, which ensures that

$$\begin{aligned} nE\omega_n &= m(z)\bar{\varepsilon}_n \sum_{k=1}^n E(\bar{\mu}_k) + m(z)\bar{\varepsilon}_n^2 \sum_{k=1}^n E\big(\bar{\mu}_k^2\big) \\ &\quad + m(z)\bar{\varepsilon}_n^3 \sum_{k=1}^n E\big(\bar{\mu}_k^3\big) + m(z)\bar{\varepsilon}_n^3 \sum_{k=1}^n E\big(\beta_k \bar{\mu}_k^4\big) \\ &\triangleq H_1 + H_2 + H_3 + H_4, \end{aligned} \tag{6.6}$$

where H_j, $j = 1, 2, 3, 4$ are defined in the obvious way. As will be seen, H_3 and H_4 are both negligible and the contribution to the limit of $nE\omega_n$ comes from H_1 and H_2. Now, we analyze H_j, $j = 1, \dots, 4$ one by one.

Consider H_4 first. It follows from (4.1) and (4.34) that

$$\bar{\varepsilon}_n = \frac{1}{z + m(z) + \mathrm{o}(1)} = -m(z) + \mathrm{o}(1). \tag{6.7}$$

By Lemma 4.2 and Lemma 4.3,

$$E\left|\beta_k \bar{\mu}_k^4\right| \le K E\left|\bar{\mu}_k^4\right| I\left(U_n^c\right) + E\left|\beta_k \bar{\mu}_k^4\right| I(U_n) \le K\left(\frac{\delta^4}{n} + \frac{n}{p^2}\right) + \mathrm{o}\left(n^{-\ell}\right) \le K\delta^4 n^{-1},$$

which, together with (6.7), further implies

$$H_4 = \mathrm{o}(1). \tag{6.8}$$

It follows from Lemma 4.3 and (6.7) that

$$H_3 = \mathrm{o}(1). \tag{6.9}$$

Consider H_1 next. We have, via (4.6) and (4.2),

$$\begin{aligned} H_1 &= m(z)\bar{\varepsilon}_n \sum_{k=1}^{n} \left(E\frac{1}{np}\operatorname{tr}\mathbf{X}\mathbf{D}^{-1}\mathbf{X}^T - E\frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(1)} \right) \\ &= \left(1 + z\sqrt{\frac{n}{p}}\right) m(z)\bar{\varepsilon}_n \frac{1}{n}\sum_{k=1}^{n} E\left(\operatorname{tr}\mathbf{D}^{-1} - \operatorname{tr}\mathbf{D}_k^{-1}\right) + \sqrt{\frac{n}{p}} m(z)\bar{\varepsilon}_n \\ &= -\left(1 + z\sqrt{\frac{n}{p}}\right) m(z)\bar{\varepsilon}_n \frac{1}{n}\sum_{k=1}^{n} E\left[\beta_k\left(1 + \frac{1}{np}\mathbf{s}_k^T\mathbf{M}_k^{(2)}\mathbf{s}_k\right)\right] + \sqrt{\frac{n}{p}} m(z)\bar{\varepsilon}_n. \end{aligned} \tag{6.10}$$

Applying (4.23), (4.26) and (4.34), it is easy to see

$$1 + \frac{1}{np}\mathbf{s}_k^T\mathbf{M}_k^{(2)}\mathbf{s}_k = 1 + \left(\frac{1}{np}\operatorname{tr}\mathbf{M}_k^{(1)}\right)' + \mathrm{o}_{L_4}(1) = 1 + m'(z) + \mathrm{o}_{L_4}(1).$$

This, together with (6.7), Lemma 4.2 and (6.3), ensures that

$$H_1 = -m^2(z)\left(1 + m'(z)\right) E m_n(z) + \mathrm{o}(1) \to -m^3(z)\left(1 + m'(z)\right). \tag{6.11}$$

Consider H_2 now. By the previous estimation of $E\bar{\mu}_k$ included in H_1 we obtain

$$E\bar{\mu}_k^2 = E(\bar{\mu}_k - E\bar{\mu}_k)^2 + \mathrm{O}\left(n^{-2}\right). \tag{6.12}$$

Furthermore a direct calculation yields

$$E(\bar{\mu}_k - E\bar{\mu}_k)^2 = S_1 + S_2, \tag{6.13}$$

where

$$S_1 = \frac{1}{n} E\big(X_{11}^2 - 1\big)^2 + E\gamma_{k1}^2, \qquad S_2 = S_{21} + S_{22},$$
$$S_{21} = \frac{1}{n^2 p^2} E\big(\operatorname{tr} \mathbf{M}_k^{(1)} - E \operatorname{tr} \mathbf{M}_k^{(1)}\big)^2,$$
$$S_{22} = -\frac{2}{np\sqrt{np}} E\big[\big(\mathbf{s}_k^T \mathbf{s}_k - p\big)\big(\mathbf{s}_k^T \mathbf{M}_k^{(1)} \mathbf{s}_k - E \operatorname{tr} \mathbf{M}_k^{(1)}\big)\big].$$

We claim that

$$nS_1 \to \nu_4 - 1 + 2m'(z), \qquad nS_{21} \to 0, \qquad nS_{22} \to 0, \qquad \text{as } n \to \infty. \tag{6.14}$$

Indeed, with notation $\mathbf{M}_k^{(1)} = (a_{ij}^{(1)})_{p\times p}, i, j = 1, \dots, p$, as illustrated in (4.16), we have $\frac{1}{n^2p^2} \sum_{k=1}^n \sum_{i=1}^p E|a_{ii}^{(1)}|^2 \to 0$. Via this, (4.34) and (4.6), a simple calculation yields

$$\begin{aligned}
nE\gamma_{k1}^2 &= \frac{1}{np^2} E\left(\sum_{i\neq j} X_{ik}X_{jk}a_{ij}^{(1)} + \sum_{i=1}^p \big(X_{ik}^2 - 1\big)a_{ii}^{(1)}\right)^2 \\
&= \frac{1}{np^2} E\left(\sum_{i\neq j}\sum_{s\neq t} X_{ik}X_{jk}X_{sk}X_{tk}a_{ij}^{(1)}a_{st}^{(1)}\right) + \frac{1}{np^2}\sum_{i=1}^p E\big[\big(X_{ik}^2 - 1\big)^2\big(a_{ii}^{(1)}\big)^2\big] \\
&= \frac{2}{np^2} E\left(\sum_{i,j} a_{ij}^{(1)}a_{ji}^{(1)}\right) + \mathrm{o}(1) = \frac{2}{np^2} E \operatorname{tr}\big(\mathbf{M}_k^{(1)}\big)^2 + \mathrm{o}(1) \\
&= \frac{2}{n} E \operatorname{tr} \mathbf{D}_k^{-2} + \mathrm{o}(1) \to 2m'(z).
\end{aligned}$$

Since $E|X_{11}^2 - 1|^2 = \nu_4 - 1$, we have proved the first result of (6.14). By Burkholder's inequality, Lemma 4.4, (4.6), (4.18) and (4.23)

$$n|S_{21}| = K\left(1 + z\sqrt{\frac{n}{p}}\right)^2 \frac{1}{n} E\big|M^{(1)}(z)\big|^2 + Kn^{-1} \le Kn^{-1}. \tag{6.15}$$

Furthermore,

$$\begin{aligned}
n|S_{22}| &= \frac{2}{p\sqrt{np}} \left| E\left(\sum_{t=1}^p \big(X_{tk}^2 - 1\big)\right) \cdot \left(\sum_{i,j} X_{ik}X_{jk}a_{ij}^{(1)}\right)\right| \\
&= \frac{2}{p\sqrt{np}} \big|E\big(X_{11}^2 - 1\big)X_{11}^2 \cdot E \operatorname{tr} \mathbf{M}_k^{(1)}\big| \le K\sqrt{\frac{n}{p}} + \mathrm{o}\big(n^{-\ell}\big) \to 0.
\end{aligned}$$

Therefore, the proof of the second result of (6.14) is completed. We then conclude from (6.14), (6.12), (6.13) and (6.7) that

$$H_2 \to m^3(z)\bigl(2m'(z) + \nu_4 - 1\bigr). \tag{6.16}$$

Finally, by (6.6), (6.8), (6.9), (6.11) and (6.16), we obtain

$$nE\omega_n = m^3(z)\bigl(m'(z) + \nu_4 - 2\bigr) + \mathrm{o}(1).$$

Lemma 6.1 is thus proved. This finishes the proof of Proposition 3.1.

7. Proof of Proposition 3.2

Recall the definition of U_n below Proposition 3.2 or in Section 4. For $i = l, r$, by Lemma 4.4

$$E\left|\int_{\varsigma_i} M_n^{(1)}(z) I\bigl(U_n^c\bigr)\,\mathrm{d}z\right|^2 \le \int_{\varsigma_i} E\bigl|M_n^{(1)}(z)\bigr|^2\,\mathrm{d}z \le K\|\varsigma_i\| \to 0, \qquad \text{as } n \to \infty, v_1 \to 0.$$

Moreover,

$$\left|\int_{\varsigma_i} EM_n(z) I\bigl(U_n^c\bigr)\,\mathrm{d}z\right| \le \int_{\varsigma_i} \bigl|EM_n(z)\bigr|\,\mathrm{d}z \to 0, \qquad \text{as } n \to \infty, v_1 \to 0,$$

where the convergence follows from Proposition 6.1.

8. Calculation of the mean and covariance

To complete the proof of Theorem 1.1 and Corollary 1.1, it remains to calculate the mean function and covariance function of $Y(f)$ and $X(f)$. The computation exactly follows Bai and Yao [6] and so we omit it.

Acknowledgements

The authors would like to thank the editor, the associate editor and the referees' insightful suggestions and comments which significantly improve the quality and the exposition of the work. Particularly the current mean correction term and the calibration of the mean correction term benefit from one of the referees' suggestions. The second author would like to thank Dr. Lin Liang-Ching for his help with the real data analysis. This work was partially supported by the Ministry of Education, Singapore, under grant # ARC 14/11.

References

[1] Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D. and Levine, A.J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **96** 6745–6750.

[2] Bai, Z., Jiang, D., Yao, J.-F. and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

[3] Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175

[4] Bai, Z.D. (1993). Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.* **21** 625–648. MR1217559

[5] Bai, Z.D. and Silverstein, J.W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792

[6] Bai, Z.D. and Yao, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11** 1059–1092. MR2189081

[7] Bai, Z.D. and Yin, Y.Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. MR0929083

[8] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[9] Birke, M. and Dette, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. MR2189467

[10] 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. MR2995802

[11] Chen, S.X., Zhang, L.-X. and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

[12] Costin, O. and Lebowitz, J.L. (1995). Gaussian fluctuation in random matrices. *Phys. Rev. Lett.* **75** 69–72. MR3155254

[13] Donoho, D.L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. In *American Math. Society Conference on Math. Challenges of the 21st Century*.

[14] Fan, J. and Fan, Y. (2008). High-dimensional classification using features annealed independence rules. *Ann. Statist.* **36** 2605–2637. MR2485009

[15] Fisher, T.J. (2012). On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size. *J. Statist. Plann. Inference* **142** 312–326. MR2827150

[16] Fisher, T.J., Sun, X. and Gallagher, C.M. (2010). A new test for sphericity of the covariance matrix for high dimensional data. *J. Multivariate Anal.* **101** 2554–2570. MR2719881

[17] Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D. and Lander, E.S. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286** 531–537.

[18] Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** 151–204. MR1487983

[19] John, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568

[20] Karoui, E.N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity. Available at arXiv:math/0309355.

[21] Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

[22] Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

[23] Pan, G.M. (2014). Comparison between two types of large sample covariance matrices. *Ann*. *Inst*. *Henri Poincare Probab*. *Stat*. **50** 315–713.

[24] Pan, G.M. and Gao, J.T. (2012). Asymptotic theorey for sample covariance matrix under cross-sectional dependence. Preprint.

[25] Pan, G.M. and Zhou, W. (2011). Central limit theorem for Hotelling's T^2 statistic under large dimension. *Ann*. *Appl*. *Probab*. **21** 1860–1910. MR2884053

[26] Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J*. *Japan Statist*. *Soc*. **35** 251–272. MR2328427

[27] Srivastava, M.S., Kollo, T. and von Rosen, D. (2011). Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *J*. *Multivariate Anal*. **102** 1090–1103. MR2793878

[28] Titchmarsh, E.C. (1939). *The Theory of Functions*, 2nd ed. London: Oxford Univ. Press.

Received January 2012 and revised January 2014