The Annals of Statistics
2023, Vol. 51, No. 3, 1427–1451
https://doi.org/10.1214/23-AOS2300

ASYMPTOTIC NORMALITY FOR EIGENVALUE STATISTICS OF A GENERAL SAMPLE COVARIANCE MATRIX WHEN $p/n \to \infty$ AND APPLICATIONS

BY JIAXIN QIU[1,a], ZENG LI[2,b] AND JIANFENG YAO[3,c]

[1]*Department of Statistics and Actuarial Science, The University of Hong Kong,* [a]*qiujx@connect.hku.hk*
[2]*Department of Statistics and Data Science, Southern University of Science and Technology,* [b]*liz9@sustech.edu.cn*
[3]*School of Data Science, The Chinese University of Hong Kong (Shenzhen),* [c]*jeffyao@cuhk.edu.cn*

The asymptotic normality for a large family of eigenvalue statistics of a general sample covariance matrix is derived under the ultrahigh-dimensional setting, that is, when the dimension to sample size ratio $p/n \to \infty$. Based on this CLT result, we extend the covariance matrix test problem to the new ultra-high-dimensional context, and apply it to test a matrix-valued white noise. Simulation experiments are conducted for the investigation of finite-sample properties of the general asymptotic normality of eigenvalue statistics, as well as the two developed tests.

1. Introduction. Let $\mathbf{y} \in \mathbb{R}^p$ be a population of the form $\mathbf{y} = \boldsymbol{\Sigma}_p^{1/2}\mathbf{x}$ where $\boldsymbol{\Sigma}_p$ is a $p \times p$ positive definite matrix, $\mathbf{x} \in \mathbb{R}^p$ a p-dimensional random vector with independent and identically distributed (i.i.d.) components with zero mean and unit variance. Given a sample $\{\mathbf{y}_j = \boldsymbol{\Sigma}_p^{1/2}\mathbf{x}_j\}_{1\le j\le n}$, of $\mathbf{y}$, the sample covariance matrix is $\mathbf{S}_n = \frac{1}{n}\sum_{j=1}^n \mathbf{y}_j\mathbf{y}_j' = \frac{1}{n}\boldsymbol{\Sigma}_p^{1/2}\mathbf{X}\mathbf{X}'\boldsymbol{\Sigma}_p^{1/2}$, where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$. We consider the *ultrahigh-dimensional setting* where $n \to \infty$, $p = p(n) \to \infty$ such that $p/n \to \infty$. The $p \times p$ matrix $\mathbf{S}_n$ has only a small number of nonzero eigenvalues, which are the same as those of its $n \times n$ companion matrix $\underline{\mathbf{S}}_n = \frac{1}{n}\mathbf{X}'\boldsymbol{\Sigma}_p\mathbf{X}$. The limiting distribution of these nonzero eigenvalues is known (see Bai and Yin (1988), Wang and Paul (2014)). Precisely, consider the renormalized sample covariance matrix

$$\mathbf{A}_n = \frac{1}{\sqrt{npb_p}}(\mathbf{X}'\boldsymbol{\Sigma}_p\mathbf{X} - pa_p\mathbf{I}_n), \tag{1}$$

where $\mathbf{I}_n$ is the identity matrix of order n, $a_p = p^{-1}\operatorname{tr}(\boldsymbol{\Sigma}_p)$, $b_p = p^{-1}\operatorname{tr}(\boldsymbol{\Sigma}_p^2)$. Denote the eigenvalues of $\mathbf{A}_n$ as $\lambda_1, \dots, \lambda_n$. According to Wang and Paul (2014), under the condition that $\sup_p \|\boldsymbol{\Sigma}_p\| < \infty$, the eigenvalue distribution of $\mathbf{A}_n$, $F^{\mathbf{A}_n} = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}$, converges to the celebrated semi-circle law. In this paper, we focus on the so-called linear spectral statistics (LSS) of $\mathbf{A}_n$, that is, $\frac{1}{n}\sum_{i=1}^n f(\lambda_i)$ where $f(\cdot)$ is a smooth function we are interested in. The main contribution of this paper is to establish the central limit theorem (CLT) for such LSS of $\mathbf{A}_n$ under the ultrahigh-dimensional setting. The study of asymptotic normality of LSS for different types of random matrix models has received extensive attention in the past decades; see the monographs Bai and Silverstein (2010), Couillet and Debbah (2011), Yao, Zheng and Bai (2015). It plays a very important role in high-dimensional data analysis because many well-established statistics can be represented as LSS of sample covariance or correlation matrix. In facing the curse of dimensionality, most asymptotic results are discussed under the *Marchenko–Pastur* asymptotic regime, where $p/n \to c \in (0, \infty)$. However, this does not fit

Received October 2021; revised June 2023.
MSC2020 subject classifications. Primary 62H10; secondary 62H15.
Key words and phrases. Asymptotic normality, linear spectral statistics, general sample covariance matrix, ultra-dimension, matrix white noise, separable covariance.

the case of ultrahigh dimension when $p \gg n$. Hence, in this paper, we study the asymptotic behavior of LSS of $\mathbf{A}_n$ when $n \to \infty$, $p = p(n) \to \infty$ such that $p/n \to \infty$.

A special version of $\mathbf{A}_n$ for the case where $\mathbf{\Sigma}_p = \mathbf{I}_p$ has already been studied in the literature. The matrix becomes

$$\mathbf{A}_n^{\mathsf{iden}} = \frac{1}{\sqrt{np}}(\mathbf{X}'\mathbf{X} - p\mathbf{I}_n).$$

Bai and Yin (1988) is the first to study this matrix. They proved that the ultrahigh- dimensional limiting eigenvalue distribution of $\mathbf{A}_n^{\mathsf{iden}}$ is the semicircle law. Chen and Pan (2012) studied the behavior of the largest eigenvalue of $\mathbf{A}_n^{\mathsf{iden}}$. Chen and Pan (2015) and Bao (2015) independently established the CLT for LSS of $\mathbf{A}_n^{\mathsf{iden}}$, the limiting variance function of which coincides with that of a Wigner matrix given in Bai and Yao (2005). However, the asymptotic distribution of LSS of $\mathbf{A}_n$ is quite different and worth further investigation.

In this paper, we establish the CLT for LSS of $\mathbf{A}_n$. The general strategy of the proof follows that of Bai and Yao (2005) for the CLT for LSS of a large Wigner matrix. However, the calculations are more involved here as the matrix $\mathbf{A}_n$ is a quadratic function of the independent entries (X_{ij}) while a Wigner matrix in Bai and Yao (2005) is a linear function of its entries. Similar to Chen and Pan (2015), a key step is to establish the CLT for some smooth integral of the Stieltjes transform $M_n(z)$ of $\mathbf{A}_n$; see Proposition 6.1. To derive the limiting mean and covariance functions, we divide $M_n(z)$ into two parts: a nonrandom part and a random part. Our approaches to handling these two parts are technically different from the existing literature. For the random part, the method in Chen and Pan (2015) depends heavily on an explicit expression for $\mathsf{tr}(\mathbf{M}_k^{(1)})/(npb_p)$ (see Section S2.16 for more details). This explicit expression does not exist in our matrix model, so we need to provide a first-order approximation for it, which is given in Lemma S1.21. For the nonrandom part, we utilize the generalized Stein's equation to find the asymptotic expansion of the expectation of Stieltjes transform, which provides some new enlightenment for conventional procedures.

To demonstrate the potential of our newly established CLT, we further studied two hypothesis testing problems about population covariance matrices. First, we examine the identity hypothesis $H_0: \mathbf{\Sigma}_p = \mathbf{I}_p$ under the ultrahigh-dimensional setting and compare it with cases of less high dimensions. Then we use this result to test whether a matrix-valued noise has a prespecified separable covariance matrix. Matrix-valued data are now becoming increasingly important in several fields, for example, in time-series analysis. For a sequence of i.i.d. $p_1 \times p_2$ matrices $\{\mathbf{E}_t\}_{1 \le t \le T}$, we adopt a Frobenius-norm-type statistic to test whether the covariance matrix of $\mathsf{vec}(\mathbf{E}_t)$ equals a prespecified separable covariance, that is, $\mathrm{Cov}(\mathsf{vec}(\mathbf{E}_t)) = \mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2$, where $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are two given $p_1 \times p_1$ and $p_2 \times p_2$ nonnegative definite matrices. Here, p_1, p_2 and T are of comparable magnitude. Our test statistic can be represented as an eigenvalue statistic of the sample covariance matrix with dimension $p_1 p_2$ much larger than the sample size T. Therefore, our CLT can be employed to derive the asymptotic null distribution and perform power analysis of the test. Second, we discuss applying our CLT to the well-known sphericity test under an ultrahigh-dimensional setting. Good numerical performance lends full support to the correctness of our CLT results.

The paper is organized as follows. Section 2 provides preliminary knowledge of some technical tools. Section 3 establishes our main CLT for LSS of $\mathbf{A}_n$. Section 4 contains two hypothesis testing applications. Section 5 reports numerical studies. Technical proofs and lemmas are relegated to Section 6 and the Supplementary Material (Qiu, Li and Yao (2023)).

Throughout the paper, we reserve boldfaced symbols for vectors and matrices. For any matrix $\mathbf{A}$, we let A_{ij}, $\lambda_j^{\mathbf{A}}$, $\mathbf{A}'$, $\mathsf{tr}(\mathbf{A})$, and $\|\mathbf{A}\|$ represent, respectively, its (i, j)th element, its jth largest eigenvalue, its transpose, its trace and its spectral norm (i.e., the largest singular value of $\mathbf{A}$). The notation $\mathbb{1}_{\{\cdot\}}$ stands for the indicator function. For the random variable X, we denote its ath moment by ν_a and its ath cumulant by κ_a. We use K to denote constants that may vary from line to line. For simplicity, we sometimes omit the variable z when representing some matrices and functions (e.g., Stieltjes transforms) of z, provided that it does not lead to confusion.

2. Preliminaries.

In this section, we introduce some useful preliminary results. For any $n \times n$ Hermitian matrix $\mathbf{B}_n$, its empirical spectral distribution (ESD) is defined by

$$F^{\mathbf{B}_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\lambda_i^{\mathbf{B}_n} \le x\}}.$$

If $F^{\mathbf{B}_n}(x)$ converges to a nonrandom limit $F(x)$ as $n \to \infty$, we call $F(x)$ the limiting spectral distribution (LSD) of $\mathbf{B}_n$.

As for the LSD of $\mathbf{A}_n$ defined in (1), Wang and Paul (2014) derived the LSD of renormalized sample covariance matrices with the more general form

$$\mathbf{C}_n = \sqrt{\frac{p}{n}}\left(\frac{1}{p}\mathbf{T}_n^{1/2}\mathbf{X}_n^*\mathbf{\Sigma}_p\mathbf{X}_n\mathbf{T}_n^{1/2} - \frac{1}{p}\operatorname{tr}(\mathbf{\Sigma}_p)\mathbf{T}_n\right), \tag{2}$$

where $\mathbf{X}_n$ and $\mathbf{\Sigma}_p$ are the same as those in (1). Here, $\mathbf{T}_n$ is a $n \times n$ nonnegative definite Hermitian matrix, whose ESD, $F^{\mathbf{T}_n}$, converges weakly to H, a nonrandom distribution function on $\mathbb{R}^+$, which does not degenerate to zero. The LSD of $\mathbf{C}_n$ is described in terms of its Stieltjes transform. The Stieltjes transform of any cumulative distribution function G is defined by

$$m_G(z) = \int \frac{1}{\lambda - z}\,\mathrm{d}G(\lambda), \quad z \in \mathbb{C}^+ := \{u + iv,\ u \in \mathbb{R}, v > 0\}.$$

Wang and Paul (2014) proved that, when $p \wedge n \to \infty$ and $p/n \to \infty$, $F^{\mathbf{C}_n}$ almost surely converges to a nonrandom distribution, whose Stieltjes transform $m_{\mathbf{C}}(z)$ satisfies the following system of equations:

$$\begin{cases} m_{\mathbf{C}}(z) = -\displaystyle\int \frac{\mathrm{d}H(x)}{z + x\theta g(z)}, \\ g(z) = -\displaystyle\int \frac{x\,\mathrm{d}H(x)}{z + x\theta g(z)}, \end{cases} \tag{3}$$

for any $z \in \mathbb{C}^+$, where $\theta = \lim_{p\to\infty}(1/p)\operatorname{tr}(\mathbf{\Sigma}_p^2)$.

Note that $\mathbf{A}_n$ is a special case of $\mathbf{C}_n$ with $\mathbf{T}_n = \mathbf{I}_n$. By (3), we can easily show that the Stieltjes transform $m(z)$ of LSD of $\mathbf{A}_n$ satisfies

$$m(z) = -\frac{1}{z + m(z)}, \tag{4}$$

which is exactly the Stieltjes transform of the semicircle law with density function given by

$$F'(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\,\mathbb{1}_{\{|x| \le 2\}}.$$

3. Main results.

Let $\mathcal{U}$ denote any open region on the complex plane that includes $[-\eta, \eta]$, where $\eta = 2\limsup_p \|\mathbf{\Sigma}_p\|/\sqrt{\theta}$ with $\theta = \lim_p \operatorname{tr}(\mathbf{\Sigma}_p^2)/p$, and $\mathcal{M}$ be the set of analytic functions defined on $\mathcal{U}$. For any $f \in \mathcal{M}$, we consider a LSS of $\mathbf{A}_n$ of the form:

$$\int f(x)\,\mathrm{d}F^{\mathbf{A}_n}(x) = \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i^{\mathbf{A}_n}).$$

Since $F^{\mathbf{A}_n}$ converges to F almost surely, we have

$$\int f(x)\,\mathrm{d}F^{\mathbf{A}_n}(x) \to \int f(x)\,\mathrm{d}F(x).$$

A question naturally arises: how fast does $\int f(x)\,\mathrm{d}\{F^{\mathbf{A}_n}(x) - F(x)\}$ converge to zero?

To answer this question, we consider a renormalized functional:

$$\begin{aligned} G_n(f) &= n \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}\{F^{\mathbf{A}_n}(x) - F(x)\} \\ &\quad - \frac{n}{2\pi i} \oint_{|m|=\rho} f(-m - m^{-1}) \mathcal{X}_n(m) \frac{1 - m^2}{m^2}\,\mathrm{d}m, \end{aligned} \tag{5}$$

where $\rho < 1$,

$$\begin{aligned} \mathcal{X}_n(m) &= \frac{-\mathcal{B}_n(m) + \sqrt{\mathcal{B}_n(m)^2 - 4\mathcal{A}_n(m)\mathcal{C}_n(m)}}{2\mathcal{A}_n(m)}, \\ \mathcal{A}_n(m) &= m - \sqrt{\frac{n}{p}} \frac{c_p}{b_p\sqrt{b_p}}(1 + m^2), \qquad \mathcal{B}_n(m) = m^2 - 1 - \sqrt{\frac{n}{p}} \frac{c_p}{b_p\sqrt{b_p}} m(1 + 2m^2), \\ \mathcal{C}_n(m) &= \frac{m^3}{n}\left\{\frac{1}{1 - m^2} + \frac{(\nu_4 - 3)\tilde{b}_p}{b_p}\right\} - \sqrt{\frac{n}{p}} \frac{c_p}{b_p\sqrt{b_p}} m^4 + \frac{n}{p}\left(-\frac{c_p^2}{b_p^3} + \frac{d_p}{b_p^2}\right) m^5, \\ b_p &= \frac{1}{p}\operatorname{tr}(\mathbf{\Sigma}_p^2), \qquad \tilde{b}_p = \frac{1}{p}\sum_{i=1}^{p}(\mathbf{\Sigma}_p)_{ii}^2, \qquad c_p = \frac{1}{p}\operatorname{tr}(\mathbf{\Sigma}_p^3), \qquad d_p = \frac{1}{p}\operatorname{tr}(\mathbf{\Sigma}_p^4), \end{aligned} \tag{6}$$

where $(\mathbf{\Sigma}_p)_{ii}$ is the ith diagonal element of $\mathbf{\Sigma}_p$, and $\sqrt{\mathcal{B}_n(m)^2 - 4\mathcal{A}_n(m)\mathcal{C}_n(m)}$ is a complex number whose imaginary part has the same sign as that of $\mathcal{B}_n(m)$. The contour integral in (5) is a technical correction to the mean of the statistic that is necessary for describing its asymptotic normality. The main result is formulated in the theorem below.

THEOREM 3.1. *Suppose that*

(A) $\mathbf{X} = (X_{ij})_{p\times n}$ *where* $\{X_{ij},\ 1 \le i \le p,\ 1 \le j \le n\}$ *are i.i.d. real random variables with* $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = 1$, $\mathbb{E}X_{ij}^4 = \nu_4$ *and* $\mathbb{E}|X_{ij}|^{6+\varepsilon_0} < \infty$ *for some positive* ε_0;

(B) $\{\mathbf{\Sigma}_p,\ p \ge 1\}$ *is a sequence of nonnegative definite matrices, bounded in spectral norm, such that the following limits exist*:

- $\gamma = \lim_{p\to\infty} \frac{1}{p}\operatorname{tr}(\mathbf{\Sigma}_p)$,
- $\theta = \lim_{p\to\infty} \frac{1}{p}\operatorname{tr}(\mathbf{\Sigma}_p^2)$,
- $\omega = \lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p}(\mathbf{\Sigma}_p)_{ii}^2$;

(C1) $p \wedge n \to \infty$ *and* $n^2/p = O(1)$.

Then, for any $k \ge 1$ *and* $f_1, \dots, f_k \in \mathcal{M}$, *the* k-*dimensional 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* $\mathbb{E}Y(f) = 0$ *and covariance function*

$$\operatorname{Cov}(Y(f_1), Y(f_2)) = \frac{\omega}{\theta}(\nu_4 - 3)\Psi_1(f_1)\Psi_1(f_2) + 2\sum_{k=1}^{\infty} k\Psi_k(f_1)\Psi_k(f_2) \tag{7}$$

$$= \frac{1}{4\pi^2}\int_{-2}^{2}\int_{-2}^{2} f_1'(x) f_2'(y) H(x, y)\,\mathrm{d}x\,\mathrm{d}y, \tag{8}$$

where

$$\Psi_k(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos x)e^{ikx}\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos x)\cos(kx)\,\mathrm{d}x, \tag{9}$$

$$H(x,y) = \frac{\omega}{\theta}(\nu_4-3)\sqrt{4-x^2}\sqrt{4-y^2} + 2\log\left(\frac{4-xy+\sqrt{(4-x^2)(4-y^2)}}{4-xy-\sqrt{(4-x^2)(4-y^2)}}\right). \tag{10}$$

The proof of Theorem 3.1 is postponed to Section 6.

REMARK 3.1. The appearance of the parameter ω in the limiting covariance function of Theorem 3.1 may surprise. The reason is that for the Gaussian population, one would expect the distribution of linear spectral statistics to depend on $\mathbf{\Sigma}_p$ only through its eigenvalues. This is indeed true since in this case $\nu_4 = 3$ the asymptotic covariances do not depend on ω. However, it does depend on it for non-Gaussian populations. Precisely, this dependence originates from the following identity about covariance of quadratic forms of a p-dimensional isotropic population $\mathbf{x}_1 = (X_{11},\dots,X_{1p})$ with i.i.d. coordinates (Bai and Silverstein (2004), equation (1.15)):

$$\begin{aligned}&\mathbb{E}(\mathbf{x}_1'\mathbf{A}\mathbf{x}_1 - \operatorname{tr}\mathbf{A})(\mathbf{x}_1'\mathbf{B}\mathbf{x}_1 - \operatorname{tr}\mathbf{B})\\ &\quad = \{\mathbb{E}|X_{11}|^4 - |\mathbb{E}X_{11}^2|^2 - 2\}\sum_i A_{ii}B_{ii} + |\mathbb{E}X_{11}^2|^2\operatorname{tr}\mathbf{A}\mathbf{B}' + \operatorname{tr}\mathbf{A}\mathbf{B}\end{aligned}$$

for $p\times p$ symmetric matrices $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$. This identity is used to obtain the explicit expression of the covariance function in Theorem 3.1: precisely, we used this identity with $\mathbf{A} = \mathbf{B} = \mathbf{\Sigma}_p$ to calculate $(npb_p)^{-1}\mathbb{E}(\mathbf{x}_k'\mathbf{\Sigma}_p\mathbf{x}_k - pa_p)^2$; see equation (S2.60). Hence, the first term in the above identity is deterministic and contributes to the parameter ω. More specifically, we have

$$\frac{1}{npb_p}\mathbb{E}(\mathbf{x}_k'\mathbf{\Sigma}_p\mathbf{x}_k - pa_p)^2 = \frac{1}{npb_p}\left\{(\nu_4-3)\sum_i(\mathbf{\Sigma}_p)_{ii}^2 + 2\operatorname{tr}(\mathbf{\Sigma}_p^2)\right\} = \frac{1}{n}\left\{\frac{\widetilde{b}_p}{b_p}(\nu_4-3)+2\right\},$$

where $\widetilde{b}_p = p^{-1}\sum_i(\mathbf{\Sigma}_p)_{ii}^2 \to \omega$ and $b_p = p^{-1}\operatorname{tr}(\mathbf{\Sigma}_p^2) \to \theta$ as $p\to\infty$. In particular, for a Gaussian population $\mathbf{x}_1$, $\mathbb{E}|X_{11}|^4 - |\mathbb{E}X_{11}^2|^2 - 2 = 0$ and the parameter ω disappears in the limiting covariance function.

REMARK 3.2. Note that we require $p \ge Kn^2$ asymptotically in Assumption (C1). In fact, we can relax this assumption to cover the whole range of $n \ll p \ll n^2$ for the dimension p. However, we cannot obtain a closed-form formula for $\mathcal{X}_n(m)$ when $n \ll p \ll n^2$ and this presents a problem in practical applications. More details on this problem are provided in Remark 6.1.

REMARK 3.3. If $\mathbf{\Sigma}_p = \mathbf{I}_p$, we have $a_p = b_p = \tilde{b}_p = c_p = d_p = 1$ and $\gamma = \theta = \omega = 1$, then our Theorem 3.1 reduces to the CLT derived in Chen and Pan (2015).

Applying Theorem 3.1 to three polynomial functions, we obtain the following corollary.

COROLLARY 3.1. *With the same notation and assumptions given in Theorem 3.1, consider three analytic functions $f_1(x) = x$, $f_2(x) = x^2$, $f_3(x) = x^3$. We have*

$$G_n(f_1) = \operatorname{tr}(\mathbf{A}_n) \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega}{\theta}(\nu_4-3)+2\right);$$

$$G_n(f_2) = \mathrm{tr}(\mathbf{A}_n^2) - n - \left\{\frac{\tilde{b}_p}{b_p}(\nu_4 - 3) + 1\right\} \xrightarrow{d} \mathcal{N}(0, 4);$$

$$G_n(f_3) = \mathrm{tr}(\mathbf{A}_n^3) - \frac{c_p}{b_p\sqrt{b_p}}\sqrt{\frac{n}{p}}\left\{n + 1 + \frac{\tilde{b}_p}{b_p}(\nu_4 - 3)\right\} \xrightarrow{d} \mathcal{N}\left(0, \frac{9\omega}{\theta}(\nu_4 - 3) + 24\right).$$

The calculations in these applications are elementary, and thus omitted. Note that the mean correction terms for $G_n(f_1)$, $G_n(f_2)$ and $G_n(f_3)$ are 0, $\frac{\tilde{b}_p}{b_p}(\nu_4 - 3) + 1$, and $\frac{c_p}{b_p\sqrt{b_p}}\sqrt{\frac{n}{p}}\{n + 1 + \frac{\tilde{b}_p}{b_p}(\nu_4 - 3)\}$, respectively.

3.1. *Case of $p \geq Kn^3$.* When $p \geq Kn^3$, the mean correction term in (5) can be further simplified as

$$\begin{aligned}(11)\quad & -\frac{n}{2\pi i}\oint_{|m|=\rho} f(-m - m^{-1})\mathcal{X}_n(m)\frac{1 - m^2}{m^2}\,\mathrm{d}m \\ & = -\left[\frac{1}{4}\{f(2) + f(-2)\} - \frac{1}{2}\Psi_0(f) + \frac{\tilde{b}_p}{b_p}(\nu_4 - 3)\Psi_2(f)\right] - \sqrt{\frac{n^3}{p}}\frac{c_p\Psi_3(f)}{b_p\sqrt{b_p}} + o(1).\end{aligned}$$

For any function $f \in \mathcal{M}$, we define a new normalization of the LSS:

$$Q_n(f) = n\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}\{F^{\mathbf{A}_n}(x) - F(x)\} - \sqrt{\frac{n^3}{p}}\frac{c_p}{b_p\sqrt{b_p}}\Psi_3(f). \tag{12}$$

Note that the last term in (12) makes no contribution if the function f is even ($\Psi_3(f) = 0$) or $n^3/p = o(1)$. Combining (11) and Theorem 3.1, we obtain the following CLT for $Q_n(f)$.

COROLLARY 3.2. *Suppose the assumptions* (*A*) *and* (*B*) *in Theorem* 3.1 *hold and*

(C2) $p \wedge n \to \infty$ *and* $n^3/p = O(1)$.

Then, for any $k \geq 1$ and $f_1, \dots, f_k \in \mathcal{M}$, the k-dimensional vector $(Q_n(f_1), \dots, Q_n(f_k))$ *converges weakly to a Gaussian vector* $(Y(f_1), \dots, Y(f_k))$ *with mean function*

$$\mathbb{E}Y(f_k) = \frac{1}{4}\{f_k(2) + f_k(-2)\} - \frac{1}{2}\Psi_0(f_k) + \frac{\omega}{\theta}(\nu_4 - 3)\Psi_2(f_k)$$

and covariance function given in (7).

4. Applications to hypothesis testing about large covariance matrices.

4.1. *The identity hypothesis "$\Sigma_p = \mathbf{I}_p$."* Let $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ be a $p \times n$ data matrix with n i.i.d. p-dimensional random vectors $\{\mathbf{y}_i = \mathbf{\Sigma}_p^{1/2}\mathbf{x}_i\}_{1\leq i\leq n}$ with covariance matrix $\mathbf{\Sigma}_p = \mathrm{Var}(\mathbf{y}_i)$ and $\mathbf{x}_i$ has p i.i.d. components $\{X_{ij}\}_{1\leq j\leq p}$ satisfying $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = 1$, $\mathbb{E}X_{ij}^4 = \nu_4$. We consider the identity testing problem:

$$H_0 : \mathbf{\Sigma}_p = \mathbf{I}_p \quad \text{vs.} \quad H_1 : \mathbf{\Sigma}_p \neq \mathbf{I}_p, \tag{13}$$

under two different asymptotic regimes: high-dimensional regime, "$p \wedge n \to \infty$, $p/n \to c \in (0, \infty)$" and ultrahigh-dimensional regime, "$p \wedge n \to \infty$, $p/n \to \infty$." We will consider two well-known test statistics and discuss their limiting distributions under both regimes.

For the identity testing problem (13), Nagao (1973) proposed a statistic based on the Frobenius norm:

$$V = \frac{1}{p}\mathrm{tr}\{(\mathbf{S}_n - \mathbf{I}_p)^2\},$$

where $\mathbf{S}_n = \frac{1}{n}\mathbf{Y}\mathbf{Y}'$ is the sample covariance matrix. Nagao's test based on V performs well when n tends to infinity while p remains fixed and small. However, Ledoit and Wolf (2002) showed that Nagao's test has poor properties when p is large. They proposed a modification of the form

$$W = \frac{1}{p}\operatorname{tr}\{(\mathbf{S}_n - \mathbf{I}_p)^2\} - \frac{1}{np}\{\operatorname{tr}(\mathbf{S}_n)\}^2 + \frac{p}{n}. \tag{14}$$

When $p \wedge n \to \infty$, $p/n = c_n \to c \in (0, \infty)$, under normality assumption, Ledoit and Wolf (2002) proved that the limiting distribution of W under H_0 is

$$nW - p - 1 \xrightarrow{d} \mathcal{N}(0, 4).$$

By Lemma 2.2 in Wang and Yao (2013) and the delta method, we can further remove the normality assumption and show that under H_0, when $p \wedge n \to \infty$, $p/n = c_n \to c \in (0, \infty)$,

$$nW - p - (\nu_4 - 2) \xrightarrow{d} \mathcal{N}(0, 4). \tag{15}$$

The proof of this result is provided in Section S4.1 of the Supplementary Material.

Now we derive the limiting distribution of W under both H_0 and H_1 when $p/n \to \infty$. We will show that the test based on W is consistent under the ultrahigh-dimensional setting. The main results of the test based on W is as follows.

THEOREM 4.1. *Assume that $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ is a $p \times n$ data matrix with n i.i.d. p-dimensional random vectors $\{\mathbf{y}_i = \mathbf{\Sigma}_p^{1/2}\mathbf{x}_i\}_{1\le i\le n}$ with covariance matrix $\mathbf{\Sigma}_p = \operatorname{Var}(\mathbf{y}_i)$ and $\mathbf{x}_i$ has p i.i.d. components $\{X_{ij}\}_{1\le j\le p}$ satisfying $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = 1$, $\mathbb{E}X_{ij}^4 = \nu_4$ and $\mathbb{E}|X_{ij}|^{6+\varepsilon_0} < \infty$ for some positive ε_0. Then under H_0, when $p \wedge n \to \infty$ and $n^2/p = O(1)$,*

$$nW - p - (\nu_4 - 2) \xrightarrow{d} \mathcal{N}(0, 4). \tag{16}$$

Note that the asymptotic distribution (16) coincides with (15), which means W has the same limiting null distribution in both high-dimensional and ultrahigh-dimensional settings. Therefore, W can be used to test (13) under the ultrahigh-dimensional setting. For nominal level α, the corresponding rejection rule is

$$\frac{1}{2}\{nW - p - (\nu_4 - 2)\} \ge z_\alpha,$$

where z_α is the α upper quantile of standard normal distribution.

Under the alternative hypothesis H_1 where $\mathbf{\Sigma}_p \ne \mathbf{I}_p$, we have the following.

THEOREM 4.2. *Under the same assumptions as in Theorem 4.1, further assume that $\{\mathbf{\Sigma}_p,\ p \ge 1\}$ is a sequence of nonnegative definite matrices, bounded in spectral norm such that the following limits exist*:

$$\gamma = \lim_{p\to\infty}\frac{1}{p}\operatorname{tr}(\mathbf{\Sigma}_p), \qquad \theta = \lim_{p\to\infty}\frac{1}{p}\operatorname{tr}(\mathbf{\Sigma}_p^2), \qquad \omega = \lim_{p\to\infty}\frac{1}{p}\sum_{i=1}^{p}(\mathbf{\Sigma}_p)_{ii}^2.$$

Then when $p \wedge n \to \infty$ and $n^2/p = O(1)$,

$$nW - p - \theta\left\{\frac{\omega}{\theta}(\nu_4 - 3) + 1\right\} + n(2\gamma - 1 - \theta) \xrightarrow{d} \mathcal{N}(0, 4\theta^2).$$

The proof of Theorem 4.2 is given in Section S4.2 of the Supplementary Material. Note that Theorem 4.2 contains Theorem 4.1 as a particular case. Indeed, when $\mathbf{\Sigma}_p = \mathbf{I}_p$, $\gamma = \theta = \omega = 1$ and Theorem 4.2 reduces to Theorem 4.1, which states the limiting null distribution of W. With Theorem 4.2, the asymptotic power of W can be derived.

PROPOSITION 4.1. *With the same assumptions as in Theorem 4.2, when $p \wedge n \to \infty$ and $n^2/p = O(1)$, the testing power of W for (13) satisfies*

$$\beta(H_1) \to 1 - \Phi\left(\frac{1}{2\theta}\{2z_\alpha - \omega(\nu_4 - 3) - \theta + n(2\gamma - 1 - \theta) + (\nu_4 - 2)\}\right).$$

Thus, if $\gamma = \theta = 1$, then $\beta(H_1) \to 1 - \Phi(z_\alpha - \frac{\omega-1}{2}(\nu_4 - 3))$; otherwise, $\beta(H_1) \to 1$.

The second test statistic for (13) we consider is the likelihood ratio test (LRT) statistic studied in Bai et al. (2009). They assumed that $\nu_4 = 3$. The LRT statistic is defined as

$$\mathcal{L}_0 = \operatorname{tr}(\mathbf{S}_n) - \log|\mathbf{S}_n| - p. \tag{17}$$

Bai et al. (2009) derived the limiting null distribution of $\mathcal{L}_0$ when $p \wedge n \to \infty$, $p/n \to c \in (0, 1)$. However, this LRT statistic is degenerate and not applicable when $p > n$ because $|\mathbf{S}_n| = 0$. Thus, for $p > n$, we introduce a quasi-LRT test statistic

$$\mathcal{L} = \operatorname{tr}(\widehat{\mathbf{S}}_n) - \log|\widehat{\mathbf{S}}_n| - n,$$

where $\widehat{\mathbf{S}}_n = \frac{1}{p}\mathbf{Y}'\mathbf{Y}$ is now of full rank. When $p \wedge n \to \infty$, $p/n = c_n \to c \in (1, \infty)$, the limiting null distribution of $\mathcal{L}$ is

$$\mathcal{L}^* := \frac{\mathcal{L} - nF_1(c_n) - \mu_1}{\sigma_1} \xrightarrow{d} \mathcal{N}(0, 1), \tag{18}$$

where $F_1(c_n) = 1 - (1 - c_n)\log(1 - 1/c_n)$, $\mu_1 = -\frac{1}{2}\log(1 - 1/c_n)$ and $\sigma_1^2 = -2\log(1 - 1/c_n) - 2/c_n$.

Now we will show that this asymptotic distribution (18) still holds in the ultrahigh-dimensional setting. Note that σ_1 in the limit (18) satisfies

$$\sigma_1 = \sqrt{-2\log\left(1 - \frac{1}{c_n}\right) - \frac{2}{c_n}} = \sqrt{\frac{1}{c_n^2} + \frac{2}{3c_n^3} + o\left(\frac{1}{c_n^3}\right)} = \frac{1}{c_n} + \frac{1}{3c_n^2} + o\left(\frac{1}{c_n^2}\right),$$

which implies that

$$\frac{1}{\sigma_1} = c_n - \frac{1}{3} + o(1). \tag{19}$$

First, we consider the random part of $\mathcal{L}^*$. Let $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_n$ be the eigenvalues of $\widehat{\mathbf{S}}_n$ and $\widetilde{\lambda}_1 \geq \cdots \geq \widetilde{\lambda}_n$ be the eigenvalues of $\widetilde{\mathbf{S}}_n = \sqrt{\frac{n}{p}}(\frac{1}{n}\mathbf{X}'\mathbf{X} - \frac{p}{n}\mathbf{I}_n)$. By using the basic identity $\widehat{\lambda}_i = \frac{\widetilde{\lambda}_i}{\sqrt{c_n}} + 1$, we have

$$\begin{aligned}
\mathcal{L} &= \sum_{i=1}^n \widehat{\lambda}_i - n - \sum_{i=1}^n \log(\widehat{\lambda}_i) = \sum_{i=1}^n \frac{\widetilde{\lambda}_i}{\sqrt{c_n}} - \sum_{i=1}^n \log\left(1 + \frac{\widetilde{\lambda}_i}{\sqrt{c_n}}\right) \\
&= \sum_{i=1}^n \frac{\widetilde{\lambda}_i}{\sqrt{c_n}} - \sum_{i=1}^n \left\{\frac{\widetilde{\lambda}_i}{\sqrt{c_n}} - \frac{1}{2}\frac{\widetilde{\lambda}_i^2}{c_n} + \frac{1}{3}\frac{\widetilde{\lambda}_i^3}{c_n\sqrt{c_n}} - \frac{1}{4}\frac{\widetilde{\lambda}_i^4}{c_n^2} + o_P\left(\frac{1}{c_n^2}\right)\right\} \\
&= \frac{1}{2c_n}\operatorname{tr}(\widetilde{\mathbf{S}}_n^2) - \frac{1}{3c_n\sqrt{c_n}}\operatorname{tr}(\widetilde{\mathbf{S}}_n^3) + \frac{1}{4c_n^2}\operatorname{tr}(\widetilde{\mathbf{S}}_n^4) + o_P\left(\frac{n}{c_n^2}\right),
\end{aligned} \tag{20}$$

where we use the fact that $\{\widetilde{\lambda}_i\}_{1 \leq i \leq n}$ are bounded in probability; see Lemma 6.1. Taking $\nu_4 = 3$ (the assumption in Bai et al. (2009)) and $\mathbf{\Sigma}_p = \mathbf{I}_p$ in Corollary 3.1, we obtain that, under H_0,

$$\operatorname{tr}(\widetilde{\mathbf{S}}_n^2) - n - 1 \xrightarrow{d} \mathcal{N}(0,4), \qquad \operatorname{tr}(\widetilde{\mathbf{S}}_n^3) - \frac{n+1}{\sqrt{c_n}} \xrightarrow{d} \mathcal{N}(0,24), \tag{21}$$

$$\operatorname{tr}(\widetilde{\mathbf{S}}_n^4) - 2n - \left(\frac{n}{c_n} + \frac{1}{c_n} + 5\right) \xrightarrow{d} \mathcal{N}(0,72). \tag{22}$$

Combining (19)–(22) gives us that

$$\frac{\mathcal{L}}{\sigma_1} = \frac{1}{2}\operatorname{tr}(\widetilde{\mathbf{S}}_n^2) + o_P\left(\frac{n^2}{p}\right). \tag{23}$$

Second, we consider the deterministic part of $\mathcal{L}^*$. Note that

$$\begin{aligned} nF_1(c_n) + \mu &= n - \left\{n(1-c_n) + \frac{1}{2}\right\}\log\left(1 - \frac{1}{c_n}\right) \\ &= n - \left\{n(1-c_n) + \frac{1}{2}\right\} \cdot \left\{-\frac{1}{c_n} - \frac{1}{2c_n^2} - \frac{1}{3c_n^3} + o\left(\frac{1}{c_n^3}\right)\right\} \\ &= \frac{n}{2c_n} + \frac{1}{2c_n} + \frac{n}{6c_n^2} + o\left(\frac{n}{c_n^2}\right). \end{aligned}$$

Together with (19), we find

$$\frac{nF_1(c_n) + \mu_1}{\sigma_1} = \frac{n+1}{2} + o_P\left(\frac{n^2}{p}\right). \tag{24}$$

Therefore, from (21), (23) and (24), we conclude that, under H_0, as $p \wedge n \to \infty$, $n^2/p = O(1)$,

$$\mathcal{L}^* = \frac{\mathcal{L}}{\sigma_1} - \frac{nF_1(c_n) + \mu_1}{\sigma_1} = \frac{1}{2}\{\operatorname{tr}(\widetilde{\mathbf{S}}_n^2) - n - 1\} + o_P(1) \xrightarrow{d} \mathcal{N}(0,1),$$

which is the same as the limiting distribution (18) derived under the scheme $p \wedge n \to \infty$, $p/n = c_n \to c \in (1,\infty)$. Finally, we summarize the discussion above in the following proposition.

PROPOSITION 4.2. (1) (*Bai et al.* (2009)) *Assume that* $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ *is a* $p \times n$ *data matrix with n i.i.d. p-dimensional random vectors* $\{\mathbf{y}_i = \mathbf{\Sigma}_p^{1/2}\mathbf{x}_i\}_{1\le i\le n}$ *with covariance matrix* $\mathbf{\Sigma}_p = \operatorname{Var}(\mathbf{y}_i)$ *and* $\mathbf{x}_i$ *has p i.i.d. components* $\{X_{ij}\}_{1\le j\le p}$ *satisfying* $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = 1$, $\mathbb{E}X_{ij}^4 = \nu_4 = 3$. $\mathcal{L}_0$ *is defined as* (17). *Then under* H_0, *when* $p \wedge n \to \infty$, $p/n \to c \in (0,1)$, *we have*

$$\frac{\mathcal{L}_0 - nF_0(c_n) - \mu_0}{\sigma_0} \xrightarrow{d} \mathcal{N}(0,1),$$

where $c_n = p/n$ *and*

$$F_0(c_n) = 1 - \frac{c_n - 1}{c_n}\log(1 - c_n), \qquad \mu_0 = -\frac{\log(1-c_n)}{2}, \qquad \sigma_0^2 = -2\log(1-c_n) - 2c_n.$$

(2) *Under the assumptions in* (1), *consider the normalized quasi-LRT statistic* $\mathcal{L}^*$ *defined in* (18). *Under* H_0, *when* $p \wedge n \to \infty$, $p/n \to c \in (1,\infty)$, *we have*

$$\mathcal{L}^* \xrightarrow{d} \mathcal{N}(0,1).$$

(3) *Under the assumptions in* (1), *consider the normalized quasi-LRT statistic* $\mathcal{L}^*$ *defined in* (18). *Under* H_0, *when* $p \wedge n \to \infty$ *and* $n^2/p = O(1)$, *we have*

$$\mathcal{L}^* \xrightarrow{d} \mathcal{N}(0, 1).$$

Note that the results (2) and (3) in this proposition are new.

By using our Theorems 4.1–4.2 and Proposition 4.1, we now consider an application to testing whether the population covariance matrix of the matrix-valued white noise is equal to a prespecified separable matrix, where the asymptotic regime $n^2/p = O(1)$ arises in a very natural manner.

EXAMPLE 4.1. Chen, Xiao and Yang (2021) proposed a matrix autoregressive model with the form

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}' + \mathbf{E}_t, \quad t = 1, \dots, T,$$

where $\mathbf{X}_t$ is a $p_1 \times p_2$ random matrix observed at time t, $\mathbf{A}$ and $\mathbf{B}$ are $p_1 \times p_1$ and $p_2 \times p_2$ deterministic autoregressive coefficient matrices, $\mathbf{E}_t = (e_{t,ij})$ is a $p_1 \times p_2$ matrix-valued white noise. It is assumed that the error white noise matrix $\mathbf{E}_t$ has a specific covariance structure

$$\mathrm{Cov}\{\mathsf{vec}(\mathbf{E}_t)\} = \mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2, \tag{25}$$

where $\mathsf{vec}(\cdot)$ denotes the vectorization, $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are $p_1 \times p_1$ and $p_2 \times p_2$ nonnegative definite matrices. In other words, the noise $\mathbf{E}_t$ has a separable covariance matrix.

For any observed matrix-valued noise sequence $\{\mathbf{E}_t\}_{1 \le t \le T}$, we aim to test whether its covariance matrix is equal to a prespecified separable matrix as in (25). Specifically, suppose that $\{\mathbf{E}_t\}_{1 \le t \le T}$ is an observed i.i.d. sequence of $p_1 \times p_2$ matrices and p_1, p_2, T are of comparable magnitude as follows:

$$T \to \infty, \qquad \frac{p_1}{T} = \frac{p_1(T)}{T} \to d_1 \in (0, \infty), \qquad \frac{p_2}{T} = \frac{p_2(T)}{T} \to d_2 \in (0, \infty), \tag{26}$$

we aim to test

$$H_0 : \mathrm{Cov}\{\mathsf{vec}(\mathbf{E}_t)\} = \mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2 \quad \text{vs.} \quad H_1 : \mathrm{Cov}\{\mathsf{vec}(\mathbf{E}_t)\} \neq \mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2, \tag{27}$$

where $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are two prespecified $p_1 \times p_1$ and $p_2 \times p_2$ nonnegative definite matrices. Testing $H_0 : \mathrm{Cov}\{\mathsf{vec}(\mathbf{E}_t)\} = \mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2$ is equivalent to testing

$$H_0' : \mathrm{Cov}\{(\mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2)^{-1/2}\mathsf{vec}(\mathbf{E}_t)\} = \mathbf{I}_{p_1 p_2}.$$

To this end, we define a test statistic following (14):

$$W^* = \frac{1}{p_1 p_2} \mathrm{tr}\{(\mathbf{S}_T - \mathbf{I}_{p_1 p_2})^2\} - \frac{1}{p_1 p_2 T}\{\mathrm{tr}(\mathbf{S}_T)\}^2 + \frac{p_1 p_2}{T},$$

where

$$\mathbf{S}_T = \frac{1}{T}\mathbf{Y}_T\mathbf{Y}_T', \qquad \mathbf{Y}_T = (\mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2)^{-1/2}\{\mathsf{vec}(\mathbf{E}_1), \dots, \mathsf{vec}(\mathbf{E}_T)\} := (Y_{ij})_{p_1 p_2 \times T}. \tag{28}$$

Note that W^* measures the distance between sample covariance matrix of $\mathsf{vec}(\mathbf{E}_t)$ and $\mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2$. Naturally, we reject H_0 when W^* is too large and the critical value is determined by the limiting null distribution of W^*.

The asymptotic null distribution of the test statistic W^* can be derived by Theorem 4.1. Assume that $\{\mathbf{E}_t = (e_{t,ij})_{p_1 \times p_2}\}_{1 \le t \le T}$ is a sequence of i.i.d. sample matrices satisfying $\mathsf{vec}(\mathbf{E}_t) = (\mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2)^{1/2}\mathsf{vec}(\mathbf{Z}_t)$, where $\mathbf{Z}_t = (Z_{t,ij})_{p_1 \times p_2}$ is a $p_1 \times p_2$ matrix with i.i.d. real entries $Z_{t,ij}$ satisfying $\mathbb{E}Z_{t,ij} = 0$, $\mathbb{E}Z_{t,ij}^2 = 1$, $\mathbb{E}Z_{t,ij}^4 = \nu_4$ and $\mathbb{E}|Z_{t,ij}|^{6+\varepsilon_0} < \infty$ for some positive ε_0. Then under the null hypothesis, we have

$$TW^* - p_1 p_2 - (\nu_4 - 2) \xrightarrow{d} \mathcal{N}(0, 4). \tag{29}$$

Accordingly, we reject H_0 at nominal level α if

$$\frac{1}{2}\{TW^* - p_1 p_2 - (\nu_4 - 2)\} \geq z_\alpha.$$

Moreover, the asymptotic power of the proposed test for (27) can be derived by Theorem 4.2. Suppose that H_1 in (27) is true and the population covariance matrix of $\mathsf{vec}(\mathbf{E}_t)$ is $(\widetilde{\boldsymbol{\Sigma}}_1 \otimes \widetilde{\boldsymbol{\Sigma}}_2)^{1/2}\mathsf{vec}(\mathbf{Z}_t)$, where $\widetilde{\boldsymbol{\Sigma}}_1$ and $\widetilde{\boldsymbol{\Sigma}}_2$ are two $p_1 \times p_1$ and $p_2 \times p_2$ nonnegative definite matrices with bounded spectral norm. Let $\widetilde{\boldsymbol{\Sigma}} := (\widetilde{\boldsymbol{\Sigma}}_1 \otimes \widetilde{\boldsymbol{\Sigma}}_2)^{1/2}(\boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2)^{-1}(\widetilde{\boldsymbol{\Sigma}}_1 \otimes \widetilde{\boldsymbol{\Sigma}}_2)^{1/2}$ and suppose that the following limits exist:

$$\gamma = \lim_{T\to\infty} \frac{1}{p_1 p_2}\mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}), \qquad \theta = \lim_{T\to\infty} \frac{1}{p_1 p_2}\mathsf{tr}(\widetilde{\boldsymbol{\Sigma}}^2), \qquad \omega = \lim_{T\to\infty} \frac{1}{p_1 p_2}\sum_{i=1}^{p_1 p_2}(\widetilde{\boldsymbol{\Sigma}})_{ii}^2.$$

Then when p_1, p_2, T tend to infinity as in (26), the testing power of W^* for (27) satisfies

$$\beta(H_1) \to 1 - \Phi\left(\frac{1}{2\theta}\left[2z_\alpha - \omega(\nu_4 - 3) - \theta + n(2\gamma - 1 - \theta) + (\nu_4 - 2)\right]\right). \tag{30}$$

If $\gamma = \theta = 1$, then $\beta(H_1) \to 1 - \Phi(z_\alpha - \frac{\omega-1}{2}(\nu_4 - 3))$; otherwise, $\beta(H_1) \to 1$.

4.2. *Discussion on sphericity test "$\Sigma_p = \sigma^2\mathbf{I}_p$."* Corollary 3.2 is used in Li and Yao (2016) to derive the asymptotic power of two sphericity tests, John's invariant test and Quasi-likelihood ratio test (QLRT), under the setting $p \wedge n \to \infty$ and $n^3/p = O(1)$. Specifically, let $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ be a $p \times n$ data matrix with n i.i.d. p-dimensional random vectors $\{\mathbf{y}_i\}_{1\leq i\leq n}$ with covariance matrix $\boldsymbol{\Sigma}$. The goal is to test

$$H_0 : \boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p \quad \text{vs.} \quad H_1 : \boldsymbol{\Sigma} \neq \sigma^2\mathbf{I}_p,$$

where σ^2 is an unknown positive constant. John's test statistic is defined as

$$U_n = \frac{1}{p}\mathsf{tr}\left[\left\{\frac{\mathbf{S}_n}{\mathsf{tr}(\mathbf{S}_n)/p} - \mathbf{I}_p\right\}^2\right] = \frac{p^{-1}\sum_{i=1}^p (l_i - \bar{l})^2}{\bar{l}^2},$$

where $\{l_i\}_{1\leq i\leq p}$ are eigenvalues of p-dimensional sample covariance matrix $\mathbf{S}_n = \frac{1}{n}\mathbf{Y}\mathbf{Y}'$ and $\bar{l} = \frac{1}{p}\sum_{i=1}^p l_i$. The QLRT statistic is defined as

$$L_n = \frac{p}{n}\log\frac{(n^{-1}\sum_{i=1}^n \tilde{l}_i)^n}{\prod_{i=1}^n \tilde{l}_i},$$

where $\{\tilde{l}_i\}_{1\leq i\leq n}$ are the eigenvalues of the $n \times n$ companion matrix $\bar{\mathbf{S}}_n = \frac{1}{p}\mathbf{Y}'\mathbf{Y}$.

Assume that the data matrix $\mathbf{Y}$ has the structure $\mathbf{Y} = \boldsymbol{\Sigma}_p^{1/2}\mathbf{X}$, where $\mathbf{X}$ satisfies assumption (A) in Theorem 3.1. The eigenvalues l_i of $\mathbf{S}_n = \boldsymbol{\Sigma}_p^{1/2}\mathbf{X}\mathbf{X}'\boldsymbol{\Sigma}_p^{1/2}/n$, $\tilde{l}_i$ of $\bar{\mathbf{S}}_n = \mathbf{X}'\boldsymbol{\Sigma}_p\mathbf{X}/p$ and $\lambda_i^{\mathbf{A}_n}$ of $\mathbf{A}_n = (\mathbf{X}'\boldsymbol{\Sigma}_p\mathbf{X} - pa_p\mathbf{I}_n)/\sqrt{npb_p}$ satisfy

$$l_i = \sqrt{\frac{pb_p}{n}}\lambda_i^{\mathbf{A}_n} + \frac{pa_p}{n}, \qquad \tilde{l}_i = \sqrt{\frac{nb_p}{p}}\lambda_i^{\mathbf{A}_n} + a_p, \tag{31}$$

for $1 \leq i \leq n$. The remaining $p - n$ eigenvalues of $\mathbf{S}_n$ are all zero. Therefore, both U_n and L_n can be expressed as eigenvalue statistics of $\mathbf{A}_n$ by (31). Then we can utilize our Corollary 3.2 to derive the asymptotic distributions of U_n and L_n under both the null and alternative hypotheses. Their power functions are proven to converge to 1 under the assumption $n^3/p = O(1)$. More details can be found in Li and Yao (2016).

5. Simulation results. In this section, we implement some simulation studies to examine:

(1) finite-sample properties of some LSS for $\mathbf{A}_n$ by comparing their empirical means and variances with theoretical limiting values;
(2) finite-sample performance of the covariance test for matrix-valued noise in Example 4.1.

5.1. *LSS of* $\mathbf{A}_n$. First, we compare the empirical mean and variance of normalized $\{G_n(f_i) = \mathrm{tr}(\mathbf{A}_n^i),\ i = 1, 2, 3\}$ with their theoretical limits in Corollary 3.1, where $f_i(x) = x^i$, $i = 1, 2, 3$. Define

$$\overline{G}_n(f_1) := \frac{G_n(f_1)}{\sqrt{\mathrm{Var}(Y(f_1))}} = \frac{\mathrm{tr}(\mathbf{A}_n)}{\sqrt{\frac{\omega}{\theta}(\nu_4 - 3) + 2}},$$

$$\overline{G}_n(f_2) := \frac{G_n(f_2)}{\sqrt{\mathrm{Var}(Y(f_2))}} = \frac{1}{2}\left[\mathrm{tr}(\mathbf{A}_n^2) - n - \left\{\frac{\tilde{b}_p}{b_p}(\nu_4 - 3) + 1\right\}\right],$$

$$\overline{G}_n(f_3) := \frac{G_n(f_3)}{\sqrt{\mathrm{Var}(Y(f_3))}} = \frac{\mathrm{tr}(\mathbf{A}_n^3) - \frac{c_p}{b_p\sqrt{b_p}}\sqrt{\frac{n}{p}}\{n + 1 + \frac{\tilde{b}_p}{b_p}(\nu_4 - 3)\}}{\sqrt{\frac{9\omega}{\theta}(\nu_4 - 3) + 24}}.$$

According to Corollary 3.1, $\overline{G}_n(f_i) \xrightarrow{d} \mathcal{N}(0, 1)$, $i = 1, 2, 3$. Hence, we directly compare the empirical distribution of $\overline{G}_n(f_i)$ with $\mathcal{N}(0, 1)$ under different scenarios. Specifically, we consider two data distributions of $\{X_{ij}\}$, that is,

(1) *Gaussian data*: $\{X_{ij},\ 1 \le i \le p,\ 1 \le j \le n\}$ i.i.d. $\mathcal{N}(0, 1)$, with $\mathbb{E}X_{ij}^4 = \nu_4 = 3$.
(2) *Non-Gaussian data*: $\{X_{ij},\ 1 \le i \le p,\ 1 \le j \le n\}$ i.i.d. $\mathrm{Gamma}(4, 2) - 2$, with $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = 1$, $\mathbb{E}X_{ij}^4 = 4.5$.

As for the covariance matrix $\mathbf{\Sigma}_p$, we consider four distinct types:

(A) $\mathbf{\Sigma}_A = \mathbf{I}_p$;
(B) $\mathbf{\Sigma}_B$ is diagonal, $1/4$ of its diagonal elements are 0.5, and $3/4$ are 1.
(C) $\mathbf{\Sigma}_C$ is diagonal, one-half of its diagonal elements are 0.5, and one-half are 1.
(D) $\mathbf{\Sigma}_D$ is tridiagonal, all of its main diagonal elements are 2, all of its subdiagonal and superdiagonal elements are 1.

Empirical mean and variance of $\overline{G}_n(f_i)$ are calculated for various combinations of (p, n) under different model settings. For each pair of (p, n), 5000 independent replications are used to obtain the empirical mean and variance. Table 1 reports the empirical values of $\overline{G}_n(f_i)$ under the setting $p = n^2$. As shown in Table 1, the empirical mean and variance of $\overline{G}_n(f_i)$ closely match their theoretical limits 0 and 1 under all scenarios, including all four types of $\mathbf{\Sigma}_p$, and for both Gaussian and non-Gaussian data. Histograms of $\overline{G}_n(f_i)$ for the case (D) are shown in Figure 1: we see that the empirical densities match well with standard normal distribution. The histograms for the cases (A), (B) and (C) are similar, and thus omitted.

Additional simulation results for the case $p = n^{2.5}$ are shown in the Supplementary Material. The numerical performance is similar.

5.2. *Covariance testing for matrix-valued noise.* Empirical size and power of the covariance testing for matrix-valued noise in Example 4.1 are examined to testify the asymptotic testing power of W^* given in (30). We compare the empirical power of W^* with its limits under various model settings. Specifically, the vectorization of data matrix $\mathbf{E}_t$ is $\mathrm{vec}(\mathbf{E}_t) = (\mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2)^{1/2}\mathrm{vec}(\mathbf{Z}_t)$. We consider two data distributions of $\mathbf{Z}_t = \{Z_{t,ij}\}$.

TABLE 1

Empirical mean and variance of $\overline{G}_n(f_i)$, $i = 1, 2, 3$ from 5000 *replications. Theoretical mean and variance are* 0 *and* 1, *respectively. Dimension $p = n^2$*

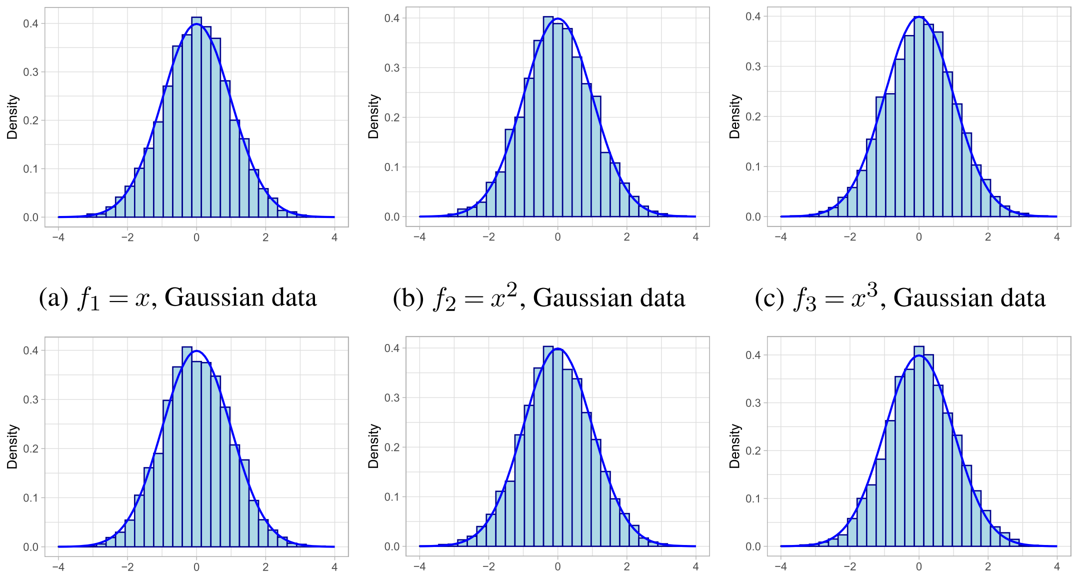
	$\mathbf{\Sigma}_p = \mathbf{\Sigma}_A$		$\mathbf{\Sigma}_p = \mathbf{\Sigma}_B$		$\mathbf{\Sigma}_p = \mathbf{\Sigma}_C$		$\mathbf{\Sigma}_p = \mathbf{\Sigma}_D$	
n	mean	var	mean	var	mean	var	mean	var
50	0.0134	1.0207	−0.0206	0.9929	0.0098	1.0035	−0.0097	0.9964
100	0.004	0.98	−0.0009	0.9985	−0.0144	1.0176	0.0298	1.0017
150	0.0106	0.9926	−0.003	1.0163	−0.0131	0.9862	0.0225	0.9948
200	0.013	0.9889	−0.0223	1.0221	−0.0061	0.9888	0.0073	0.9877
$\overline{G}_n(f_1)$	*Gaussian*							
50	0.0133	1.0239	0.0214	1.0158	−0.0299	0.9919	−0.0137	0.9532
100	0.006	1.0123	−0.0091	0.9856	0.023	0.985	−0.0198	1.0322
150	0.0042	1.0212	0.0248	1.0106	−0.0119	1.0033	0.0078	0.9946
200	0.0019	0.9731	0.0171	0.9987	−0.0195	0.9955	0.004	0.9645
$\overline{G}_n(f_1)$	*Non-Gaussian*							
50	−0.0029	1.0364	−0.0003	1.0575	0.0055	1.0979	0.0073	1.076
100	0.0016	1.0204	−0.0259	0.9847	−0.0033	1.0442	0.0276	1.0319
150	−0.0122	1.0082	−0.0097	1.0387	−0.0099	1.024	−0.0185	1.0485
200	−0.0084	1	−0.0051	1.027	−0.0116	1.0132	0.0033	1.0305
$\overline{G}_n(f_2)$	*Gaussian*							
50	−0.0151	1.2039	−0.0021	1.1874	−0.0142	1.2068	0.0044	1.2069
100	−0.0126	1.0698	0.0148	1.1401	0.0319	1.1135	0.035	1.1151
150	−0.0075	1.1044	−0.0278	1.0692	−0.0084	1.052	−0.008	1.0121
200	−0.0074	1.0332	0.0001	1.0518	−0.0149	1.0342	−0.0171	1.0253
$\overline{G}_n(f_2)$	*Non-Gaussian*							
50	0.0578	1.1324	0.0627	1.1385	0.0681	1.1414	0.0854	1.1823
100	0.0216	1.0325	0.0401	1.061	0.0382	1.0332	0.082	1.0587
150	0.0592	1.0869	0.0236	1.0127	0.0261	1.0543	0.0819	1.0423
200	0.0336	0.9989	0.0248	1.0211	0.0182	1.0633	0.0406	1.0342
$\overline{G}_n(f_3)$	*Gaussian*							
50	0.1529	1.2265	0.1359	1.2057	0.138	1.1964	0.1347	1.1474
100	0.0884	1.0941	0.0953	1.0988	0.1346	1.094	0.0791	1.1134
150	0.0818	1.1206	0.107	1.0683	0.0667	1.0739	0.069	1.0368
200	0.0677	1.0203	0.0807	1.051	0.0468	1.0381	0.0698	1.0095
$\overline{G}_n(f_3)$	*Non-Gaussian*							

(1) *Gaussian matrix white noise*: $\{Z_{t,ij},\ 1 \le i \le p,\ 1 \le j \le n\}$ i.i.d. $\mathcal{N}(0,1)$, with $\nu_4 = \mathbb{E}Z_{t,ij}^4 = 3$.

(2) *Non-Gaussian matrix white noise*: $\{Z_{t,ij},\ 1 \le i \le p,\ 1 \le j \le n\}$ i.i.d. $\mathsf{Gamma}(4,2) - 2$, with $\mathbb{E}Z_{t,ij} = 0$, $\mathbb{E}Z_{t,ij}^2 = 1$, $\nu_4 = \mathbb{E}Z_{t,ij}^4 = 4.5$.

As for covariance matrix $\mathbf{\Sigma}_1 \otimes \mathbf{\Sigma}_2$, we set $\mathbf{\Sigma}_1$ as a $p_1 \times p_1$ tridiagonal matrix, and $\mathbf{\Sigma}_2$ as a $p_2 \times p_2$ symmetric Toeplitz matrix. More specifically,

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 2 & 1 & & & \\ 1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 2 & 1 \\ & & & 1 & 2 \end{pmatrix}_{p_1 \times p_1},$$



(a) $f_1 = x$, Gaussian data (b) $f_2 = x^2$, Gaussian data (c) $f_3 = x^3$, Gaussian data

(d) $f_1 = x$, non-Gaussian data (e) $f_2 = x^2$, non-Gaussian data (f) $f_3 = x^3$, non-Gaussian data

FIG. 1. *Histograms of* $\overline{G}_n(f_i)$, $i = 1, 2, 3$ *from* 5000 *replications under the case* (D) *with* $(p, n) = (200^2, 200)$. *The curves are density functions of standard normal distribution.*

and $\mathbf{\Sigma}_2 = (\rho^{|i-j|})_{p_2 \times p_2}$ with $|\rho| < 1$. We set $\rho = 0.45$, $p_1 = p_2 = T = 40, 60, 80, 100, 120$. The nominal level of the test is $\alpha = 0.05$. To obtain the empirical power, we keep $\mathbf{\Sigma}_1$ unchanged and replace ρ in $\mathbf{\Sigma}_2$ with $\rho(1 + \lambda)$ satisfying $|\rho(1 + \lambda)| < 1$. We vary $\lambda = 0, 0.2, 0.3, 0.4, 0.5$ to obtain different levels of testing power. For each pair of (p_1, p_2, T), 5000 independent replications are used to obtain the empirical size and power. Empirical values and theoretical limits are compared in Table 2. As shown in Table 2, the empirical power tends to 1 when either p_1, p_2, T or λ increases. Most importantly, the empirical power value is consistent with its theoretical limit under all scenarios.

TABLE 2
Empirical (*Emp*) *and Theoretical* (*Theo*) *size* ($\lambda = 0$) *and power of the covariance testing for matrix-valued noise in Example* 4.1 *with* 5000 *replications*

			$\lambda = 0$		$\lambda = 0.2$		$\lambda = 0.3$		$\lambda = 0.4$		$\lambda = 0.5$	
p_1	p_2	T	Emp	Theo	Emp	Theo	Emp	Theo	Emp	Theo	Emp	Theo
40	40	40	0.0490	0.05	0.0950	0.0880	0.2856	0.3087	0.8230	0.8354	0.9992	0.9992
60	60	60	0.0554	0.05	0.1650	0.1625	0.6484	0.6606	0.9974	0.9969	1	1
80	80	80	0.0520	0.05	0.2600	0.2699	0.8994	0.9084	1	1	1	1
100	100	100	0.0526	0.05	0.3916	0.4049	0.9864	0.9878	1	1	1	1
120	120	120	0.0542	0.05	0.5356	0.5524	0.9986	0.9992	1	1	1	1
Gaussian												
40	40	40	0.0568	0.05	0.0716	0.0662	0.2214	0.2353	0.7008	0.7568	0.9942	0.9977
60	60	60	0.0610	0.05	0.1298	0.1277	0.5462	0.5752	0.9878	0.9930	1	1
80	80	80	0.0580	0.05	0.2202	0.2216	0.8356	0.8655	1	1	1	1
100	100	100	0.0530	0.05	0.3312	0.3464	0.9694	0.9785	1	1	1	1
120	120	120	0.0562	0.05	0.4886	0.4910	0.9974	0.9984	1	1	1	1
Non-Gaussian												

6. Proof of Theorem 3.1. In Section 6.1, we first present the preliminary step of data truncation. The general strategy of the main proof of Theorem 3.1 is explained in Section 6.2. Three major steps of the general strategy are presented in Sections 6.3–6.5.

6.1. *Truncation, centralization and rescaling.* We first truncate the elements of $\mathbf{X}$ without changing the weak limit of $G_n(f)$. Let $\{\delta_n\}$ be a sequence of positive numbers such that

$$\delta_n^{-4}\mathbb{E}|X_{11}|^4\mathbb{1}_{\{|X_{11}|\geq\delta_n\sqrt[4]{np}\}}\to 0,\qquad \delta_n\downarrow 0,\qquad \delta_n\sqrt[4]{np}\uparrow\infty, \tag{32}$$

as $n\to\infty$. Define

$$\widehat{X}_{ij}=X_{ij}\mathbb{1}_{\{|X_{11}|\leq\delta_n\sqrt[4]{np}\}},\qquad \sigma^2=\mathbb{E}|\widehat{X}_{ij}-\mathbb{E}\widehat{X}_{ij}|^2,\qquad \widehat{\mathbf{X}}=(\widehat{X}_{ij})_{p\times n},$$
$$\widetilde{X}_{ij}=(\widehat{X}_{ij}-\mathbb{E}\widehat{X}_{ij})/\sigma,\qquad \widetilde{\mathbf{X}}=(\widetilde{X}_{ij})_{p\times n},$$
$$\widehat{\mathbf{A}}_n=(\widehat{\mathbf{X}}'\mathbf{\Sigma}_p\widehat{\mathbf{X}}-pa_p\mathbf{I}_n)/\sqrt{npb_p},\qquad \widetilde{\mathbf{A}}_n=(\widetilde{\mathbf{X}}'\mathbf{\Sigma}_p\widetilde{\mathbf{X}}-pa_p\mathbf{I}_n)/\sqrt{npb_p}.$$

Define $\widehat{G}_n(f)$ and $\widetilde{G}_n(f)$ similarly by means of (5) with the matrix $\mathbf{A}_n$ replaced by $\widehat{\mathbf{A}}_n$ and $\widetilde{\mathbf{A}}_n$, respectively. First, observe that

$$\begin{aligned}\mathbb{P}\big(G_n(f)\neq\widehat{G}_n(f)\big)&\leq\mathbb{P}(\mathbf{A}_n\neq\widehat{\mathbf{A}}_n)\leq np\mathbb{P}\big(|X_{11}|>\delta_n\sqrt[4]{np}\big)\\&\leq K\delta_n^{-4}\mathbb{E}|X_{11}|^4\mathbb{1}_{\{|X_{11}|\geq\delta_n\sqrt[4]{np}\}}=o(1).\end{aligned}$$

Now we consider the difference between $\widehat{G}_n(f)$ and $\widetilde{G}_n(f)$. For any analytic function f on $\mathcal{U}$, we have

$$\begin{aligned}&\mathbb{E}\big|\widehat{G}_n(f)-\widetilde{G}_n(f)\big|\\&\quad\leq\mathbb{E}\sum_{j=1}^n\big|f\big(\lambda_j^{\widehat{\mathbf{A}}_n}\big)-f\big(\lambda_j^{\widetilde{\mathbf{A}}_n}\big)\big|\leq\frac{K_f}{\sqrt{npb_p}}\mathbb{E}\sum_{j=1}^n\big|\lambda_j^{\widehat{\mathbf{X}}'\mathbf{\Sigma}_p\widehat{\mathbf{X}}}-\lambda_j^{\widetilde{\mathbf{X}}'\mathbf{\Sigma}_p\widetilde{\mathbf{X}}}\big|\\&\quad\leq\frac{K_f}{\sqrt{npb_p}}\mathbb{E}\big|\mathrm{tr}(\widehat{\mathbf{X}}-\widetilde{\mathbf{X}})'\mathbf{\Sigma}_p(\widehat{\mathbf{X}}-\widetilde{\mathbf{X}})\cdot 2\big\{\mathrm{tr}\big(\widehat{\mathbf{X}}'\mathbf{\Sigma}_p\widehat{\mathbf{X}}\big)+\mathrm{tr}\big(\widetilde{\mathbf{X}}'\mathbf{\Sigma}_p\widetilde{\mathbf{X}}\big)\big\}\big|^{1/2}\\&\quad\leq\frac{2K_f}{\sqrt{npb_p}}\big|\mathbb{E}\,\mathrm{tr}(\widehat{\mathbf{X}}-\widetilde{\mathbf{X}})'\mathbf{\Sigma}_p(\widehat{\mathbf{X}}-\widetilde{\mathbf{X}})\big|^{1/2}\big|\mathbb{E}\,\mathrm{tr}\big(\widehat{\mathbf{X}}'\mathbf{\Sigma}_p\widehat{\mathbf{X}}\big)+\mathbb{E}\,\mathrm{tr}\big(\widetilde{\mathbf{X}}'\mathbf{\Sigma}_p\widetilde{\mathbf{X}}\big)\big|^{1/2},\end{aligned}$$

where K_f is a bound for $|f'(x)|$ by Lemma S1.6.

It follows from (32) that

$$\big|\sigma^2-1\big|\leq 2\mathbb{E}X_{11}^2\mathbb{1}_{\{|X_{11}|\geq\delta_n\sqrt[4]{np}\}}\leq\frac{2}{\delta_n^2\sqrt{np}}\mathbb{E}|X_{11}|^4\mathbb{1}_{\{|X_{11}|\geq\delta_n\sqrt[4]{np}\}}=o\big((np)^{-1/2}\big),$$

and

$$\begin{aligned}|\mathbb{E}\widehat{X}_{11}|&=|\mathbb{E}X_{11}\mathbb{1}_{\{|X_{11}|\geq\delta_n\sqrt[4]{np}\}}|\leq\mathbb{E}|X_{11}|\mathbb{1}_{\{|X_{11}|\geq\delta_n\sqrt[4]{np}\}}\\&\leq\frac{1}{\delta_n^3(np)^{3/4}}\mathbb{E}|X_{11}|^4\mathbb{1}_{\{|X_{11}|\geq\delta_n\sqrt[4]{np}\}}=o\big((np)^{-3/4}\big).\end{aligned}$$

These give us

$$\begin{aligned}\frac{1}{\sqrt{np}}\big\{\mathrm{tr}(\widehat{\mathbf{X}}-\widetilde{\mathbf{X}})'\mathbf{\Sigma}_p(\widehat{\mathbf{X}}-\widetilde{\mathbf{X}})\big\}^{1/2}&\leq\sum_{i,j}(\mathbf{\Sigma}_p)_{ii}\mathbb{E}\bigg|\frac{\sigma-1}{\sigma}\widehat{X}_{ij}+\frac{\mathbb{E}\widehat{X}_{ij}}{\sigma}\bigg|^2\\&\leq Kpn\bigg\{\frac{(1-\sigma)^2}{\sigma^2}\mathbb{E}|\widehat{X}_{11}|^2+\frac{1}{\sigma^2}\mathbb{E}|\widehat{X}_{11}|^2\bigg\}=o(1),\end{aligned}$$

and

$$\mathbb{E}\operatorname{tr}(\widehat{\mathbf{X}}'\mathbf{\Sigma}_p\widehat{\mathbf{X}}) \le \sum_{i,j}(\mathbf{\Sigma}_p)_{ii}\mathbb{E}|\widehat{X}_{ij}|^2 \le Knp, \qquad \mathbb{E}\operatorname{tr}(\widetilde{\mathbf{X}}'\mathbf{\Sigma}_p\widetilde{\mathbf{X}}) \le \sum_{i,j}(\mathbf{\Sigma}_p)_{ii}\mathbb{E}|\widetilde{X}_{ij}|^2 \le Knp.$$

From the above estimates, we obtain

$$G_n(f) = \widetilde{G}_n(f) + o_P(1).$$

Thus, we only need to find the limit distribution of $\{\widetilde{G}_n(f_j), j = 1, \dots, k\}$. Hence, in what follows, we assume that the underlying variables are truncated at $\delta_n \sqrt[4]{np}$, centralized and renormalized. For convenience, we shall suppress the superscript on the variables, and assume that, for any $1 \le i \le p$ and $1 \le j \le n$,

$$|X_{ij}| \le \delta_n \sqrt[4]{np}, \qquad \mathbb{E}X_{ij} = 0, \qquad \mathbb{E}X_{ij}^2 = 1,$$
$$\mathbb{E}X_{ij}^a = \nu_a + o(1), \qquad a = 4, 5, \qquad \mathbb{E}|X_{ij}|^{6+\varepsilon_0} < \infty,$$

where δ_n satisfies the condition (32).

6.2. *Strategy of the proof.* The general strategy of the proof follows the method established in Bai and Silverstein (2004) and Bai and Yao (2005).

Let $\mathcal{C}$ be the closed contour formed by the boundary of the rectangle with $(\pm u_1, \pm i v_1)$ where $u_1 > \eta = 2 \limsup_p \|\mathbf{\Sigma}_p\|/\sqrt{\theta}$ with $\theta = \lim_{p\to\infty} \operatorname{tr}(\mathbf{\Sigma}_p^2)/p$, $0 < v_1 \le 1$. Assume that u_1 and v_1 are fixed and sufficiently small such that $\mathcal{C} \subset \mathcal{U}$. Note that the contour $\mathcal{C}$ encloses the support of $F^{\mathbf{A}_n}(x)$ and $F(x)$. Then, for any $x \in (-u_1, u_1)$, by Cauchy's integral theorem, we have

$$f(x) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f(z)}{z - x}\,\mathrm{d}z.$$

By this formula, with probability one, we can rewrite $G_n(f)$ as

$$\begin{aligned}
G_n(f) &= \underbrace{n\int_{-\infty}^{+\infty} f(x)\{F^{\mathbf{A}_n}(x) - F(x)\}\,\mathrm{d}x}_{\text{use Cauchy's integral theorem}} - \underbrace{\frac{n}{2\pi i}\oint_{|m|=\rho} f\left(-m - \frac{1}{m}\right)\mathcal{X}_n(m)\frac{1-m^2}{m^2}\,\mathrm{d}m}_{\text{change of variable, let } z = -m - m^{-1}} \\
&= \frac{n}{2\pi i}\oint_{\mathcal{C}} f(z)\left\{\int_{-\infty}^{+\infty} \frac{F^{\mathbf{A}_n}(x) - F(x)}{z - x}\,\mathrm{d}x\right\}\mathrm{d}z + \frac{n}{2\pi i}\oint_{\mathcal{C}} f(z)\mathcal{X}_n(m(z))\,\mathrm{d}z \\
&= -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z) n\{m_n(z) - m(z) - \mathcal{X}_n(m(z))\}\,\mathrm{d}z,
\end{aligned}$$

where $m_n(z)$ and $m(z)$ are the Stieltjes transforms of $F^{\mathbf{A}_n}$ and F, respectively.

Although this equality may not be correct when some eigenvalues of $\mathbf{A}_n$ run outside the contour, the probability of this event decays rapidly to zero. A corrected version of $G_n(f)$ is

$$G_n(f)\mathbb{1}_{U_n} = -\frac{\mathbb{1}_{U_n}}{2\pi i}\oint_{\mathcal{C}} f(z) n\{m_n(z) - m(z) - \mathcal{X}_n(m(z))\}\,\mathrm{d}z,$$

where $U_n = \{\max_{1\le j\le n} |\lambda_j^{\mathbf{A}_n}| < \eta + \varepsilon\}$. The quantity $G_n(f)\mathbb{1}_{U_n^c}$ will not matter in our proof due to the bound for eigenvalues of $\mathbf{A}_n$ established in Lemma 6.1, and thus we need only to consider $G_n(f)\mathbb{1}_{U_n}$. Therefore, the problem of finding the limit distribution of $G_n(f)$ reduces to the study of

$$M_n(z) = n\{m_n(z) - m(z) - \mathcal{X}_n(m(z))\}.$$

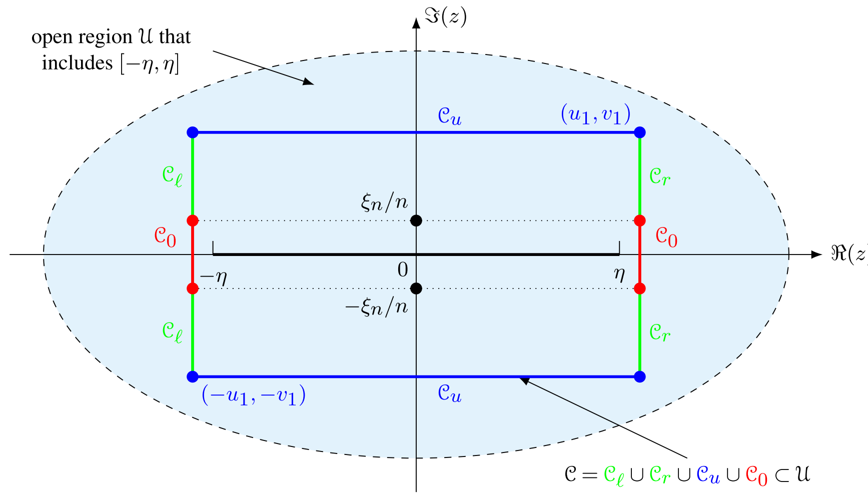


FIG. 2. *Open region $\mathcal{U}$ and decomposition of the closed contour $\mathcal{C}$.*

LEMMA 6.1. *Suppose the assumptions* (A)–(C1) *in Theorem* 3.1 *hold, then for any* $\varepsilon > 0$, *we have*

$$\mathbb{P}\Big(\max_{1\le j\le n} |\lambda_j^{\mathbf{A}_n}| \ge \eta + \varepsilon\Big) = o(n^{-1}),$$

where $\eta = 2\limsup_p \|\mathbf{\Sigma}_p\|/\sqrt{\theta}$.

The proof of this lemma is given in Section S2 of the Supplementary Material.

Throughout the paper, we set $\mathbb{C}_1 = \{z : z = u + iv, u \in [-u_1, u_1], \ |v| \ge v_1\}$. The limiting process of $M_n(z)$ on $\mathbb{C}_1$ is stated in the following proposition. Most of the remaining work will deal with proving this proposition.

PROPOSITION 6.1. *Under the assumption* $p \wedge n \to \infty$, $n^2/p = O(1)$ *and after truncation of the data, the empirical process* $\{M_n(z), z \in \mathbb{C}_1\}$ *converges weakly to a centered Gaussian process* $\{M(z), z \in \mathbb{C}_1\}$ *with the covariance function*

$$\Lambda(z_1, z_2) = m'(z_1)m'(z_2)\bigg[\frac{\omega}{\theta}(\nu_4 - 3) + 2\{1 - m(z_1)m(z_2)\}^{-2}\bigg]. \tag{33}$$

Now we explain how Proposition 6.1 implies Theorem 3.1. Write the contour $\mathcal{C}$ as $\mathcal{C} = \mathcal{C}_\ell \cup \mathcal{C}_r \cup \mathcal{C}_u \cup \mathcal{C}_0$ (see Figure 2), where

$$\mathcal{C}_\ell = \{z = -u_1 + iv, \xi_n/n < |v| < v_1\},$$
$$\mathcal{C}_r = \{z = u_1 + iv, \xi_n/n < |v| < v_1\},$$
$$\mathcal{C}_0 = \{z = \pm u_1 + iv, |v| \le \xi_n/n\},$$
$$\mathcal{C}_u = \{z = u \pm iv_1, |u| \le u_1\},$$

and $\{\xi_n\}$ is a slowly varying sequence of positive constants; recall that v_1 is a positive constant, which is independent of n. By $\mathcal{C}_\ell \cup \mathcal{C}_0 \cup \mathcal{C}_r = \mathcal{C} \setminus \mathbb{C}_1$, we can write

$$G_n(f)\mathbb{1}_{U_n} = -\frac{1}{2\pi i}\int_{\mathcal{C}_u} f(z)M_n(z)\mathbb{1}_{U_n}\,\mathrm{d}z - \frac{1}{2\pi i}\int_{\mathcal{C}\setminus\mathbb{C}_1} f(z)M_n(z)\mathbb{1}_{U_n}\,\mathrm{d}z. \tag{34}$$

To prove Theorem 3.1, we need to show that, for $j = \ell, r, 0$,

$$\lim_{v_1 \downarrow 0}\limsup_{n\to\infty}\int_{\mathcal{C}_j} \mathbb{E}|M_n(z)\mathbb{1}_{U_n}|^2\,\mathrm{d}z = 0 \tag{35}$$

and

$$\lim_{v_1 \downarrow 0} \int_{\mathcal{C}_j} \mathbb{E}|M(z)|^2 \,\mathrm{d}z = 0. \tag{36}$$

The proofs of (35) and (36) are provided in Section S3 of the Supplementary Material. By (35) and the fact that f is bounded over the bounded open region $\mathcal{U}$, for sufficiently small v_1, we have

$$\frac{1}{2\pi i} \int_{\mathcal{C} \backslash \mathbb{C}_1} f(z) M_n(z) \mathbb{1}_{U_n} \,\mathrm{d}z = o_P(1),$$

which, together with (34), (36) and Proposition 6.1, implies that

$$G_n(f)\mathbb{1}_{U_n} \xrightarrow{d} -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) M(z) \,\mathrm{d}z =: Y(f).$$

The calculation of the limiting covariance function of $Y(f)$ (see (7) and (8)) is quite similar to that given in Section 5 of Bai and Yao (2005); it is then omitted. This completes the proof of Theorem 3.1.

To prove Proposition 6.1, we decompose $M_n(z)$ into a random part $M_n^{(1)}(z)$ and a deterministic part $M_n^{(2)}(z)$ for $z \in \mathcal{C}$, where

$$M_n^{(1)}(z) = n\{m_n(z) - \mathbb{E}m_n(z)\}, \qquad M_n^{(2)}(z) = n\{\mathbb{E}m_n(z) - m(z) - \mathcal{X}_n(m(z))\}.$$

The random part contributes to the covariance function and the deterministic part contributes to the mean function. By Theorem 8.1 in Billingsley (1968), the proof of Proposition 6.1 is then complete if we can verify the following three steps:

Step 1. Finite-dimensional convergence of $M_n^{(1)}(z)$ in distribution on $\mathbb{C}_1$ to a centered multivariate Gaussian random vector with covariance function given by (33).

Step 2. Tightness of the $M_n^{(1)}(z)$ for $z \in \mathbb{C}_1$.

Step 3. Convergence of the nonrandom part $M_n^{(2)}(z)$ to zero on $\mathbb{C}_1$.

The proofs of these steps are presented in the coming sections.

6.3. *Finite-dimensional convergence of $M_n^{(1)}(z)$ in distribution.*

In this section, we consider the finite-dimensional convergence of the random part $M_n^{(1)}(z)$ under the assumption $p/n \to \infty$ (which is implied by $n^2/p = O(1)$).

LEMMA 6.2. *Suppose the assumptions* (*A*) *and* (*B*) *in Theorem* 3.1 *hold and* $p/n \to \infty$ *as* $p \wedge n \to \infty$, *then for any set of* r *points* $\{z_1, z_2, \dots, z_r\} \subseteq \mathbb{C}_1$, *the random vector* $(M_n^{(1)}(z_1), \dots, M_n^{(1)}(z_r))$ *converges weakly to a* r-*dimensional centered Gaussian distribution with covariance matrix given by* $\Lambda(z_i, z_j)$ *defined in* (33), *where* $1 \le i, j \le r$.

We now explain the sketch of proof of Lemma 6.2, and the technical details are provided in the Supplementary Material. By the fact that a random vector is multivariate normally distributed if and only if every linear combination of its components is normally distributed, we need only show that for any positive integer r and any complex sequence $\{a_j\}$, the sum $\sum_{j=1}^r a_j M_n^{(1)}(z_j)$ converges weakly to a Gaussian random variable. To this end, we first decompose the random part $M_n^{(1)}(z)$ as a sum of martingale difference. Then we apply the martingale CLT (Lemma S1.4) to obtain the asymptotic distribution of $M_n^{(1)}(z)$. Details of these two steps are provided in the Supplementary Material.

6.4. *Proof of tightness of $M_n^{(1)}(z)$.* This subsection is to verify the tightness of $M_n^{(1)}(z)$ for $z \in \mathbb{C}_1$ by using Theorem 12.3 of Billingsley (1968) (see Lemma S1.5). From Lemma S1.10 and Lemma S1.11, we can show that the condition (i) of Lemma S1.5 holds. Condition (ii) of Lemma S1.5 will be verified by showing

$$\frac{\mathbb{E}|M_n^{(1)}(z_1) - M_n^{(1)}(z_2)|^2}{|z_1 - z_2|^2} \le K, \quad z_1, z_2 \in \mathbb{C}_1. \tag{37}$$

The proof of (37) exactly follow Chen and Pan (2015); it is then omitted.

6.5. *Convergence of $M_n^{(2)}(z)$.* In this section, we obtain the asymptotic expansion of $n(\mathbb{E}m_n(z) - m(z))$ for $z \in \mathbb{C}_1$ (see definition of $\mathbb{C}_1$ in Section 6.2) and the result is stated in Lemma 6.3. This lemma, together with the finite-dimensional convergence (see Section 6.3) and the tightness of $M_n^{(1)}(z)$ (see Section 6.4), implies Proposition 6.1. To prove Lemma 6.3, we will follow the strategy in Khorunzhy, Khoruzhenko and Pastur (1996) and Bao (2015). The main tool is the generalized Stein's equation (see Lemma 6.4).

LEMMA 6.3. *With the same notation as in the previous sections*:

(1) *if $p \wedge n \to \infty$ and $n^2/p = O(1)$, we have*

$$M_n^{(2)} = n\{\mathbb{E}m_n(z) - m(z) - \mathcal{X}_n(m(z))\} = o(1), \tag{38}$$

uniformly for $z \in \mathbb{C}_1$, where $\mathcal{X}_n(m)$ is defined by (6);

(2) *if $p \wedge n \to \infty$ and $n^3/p = O(1)$, we have*

$$\begin{aligned} &n\left\{\mathbb{E}m_n(z) - m(z) + \sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}\frac{m^4}{1-m^2}\right\} \\ &\quad = \frac{m^3}{1-m^2}\left\{\frac{m^2}{1-m^2} + \frac{\tilde{b}_p}{b_p}(\nu_4 - 3) + 1\right\} + o(1) \end{aligned} \tag{39}$$

uniformly for $z \in \mathbb{C}_1$.

PROOF. Let $\mathbf{Y} = (npb_p)^{-1/4}\mathbf{X}$, then $\mathbf{A}_n = \mathbf{Y}'\mathbf{\Sigma}_p\mathbf{Y} - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_n$. To simplify notation, we denote

$$\mathbf{D} := (\mathbf{A}_n - z\mathbf{I}_n)^{-1}, \qquad \mathbf{E} := \mathbf{\Sigma}_p\mathbf{Y}\mathbf{D}\mathbf{Y}'\mathbf{\Sigma}_p = (E_{ij})_{p\times p}, \qquad \mathbf{F} := \mathbf{\Sigma}_p\mathbf{Y}\mathbf{D} = (F_{ij})_{p\times n}.$$

By the basic identity,

$$\mathbf{D} = -\frac{1}{z}\mathbf{I}_n + \frac{1}{z}\mathbf{D}\mathbf{A}_n = -\frac{1}{z}\mathbf{I}_n + \frac{1}{z}\left(\mathbf{D}\mathbf{Y}'\mathbf{\Sigma}_p\mathbf{Y} - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{D}\right),$$

we have

$$\begin{aligned} \mathbb{E}m_n(z) &= -\frac{1}{z} + \frac{1}{z}\cdot\frac{1}{n}\mathbb{E}\,\mathrm{tr}(\mathbf{D}\mathbf{A}_n) \\ &= -\frac{1}{z} - \frac{1}{z}\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbb{E}\left(\frac{1}{n}\mathrm{tr}\,\mathbf{D}\right) + \frac{1}{zn}\mathbb{E}\ \mathrm{tr}(\mathbf{Y}'\mathbf{\Sigma}_p\mathbf{Y}\mathbf{D}) \\ &= -\frac{1}{z} - \frac{1}{z}\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbb{E}m_n(z) + \frac{1}{zn}\sum_{j,k}\mathbb{E}(Y_{jk}F_{jk}). \end{aligned} \tag{40}$$

The basic idea of the following derivation is regarding $F_{jk} := (\mathbf{\Sigma}_p\mathbf{Y}\mathbf{D})_{jk}$ as an analytic function of Y_{jk}, and then use the generalized Stein's equation (Lemma 6.4 below) to expand $\mathbb{E}(Y_{jk}F_{jk})$ in (40).

LEMMA 6.4 (Generalized Stein's equation, Khorunzhy, Khoruzhenko and Pastur (1996)). *For any real-valued random variable* ξ *with* $\mathbb{E}|\xi|^{k+2} < \infty$ *and complex-valued function* $g(t)$ *with continuous and bounded* $k+1$ *derivatives, we have*

$$\mathbb{E}\{\xi g(\xi)\} = \sum_{a=0}^{k} \frac{\kappa_{a+1}}{a!} \mathbb{E}\{g^{(a)}(\xi)\} + \varepsilon,$$

where κ_a *is the* a*th cumulant of* ξ, *and*

$$|\varepsilon| \le C \sup_t \left|g^{(k+1)}(t)\right| \mathbb{E}\{|\xi|^{k+2}\},$$

where the positive constant C *depends on* k.

Applying Lemma 6.4 to the last term in (40), we obtain the following expansion:

$$\begin{aligned} \mathbb{E}m_n(z) &= -\frac{1}{z} - \frac{1}{z}\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbb{E}m_n(z) \\ &\quad + \frac{1}{zn}\sum_{a=0}^{4} \frac{1}{(npb_p)^{(a+1)/4}} \sum_{j,k} \frac{\kappa_{a+1}}{a!} \mathbb{E}\left(\frac{\partial^a F_{jk}}{\partial Y_{jk}^a}\right) + \varepsilon_n, \end{aligned} \tag{41}$$

where κ_a is the ath cumulant of Y_{jk}, $\frac{\partial^a F_{jk}}{\partial Y_{jk}^a}$ denotes the ath order derivative of F_{jk} with respect to Y_{jk}, and

$$|\varepsilon_n| \le \frac{K}{n} \frac{1}{(npb_p)^{6/4}} \sum_{j,k} \sup_{j,k} \mathbb{E}_{jk} \left|\frac{\partial^5 F_{jk}}{\partial Y_{jk}^5}\right|. \tag{42}$$

The explicit formula of the derivatives of F_{jk} are provided in Lemma S1.15.

From Lemma S1.14 and the identity $\mathbf{DX}'\mathbf{\Sigma}_p\mathbf{X} = pa_p\mathbf{D} + \sqrt{npb_p}(\mathbf{I}_n + z\mathbf{D})$, it is not difficult to obtain the following estimates:

$$\begin{aligned} &D_{kk}^{a_1} F_{jk}^{a_2} E_{jj}^{a_3} \le K n^{a_3/2} \left[\sum_{\alpha} \{(\mathbf{\Sigma Y})_{j\alpha}\}^2\right]^{(a_2+2a_3)/2} \qquad (a_1, a_2, a_3 \ge 0), \\ &\mathbb{E}\left|(\mathbf{\Sigma}_p^{-1/2}\mathbf{E}\mathbf{\Sigma}_p^{-1/2})_{jj} - \frac{\mathbb{E}m_n}{a_p\sqrt{p/(nb_p)} + z + \mathbb{E}m_n}\right| = O\left(\left(\frac{n}{p}\right)^2\right) + O\left(\frac{1}{p}\right), \\ &\left|\sum_{j,k} F_{jk}\right| = O\big((np)^{3/4}\big), \\ &\left|\sum_{j,k} F_{jk}^{a_2}\right| = O\big(p^{a_2/4} n^{1-a_2/4}\big) \quad (a_2 \ge 2). \end{aligned} \tag{43}$$

By (42), (43) and Lemma S1.15, we obtain $|\varepsilon_n| = o(1/n)$. By the fact

$$\mathbb{E}\left|D_{kk} + \frac{1}{z + \mathbb{E}m_n}\right|^2 = O\left(\frac{1}{n}\right) + O\left(\frac{n}{p}\right), \quad k = 1, \dots, n, \tag{44}$$

which is verified in Lemma S1.13, and the estimates above, we can extract the leading order terms in (41) to obtain

$$
\begin{aligned}
\mathbb{E}m_n(z) = &-\frac{1}{z} - \frac{1}{z}\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbb{E}m_n(z) \\
&+ \frac{1}{zn}\frac{1}{\sqrt{npb_p}}\sum_{j,k}\mathbb{E}\{(\boldsymbol{\Sigma}_p)_{jj}D_{kk} - E_{jj}D_{kk} - F_{jk}^2\} \\
&- \frac{1}{zn}\frac{\nu_4 - 3}{npb_p}\sum_{j,k}\mathbb{E}\{(\boldsymbol{\Sigma}_p)_{jj}^2 D_{kk}^2\} + o\left(\frac{1}{n}\right) \\
= &-\frac{1}{z} - \frac{1}{zn}\frac{1}{\sqrt{npb_p}}\mathbb{E}\{\operatorname{tr}(\mathbf{E})\operatorname{tr}(\mathbf{D})\} - \frac{1}{zn}\frac{1}{\sqrt{npb_p}}\mathbb{E}\{\operatorname{tr}(\mathbf{F}\mathbf{F}')\} \\
&- \frac{(\nu_4 - 3)\tilde{b}_p}{zn^2 b_p}\mathbb{E}\left(\sum_k D_{kk}^2\right) + o\left(\frac{1}{n}\right).
\end{aligned} \tag{45}
$$

Using the same argument as in the proof of Lemma S1.21, if $n^2/p = O(1)$, we can show that

$$
\mathbb{E}\left|\frac{1}{\sqrt{npb_p}}\operatorname{tr}\mathbf{E} - m(z)\right|^2 = O\left(\frac{1}{n}\right).
$$

This, together with c_r-inequality, implies that

$$
\mathbb{E}\left|\frac{1}{\sqrt{npb_p}}\operatorname{tr}\mathbf{E} - \mathbb{E}\frac{1}{\sqrt{npb_p}}\operatorname{tr}\mathbf{E}\right|^2 = o(1). \tag{46}
$$

Together with the fact that $\operatorname{Var}(m_n) = O(n^{-2})$ (see Lemma S1.12), we obtain

$$
\operatorname{Cov}\left(\frac{1}{\sqrt{npb_p}}\operatorname{tr}\mathbf{E}, \frac{1}{n}\operatorname{tr}\mathbf{D}\right) \le \sqrt{(46)}\cdot\sqrt{\operatorname{Var}(m_n)} = o\left(\frac{1}{n}\right). \tag{47}
$$

Note that

$$
\operatorname{tr}(\mathbf{F}\mathbf{F}') = \operatorname{tr}(\boldsymbol{\Sigma}_p\mathbf{Y}\mathbf{D}^2\mathbf{Y}'\boldsymbol{\Sigma}_p) = \frac{\partial}{\partial z}\operatorname{tr}(\boldsymbol{\Sigma}_p\mathbf{Y}\mathbf{D}\mathbf{Y}'\boldsymbol{\Sigma}_p) = \frac{\partial}{\partial z}\operatorname{tr}\mathbf{E}. \tag{48}
$$

Applying (44), (47) and (48) to (45), we have

$$
\begin{aligned}
\mathbb{E}m_n(z) = &-\frac{1}{z} - \frac{1}{z}\cdot\frac{1}{\sqrt{npb_p}}\mathbb{E}(\operatorname{tr}\mathbf{E})\cdot\frac{1}{n}\mathbb{E}(\operatorname{tr}\mathbf{D}) - \frac{1}{zn}\frac{1}{\sqrt{npb_p}}\mathbb{E}\left(\frac{\partial}{\partial z}\operatorname{tr}\mathbf{E}\right) \\
&- \frac{\nu_4 - 3}{zn}\frac{\tilde{b}_p}{b_p}\{\mathbb{E}m_n(z)\}^2 + o\left(\frac{1}{n}\right).
\end{aligned} \tag{49}
$$

The problem reduces to estimate $(1/\sqrt{npb_p})\mathbb{E}(\operatorname{tr}\mathbf{E})$. To this end, we apply Lemma 6.4 again to the term $(1/\sqrt{npb_p})\mathbb{E}(\operatorname{tr}\mathbf{E})$ to find its expansion. We denote

$$
\widehat{\mathbf{E}} := \boldsymbol{\Sigma}_p\mathbf{Y}\mathbf{D}\mathbf{Y}'\boldsymbol{\Sigma}_p^2, \qquad \widehat{\mathbf{F}} := \boldsymbol{\Sigma}_p^2\mathbf{Y}\mathbf{D},
$$

and write

$$
\frac{1}{\sqrt{npb_p}}\mathbb{E}(\operatorname{tr}\mathbf{E}) = \frac{1}{\sqrt{npb_p}}\sum_{j,k}\mathbb{E}(Y_{jk}\widehat{F}_{jk}). \tag{50}
$$

The first four derivatives of $\widehat{F}_{jk}$ with respect to Y_{jk} are presented in Lemma S1.16. Applying generalized Stein's equation with the derivatives of $\widehat{F}_{jk}$ to the last term in (50), and using the similar estimates above, gives us

$$\begin{aligned}
\frac{1}{\sqrt{npb_p}}\mathbb{E}(\operatorname{tr}\mathbf{E}) &= \frac{1}{\sqrt{npb_p}}\sum_{a=0}^{3}\frac{1}{(npb_p)^{(a+1)/4}}\sum_{j,k}\frac{\kappa_{a+1}}{a!}\mathbb{E}\left(\frac{\partial^a \widehat{F}_{jk}}{\partial Y_{jk}^a}\right)+\widetilde{\varepsilon}_n \\
&= \frac{1}{npb_p}\sum_{j,k}\mathbb{E}\{(\mathbf{\Sigma}_p^2)_{jj}D_{kk}-\widehat{E}_{jj}D_{kk}-F_{jk}\widehat{F}_{jk}\} \\
&\quad+\frac{1}{\sqrt{npb_p}}\frac{\nu_4-3}{npb_p}\sum_{j,k}\mathbb{E}\{-(\mathbf{\Sigma}_p^2)_{jj}(\mathbf{\Sigma}_p)_{jj}D_{kk}^2\}+o\left(\frac{1}{n}\right) \\
&= \mathbb{E}m_n-\frac{1}{npb_p}\mathbb{E}\{\operatorname{tr}(\widehat{\mathbf{E}})\operatorname{tr}(\mathbf{D})\}-\frac{1}{npb_p}\mathbb{E}\{\operatorname{tr}(\mathbf{F}\widehat{\mathbf{F}}')\} \\
&\quad-\frac{1}{\sqrt{npb_p}}\frac{\nu_4-3}{npb_p}\mathbb{E}\left\{\sum_j(\mathbf{\Sigma}_p^2)_{jj}(\mathbf{\Sigma}_p)_{jj}\right\}\left(\sum_k D_{kk}^2\right)+o\left(\frac{1}{n}\right)
\end{aligned}$$

$$= \mathbb{E}m_n-\sqrt{\frac{n}{p}}\frac{1}{\sqrt{npb_p}}\mathbb{E}(\operatorname{tr}\widehat{\mathbf{E}})\cdot\frac{1}{n}\mathbb{E}(\operatorname{tr}\mathbf{D})+o\left(\frac{1}{n}\right) \tag{51}$$

$$= \mathbb{E}m_n-\sqrt{\frac{n}{p}}\left\{\frac{c_p}{b_p\sqrt{b_p}}\mathbb{E}m_n+O\left(\sqrt{\frac{n}{p}}\right)\right\}\cdot\mathbb{E}m_n+o\left(\frac{1}{n}\right) \tag{52}$$

$$= \mathbb{E}m_n-\sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(\mathbb{E}m_n)^2+o\left(\sqrt{\frac{n}{p}}\right)+o\left(\frac{1}{n}\right), \tag{53}$$

where (52) comes from similar arguments in the proof of Lemma S1.21. Plugging (53) into (49), we have

$$\begin{aligned}
\mathbb{E}m_n &= -\frac{1}{z}-\frac{1}{z}\left\{\mathbb{E}m_n-\sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(\mathbb{E}m_n)^2\right\}\mathbb{E}m_n \\
&\quad-\frac{1}{zn}\cdot\frac{\partial}{\partial z}\left\{\mathbb{E}m_n-\sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}\cdot(\mathbb{E}m_n)^2\right\}-\frac{\nu_4-3}{zn}\frac{\tilde{b}_p}{b_p}(\mathbb{E}m_n)^2 \\
&\quad+o\left(\sqrt{\frac{n}{p}}\right)+o\left(\frac{1}{n}\right) \\
&= -\frac{1}{z}-\frac{(\mathbb{E}m_n)^2}{z}+\frac{1}{z}\sqrt{\frac{n}{p}}\frac{c_p m^3}{b_p\sqrt{b_p}}-\frac{1}{zn}\left\{\frac{m^2}{1-m^2}+\frac{(\nu_4-3)\tilde{b}_p}{b_p}m^2\right\} \\
&\quad+o\left(\sqrt{\frac{n}{p}}\right)+o\left(\frac{1}{n}\right).
\end{aligned}$$

Solving this equation yields that

$$\begin{aligned}
n(\mathbb{E}m_n-m) &= \frac{m^3}{1-m^2}\left\{\frac{m^2}{1-m^2}+\frac{\tilde{b}_p}{b_p}(\nu_4-3)+1\right\} \\
&\quad-\sqrt{\frac{n^3}{p}}\frac{c_p}{b_p\sqrt{b_p}}\frac{m^4}{1-m^2}++o\left(\sqrt{\frac{n^3}{p}}\right)+o(1).
\end{aligned} \tag{54}$$

This implies (39) under the assumption $n^3/p = O(1)$.

Moreover, to obtain (38) under the assumption $n^2/p = O(1)$, we need to figure out the remainder term $o(\sqrt{n/p})$ in (53) more carefully. Indeed, this remainder term comes from the estimate of $\mathbb{E}(\operatorname{tr}\widehat{\mathbf{E}})/(b_p\sqrt{np})$ in (51). To get a more precise estimation, we use the similar argument above for calculating the asymptotic expansion of $\mathbb{E}(\operatorname{tr}\widehat{\mathbf{E}})/(b_p\sqrt{np})$:

$$\frac{1}{\sqrt{np}b_p}\mathbb{E}(\operatorname{tr}\widehat{\mathbf{E}}) = \frac{c_p}{b_p\sqrt{b_p}}\mathbb{E}m_n - \sqrt{\frac{n}{p}}\frac{d_p}{b_p^2}(\mathbb{E}m_n)^2 + O\left(\frac{n}{p}\right). \tag{55}$$

By substituting this expression into (51), we obtain a more precise estimation than (53):

$$\frac{1}{\sqrt{npb_p}}\mathbb{E}(\operatorname{tr}\mathbf{E}) = \mathbb{E}m_n - \sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(\mathbb{E}m_n)^2 + \frac{n}{p}\frac{d_p}{b_p^2}(\mathbb{E}m_n)^3 + o\left(\frac{1}{n}\right) + o\left(\frac{n}{p}\right). \tag{56}$$

Plugging (56) into (49), we have

$$\begin{aligned}
\mathbb{E}m_n &= -\frac{1}{z} - \frac{1}{z}\left\{\mathbb{E}m_n - \sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(\mathbb{E}m_n)^2 + \frac{n}{p}\frac{d_p}{b_p^2}(\mathbb{E}m_n)^3\right\}\mathbb{E}m_n \\
&\quad - \frac{1}{zn}\cdot\frac{\partial}{\partial z}\left\{\mathbb{E}m_n - \sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(\mathbb{E}m_n)^2\right\} - \frac{\nu_4 - 3}{zn}\frac{\tilde{b}_p}{b_p}(\mathbb{E}m_n)^2 + o\left(\frac{1}{n}\right) + o\left(\frac{n}{p}\right) \\
&= -\frac{1}{z} - \frac{1}{z}(\mathbb{E}m_n)^2 + \frac{1}{z}\sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(\mathbb{E}m_n)^3 - \frac{1}{z}\frac{n}{p}\frac{d_p}{b_p^2}m^4 \\
&\quad - \frac{1}{zn}\left\{\frac{m^2}{1-m^2} + \frac{(\nu_4-3)\tilde{b}_p}{b_p}m^2\right\} + o\left(\frac{1}{n}\right) + o\left(\frac{n}{p}\right).
\end{aligned}$$

Multiplying $-z$ on both sides, we have

$$\begin{aligned}
-z\mathbb{E}m_n &= 1 + (\mathbb{E}m_n)^2 - \sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(\mathbb{E}m_n)^2\cdot\mathbb{E}m_n + \frac{n}{p}\frac{d_p}{b_p^2}m^4 \\
&\quad + \frac{1}{n}\left\{\frac{m^2}{1-m^2} + \frac{(\nu_4-3)\tilde{b}_p}{b_p}m^2\right\} + o\left(\frac{1}{n}\right) + o\left(\frac{n}{p}\right).
\end{aligned} \tag{57}$$

This implies that

$$(\mathbb{E}m_n)^2 = -1 - z\mathbb{E}m_n + \sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(\mathbb{E}m_n)^3 + O\left(\frac{1}{n}\right) + O\left(\frac{n}{p}\right). \tag{58}$$

Plugging (58) into (57) yields that

$$\begin{aligned}
-z\mathbb{E}m_n &= 1 + (\mathbb{E}m_n)^2 + \sqrt{\frac{n}{p}}\frac{c_p}{b_p\sqrt{b_p}}(1 + z\mathbb{E}m_n)\mathbb{E}m_n - \frac{n}{p}\frac{c_p^2}{b_p^3}m^4 \\
&\quad + \frac{n}{p}\frac{d_p}{b_p^2}m^4 + \frac{1}{n}\left\{\frac{m^2}{1-m^2} + \frac{(\nu_4-3)\tilde{b}_p}{b_p}m^2\right\} + o\left(\frac{1}{n}\right) + o\left(\frac{n}{p}\right).
\end{aligned}$$

This equation can be written as a quadratic equation of $\mathbb{E}m_n - m$:

$$0 = \mathcal{A}_n(m)(\mathbb{E}m_n - m)^2 + \mathcal{B}_n(m)(\mathbb{E}m_n - m) + \mathcal{C}_n(m) + o\left(\frac{1}{n}\right) + o\left(\frac{n}{p}\right),$$

where $\mathcal{A}_n(m)$, $\mathcal{B}_n(m)$ and $\mathcal{C}_n(m)$ are defined in (6). The equation has two solutions:

$$x_1 = \frac{-\mathcal{B}_n(m) + \sqrt{\mathcal{B}_n(m)^2 - 4\mathcal{A}_n(m)\mathcal{C}_n(m)}}{2\mathcal{A}_n(m)} + o\left(\frac{1}{n}\right) + o\left(\frac{n}{p}\right),$$

$$x_2 = \frac{-\mathcal{B}_n(m) - \sqrt{\mathcal{B}_n(m)^2 - 4\mathcal{A}_n(m)\mathcal{C}_n(m)}}{2\mathcal{A}_n(m)} + o\left(\frac{1}{n}\right) + o\left(\frac{n}{p}\right).$$

When $n^2/p = O(1)$, from the definition of $\mathcal{A}_n(m)$, $\mathcal{B}_n(m)$ and $\mathcal{C}_n(m)$, we can verify that $x_1 = o(1)$ while $x_2 = (1-m^2)/m^2 + o(1)$. Since $\mathbb{E}m_n - m = o(1)$, we choose x_1 to be the expression of $\mathbb{E}m_n - m$, and

$$n(\mathbb{E}m_n - m) = n \cdot \frac{-\mathcal{B}_n(m) + \sqrt{\mathcal{B}_n(m)^2 - 4\mathcal{A}_n(m)\mathcal{C}_n(m)}}{2\mathcal{A}_n(m)} + o(1) + o\left(\frac{n^2}{p}\right). \tag{59}$$

This implies (38) under the assumption $n^2/p = O(1)$. Though the constants that appear in successive estimations used in this proof are implicitly given, they are all independent of $z \in \mathbb{C}_1$. Thus, the convergence of $M_n^{(1)}(z)$ is uniform over $z \in \mathbb{C}_1$. □

REMARK 6.1 (Weakening of the assumption $n^2/p = O(1)$). It is indeed possible to relax the assumption $n^2/p = O(1)$, that is, $p \geq Kn^2$ for some constant K and large n, to cover also the range $n \ll p \ll n^2$ for the dimension p by using an approximation for $\mathbb{E}(\operatorname{tr}\widehat{\mathbf{E}})/(b_p\sqrt{np})$ of a higher order than in (55). By iterating the expansion process with Lemma 6.4, we can expand further the $O(n/p)$ term in (55) and obtain

$$\frac{1}{\sqrt{np}b_p}\mathbb{E}(\operatorname{tr}\widehat{\mathbf{E}}) = \frac{c_p}{b_p\sqrt{b_p}}\mathbb{E}m_n - \sqrt{\frac{n}{p}}\frac{d_p}{b_p^2}(\mathbb{E}m_n)^2 + \frac{n}{p}\frac{e_p}{b_p^{5/2}}(\mathbb{E}m_n)^3 + O\left(\left(\frac{n}{p}\right)^{3/2}\right), \tag{60}$$

where $e_p = \operatorname{tr}(\mathbf{\Sigma}_p^5)/p$. Substituting this expression into (51) leads to a cubic equation for $n(\mathbb{E}m_n - m)$ with reminder term $n \cdot \sqrt{n/p} \cdot O((n/p)^{3/2}) = o(n^{5/2}/p^{3/2})$, which comes from (51) and (60). This reminder term is $o(1)$ under the assumption $p \geq Kn^{5/3}$, which enlarges the previous range of $p \geq Kn^2$. The other derivations for our CLT (Theorem 3.1) remain valid under the scenario $p \geq Kn^{5/3}$; so, the theorem still holds.

In fact, we can iterate this expansion process to cover the whole range of $n \ll p \ll n^2$. However, this leads higher and higher degree equations for $n(\mathbb{E}m_n - m)$ and we cannot obtain a closed-form formula for $\mathcal{X}_n(m)$, which is the approximation of $n(\mathbb{E}m_n - m)$ that appears in the spectral statistic $G_n(f)$ as a correction term. This presents a problem in practical applications and we do not pursue this direction in the paper.

Acknowledgments The authors are grateful to the Editor, the Associate Editor and the referees whose numerous comments on both technical and expository aspects of the paper have much contributed to its improvement.

Zeng Li is the corresponding author.

Funding Zeng Li's research is supported by NSFC (National Natural Science Foundation of China) Grant No. 12101292, NSFC Grant No. 12031005 and Shenzhen Fundamental Research Program JCYJ20220818100602005.

SUPPLEMENTARY MATERIAL

Supplement to "Asymptotic normality for eigenvalue statistics of a general sample covariance matrix when $p/n \to \infty$ and applications" (DOI: 10.1214/23-AOS2300SUPP; .pdf). This supplementary document contains some technical lemmas and their proofs, proofs of Lemmas 6.1–6.2, equations (15), (35), (36), Theorem 4.2 and Proposition 4.1. We also report some additional simulation results in this document.

REFERENCES

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 https://doi.org/10.1214/09-AOS694

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

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 https://doi.org/10.1214/aop/1078415845

BAI, Z. D. and YAO, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11** 1059–1092. MR2189081 https://doi.org/10.3150/bj/1137421640

BAI, Z. D. and YIN, Y. Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. MR0929083

BAO, Z. (2015). On asymptotic expansion and central limit theorem of linear eigenvalue statistics for sample covariance matrices when $N/M \to 0$. *Theory Probab. Appl.* **59** 185–207. MR3416046 https://doi.org/10.1137/S0040585X97T987089

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

CHEN, B. and PAN, G. (2015). CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli* **21** 1089–1133. MR3338658 https://doi.org/10.3150/14-BEJ599

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 https://doi.org/10.3150/11-BEJ381

CHEN, R., XIAO, H. and YANG, D. (2021). Autoregressive models for matrix-valued time series. *J. Econometrics* **222** 539–560. MR4234832 https://doi.org/10.1016/j.jeconom.2020.07.015

COUILLET, R. and DEBBAH, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge Univ. Press, Cambridge. MR2884783 https://doi.org/10.1017/CBO9780511994746

KHORUNZHY, A. M., KHORUZHENKO, B. A. and PASTUR, L. A. (1996). Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.* **37** 5033–5060. MR1411619 https://doi.org/10.1063/1.531589

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 https://doi.org/10.1214/aos/1031689018

LI, Z. and YAO, J. (2016). Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size. *Electron. J. Stat.* **10** 2973–3010. MR3567239 https://doi.org/10.1214/16-EJS1199

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

QIU, J., LI, Z. and YAO, J. (2023). Supplement to "Asymptotic normality for eigenvalue statistics of a general sample covariance matrix when $p/n \to \infty$ and applications." https://doi.org/10.1214/23-AOS2300SUPP

WANG, L. and PAUL, D. (2014). Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.* **126** 25–52. MR3173080 https://doi.org/10.1016/j.jmva.2013.12.015

WANG, Q. and YAO, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.* **7** 2164–2192. MR3104916 https://doi.org/10.1214/13-EJS842

YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 https://doi.org/10.1017/CBO9781107588080

Submitted to the Annals of Statistics

SUPPLEMENT TO "ASYMPTOTIC NORMALITY FOR EIGENVALUE STATISTICS OF A GENERAL SAMPLE COVARIANCE MATRIX WHEN $P/N \to \infty$ AND APPLICATIONS"

BY JIAXIN QIU[1,a], ZENG LI[2,b] AND JIANFENG YAO[3,c]

[1]*Department of Statistics and Actuarial Science, The University of Hong Kong,* [a]*qiujx@connect.hku.hk*

[2]*Department of Statistics and Data Science, Southern University of Science and Technology,* [b]*liz9@sustech.edu.cn*

[3]*School of Data Science, The Chinese University of Hong Kong (Shenzhen),* [c]*jeffyao@cuhk.edu.cn*

CONTENTS

This supplementary document contains some technical lemmas and their proofs, proofs of Lemmas 6.1 – 6.2, equations (15), (35), (36), Theorem 4.2, and Proposition 4.1. We also report some additional simulation results in this document.

S1. Some technical lemmas.

LEMMA S1.1 (Bai and Silverstein (2010), Lemma B.26). *Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ non-random matrix and $\mathbf{x} = (X_1, \dots, X_n)'$ be a random vector of independent entries. Assume that $\mathbb{E}X_i = 0$, $\mathbb{E}|X_i|^2 = 1$ and $\mathbb{E}|X_i|^\ell \leqslant \nu_\ell$. Then, for any $k \geqslant 1$,*

$$\mathbb{E}|\mathbf{x}^*\mathbf{A}\mathbf{x} - \mathrm{tr}\mathbf{A}|^k \leqslant C_k\Big[\big\{\nu_4\mathrm{tr}(\mathbf{A}\mathbf{A}^*)\big\}^{k/2} + \nu_{2k}\mathrm{tr}(\mathbf{A}\mathbf{A}^*)^{k/2}\Big],$$

where C_k is a constant depending on k only.

LEMMA S1.2 (Pan and Zhou (2011), Lemma 5). *Let* $\mathbf{A}$ *be a* $p \times p$ *deterministic complex matrix with zero diagonal elements. Let* $\mathbf{x} = (X_1, \ldots, X_p)'$ *be a random vector with i.i.d. real entries. Assume that* $\mathbb{E}X_i = 0$, $\mathbb{E}|X_i|^2 = 1$. *Then, for any* $k \geqslant 2$,

$$\mathbb{E}|\mathbf{x}'\mathbf{A}\mathbf{x}|^k \leqslant C_k \big\{\mathbb{E}|X_1|^k\big\}^2 \big(\mathrm{tr}\mathbf{A}\mathbf{A}^*\big)^{k/2}, \tag{S1.1}$$

where C_k *is a constant depending on* k *only.*

LEMMA S1.3 (Burkholder's inequality, Burkholder (1973)). *Let* $\{X_i\}$ *be a complex martingale difference sequence withe respect to the increasing* σ*-field* $\{\mathcal{F}_i\}$. *Then for* $k \geqslant 2$, *the following inequality*

$$\mathbb{E}\Big|\sum_i X_i\Big|^k \leqslant C_k \mathbb{E}\Big[\sum_i \mathbb{E}\big\{|X_i|^2 \mid \mathcal{F}_{i-1}\big\}\Big]^{k/2} + C_k \mathbb{E}\sum_i |X_i|^k$$

holds, where C_k *is a constant depending on* k *only.*

LEMMA S1.4 (Martingale CLT, Billingsley (2008)). *Suppose for each* n, $\{Y_{nk}\}_{1\leqslant k\leqslant r_n}$ *is a real martingale difference sequence with respect to the* σ*-field* $\{\mathcal{F}_{nk}\}$ *having second moments. If as* $n \to \infty$,

$$\sum_{k=1}^{r_n} \mathbb{E}(Y_{nk}^2 \mid \mathcal{F}_{n,k-1}) \xrightarrow{p} \sigma^2,$$

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

$$\sum_{k=1}^{r_n} \mathbb{E}\Big(Y_{nk}^2 \mathbb{1}_{\{|Y_{nk}|\geqslant\varepsilon\}}\Big) \to 0,$$

then

$$\sum_{k=1}^{r_n} Y_{nk} \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

LEMMA S1.5 (Billingsley (1968), Theorem 12.3). *The sequence* $\{X_n\}$ *is tight if it satisfies these two conditions:*

(i) *The sequence* $\{X_n(0)\}$ *is tight.*
(ii) *There exist constants* $\gamma \geqslant 0$ *and* $\alpha > 1$ *and a non-decreasing, continuous function* F *on* $[0,1]$ *such that*

$$\mathbb{P}\Big(\big|X_n(t_2) - X_n(t_1)\big| \geqslant \lambda\Big) \leqslant \frac{1}{\lambda^\gamma}\big|F(t_2) - F(t_1)\big|^\alpha$$

holds for all t_1, t_2 *and* n *and all positive* λ.

LEMMA S1.6 (Bai and Silverstein (2004), Lemma 2.3). *Let* f *be analytic in* D, *a connected open set of* $\mathbb{C}$, *satisfying* $|f(z)| \leqslant M$ *for any* $z \in D$, *then, on any set bounded by a contour interior to* D, $f'(z)$ *is bounded.*

Lemma S1.7 is about the asymptotic expression of $\mathbb{Z}_k$, which is used in Section S2.16 to derive the finite-dimensional convergence of $M_n^{(1)}(z)$.

LEMMA S1.7. *For* $z_1, z_2 \in \mathbb{C}^+$,

$$\mathbb{Z}_k := \frac{1}{n(pb_p)^2}\mathrm{tr}\Big\{\mathbb{E}_k\mathbf{M}_k^{(1)}(z_1)\cdot\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\Big\} = \frac{\frac{k}{n}m(z_1)m(z_2)}{1-\frac{k}{n}m(z_1)m(z_2)} + o_{L_1}(1).$$

Lemmas S1.8 and S1.9 are used in the proof of Lemma S1.7.

LEMMA S1.8. *For* $\vartheta_i(z)$ *and* $\zeta_i(z)$ *defined in Lemma S1.7, we have*

$$\mathbb{E}\Big|\vartheta_i(z) - \frac{m(z)}{z}\Big|^4 \to 0, \qquad \mathbb{E}\Big|\zeta_i(z) + zm(z)\Big|^4 \to 0, \qquad \text{as } n\to\infty.$$

LEMMA S1.9. *Let* $\mathbf{B}$ *be any matrix independent of* $\mathbf{x}_i$.

$$\mathbb{E}\big|\mathbf{x}_i'\mathbf{M}_{ki}\mathbf{B}\mathbf{x}_i\big|^2 \leqslant Kp^2n^2\mathbb{E}\|\mathbf{B}\|^2, \tag{S1.2}$$

$$\mathbb{E}\big|\mathbf{x}_i'\mathbf{\Sigma}_p\mathbf{M}_{ki}\mathbf{B}\mathbf{\Sigma}_p\mathbf{x}_i\big|^2 \leqslant Kp^2n^2\mathbb{E}\|\mathbf{B}\|^2. \tag{S1.3}$$

Lemmas S1.10 and S1.11 are used in Sections S2.16 and 6.4.

LEMMA S1.10. *For* $z\in\mathbb{C}_1$, *we have*

$$|\beta_k(z)| \leqslant 1/v_1, \qquad |\beta_k^{\mathrm{tr}}(z)| \leqslant 1/v_1,$$

$$\Big|1 + \frac{1}{npb_p}\mathrm{tr}\mathbf{M}_k^{(s)}(z)\Big| \leqslant 1 + \frac{1}{v_1^s}, \qquad s=1,2,$$

$$\big|\beta_k\{1+\mathbf{q}_k'\mathbf{D}_k^2(z)\mathbf{q}_k\}\big| \leqslant \frac{1}{v_1}. \tag{S1.4}$$

LEMMA S1.11. *Under the assumption* $p\wedge n\to\infty,\ p/n\to\infty$ *and truncation, for* $z\in\mathbb{C}_1$,

$$\mathbb{E}|\gamma_{ks}|^2 \leqslant \frac{K}{n}, \qquad \mathbb{E}|\gamma_{ks}|^4 \leqslant K\left(\frac{1}{n^2}+\frac{n}{p^2}\right),$$

$$\mathbb{E}|\eta_k|^2 \leqslant \frac{K}{n}, \qquad \mathbb{E}|\eta_k|^4 \leqslant K\frac{\delta_n^4}{n} + K\left(\frac{1}{n^2}+\frac{n}{p^2}\right).$$

Lemmas S1.12 , S1.13 and S1.14 are used in Section 6.5 to derive the convergence of the non-random part $M_n^{(2)}(z)$. They are proved following the strategy in Bao (2015).

LEMMA S1.12. *Under the assumption* $p\wedge n\to\infty,\ p/n\to\infty$, *for* $z\in\mathbb{C}_1$, *we have*

$$\mathrm{Var}(m_n) = O\left(\frac{1}{n^2}\right). \tag{S1.5}$$

LEMMA S1.13. *Under the assumption* $p\wedge n\to\infty,\ p/n\to\infty$, *for* $z\in\mathbb{C}_1$ *and* $1\leqslant\ell\leqslant n$,

$$\mathbb{E}\Big|D_{\ell\ell}+\frac{1}{z+\mathbb{E}m_n}\Big|^2 = O\left(\frac{1}{n}\right)+O\left(\frac{n}{p}\right).$$

LEMMA S1.14. *Under the assumption* $p\wedge n\to\infty,\ p/n\to\infty$, *for* $z\in\mathbb{C}_1$ *and* $1\leqslant\ell\leqslant p$,

$$\mathbb{E}\Big|\widetilde{D}_{\ell\ell}+\frac{1}{a_p\sqrt{p/(nb_p)}+z+\mathbb{E}m_n}\Big|^2 = O\left(\left(\frac{n}{p}\right)^3\right)+O\left(\frac{n}{p^2}\right), \tag{S1.6}$$

where $D_{\ell\ell}$ is the ℓ-th diagonal entry of the matrix

$$\widetilde{\mathbf{D}} = (\widetilde{D}_{ij})_{p\times p} = \left(\mathbf{\Sigma}_p^{1/2}\mathbf{Y}\mathbf{Y}'\mathbf{\Sigma}_p^{1/2} - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_p - z\mathbf{I}_p\right)^{-1}.$$

Lemmas S1.15 and S1.16 provide some derivatives of F_{jk} and $\widehat{F}_{jk}$ for applying the generalized Stein's equation in Section 6.5.

Recall that

$$\mathbf{D} := (\mathbf{A}_n - z\mathbf{I}_n)^{-1}, \qquad \mathbf{E} := \mathbf{\Sigma}_p\mathbf{Y}\mathbf{D}\mathbf{Y}'\mathbf{\Sigma}_p = (E_{ij})_{p\times p}, \qquad \mathbf{F} := \mathbf{\Sigma}_p\mathbf{Y}\mathbf{D} = (F_{ij})_{p\times n}.$$

LEMMA S1.15 (Derivatives of F_{jk}).

$$\frac{\partial F_{jk}}{\partial Y_{jk}} = (\mathbf{\Sigma}_p)_{jj}D_{kk} - E_{jj}D_{kk} - F_{jk}^2;$$

$$\frac{\partial^2 F_{jk}}{\partial Y_{jk}^2} = -6(\mathbf{\Sigma}_p)_{jj}F_{jk}D_{kk} + 6E_{jj}F_{jk}D_{kk} + 2F_{jk}^3;$$

$$\frac{\partial^3 F_{jk}}{\partial Y_{jk}^3} = -6(\mathbf{\Sigma}_p)_{jj}^2D_{kk}^2 + 36(\mathbf{\Sigma}_p)_{jj}F_{jk}^2D_{kk} + 12(\mathbf{\Sigma}_p)_{jj}E_{jj}D_{kk}^2 - 36E_{jj}F_{jk}^2D_{kk} - 6E_{jj}^2D_{kk}^2 - 6F_{jk}^4;$$

$$\begin{aligned}\frac{\partial^4 F_{jk}}{\partial Y_{jk}^4} &= 120(\mathbf{\Sigma}_p)_{jj}^2F_{jk}D_{kk}^2 - 240(\mathbf{\Sigma}_p)_{jj}F_{jk}^3D_{kk} - 240(\mathbf{\Sigma}_p)_{jj}E_{jj}F_{jk}D_{kk}^2 + 240E_{jj}F_{jk}^3D_{kk}\\ &\quad + 120E_{jj}^2F_{jk}D_{kk}^2 + 24F_{jk}^5;\end{aligned}$$

$$\begin{aligned}\frac{\partial^5 F_{jk}}{\partial Y_{jk}^5} &= -120F_{jk}^6 - 1800E_{jj}F_{jk}^4D_{kk} - 1800E_{jj}^2F_{jk}^2D_{kk}^2 - 120E_{jj}^3D_{kk}^3 + 1800(\mathbf{\Sigma}_p)_{jj}F_{jk}^4D_{kk}\\ &\quad + 3600(\mathbf{\Sigma}_p)_{jj}E_{jj}F_{jk}^2D_{kk}^2 + 360(\mathbf{\Sigma}_p)_{jj}E_{jj}^2D_{kk}^3 - 1800(\mathbf{\Sigma}_p)_{jj}^2F_{jk}^2D_{kk}^2\\ &\quad - 360(\mathbf{\Sigma}_p)_{jj}^2E_{jj}D_{kk}^3 + 120(\mathbf{\Sigma}_p)_{jj}^3D_{kk}^3.\end{aligned}$$

Recall that

$$\widehat{\mathbf{E}} = \mathbf{\Sigma}_p\mathbf{Y}\mathbf{D}\mathbf{Y}'\mathbf{\Sigma}_p^2 = (\widehat{E}_{ij})_{p\times p}, \qquad \widehat{\mathbf{F}} := \mathbf{\Sigma}_p^2\mathbf{Y}\mathbf{D} = (\widehat{F}_{ij})_{p\times n}.$$

LEMMA S1.16 (Derivatives of $\widehat{F}_{jk}$).

$$\frac{\partial \widehat{F}_{jk}}{\partial Y_{jk}} = (\mathbf{\Sigma}_p^2)_{jj}D_{kk} - \widehat{E}_{jj}D_{kk} - F_{jk}\widehat{F}_{jk};$$

$$\frac{\partial^2 \widehat{F}_{jk}}{\partial Y_{jk}^2} = -2(\mathbf{\Sigma}_p)_{jj}\widehat{F}_{jk}D_{kk} - 4(\mathbf{\Sigma}_p^2)_{jj}F_{jk}D_{kk} + 2F_{jk}^2\widehat{F}_{jk} + 4\widehat{E}_{jj}F_{jk}D_{kk} + 2E_{jj}\widehat{F}_{jk}D_{kk};$$

$$\begin{aligned}\frac{\partial^3 \widehat{F}_{jk}}{\partial Y_{jk}^3} &= -6(\mathbf{\Sigma}_p^2)_{jj}(\mathbf{\Sigma}_p)_{jj}D_{kk}^2 - 6F_{jk}^3\widehat{F}_{jk} - 18\widehat{E}_{jj}F_{jk}^2D_{kk} - 18E_{jj}F_{jk}\widehat{F}_{jk}D_{kk} - 6E_{jj}\widehat{E}_{jj}D_{kk}^2\\ &\quad + 18(\mathbf{\Sigma}_p)_{jj}F_{jk}\widehat{F}_{jk}D_{kk} + 6(\mathbf{\Sigma}_p)_{jj}\widehat{E}_{jj}D_{kk}^2 + 18(\mathbf{\Sigma}_p^2)_{jj}F_{jk}^2D_{kk} + 6(\mathbf{\Sigma}_p^2)_{jj}E_{jj}D_{kk}^2;\end{aligned}$$

$$\begin{aligned}\frac{\partial^4 \widehat{F}_{jk}}{\partial Y_{jk}^4} &= 24F_{jk}^4\widehat{F}_{jk} + 96\widehat{E}_{jj}F_{jk}^3D_{kk} + 144E_{jj}F_{jk}^2\widehat{F}_{jk}D_{kk} + 96E_{jj}\widehat{E}_{jj}F_{jk}D_{kk}^2 + 24E_{jj}^2\widehat{F}_{jk}D_{kk}^2\\ &\quad - 144(\mathbf{\Sigma}_p)_{jj}F_{jk}^2\widehat{F}_{jk}D_{kk} - 96(\mathbf{\Sigma}_p)_{jj}\widehat{E}_{jj}F_{jk}D_{kk}^2 - 48(\mathbf{\Sigma}_p)_{jj}E_{jj}\widehat{F}_{jk}D_{kk}^2 + 24(\mathbf{\Sigma}_p)_{jj}^2\widehat{F}_{jk}D_{kk}^2\\ &\quad - 96(\mathbf{\Sigma}_p^2)_{jj}F_{jk}^3D_{kk} - 96(\mathbf{\Sigma}_p^2)_{jj}E_{jj}F_{jk}D_{kk}^2 + 96(\mathbf{\Sigma}_p)_{jj}(\mathbf{\Sigma}_p^2)_{jj}F_{jk}D_{kk}^2.\end{aligned}$$

Lemmas S1.17 and S1.18 provide the derivatives of some quantities with respect to Y_{jk}, which can be used to obtain the derivatives of F_{jk} (Lemma S1.15) and $\widehat{F}_{jk}$ (Lemma S1.16).

LEMMA S1.17. *For any $\alpha, j \in \{1, 2, \ldots, p\}$ and $\beta, k \in \{1, 2, \ldots, n\}$, we have*

$$\frac{\partial D_{\alpha\beta}}{\partial Y_{jk}} = -F_{j\alpha}D_{\beta k} - F_{j\beta}D_{\alpha k};$$

$$\frac{\partial F_{\alpha\beta}}{\partial Y_{jk}} = (\mathbf{\Sigma}_p)_{\alpha j}D_{k\beta} - E_{j\alpha}D_{\beta k} - F_{j\beta}F_{\alpha k};$$

$$\frac{\partial (E_{jj}D_{kk})}{\partial Y_{jk}} = 2(\mathbf{\Sigma}_p)_{jj}F_{jk}D_{kk} - 4E_{jj}F_{jk}D_{kk}.$$

LEMMA S1.18. *For any $\alpha, j \in \{1, 2, \ldots, p\}$ and $\beta, k \in \{1, 2, \ldots, n\}$, we have*

$$\frac{\partial \widehat{F}_{\alpha\beta}}{\partial Y_{jk}} = (\mathbf{\Sigma}_p^2)_{\alpha j}D_{k\beta} - \widehat{E}_{j\alpha}D_{\beta k} - F_{j\beta}\widehat{F}_{\alpha k};$$

$$\frac{\partial (\widehat{E}_{jj}D_{kk})}{\partial Y_{jk}} = (\mathbf{\Sigma}_p)_{jj}\widehat{F}_{jk}D_{kk} + (\mathbf{\Sigma}_p^2)_{jj}F_{jk}D_{kk} - E_{jj}\widehat{F}_{jk}D_{kk} - 3\widehat{E}_{jj}F_{jk}D_{kk}.$$

Lemmas S1.19 and S1.20 are used in Section S3 to prove equations (35) and (36).

LEMMA S1.19. *For $z \in \mathcal{C}_\ell \cup \mathcal{C}_r$, we have*

$$|\beta_k|\mathbb{1}_{U_n} \leqslant K, \qquad |\varepsilon_k| \leqslant K, \qquad \mathbb{E}|\gamma_{k2}|^4\mathbb{1}_{U_n} = O(n^{-2}), \qquad \mathbb{E}|\mu_k|^4\mathbb{1}_{U_n} = O(n^{-1}).$$

LEMMA S1.20. *For $z \in \mathcal{C}_\ell \cup \mathcal{C}_r$, we have*

$$\mathbb{E}\left|M_n^{(1)}(z)\mathbb{1}_{U_n}\right|^2 \leqslant K.$$

Lemma S1.21 is used in the proof of Lemma 6.2.

LEMMA S1.21. *Under the assumption $p \wedge n \to \infty$, $p/n \to \infty$, for $z \in \mathbb{C}_1$, we have*

$$\mathbb{E}\left|\frac{1}{npb_p}\mathrm{tr}\left\{\mathbf{M}_k^{(1)}(z)\right\} - m(z)\right|^2 \leqslant \frac{Kn}{p} + \frac{K}{n^2}.$$

S2. Proofs of lemmas. This section contains proofs of Lemmas S1.7 – S1.21, Lemmas 6.1 – 6.2.

S2.1. *Proof of Lemma S1.7.*

PROOF. Let $\{\mathbf{e}_i, i = 1\ldots, k-1, k+1, \ldots, n\}$ be the $(n-1)$-dimensional unit vectors with the i-th (or $(i-1)$-th) element equal to one and the remaining equal to zero according as $i < k$ (or $i > k$). Write $\mathbf{X}_k = \mathbf{X}_{ki} + \mathbf{x}_i\mathbf{e}_i'$. Let $\mathbf{I}_{(i)}$ be $n \times n$ diagonal matrix with all 1's on the diagonal except the i-th element being zero, and

$$\mathbf{D}_{ki,r}^{-1} = \mathbf{D}_k^{-1} - \mathbf{e}_i\mathbf{h}_i' = \frac{1}{\sqrt{npb_p}}\Big(\mathbf{X}_{ki}'\boldsymbol{\Sigma}_p\mathbf{X}_k - pa_p\mathbf{I}_{(i)}\Big) - z\mathbf{I}_{n-1},$$

$$\mathbf{D}_{ki}^{-1} = \mathbf{D}_k^{-1} - \mathbf{e}_i\mathbf{h}_i' - \mathbf{r}_i\mathbf{e}_i' = \frac{1}{\sqrt{npb_p}}\Big(\mathbf{X}_{ki}'\boldsymbol{\Sigma}_p\mathbf{X}_{ki} - pa_p\mathbf{I}_{(i)}\Big) - z\mathbf{I}_{n-1},$$

$$\mathbf{h}_i' = \frac{1}{\sqrt{npb_p}}\mathbf{x}_i'\boldsymbol{\Sigma}_p\mathbf{X}_{ki} + \frac{1}{\sqrt{npb_p}}\Big(\mathbf{x}_i'\boldsymbol{\Sigma}_p\mathbf{x}_i - pa_p\Big)\mathbf{e}_i', \qquad \mathbf{r}_i = \frac{1}{\sqrt{npb_p}}\mathbf{X}_{ki}'\boldsymbol{\Sigma}_p\mathbf{x}_i,$$

$$\zeta_i = \frac{1}{1+\vartheta_i}, \qquad \vartheta_i = \mathbf{h}_i'\mathbf{D}_{ki,r}(z)\mathbf{e}_i, \qquad \mathbf{M}_{ki} = \boldsymbol{\Sigma}_p\mathbf{X}_{ki}\mathbf{D}_{ki}(z)\mathbf{X}_{ki}'\boldsymbol{\Sigma}_p.$$

We have some crucial identities,

$$\mathbf{X}_{ki}\mathbf{e}_i = \mathbf{0}, \qquad \mathbf{e}_i'\mathbf{D}_{ki,r} = \mathbf{e}_i'\mathbf{D}_{ki} = -\frac{\mathbf{e}_i'}{z}, \tag{S2.1}$$

where $\mathbf{0}$ is a p-dimensional vector with all the elements equal to 0. By using (S2.1) and some frequently used formulas about the inverse of matrices, we obtain two useful identities:

$$\begin{aligned}\mathbf{D}_k - \mathbf{D}_{ki,r} &= -\mathbf{D}_{ki,r}(\mathbf{D}_k^{-1} - \mathbf{D}_{ki,r}^{-1})\mathbf{D}_k = -\mathbf{D}_{ki,r}(\mathbf{e}_i\mathbf{h}_i')\mathbf{D}_k \\ &= -\mathbf{D}_{ki,r}(\mathbf{e}_i\mathbf{h}_i')(\zeta_i\mathbf{D}_{ki,r}) = -\zeta_i\mathbf{D}_{ki,r}(\mathbf{e}_i\mathbf{h}_i')\mathbf{D}_{ki,r}\end{aligned} \tag{S2.2}$$

and

$$\begin{aligned}\mathbf{D}_{ki,r} - \mathbf{D}_{ki} &= -\mathbf{D}_{ki}(\mathbf{D}_{ki,r}^{-1} - \mathbf{D}_{ki}^{-1})\mathbf{D}_{ki,r} = -\mathbf{D}_{ki}(\mathbf{r}_i\mathbf{e}_i')\mathbf{D}_{ki,r} \\ &= -\mathbf{D}_{ki}\Big(\frac{1}{\sqrt{npb_p}}\mathbf{X}_{ki}'\boldsymbol{\Sigma}_p\mathbf{x}_i\mathbf{e}_i'\Big)\mathbf{D}_{ki} = \frac{1}{z\sqrt{npb_p}}\mathbf{D}_{ki}\mathbf{X}_{ki}'\boldsymbol{\Sigma}_p\mathbf{x}_i\mathbf{e}_i'.\end{aligned} \tag{S2.3}$$

Using (S2.2) and (S2.3), for $i < k$, we obtain the following decomposition of $\mathbb{E}_k\mathbf{M}_k^{(1)}(z)$,

$$\begin{aligned}\mathbb{E}_k\mathbf{M}_k^{(1)}(z) &= \mathbb{E}_k\Big\{\boldsymbol{\Sigma}_p(\mathbf{X}_{ki} + \mathbf{x}_i\mathbf{e}_i')\mathbf{D}_k(\mathbf{X}_{ki} + \mathbf{x}_i\mathbf{e}_i')'\boldsymbol{\Sigma}_p\Big\} \\ &= \mathbb{E}_k\Big(\boldsymbol{\Sigma}_p\mathbf{X}_{ki}\mathbf{D}_k\mathbf{X}_{ki}'\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_p\mathbf{X}_{ki}\mathbf{D}_k\mathbf{e}_i\mathbf{x}_i'\boldsymbol{\Sigma}_p \\ &\qquad + \boldsymbol{\Sigma}_p\mathbf{x}_i\mathbf{e}_i'\mathbf{D}_k\mathbf{X}_{ki}'\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_p\mathbf{x}_i\mathbf{e}_i'\mathbf{D}_k\mathbf{e}_i\mathbf{x}_i'\boldsymbol{\Sigma}_p\Big) \\ &= \mathbb{E}_k\mathbf{M}_{ki} - \mathbb{E}_k\left\{\frac{\zeta_i(z)}{znpb_p}\mathbf{M}_{ki}\mathbf{x}_i\mathbf{x}_i'\mathbf{M}_{ki}\right\} + \mathbb{E}_k\left\{\frac{\zeta_i(z)}{z\sqrt{npb_p}}\mathbf{M}_{ki}\right\}\mathbf{x}_i\mathbf{x}_i'\boldsymbol{\Sigma}_p \\ &\qquad + \boldsymbol{\Sigma}_p\mathbf{x}_i\mathbf{x}_i'\mathbb{E}_k\left\{\frac{\zeta_i(z)}{z\sqrt{npb_p}}\mathbf{M}_{ki}\right\} - \mathbb{E}_k\{\zeta_i(z)/z\}\boldsymbol{\Sigma}_p\mathbf{x}_i\mathbf{x}_i'\boldsymbol{\Sigma}_p \\ &:= \mathbf{B}_1(z) + \mathbf{B}_2(z) + \mathbf{B}_3(z) + \mathbf{B}_4(z) + \mathbf{B}_5(z).\end{aligned} \tag{S2.4}$$

Write

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

Multiplying $\mathbf{D}_k$ on the right-hand side, we have

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

Multiplying $\mathbf{\Sigma}_p \mathbf{X}_k$ on the left-hand side, $\mathbf{X}_k' \mathbf{\Sigma}_p$ on the right-hand side, we get

$$z\mathbf{M}_k^{(1)}(z) = -\mathbf{\Sigma}_p \mathbf{X}_k \mathbf{X}_k' \mathbf{\Sigma}_p + \sum_{i=1(\neq k)}^{n} \mathbf{\Sigma}_p \mathbf{X}_k \mathbf{e}_i \mathbf{h}_i' \mathbf{D}_k \mathbf{X}_k' \mathbf{\Sigma}_p.$$

Thus,

$$\begin{aligned}
z\mathbb{E}_k\big(\mathbf{M}_k^{(1)}(z)\big) &= -\mathbb{E}_k\big(\mathbf{\Sigma}_p \mathbf{X}_k \mathbf{X}_k' \mathbf{\Sigma}_p\big) + \sum_{i=1(\neq k)}^{n} \mathbb{E}_k(\mathbf{\Sigma}_p \mathbf{X}_k \mathbf{e}_i \mathbf{h}_i' \mathbf{D}_k \mathbf{X}_k' \mathbf{\Sigma}_p) \\
&= -\mathbf{\Sigma}_p \mathbb{E}_k\bigg(\sum_{i=1(\neq k)}^{n} \mathbf{x}_i \mathbf{x}_i'\bigg)\mathbf{\Sigma}_p + \sum_{i=1(\neq k)}^{n} \mathbb{E}_k\Big\{\zeta_i \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{h}_i' \mathbf{D}_{ki,r}(\mathbf{X}_{ki}' + \mathbf{e}_i \mathbf{x}_i')\mathbf{\Sigma}_p\Big\} \\
&= -(n-k)\mathbf{\Sigma}_p^2 - \sum_{i<k}\Big(\mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p\Big) \\
&\quad + \sum_{i=1(\neq k)}^{n} \mathbb{E}_k\bigg(\frac{\zeta_i}{\sqrt{npb_p}} \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p \mathbf{X}_{ki} \mathbf{D}_{ki,r} \mathbf{X}_{ki}' \mathbf{\Sigma}_p\bigg) \\
&\quad + \sum_{i=1(\neq k)}^{n} \mathbb{E}_k\Big(\zeta_i \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{h}_i' \mathbf{D}_{ki,r} \mathbf{e}_i \mathbf{x}_i' \mathbf{\Sigma}_p\Big) \\
&= -(n-k)\mathbf{\Sigma}_p^2 - \sum_{i<k}\Big(\mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p\Big) + \sum_{i=1(\neq k)}^{n} \mathbb{E}_k\bigg(\frac{\zeta_i}{\sqrt{npb_p}} \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{M}_{ki}\bigg) \\
&\quad + \sum_{i=1(\neq k)}^{n} \mathbb{E}_k\Big(\zeta_i \vartheta_i \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p\Big). \qquad \text{(S2.5)}
\end{aligned}$$

Applying (S2.4) and (S2.5) to $\mathbb{E}_k \mathbf{M}_k^{(1)}(z_2)$ (for $i < k$) and $z_1 \mathbb{E}_k \mathbf{M}_k^{(1)}(z_1)$, we get the following decomposition:

$$\begin{aligned}
z_1 \mathbb{Z}_k &= \frac{z_1}{n(pb_p)^2} \mathsf{tr}\Big\{\mathbb{E}_k \mathbf{M}_k^{(1)}(z_1) \cdot \mathbb{E}_k \mathbf{M}_k^{(1)}(z_2)\Big\} \\
&= \frac{1}{n(pb_p)^2} \mathsf{tr}\bigg[\bigg\{-(n-k)\mathbf{\Sigma}_p^2 - \sum_{i<k}\Big(\mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p\Big) + \sum_{i=1(\neq k)}^{n} \mathbb{E}_k\bigg(\frac{\zeta_i}{\sqrt{npb_p}} \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{M}_{ki}\bigg) \\
&\quad + \sum_{i=1(\neq k)}^{n} \mathbb{E}_k\Big(\zeta_i \vartheta_i \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p\Big)\bigg\} \times \mathbb{E}_k \mathbf{M}_k^{(1)}(z_2)\bigg] \\
&= C_1(z_1, z_2) + C_2(z_1, z_2) + C_3(z_1, z_2) + C_4(z_1, z_2), \qquad \text{(S2.6)}
\end{aligned}$$

where

$$C_1(z_1,z_2) = -\frac{n-k}{n(pb_p)^2}\mathrm{tr}\Big\{\mathbf{\Sigma}_p^2\cdot\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\Big\},$$

(S2.7)

$$C_2(z_1,z_2) = -\frac{1}{n(pb_p)^2}\sum_{i<k}\mathbf{x}_i'\mathbf{\Sigma}_p\Big\{\sum_{j=1}^{5}\mathbf{B}_j(z_2)\Big\}\mathbf{\Sigma}_p\mathbf{x}_i = \sum_{j=1}^{5}C_{2j},$$

$$C_3(z_1,z_2) = \frac{1}{n(pb_p)^2}\sum_{i<k}\mathbb{E}_k\left[\frac{\zeta_i(z_1)}{\sqrt{npb_p}}\mathbf{x}_i'\mathbf{M}_{ki}(z_1)\Big\{\sum_{j=1}^{5}\mathbf{B}_j(z_2)\Big\}\mathbf{\Sigma}_p\mathbf{x}_i\right]$$

$$+\frac{1}{n(pb_p)^2}\sum_{i>k}\mathbb{E}_k\left[\frac{\zeta_i(z_1)}{\sqrt{npb_p}}\mathbf{x}_i'\mathbf{M}_{ki}(z_1)\Big\{\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\Big\}\mathbf{\Sigma}_p\mathbf{x}_i\right] = \sum_{j=1}^{6}C_{3j}, \tag{S2.8}$$

$$C_4(z_1,z_2) = \frac{1}{n(pb_p)^2}\sum_{i<k}\mathbb{E}_k\left[\zeta_i(z_1)\vartheta_i(z_1)\mathbf{x}_i'\mathbf{\Sigma}_p\Big\{\sum_{j=1}^{5}\mathbf{B}_j(z_2)\Big\}\mathbf{\Sigma}_p\mathbf{x}_i\right]$$

$$+\frac{1}{n(pb_p)^2}\sum_{i>k}\mathbb{E}_k\left[\zeta_i(z_1)\vartheta_i(z_1)\mathbf{x}_i'\mathbf{\Sigma}_p\Big\{\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\Big\}\mathbf{\Sigma}_p\mathbf{x}_i\right] = \sum_{j=1}^{6}C_{4j}. \tag{S2.9}$$

Now we estimate all the terms in (S2.6). We will show that these terms are negligible as $n\to\infty$, expect C_{25}, C_{33}, C_{45} defined in (S2.7) – (S2.9).

For $C_1(z_1,z_2)$, we have

$$\mathbb{E}|C_1(z_1,z_2)| = \frac{n-k}{n(pb_p)^2}\Big|\mathrm{tr}\Big\{\mathbf{\Sigma}_p^2\cdot\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\Big\}\Big| = O\Big(\frac{1}{p^2}\Big)\cdot O(np) = O\Big(\frac{n}{p}\Big),$$

where the second equality follows from the fact $\big|\mathrm{tr}\big\{\mathbf{\Sigma}_p^2\cdot\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\big\}\big| = O(np)$, which can be verified by using the similar argument in the proof of Lemma S1.21.

Applying Lemma S1.8 and inequality (S1.3) with $\mathbf{B}=\mathbf{I}_p$, we have

$$\begin{aligned}\mathbb{E}|C_{21}| &\leqslant \frac{1}{n(pb_p)^2}\sum_{i<k}\mathbb{E}\big|\mathbf{x}_i'\mathbf{\Sigma}_p\cdot\mathbb{E}_k\mathbf{M}_{ki}(z_2)\cdot\mathbf{\Sigma}_p\mathbf{x}_i\big|\\ &\leqslant \frac{1}{n(pb_p)^2}\sum_{i<k}\Big(\mathbb{E}\big|\mathbf{x}_i'\mathbf{\Sigma}_p\cdot\mathbb{E}_k\mathbf{M}_{ki}(z_2)\cdot\mathbf{\Sigma}_p\mathbf{x}_i\big|^2\Big)^{1/2} \leqslant \frac{Kn}{p}.\end{aligned}$$

Applying Lemma S1.8 and inequality (S1.3) with $\mathbf{B}=\mathbf{\Sigma}_p$, we have

$$\begin{aligned}\mathbb{E}|C_{22}| &\leqslant \frac{1}{n(pb_p)^2}\sum_{i<k}\mathbb{E}\left|\mathbf{x}_i'\mathbf{\Sigma}_p\cdot\mathbb{E}_k\Big\{\frac{\zeta_i(z_2)}{z_2 npb_p}\mathbf{M}_{ki}(z_2)\mathbf{x}_i\mathbf{x}_i'\mathbf{M}_{ki}(z_2)\Big\}\cdot\mathbf{\Sigma}_p\mathbf{x}_i\right|\\ &= \frac{K}{n^2(pb_p)^3}\sum_{i<k}\mathbb{E}\big|\mathbf{x}_i'\mathbf{M}_{ki}(z_2)\mathbf{\Sigma}_p\mathbf{x}_i\big|^2 \leqslant \frac{Kn}{p}.\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}\mathbb{E}|C_{23}| = \mathbb{E}|C_{24}| &\leqslant \frac{1}{n(pb_p)^2}\sum_{i<k}\mathbb{E}\left|\mathbf{x}_i'\mathbf{\Sigma}_p\cdot\mathbb{E}_k\Big\{\frac{\zeta_i(z_2)}{z_2\sqrt{npb_p}}\mathbf{M}_{ki}(z_2)\Big\}\mathbf{x}_i\mathbf{x}_i'\mathbf{\Sigma}_p\cdot\mathbf{\Sigma}_p\mathbf{x}_i\right|\\ &\leqslant \frac{K}{np^2\sqrt{np}}\sum_{i<k}\mathbb{E}\Big|\mathbf{x}_i'\mathbf{\Sigma}_p\mathbf{M}_{ki}(z_2)\mathbf{x}_i\cdot\mathbf{x}_i'\mathbf{\Sigma}_p^2\mathbf{x}_i\Big|\end{aligned}$$

$$\leqslant \frac{K}{np^2\sqrt{np}} \sum_{i<k} \left\{ \mathbb{E}\left| \mathbf{x}_i' \mathbf{\Sigma}_p \mathbf{M}_{ki}(z_2) \mathbf{x}_i \right|^2 \right\}^{1/2} \cdot \left\{ \mathbb{E}\left| \mathbf{x}_i' \mathbf{\Sigma}_p^2 \mathbf{x}_i \right|^2 \right\}^{1/2} \leqslant K\sqrt{\frac{n}{p}}.$$

Applying Lemma S1.8 and inequality (S1.2) with $\mathbf{B} = \mathbb{E}_k \mathbf{M}_{ki}(z_2)\mathbf{\Sigma}_p$, we have

$$\begin{aligned}
\mathbb{E}|C_{31}| &= \frac{1}{n(pb_p)^2} \sum_{i<k} \mathbb{E}\left| \mathbb{E}_k \left\{ \frac{\zeta_i(z_1)}{\sqrt{npb_p}} \mathbf{x}_i' \mathbf{M}_{ki}(z_1) \cdot \mathbb{E}_k \mathbf{M}_{ki}(z_2) \cdot \mathbf{\Sigma}_p \mathbf{x}_i \right\} \right| \\
&\leqslant \frac{K}{np^2\sqrt{np}} \sum_{i<k} \mathbb{E}\left| \mathbf{x}_i' \mathbf{M}_{ki}(z_1) \cdot \mathbb{E}_k \mathbf{M}_{ki}(z_2) \cdot \mathbf{\Sigma}_p \mathbf{x}_i \right| \leqslant K\sqrt{\frac{n}{p}}.
\end{aligned}$$

We define $\tilde{\zeta}_i(z)$ and $\widetilde{\mathbf{M}}_{ki}(z)$ as the analogues of $\zeta_i(z)$ and $\mathbf{M}_{ki}(z)$, respectively, using $\{\mathbf{x}_1, \ldots, \mathbf{x}_k, \tilde{\mathbf{x}}_{k+1}, \ldots, \tilde{\mathbf{x}}_n\}$, where $\tilde{\mathbf{x}}_{k+1}, \ldots, \tilde{\mathbf{x}}_n$ are i.i.d. copies of $\mathbf{x}_{k+1}, \ldots, \mathbf{x}_n$ and independent of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then,

$$\begin{aligned}
&\mathbb{E}|C_{32}| \\
&= \frac{1}{n(pb_p)^2} \sum_{i<k} \mathbb{E}\left| \mathbb{E}_k \left[\frac{\zeta_i(z_1)}{\sqrt{npb_p}} \mathbf{x}_i' \mathbf{M}_{ki}(z_1) \cdot \mathbb{E}_k \left\{ \frac{\zeta_i(z_2)}{z_2 npb_p} \mathbf{M}_{ki}(z_2) \mathbf{x}_i \mathbf{x}_i' \mathbf{M}_{ki}(z_2) \right\} \cdot \mathbf{\Sigma}_p \mathbf{x}_i \right] \right| \\
&= \frac{1}{n(pb_p)^2} \sum_{i<k} \mathbb{E}\left| \mathbb{E}_k \left[\frac{\zeta_i(z_1)}{\sqrt{npb_p}} \mathbf{x}_i' \mathbf{M}_{ki}(z_1) \cdot \mathbb{E}_k \left\{ \frac{\tilde{\zeta}_i(z_2)}{z_2 npb_p} \widetilde{\mathbf{M}}_{ki}(z_2) \mathbf{x}_i \mathbf{x}_i' \widetilde{\mathbf{M}}_{ki}(z_2) \right\} \cdot \mathbf{\Sigma}_p \mathbf{x}_i \right] \right| \\
&\leqslant \frac{K}{n^2p^3\sqrt{np}} \sum_{i<k} \mathbb{E}\left| \left[\mathbf{x}_i' \mathbf{M}_{ki}(z_1) \widetilde{\mathbf{M}}_{ki}(z_2) \mathbf{x}_i \cdot \mathbf{x}_i' \widetilde{\mathbf{M}}_{ki}(z_2) \mathbf{\Sigma}_p \mathbf{x}_i \right] \right| \\
&\leqslant \frac{K}{n^2p^3\sqrt{np}} \sum_{i<k} \left\{ \mathbb{E}\left| \mathbf{x}_i' \mathbf{M}_{ki}(z_1) \widetilde{\mathbf{M}}_{ki}(z_2) \mathbf{x}_i \right|^2 \right\}^{1/2} \left\{ \mathbb{E}\left| \mathbf{x}_i' \widetilde{\mathbf{M}}_{ki}(z_2) \mathbf{\Sigma}_p \mathbf{x}_i \right|^2 \right\}^{1/2} \\
&\overset{(S1.2)}{\leqslant} K\sqrt{\frac{n}{p}}.
\end{aligned}$$

Similarly, we have

$$\mathbb{E}|C_{3j}| \leqslant K\frac{n}{p}, \qquad j = 4, 5, 6.$$

Applying Lemma S1.8 and inequality (S1.3) with $\mathbf{B} = \mathbf{I}_{n-1}$, we obtain

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

Moreover, by using Lemmas S1.8 – S1.9 and Lemma S1.21, we obtain the following limits:

$$\begin{aligned}
C_{25} &= -\frac{1}{n(pb_p)^2} \sum_{i<k} \left(\mathbf{x}_i' \mathbf{\Sigma}_p \left[-\mathbb{E}_k \left\{ \frac{\zeta_i(z_2)}{z_2} \right\} \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p \right] \mathbf{\Sigma}_p \mathbf{x}_i \right) \\
&= -\frac{1}{n(pb_p)^2} m(z_2) \sum_{i<k} \left(\mathbf{x}_i' \mathbf{\Sigma}_p^2 \mathbf{x}_i \right)^2 \\
&= -\frac{k}{n} m(z_2) + o_{L_1}(1),
\end{aligned}$$

$$C_{45} = \frac{1}{n(pb_p)^2} \sum_{i<k} \mathbb{E}_k \left(\zeta_i(z_1) \vartheta_i(z_1) \mathbf{x}_i' \mathbf{\Sigma}_p \left[-\mathbb{E}_k \left\{ \frac{\zeta_i(z_2)}{z_2} \right\} \mathbf{\Sigma}_p \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p \right] \mathbf{\Sigma}_p \mathbf{x}_i \right)$$

$$= \frac{1}{n(pb_p)^2} \sum_{i<k} \mathbb{E}_k \Big[-m^2(z_1) m(z_2) \big(\mathbf{x}_i' \mathbf{\Sigma}_p^2 \mathbf{x}_i \big)^2 \Big] + o_{L_1}(1)$$

$$= -\frac{k}{n} m^2(z_1) m(z_2) + o_{L_1}(1),$$

and

$$C_{33} = \frac{1}{n(pb_p)^2} \sum_{i<k} \mathbb{E}_k \left[\frac{\zeta_i(z_1)}{\sqrt{npb_p}} \mathbf{x}_i' \mathbf{M}_{ki}(z_1) \left\{ \mathbb{E}_k \frac{\zeta_i(z_2)}{z_2 \sqrt{npb_p}} \mathbf{M}_{ki}(z_2) \right\} \mathbf{x}_i \mathbf{x}_i' \mathbf{\Sigma}_p^2 \mathbf{x}_i \right]$$

$$= \frac{1}{n^2 p^2 b_p^2} z_1 m(z_1) m(z_2) \left\{ \sum_{i<k} \mathbf{x}_i' \mathbb{E}_k \mathbf{M}_{ki}(z_1) \mathbb{E}_k \mathbf{M}_{ki}(z_2) \mathbf{x}_i \right\} + o_{L_4}(1)$$

$$= \frac{k}{n} m(z_1) m(z_2) z_1 \mathbb{Z}_k + o_{L_1}(1).$$

From above estimates, we have

$$z_1 \mathbb{Z}_k = -\frac{k}{n} m(z_2) - \frac{k}{n} m^2(z_1) m(z_2) + \frac{k}{n} m(z_1) m(z_2) z_1 \mathbb{Z}_k + o_{L_1}(1)$$

$$= \frac{k}{n} z_1 m(z_1) m(z_2) + \frac{k}{n} z_1 m(z_1) m(z_2) \mathbb{Z}_k + o_{L_1}(1),$$

which is equivalent to

$$\mathbb{Z}_k = \frac{\frac{k}{n} m(z_1) m(z_2)}{1 - \frac{k}{n} m(z_1) m(z_2)} + o_{L_1}(1).$$

□

S2.2. *Proof of Lemma S1.8.*

PROOF. This lemma can be proved by using similar arguments in Section 5.2.2 of Chen and Pan (2015). □

S2.3. *Proof of Lemma S1.9.*

PROOF. Note that $\mathbf{M}_{ki}$ and $\mathbf{x}_i$ are independent. By using Lemma S1.1, we have

$$\mathbb{E} \big| \mathbf{x}_i' \mathbf{M}_{ki} \mathbf{B} \mathbf{x}_i - \mathrm{tr} \mathbf{M}_{ki} \mathbf{B} \big|^2 \leqslant K \big\{ \nu_4 \mathbb{E} \mathrm{tr} \big(\mathbf{M}_{ki} \mathbf{B} \overline{\mathbf{B}} \overline{\mathbf{M}}_{ki} \big) \big\} \leqslant K n p^2 \|\mathbf{B}\|^2, \tag{S2.10}$$

where we use the fact that

$$\begin{aligned} \big| \mathrm{tr} \big(\mathbf{M}_{ki} \mathbf{B} \overline{\mathbf{B}} \overline{\mathbf{M}}_{ki} \big) \big| &= \big| \mathrm{tr} \big(\mathbf{\Sigma}_p \mathbf{X}_{ki} \mathbf{D}_{ki} \mathbf{X}_{ki}' \mathbf{\Sigma}_p \mathbf{B} \overline{\mathbf{B}} \mathbf{\Sigma}_p \mathbf{X}_{ki} \overline{\mathbf{D}}_{ki} \mathbf{X}_{ki}' \mathbf{\Sigma}_p \big) \big| \\ &= \big| \mathrm{tr} \big(\mathbf{D}_{ki}^{1/2} \mathbf{X}_{ki}' \mathbf{\Sigma}_p \mathbf{B} \overline{\mathbf{B}} \mathbf{\Sigma}_p \mathbf{X}_{ki} \overline{\mathbf{D}}_{ki} \mathbf{X}_{ki}' \mathbf{\Sigma}_p^2 \mathbf{X}_{ki} \mathbf{D}_{ki}^{1/2} \big) \big| \\ &\leqslant n \cdot \| \mathbf{D}_{ki}^{1/2} \mathbf{X}_{ki}' \mathbf{\Sigma}_p^{1/2} \| \cdot \| \mathbf{\Sigma}_p^{1/2} \| \cdot \| \mathbf{B} \overline{\mathbf{B}} \| \cdot \| \mathbf{\Sigma}_p^{1/2} \| \\ &\qquad \times \| \mathbf{\Sigma}_p^{1/2} \mathbf{X}_{ki} \overline{\mathbf{D}}_{ki} \mathbf{X}_{ki}' \mathbf{\Sigma}_p^{1/2} \| \cdot \| \mathbf{\Sigma}_p \| \cdot \| \mathbf{\Sigma}_p^{1/2} \mathbf{X}_{ki} \mathbf{D}_{ki}^{1/2} \| \\ &= n \cdot \| \mathbf{\Sigma}_p \|^2 \cdot \| \mathbf{B} \|^2 \cdot \| \mathbf{\Sigma}_p^{1/2} \mathbf{X}_{ki} \mathbf{D}_{ki} \mathbf{X}_{ki}' \mathbf{\Sigma}_p^{1/2} \|^2 \\ &= n \cdot \| \mathbf{\Sigma}_p \|^2 \cdot \| \mathbf{B} \|^2 \cdot \| \mathbf{D}_{ki} \mathbf{X}_{ki}' \mathbf{\Sigma}_p \mathbf{X}_{ki} \|^2 \\ &= n \cdot \| \mathbf{\Sigma}_p \|^2 \cdot \| \mathbf{B} \|^2 \cdot \| \sqrt{npb_p} (\mathbf{I}_{n-1} + z \mathbf{D}_{ki}) + p a_p \mathbf{I}_{(i)} \mathbf{D}_{ki} \|^2 \\ &\leqslant K n p^2 \| \mathbf{B} \|^2. \end{aligned} \tag{S2.11}$$

By (S2.10) and the c_r-inequality, we have

$$\mathbb{E}\left|\mathbf{x}_i'\mathbf{M}_{ki}\mathbf{B}\mathbf{x}_i\right|^2 \leqslant K\left\{\mathbb{E}\left|\mathbf{x}_i'\mathbf{M}_{ki}\mathbf{B}\mathbf{x}_i - \mathrm{tr}\mathbf{M}_{ki}\mathbf{B}\right|^2 + \mathbb{E}\left|\mathrm{tr}\mathbf{M}_{ki}\mathbf{B}\right|^2\right\} \leqslant Kp^2n^2\mathbb{E}\|\mathbf{B}\|^2,$$

which completes the proof of (S1.2). By using the same argument, we get (S1.3). □

S2.4. *Proof of Lemma S1.10.*

PROOF. The proof of Lemma S1.10 exactly follows Chen and Pan (2015), so is omitted. □

S2.5. *Proof of Lemma S1.11.*

PROOF. By Lemma S1.1 and taking $\mathbf{B} = \mathbf{I}_p$ in the inequality (S2.11), we have

$$\mathbb{E}|\gamma_{k2}|^2 \leqslant \frac{K}{n^2p^2}\mathrm{tr}\left(\mathbf{M}_k^{(s)}\overline{\mathbf{M}}_k^{(s)}\right) \leqslant \frac{K}{n}.$$

Similarly, we can prove that $\mathbb{E}|\eta_k|^2 \leqslant K/n$.

Now, we prove the bounds for the 4-th moments of γ_{ks} and η_{ks}. Let $\mathbf{H}$ be $\mathbf{M}_k^{(s)}$ with all diagonal elements replaced by zeros, then we have

$$\mathbb{E}|\mathbf{x}_k'\mathbf{H}\mathbf{x}_k|^4 \leqslant K(\mathbb{E}X_{11}^4)^2\mathbb{E}\left(\mathrm{tr}\mathbf{H}\mathbf{H}^*\right)^2 \leqslant K\mathbb{E}\left(\mathrm{tr}\mathbf{M}_k^{(s)}\overline{\mathbf{M}}_k^{(s)}\right)^2 \leqslant Kn^2p^4. \tag{S2.12}$$

The first inequality follows from Lemma S1.2, and the last inequality follows from (S2.11).

Let $\mathbb{E}_j(\cdot)$ denote the conditional expectation with respect to $(X_{1k}, X_{2k}, \ldots, X_{jk})$, and let $m_{jj}^{(s)}$ denote the j-th diagonal entry of $\mathbf{M}_k^{(s)}$, where $j = 1, 2, \ldots, p$. Since $\mathbb{E}_{j-1}(X_{jk}^2 - 1)m_{jj}^{(s)} = 0$, then $(X_{jk}^2 - 1)m_{jj}^{(s)}$ can be expressed as a martingale difference

$$(X_{jk}^2 - 1)m_{jj}^{(s)} = (\mathbb{E}_j - \mathbb{E}_{j-1})\left\{(X_{jk}^2 - 1)m_{jj}^{(s)}\right\}. \tag{S2.13}$$

Applying the Burkholder's inequality (Lemma S1.3) to (S2.13) yields that

$$\begin{aligned}
&\mathbb{E}\left|\sum_{j=1}^p (X_{jk}^2 - 1)m_{jj}^{(s)}\right|^4 \\
&\leqslant K\mathbb{E}\left(\sum_{j=1}^p \mathbb{E}_{j-1}\left|(X_{jk}^2 - 1)m_{jj}^{(s)}\right|^2\right)^2 + K\mathbb{E}\left(\sum_{j=1}^p \left|(X_{jk}^2 - 1)m_{jj}^{(s)}\right|^4\right)^2 \\
&\leqslant K\left(\sum_{j=1}^p \mathbb{E}|X_{11}|^4|m_{jj}^{(s)}|^2\right) + K\sum_{j=1}^p \mathbb{E}|X_{11}|^8\mathbb{E}|m_{jj}^{(s)}|^4 \\
&\leqslant Kn^5p^2 + Kn^3p^3, \qquad \text{(S2.14)}
\end{aligned}$$

where we use the fact that, with $\mathbf{e}_j$ be the j-th p-dimensional standard basis vector and $\mathbf{y}$ be an $(n-1)$-dimensional random vector with $\mathbb{E}y_i = 0$ and $\mathbb{E}y_i^2 = 1$,

$$\begin{aligned}
\mathbb{E}|m_{jj}^{(s)}|^4 &= \mathbb{E}\left|\mathbf{e}_j'\mathbf{\Sigma}_p\mathbf{X}_k\mathbf{D}_k^s\mathbf{X}_k'\mathbf{\Sigma}_p\mathbf{e}_j\right|^4 \\
&\leqslant v_1^{-4s}\mathbb{E}\left\|\mathbf{e}_j'\mathbf{\Sigma}_p\mathbf{X}_k\right\|^8 = v_1^{-4s}(\mathbf{\Sigma}_p^2)_{jj}^4\mathbb{E}\left\|\mathbf{y}\right\|^8 \leqslant Kn^4 + Kn^2p, \qquad \text{(S2.15)}
\end{aligned}$$

where $(\mathbf{\Sigma}_p^2)_{jj} = \sum_\ell (\mathbf{\Sigma}_p)_{j\ell}^2$ is the j-th diagonal elements of $\mathbf{\Sigma}_p^2$. By Rayleigh-Ritz Theorem, we know that $(\mathbf{\Sigma}_p^2)_{jj} \leqslant \lambda_{\max}(\mathbf{\Sigma}_p^2) \leqslant K$. Combining (S2.12) and (S2.14) yields that

$$\begin{aligned}\mathbb{E}|\gamma_{ks}|^4 &\leqslant \frac{1}{(npb_p)^4}\mathbb{E}\bigg|\sum_{j=1}^{p}(X_{jk}^2-1)m_{jj}^{(s)} + \mathbf{x}_k'\mathbf{H}\mathbf{x}_k\bigg|^4 \\ &\leqslant \frac{K}{n^4p^4}\mathbb{E}\bigg|\sum_{j=1}^{p}(X_{jk}^2-1)m_{jj}^{(s)}\bigg|^4 + \frac{K}{n^4p^4}\mathbb{E}|\mathbf{x}_k'\mathbf{H}\mathbf{x}_k|^4 \\ &\leqslant K\left(\frac{1}{n^2}+\frac{n}{p^2}\right).\end{aligned}$$

Moreover, by Lemma S1.1, we have

$$\mathbb{E}|\eta_k|^4 \leqslant \frac{K}{n^2p^2}\mathbb{E}\big|\mathbf{x}_k'\mathbf{\Sigma}_p\mathbf{x}_k - pa_p\big|^4 + K\mathbb{E}\big|\gamma_{k1}\big|^4 \leqslant \frac{K\delta_n^4}{n} + K\left(\frac{1}{n^2}+\frac{n}{p^2}\right).$$

This completes the proof of the lemma. □

S2.6. *Proof of Lemma S1.12.*

PROOF. By the identity $m_n - \mathbb{E}m_n = -\sum_{k=1}^n\big(\mathbb{E}_{k-1}m_n - \mathbb{E}_k m_n\big)$, we have

$$\mathrm{Var}(m_n) = \sum_{k=1}^n \mathbb{E}\left|\mathbb{E}_{k-1}m_n - \mathbb{E}_k m_n\right|^2 + 2\sum_{1\leqslant s<t\leqslant 1}\mathbb{E}\left(\mathbb{E}_{s-1}m_n - \mathbb{E}_s m_n\right)\left(\mathbb{E}_{t-1}m_n - \mathbb{E}_t m_n\right).$$

Since each term in the second sum on the RHS of the above identity is zero, we write

$$\begin{aligned}\mathrm{Var}(m_n) &= \sum_{k=1}^n \mathbb{E}\left|\mathbb{E}_{k-1}m_n - \mathbb{E}_k m_n\right|^2 \\ &= \sum_{k=1}^n \mathbb{E}\left|\mathbb{E}_{k-1}\left(m_n - \mathbb{E}_{(k)}m_n\right)\right|^2 \\ &\leqslant \sum_{k=1}^n \mathbb{E}\left|m_n - \mathbb{E}_{(k)}m_n\right|^2,\end{aligned}$$

where $\mathbb{E}_{(k)}(\cdot)$ denotes the expectation w.r.t. the σ-field generated by $\mathbf{x}_k$. To prove (S1.5), it suffices to show

$$\mathbb{E}\left|m_n - \mathbb{E}_{(k)}m_n\right|^2 = O\left(\frac{1}{n^3}\right), \qquad 1\leqslant k\leqslant n. \tag{S2.16}$$

Now we deal with the case $k=1$, and the remaining cases are analogous and omitted.

Denote $\widetilde{\mathbf{Y}} = (\widetilde{Y}_{ij})_{p\times n} := \mathbf{\Sigma}_p^{1/2}\mathbf{Y}$ where $\mathbf{Y} = (npb_p)^{-1/4}\mathbf{X}$, and let $\widetilde{\mathbf{y}}_k$ be the k-th column of $\widetilde{\mathbf{Y}}$. Let $\widetilde{\mathbf{Y}}_k$ be the $p\times(n-1)$ matrix extracted from $\widetilde{\mathbf{Y}}$ by removing $\widetilde{\mathbf{y}}_k$, then the matrix model (1) can be written as

$$\mathbf{A}_n = \begin{pmatrix} \widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} & (\widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{y}}_1)' \\ \widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{y}}_1 & \widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{Y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_{n-1}\end{pmatrix}.$$

With notations $\mathbf{A}_k = \widetilde{\mathbf{Y}}_k'\widetilde{\mathbf{Y}}_k - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_{n-1}$ and $\mathbf{D}_k = (\mathbf{A}_k - z\mathbf{I}_n)^{-1}$, we have

$$\begin{aligned}
&\mathsf{tr}\mathbf{D} - \mathsf{tr}\mathbf{D}_1 \\
&= \frac{1 + (\widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{y}}_1)'\Big(\widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{Y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_{n-1} - z\mathbf{I}_{n-1}\Big)^{-2}(\widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{y}}_1)}{\Big(\widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} - z\Big) - (\widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{y}}_1)'\Big(\widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{Y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_{n-1} - z\mathbf{I}_{n-1}\Big)^{-1}(\widetilde{\mathbf{Y}}_1'\widetilde{\mathbf{y}}_1)} \\
&= \frac{1 + \widetilde{\mathbf{y}}_1'\Big[\widetilde{\mathbf{Y}}_1\widetilde{\mathbf{Y}}_1'\Big(\widetilde{\mathbf{Y}}_1\widetilde{\mathbf{Y}}_1' - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_{n-1} - z\mathbf{I}_{n-1}\Big)^{-2}\Big]\widetilde{\mathbf{y}}_1}{\Big(\widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} - z\Big) - \widetilde{\mathbf{y}}_1'\Big\{\widetilde{\mathbf{Y}}_1\widetilde{\mathbf{Y}}_1'\Big(\widetilde{\mathbf{Y}}_1\widetilde{\mathbf{Y}}_1' - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_{n-1} - z\mathbf{I}_{n-1}\Big)^{-1}\Big\}\widetilde{\mathbf{y}}_1} \\
&=: \frac{1+U}{V},
\end{aligned}$$

where the second "=" comes from the identity

$$\mathbf{B}(\mathbf{AB} - \alpha\mathbf{I})^{-n}\mathbf{A} = \mathbf{BA}(\mathbf{BA} - \alpha\mathbf{I})^{-n}.$$

Moreover, with notations U and V, we can write $D_{11} = 1/V$ and

$$\begin{aligned}
\mathbb{E}\left|m_n - \mathbb{E}_{(1)} m_n\right|^2 &= \frac{1}{n^2}\mathbb{E}\left|(\mathsf{tr}\mathbf{D} - \mathsf{tr}\mathbf{D}_1) - \mathbb{E}_{(1)}(\mathsf{tr}\mathbf{D} - \mathsf{tr}\mathbf{D}_1)\right|^2 \qquad (\because \mathbb{E}_{(1)}\mathsf{tr}\mathbf{D}_1 = \mathsf{tr}\mathbf{D}_1) \\
&= \frac{1}{n^2}\mathbb{E}\left|\frac{1+U}{V} - \mathbb{E}_{(1)}\left(\frac{1+U}{V}\right)\right|^2 \\
&\leqslant \frac{2}{n^2}\left\{\mathbb{E}\left|\frac{1}{V} - \mathbb{E}_{(1)}\left(\frac{1}{V}\right)\right|^2 + \mathbb{E}\left|\frac{U}{V} - \mathbb{E}_{(1)}\left(\frac{U}{V}\right)\right|^2\right\}.
\end{aligned}$$

By the same arguments as those on Page 196 of Bao (2015), it is sufficient to prove that

$$\text{(S2.17)} \qquad \mathbb{E}_{(1)}\left|U - \mathbb{E}_{(1)} U\right|^2 = O\left(\frac{1}{n}\right), \qquad \mathbb{E}_{(1)}\left|V - \mathbb{E}_{(1)} V\right|^2 = O\left(\frac{1}{n}\right).$$

For simplicity of presentation, we define

$$\mathbf{H}^{[\ell]} = \left(H_{jk}^{[\ell]}\right)_{p\times p} := \widetilde{\mathbf{Y}}_1\widetilde{\mathbf{Y}}_1'\Big(\widetilde{\mathbf{Y}}_1\widetilde{\mathbf{Y}}_1' - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_{n-1} - z\mathbf{I}_{n-1}\Big)^{-\ell}, \qquad \ell = 1, 2.$$

Then, we write

$$\text{(S2.18)} \qquad U - \mathbb{E}_{(1)} U = \sum_{i\neq j} H_{ij}^{[2]}\widetilde{Y}_{i1}\widetilde{Y}_{j1} + \sum_{i=1}^{p} H_{ii}^{[2]}\big(\widetilde{Y}_{i1}^2 - \mathbb{E}\,\widetilde{Y}_{i1}^2\big),$$

$$\text{(S2.19)} \qquad V - \mathbb{E}_{(1)} V = \widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} - \sum_{i\neq j} H_{ij}^{[1]}\widetilde{Y}_{i1}\widetilde{Y}_{j1} - \sum_{i=1}^{p} H_{ii}^{[1]}\big(\widetilde{Y}_{i1}^2 - \mathbb{E}\,\widetilde{Y}_{i1}^2\big).$$

Now we proceed to prove (S2.17). From (S2.19), we have

$$\begin{aligned}
&\mathbb{E}_{(1)}|V - \mathbb{E}_{(1)} V|^2 \\
&\leqslant K\left\{\mathbb{E}_{(1)}\left|\widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\right|^2 + \mathbb{E}_{(1)}\left|\sum_{i\neq j} H_{ij}^{[1]}\widetilde{Y}_{i1}\widetilde{Y}_{j1}\right|^2 + \mathbb{E}_{(1)}\left|\sum_{i=1}^{p} H_{ii}^{[1]}\Big(\widetilde{Y}_{i1}^2 - \mathbb{E}\,\widetilde{Y}_{i1}^2\Big)\right|^2\right\}
\end{aligned}$$

(S2.20)

$$= K\left\{\mathbb{E}\left|\widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\right|^2 + \sum_{i\neq j}\left|H_{ij}^{[1]}\right|^2\mathbb{E}\left(\widetilde{Y}_{i1}^2\widetilde{Y}_{j1}^2\right) + \sum_{i=1}^{p}\left|H_{ii}^{[1]}\right|^2\mathbb{E}\left(\widetilde{Y}_{i1}^2 - \mathbb{E}\widetilde{Y}_{i1}^2\right)^2\right\}.$$

After some straightforward calculations, we obtain some estimates:

$$\text{(S2.21)} \quad \mathbb{E}\,\widetilde{Y}_{i1}^2 = O\left(\frac{1}{\sqrt{np}}\right), \quad \mathbb{E}\,\widetilde{Y}_{i1}^4 = O\left(\frac{1}{np}\right), \quad \mathbb{E}\left(\widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\right)^2 = O\left(\frac{1}{n}\right).$$

Combining (S2.21) and (S2.20), we obtain

$$\text{(S2.22)} \qquad \mathbb{E}_{(1)}\left|V - \mathbb{E}_{(1)}V\right|^2 \leqslant \frac{K}{n} + \frac{K}{np}\mathrm{tr}\left|\mathbf{H}^{[1]}\right|^2.$$

Similarly, we can show that

$$\text{(S2.23)} \qquad \mathbb{E}_{(1)}\left|U - \mathbb{E}_{(1)}U\right|^2 \leqslant \frac{K}{np}\mathrm{tr}\left|\mathbf{H}^{[2]}\right|^2.$$

To get (S2.17), it suffices to show that

$$\mathrm{tr}\left|\mathbf{H}^{[\ell]}\right|^2 = O(p), \qquad \ell = 1, 2.$$

Let $\{\mu_i^{(k)}, i = 1, \dots, n-1\}$ be eigenvalues of $\mathbf{A}_k$, then the eigenvalues of $\mathbf{H}^{[\ell]}$ $(\ell = 1, 2)$ are

$$\frac{\{\mu_i^{(1)} + a_p\sqrt{p/(nb_p)}\}^2}{\left|\mu_i^{(1)} - z\right|^{2\ell}}, \qquad i = 1, 2, \dots, n-1,$$

and a zero eigenvalue with algebraic multiplicity $(p - n + 1)$. Using the fact $\mu_i^{(1)} \geqslant -a_p\sqrt{p/(nb_p)}$, we conclude that

$$\mathrm{tr}\left|\mathbf{H}^{[\ell]}\right|^2 = \sum_{i=1}^{n-1}\frac{\{\mu_i^{(1)} + a_p\sqrt{p/(nb_p)}\}^2}{\left|\mu_i^{(1)} - z\right|^{2\ell}} = O(p), \qquad \ell = 1, 2.$$

This completes the proof of the lemma. □

S2.7. *Proof of Lemma S1.13.*

PROOF. We only provide the estimation of D_{11}, since others are analogous. Note that

$$D_{11} = V^{-1} = \left(\widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} - z - \widetilde{\mathbf{y}}_1'\mathbf{H}^{[1]}\widetilde{\mathbf{y}}_1\right)^{-1}.$$

Let $\mathbf{v}_i^{(1)} = \left(v_{i1}^{(1)}, \dots, v_{ip}^{(1)}\right)$, $(i = 1, 2, \dots, n-1)$ be the unit eigenvector of $\mathbf{A}_1$ corresponding to the eigenvalue $\mu_i^{(1)}$, and let

$$w_i^{(1)} = \frac{\sqrt{np}a_p}{\sqrt{b_p}}\left|\widetilde{\mathbf{y}}_1'\mathbf{v}_i^{(1)}\right|^2.$$

Applying spectral decomposition to $\mathbf{H}^{[1]}$ yields

$$D_{11} = \left\{\widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} - z - \sum_{i=1}^{n-1}\left(\frac{\mu_i^{(1)} + \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}}{\mu_i^{(1)} - z}\right)\left|\widetilde{\mathbf{y}}_1'\mathbf{v}_i^{(1)}\right|^2\right\}^{-1}$$

$$= \left\{ \widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} - z - \frac{1}{\sqrt{np}}\frac{\sqrt{b_p}}{a_p}\sum_{i=1}^{n-1}\frac{\left(\mu_i^{(1)} + \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\right)w_i^{(1)}}{\mu_i^{(1)} - z} \right\}^{-1}$$

$$=: \left(-z - m_n(z) + h_1\right)^{-1}, \tag{S2.24}$$

where

$$h_1 = \left\{ m_n - \frac{1}{n}\sum_{i=1}^{n-1}\left(\frac{\sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\mu_i^{(1)} + 1}{\mu_i^{(1)} - z}\right)\right\}$$

$$+ \left\{ \widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} - \frac{1}{n}\sum_{i=1}^{n-1}\left(\frac{\sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\mu_i^{(1)} + 1}{\mu_i^{(1)} - z}\right)\left(w_i^{(1)} - 1\right)\right\}.$$

By (S2.24), we obtain

$$\left| D_{11} + \frac{1}{z + \mathbb{E}\, m_n}\right| = \left| \frac{\mathbb{E}\, m_n - m_n + h_1}{\left(-z - m_n + h_1\right)\left(z + \mathbb{E}\, m_n\right)}\right| \leqslant K\left|(\mathbb{E}\, m_n - m_n) + h_1\right|,$$

which implies that

$$\mathbb{E}\left| D_{11} + \frac{1}{z + \mathbb{E}\, m_n}\right|^2$$

$$\leqslant K\Bigg\{ \mathbb{E}\left|\mathbb{E}\, m_n - m_n\right|^2 + \mathbb{E}\left| m_n - \left(\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\mu_i^{(1)} - z}\right) - \sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\left(\frac{1}{n}\sum_{i=1}^{n-1}\frac{\mu_i^{(1)}}{\mu_i^{(1)} - z}\right)\right|^2$$

$$+ \mathbb{E}\left| \frac{1}{n}\sum_{i=1}^{n-1}\left(\frac{\sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\mu_i^{(1)} + 1}{\mu_i^{(1)} - z}\right)\left(w_i^{(1)} - 1\right)\right|^2 + \mathbb{E}\left| \widetilde{\mathbf{y}}_1'\widetilde{\mathbf{y}}_1 - \sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\right|^2 \Bigg\}$$

$$=: K(\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV})$$

$$= O\left(\frac{1}{n^2}\right) + \left[O\left(\frac{1}{n^2}\right) + O\left(\frac{n}{p}\right)\right] + O\left(\frac{1}{n}\right) + O\left(\frac{1}{n}\right) \tag{S2.25}$$

$$= O\left(\frac{1}{n}\right) + O\left(\frac{n}{p}\right). \tag{S2.26}$$

Below we explain (S2.25) in more detail:

(I) Follows from Lemma S1.12.

(II) Use the fact

$$\sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\left|\frac{1}{n}\sum_{i=1}^{n-1}\frac{\mu_i^{(1)}}{\mu_i^{(1)} - z}\right| = O\left(\sqrt{\frac{n}{p}}\right)$$

and

$$\left| m_n - \frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\mu_i^{(1)} - z}\right| = \left|\frac{1}{n}\mathrm{tr}\mathbf{D} - \frac{1}{n}\mathrm{tr}\mathbf{D}_k\right| \overset{\text{(S2.39)}}{=} O\left(\frac{1}{n}\right).$$

(III) Use (S2.21).
(IV) Analogous to the estimation of $\mathbb{E}\left|V-\mathbb{E}_{(1)}V\right|^2$.

□

S2.8. *Proof of Lemma S1.14.*

PROOF. We only provide the estimation of $\widetilde{D}_{11}$, since the others are analogous.

Let $\widetilde{\mathbf{r}}_k'$ be k-th row of $\widetilde{\mathbf{Y}}=\mathbf{\Sigma}_p^{1/2}\mathbf{Y}$ and let $\mathbf{B}_k$ be the $(p-1)\times n$ matrix extracted from $\widetilde{\mathbf{Y}}$ by deleting $\widetilde{\mathbf{r}}_k'$.

With notations defined above, we can write

$$\widetilde{\mathbf{A}}=\begin{pmatrix}\widetilde{\mathbf{r}}_1'\widetilde{\mathbf{r}}_1-\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}} & \widetilde{\mathbf{r}}_1'\mathbf{B}_1' \\ \mathbf{B}_1\widetilde{\mathbf{r}}_1 & \mathbf{B}_1\mathbf{B}_1'-\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_{p-1}\end{pmatrix}.$$

Denote

$$\widetilde{\mathbf{A}}_k=\mathbf{B}_k'\mathbf{B}_k-\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_n,\qquad k=1,\ldots,n,$$

and

$$W=\widetilde{\mathbf{r}}_1'\mathbf{B}_1'\mathbf{B}_1\left(\mathbf{B}_1'\mathbf{B}_1-\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}\mathbf{I}_n-z\mathbf{I}_n\right)^{-1}\widetilde{\mathbf{r}}_1.$$

Let $\{\widetilde{\mu}_i^{(k)},\, i=1,\ldots,n\}$ be the eigenvalues of $\widetilde{\mathbf{A}}_k$, and let $\widetilde{\mathbf{v}}_i^{(1)}=(\widetilde{v}_{i1}^{(1)},\ldots,\widetilde{v}_{ip}^{(1)})$, $i=1,2,\ldots,n$, be the unit eigenvector of $\widetilde{\mathbf{A}}_1$ corresponding to the eigenvalue $\widetilde{\mu}_i^{(1)}$, and set

$$\widetilde{w}_i^{(1)}=\frac{\sqrt{np}a_p}{\sqrt{b_p}}\left|\widetilde{\mathbf{r}}_1'\widetilde{\mathbf{v}}_i^{(1)}\right|^2,$$

then we have

$$W=\frac{1}{n}\sum_{i=1}^n\left(\frac{\sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\widetilde{\mu}_i^{(1)}+1}{\widetilde{\mu}_i^{(1)}-z}\right)\widetilde{w}_i^{(1)},$$

and

$$\widetilde{D}_{11}=\left(\widetilde{\mathbf{r}}_1'\widetilde{\mathbf{r}}_1-\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}-z-W\right)^{-1}=:\left(-\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}-z-m_n+\widetilde{h}_1\right)^{-1},$$

where

$$\widetilde{h}_1=\widetilde{\mathbf{r}}_1'\widetilde{\mathbf{r}}_1+m_n-W$$

(S2.27)

$$=\widetilde{\mathbf{r}}_1'\widetilde{\mathbf{r}}_1+m_n-\frac{1}{n}\sum_{i=1}^n\left(\frac{\sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\widetilde{\mu}_i^{(1)}+1}{\widetilde{\mu}_i^{(1)}-z}\right)-\frac{1}{n}\sum_{i=1}^n\left(\frac{\sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\widetilde{\mu}_i^{(1)}+1}{\widetilde{\mu}_i^{(1)}-z}\right)\left(\widetilde{w}_i^{(1)}-1\right).$$

We define the set of events

$$\Omega_0=\left\{\left|\mathbb{E}\,m_n-m_n+\widetilde{h}_1\right|\geqslant\frac{1}{2}\sqrt{\frac{p}{n}}\right\},$$

then the inequality

$$\left|\left(\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z+\mathbb{E}\,m_n\right)\left(\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z+m_n-\widetilde{h}_1\right)\right|\geqslant K\frac{p}{n}$$

holds on Ω_0. Thus we obtain

$$\begin{aligned}&\mathbb{E}\left|\widetilde{D}_{11}+\frac{1}{a_p\sqrt{p/(nb_p)}+z+\mathbb{E}\,m_n}\right|^2\\&\quad\leqslant\mathbb{E}\left|\frac{\mathbb{E}\,m_n-m_n+\widetilde{h}_1}{\left(a_p\sqrt{p/(nb_p)}+z+\mathbb{E}\,m_n\right)\left(a_p\sqrt{p/(nb_p)}+z+m_n-\widetilde{h}_1\right)}\right|^2\\&\quad\leqslant K\left\{\left(\frac{n}{p}\right)^2\cdot\mathbb{P}(\Omega_0^c)+\frac{n}{p}\cdot\mathbb{P}(\Omega_0)\right\}\cdot\mathbb{E}\left|\mathbb{E}\,m_n-m_n+\widetilde{h}_1\right|^2,\end{aligned}$$

where we use the inequality

$$\left|\left(\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z+\mathbb{E}\,m_n\right)\left(\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z+m_n-\widetilde{h}_1\right)\right|\geqslant K\sqrt{\frac{p}{n}},\tag{S2.28}$$

that holds on the full set Ω. The inequality (S2.28) follows from the facts

$$\begin{aligned}&\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z+m_n-\widetilde{h}_1\\&=\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z-\widetilde{\mathbf{r}}_1'\widetilde{\mathbf{r}}_1+\widetilde{\mathbf{r}}_1'\mathbf{B}_1'\mathbf{B}_1\left(\mathbf{B}_1'\mathbf{B}_1-\sqrt{\tfrac{p}{n}}\tfrac{a_p}{\sqrt{b_p}}\mathbf{I}_n\right)^{-1}\widetilde{\mathbf{r}}_1\\&=\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z+\widetilde{\mathbf{r}}_1'\left\{\mathbf{B}_1'\mathbf{B}_1\left(\mathbf{B}_1'\mathbf{B}_1-\sqrt{\tfrac{p}{n}}\tfrac{a_p}{\sqrt{b_p}}\mathbf{I}_n\right)^{-1}-\mathbf{I}_n\right\}\widetilde{\mathbf{r}}_1\\&=\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z+\frac{1}{\sqrt{np}}\frac{\sqrt{b_p}}{a_p}\sum_{i=1}^n\left(\frac{\widetilde{\mu}_i^{(1)}+\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}}{\widetilde{\mu}_i^{(1)}-z}-1\right)\widetilde{w}_i^{(1)}\\&=\left(\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z\right)\left(1+\frac{1}{\sqrt{np}}\frac{\sqrt{b_p}}{a_p}\sum_{i=1}^n\frac{\widetilde{w}_i^{(1)}}{\widetilde{\mu}_i^{(1)}-z}\right)\\&=:\left(\sqrt{\frac{p}{n}}\frac{a_p}{\sqrt{b_p}}+z\right)(1+S),\end{aligned}$$

and

$$|1+S|\geqslant K\sqrt{\frac{n}{p}}.$$

We now proceed to complete the proof of (S1.6). Note that we have

$$\mathbb{P}(\Omega_0)\leqslant\frac{4n}{p}\mathbb{E}\left|\mathbb{E}\,m_n-m_n+\widetilde{h}_1\right|^2,$$

thus it is sufficient to prove that

$$\mathbb{E}\left|\mathbb{E}\,m_n-m_n+\widetilde{h}_1\right|^2=O\left(\frac{1}{n}\right)+O\left(\frac{n}{p}\right).\tag{S2.29}$$

Applying (S2.27) gives us

$$
\begin{aligned}
&\mathbb{E}\left|\mathbb{E}\, m_n - m_n + \widetilde{h}_1\right|^2 \\
&\quad \leqslant K\Bigg\{\mathbb{E}\left|\mathbb{E}\, m_n - m_n\right|^2 + \mathbb{E}\left| m_n - \left(\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\widetilde{\mu}_i^{(1)} - z}\right) - \sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\left(\frac{1}{n}\sum_{i=1}^{n-1}\frac{\widetilde{\mu}_i^{(1)}}{\widetilde{\mu}_i^{(1)} - z}\right)\right|^2 \\
&\qquad + \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n-1}\left(\frac{\sqrt{\frac{n}{p}}\frac{\sqrt{b_p}}{a_p}\widetilde{\mu}_i^{(1)} + 1}{\widetilde{\mu}_i^{(1)} - z}\right)\left(\widetilde{w}_i^{(1)} - 1\right)\right|^2 + \mathbb{E}\left(\widetilde{\mathbf{r}}_1'\widetilde{\mathbf{r}}_1\right)^2\Bigg\}.
\end{aligned}
\tag{S2.30}
$$

Combining the similar method used for (S2.26) with (S2.30) and the fact

$$
\begin{aligned}
\mathbb{E}\left(\widetilde{\mathbf{r}}_1'\widetilde{\mathbf{r}}_1\right)^2 &= \mathbb{E}\left\{\sum_{j=1}^{n}\left(\sum_{i=1}^{p}(\mathbf{\Sigma}_p^{1/2})_{1i}Y_{ij}\right)^2\right\}^2 \\
&= \mathbb{E}\left\{\sum_{j=1}^{n}\left(\sum_{i=1}^{p}(\mathbf{\Sigma}_p^{1/2})_{1i}Y_{ij}\right)^4 + \sum_{j_1\neq j_2}\left(\sum_{i=1}^{p}(\mathbf{\Sigma}_p^{1/2})_{1i}Y_{ij_1}\right)^2\left(\sum_{i=1}^{p}(\mathbf{\Sigma}_p^{1/2})_{1i}Y_{ij_2}\right)^2\right\} \\
&= O(n/p),
\end{aligned}
$$

we obtain (S2.29). □

S2.9. *Proofs of Lemmas S1.15 and S1.16.*

PROOF. The derivatives in these two lemmas can be derived by using the chain rule and Lemmas S1.17 and S1.18 repeatedly, and the details are omitted here. □

S2.10. *Proof of Lemma S1.17.*

TABLE S.1
Derivatives of $(Y_{rs}Y_{\ell t})$ *w.r.t.* Y_{jk}

$\frac{\partial(Y_{rs}Y_{\ell t})}{\partial Y_{jk}}$	$r=\ell=j$	$r\neq j,\ell\neq j$	$r=j,\ell\neq j$	$r\neq j,\ell=j$
$s=t=k$	$2Y_{jk}$	0	$Y_{\ell k}$	Y_{rk}
$s\neq k,t\neq k$	0	0	0	0
$s=k,t\neq k$	Y_{jt}	0	$Y_{\ell t}$	0
$s\neq k,t=k$	Y_{js}	0	0	Y_{rs}

PROOF. (1) By using the chain rule and derivatives shown in Table S.1, we have

$$
\begin{aligned}
\frac{\partial D_{\alpha\beta}}{\partial Y_{jk}} &= \sum_{1\leqslant s\leqslant t\leqslant p}\frac{\partial D_{\alpha\beta}}{\partial A_{st}}\cdot\frac{\partial A_{st}}{\partial Y_{jk}} \qquad \left[\frac{\partial A_{st}}{\partial Y_{jk}} := \frac{\partial(\mathbf{Y}'\mathbf{\Sigma}_p\mathbf{Y})_{st}}{\partial Y_{jk}}\right] \\
&= \sum_{s=1}^{p}\frac{\partial D_{\alpha\beta}}{\partial A_{ss}}\cdot\frac{\partial A_{ss}}{\partial Y_{jk}} + \sum_{1\leqslant s<t\leqslant p}\frac{\partial D_{\alpha\beta}}{\partial A_{st}}\cdot\frac{\partial A_{st}}{\partial Y_{jk}} \\
&= \sum_{s=1}^{p}\left(-D_{\alpha s}D_{t\beta}\right)\cdot\sum_{r,\ell}\left\{(\mathbf{\Sigma}_p)_{r\ell}\frac{\partial\left(Y_{rs}Y_{\ell s}\right)}{\partial Y_{jk}}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{s<t}\big(-D_{\alpha s}D_{t\beta}-D_{\alpha t}D_{s\beta}\big)\cdot\sum_{r,\ell}\bigg\{(\mathbf{\Sigma}_p)_{r\ell}\frac{\partial\big(Y_{rs}Y_{\ell t}\big)}{\partial Y_{jk}}\bigg\}\\
&=\big(-D_{\alpha k}D_{k\beta}\big)\cdot\bigg\{2(\mathbf{\Sigma}_p)_{jj}Y_{jk}+\sum_{\ell\neq j}(\mathbf{\Sigma}_p)_{j\ell}Y_{\ell k}+\sum_{r\neq j}(\mathbf{\Sigma}_p)_{rj}Y_{rk}\bigg\}\\
&\quad+\sum_{k<t}\big(-D_{\alpha k}D_{t\beta}-D_{\alpha t}D_{k\beta}\big)\cdot\bigg\{(\mathbf{\Sigma}_p)_{jj}Y_{jt}+\sum_{\ell\neq j}(\mathbf{\Sigma}_p)_{j\ell}Y_{\ell t}\bigg\}\\
&\quad+\sum_{s<k}\big(-D_{\alpha s}D_{k\beta}-D_{\alpha k}D_{s\beta}\big)\cdot\bigg\{(\mathbf{\Sigma}_p)_{jj}Y_{js}+\sum_{r\neq j}(\mathbf{\Sigma}_p)_{rj}Y_{rs}\bigg\}\\
&=\sum_{s=1}^{p}\big(-D_{\alpha s}D_{k\beta}-D_{\alpha k}D_{s\beta}\big)\bigg(\sum_{r=1}^{p}(\mathbf{\Sigma}_p)_{rj}Y_{rs}\bigg)\\
&=\sum_{s,r}\bigg\{-\Big((\mathbf{\Sigma}_p)_{jr}Y_{rs}D_{s\alpha}\Big)D_{\beta k}-\Big((\mathbf{\Sigma}_p)_{jr}Y_{rs}D_{s\beta}\Big)D_{\alpha k}\bigg\}\\
&=-F_{j\alpha}D_{\beta k}-F_{j\beta}D_{\alpha k},
\end{aligned}
$$

where the third equality follows from the formula (II. 18) in Khorunzhy, Khoruzhenko and Pastur (1996);

(2)

$$
\begin{aligned}
\frac{\partial F_{\alpha\beta}}{\partial Y_{jk}}&=\frac{\partial}{\partial Y_{jk}}\sum_{s,t}\Big((\mathbf{\Sigma}_p)_{\alpha s}Y_{st}D_{t\beta}\Big)=\sum_{s,t}(\mathbf{\Sigma}_p)_{\alpha s}\bigg(\frac{\partial Y_{st}}{\partial Y_{jk}}\cdot D_{t\beta}+Y_{st}\cdot\frac{\partial D_{t\beta}}{\partial Y_{jk}}\bigg)\\
&=(\mathbf{\Sigma}_p)_{\alpha j}D_{k\beta}-\sum_{s,t}(\mathbf{\Sigma}_p)_{\alpha s}Y_{st}\big(F_{jt}D_{\beta k}+F_{j\beta}D_{tk}\big)\\
&=(\mathbf{\Sigma}_p)_{\alpha j}D_{k\beta}-E_{j\alpha}D_{\beta k}-F_{j\beta}F_{\alpha k};
\end{aligned}
$$

(3)

$$
\begin{aligned}
\frac{\partial E_{jj}}{\partial Y_{jk}}&=\frac{\partial}{\partial Y_{jk}}\sum_{r}\big(\mathbf{\Sigma}_p\mathbf{Y}\mathbf{D}\big)_{jr}\big(\mathbf{Y}'\mathbf{\Sigma}_p\big)_{rj}\\
&=\sum_r\frac{\partial F_{jr}}{\partial Y_{jk}}\cdot\big(\mathbf{Y}'\mathbf{\Sigma}_p\big)_{rj}+\sum_r F_{jr}\cdot\frac{\partial\big(\mathbf{Y}'\mathbf{\Sigma}_p\big)_{rj}}{\partial Y_{jk}}\\
&=\sum_r\Big((\mathbf{\Sigma}_p)_{jj}D_{kr}-E_{jj}D_{rk}-F_{jr}F_{jk}\Big)\cdot\big(\mathbf{Y}'\mathbf{\Sigma}_p\big)_{rj}+(\mathbf{\Sigma}_p)_{jj}F_{jk}\\
&=2(\mathbf{\Sigma}_p)_{jj}F_{jk}-2E_{jj}F_{jk},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial(E_{jj}D_{kk})}{\partial Y_{jk}}&=\frac{\partial E_{jj}}{\partial Y_{jk}}\cdot D_{kk}+\frac{\partial D_{kk}}{\partial Y_{jk}}\cdot E_{jj}\\
&=\Big(2(\mathbf{\Sigma}_p)_{jj}F_{jk}-2E_{jj}F_{jk}\Big)\cdot D_{kk}-2F_{jk}D_{kk}\cdot E_{jj}\\
&=2(\mathbf{\Sigma}_p)_{jj}F_{jk}D_{kk}-4E_{jj}F_{jk}D_{kk}.
\end{aligned}
$$

The proof of lemma is complete. □

S2.11. *Proof of Lemma S1.18.*

PROOF. 1.

$$\frac{\partial \widehat{F}_{\alpha\beta}}{\partial Y_{jk}} = \frac{\partial}{\partial Y_{jk}} \sum_{s,t} \Big((\mathbf{\Sigma}_p^2)_{\alpha s} Y_{st} D_{t\beta} \Big) = \sum_{s,t} (\mathbf{\Sigma}_p^2)_{\alpha s} \left(\frac{\partial Y_{st}}{\partial Y_{jk}} \cdot D_{t\beta} + Y_{st} \cdot \frac{\partial D_{t\beta}}{\partial Y_{jk}} \right)$$
$$= (\mathbf{\Sigma}_p^2)_{\alpha j} D_{k\beta} - \sum_{s,t} (\mathbf{\Sigma}_p^2)_{\alpha s} Y_{st} \big(F_{jt} D_{\beta k} + F_{j\beta} D_{tk} \big)$$
$$= (\mathbf{\Sigma}_p^2)_{\alpha j} D_{k\beta} - \widehat{E}_{j\alpha} D_{\beta k} - F_{j\beta} \widehat{F}_{\alpha k};$$

2.

$$\frac{\partial E_{jr}}{\partial Y_{jk}} = \frac{\partial}{\partial Y_{jk}} \sum_{\ell} F_{j\ell} \big(\mathbf{Y}' \mathbf{\Sigma}_p \big)_{\ell r} = \sum_{\ell} \frac{\partial F_{j\ell}}{\partial Y_{jk}} \cdot \big(\mathbf{Y}' \mathbf{\Sigma}_p \big)_{\ell r} + \sum_{\ell} F_{j\ell} \cdot \frac{\partial \big(\mathbf{Y}' \mathbf{\Sigma}_p \big)_{\ell r}}{\partial Y_{jk}}$$
$$= \sum_{\ell} \Big((\mathbf{\Sigma}_p)_{jj} D_{k\ell} - E_{jj} D_{\ell k} - F_{j\ell} F_{jk} \Big) \cdot \big(\mathbf{Y}' \mathbf{\Sigma}_p \big)_{\ell r} + (\mathbf{\Sigma}_p)_{jr} F_{jk}$$
$$= (\mathbf{\Sigma}_p)_{jj} F_{rk} + (\mathbf{\Sigma}_p)_{jr} F_{jk} - E_{jj} F_{rk} - F_{jk} E_{jr},$$

$$\frac{\partial (\widehat{E}_{jj} D_{kk})}{\partial Y_{jk}} = \left(\frac{\partial}{\partial Y_{jk}} \sum_{r} E_{jr} (\mathbf{\Sigma}_p)_{rj} \right) \cdot D_{kk} + \widehat{E}_{jj} \cdot \big(-2 F_{jk} D_{kk} \big)$$
$$= D_{kk} \sum_{r} (\mathbf{\Sigma}_p)_{rj} \Big((\mathbf{\Sigma}_p)_{jj} F_{rk} + (\mathbf{\Sigma}_p)_{jr} F_{jk} - E_{jj} F_{rk} - F_{jk} E_{jr} \Big) - 2 \widehat{E}_{jj} F_{jk} D_{kk}$$
$$= (\mathbf{\Sigma}_p)_{jj} \widehat{F}_{jk} D_{kk} + (\mathbf{\Sigma}_p^2)_{jj} F_{jk} D_{kk} - E_{jj} \widehat{F}_{jk} D_{kk} - 3 \widehat{E}_{jj} F_{jk} D_{kk}.$$

The proof of the lemma is complete. □

S2.12. *Proof of Lemma S1.19.*

PROOF. The proofs of the first two inequalities are analogous with that of Lemma 4.2 and (4.35) in Chen and Pan (2015), it is then omitted. As follows we prove the remaining two inequalities.

When the event U_n happens, inequality (S2.11) holds and then the proof of the second inequality in Lemma S1.11 holds for $z \in \mathcal{C}_\ell \cup \mathcal{C}_r$, thus, we have

$$\mathbb{E}|\gamma_{k2}|^4 \mathbb{1}_{U_n} \leqslant K \left(\frac{1}{n^2} + \frac{1}{np} \right). \tag{S2.31}$$

Moreover, by (S2.51) and Burkholder's inequality (Lemma S1.3),

$$\mathbb{E} \left| \frac{1}{np b_p} \mathrm{tr} \mathbf{M}_k^{(1)} - \frac{1}{np b_p} \mathbb{E} \mathrm{tr} \mathbf{M}_k^{(1)} \right|^4 \mathbb{1}_{U_n}$$
$$\leqslant \frac{K}{n^4} \left(\frac{a_p}{b_p} + z \sqrt{\frac{n}{p b_p}} \right)^4 \mathbb{E} \left| \sum_{j=1}^{n} (\mathbb{E}_j - \mathbb{E}_{j-1}) (\mathrm{tr} \mathbf{D} - \mathrm{tr} \mathbf{D}_j) \right|^4 \mathbb{1}_{U_n}$$
$$\leqslant \frac{K}{n^4} \left(\frac{a_p}{b_p} + z \sqrt{\frac{n}{p b_p}} \right)^4 \left\{ \sum_{j=1}^{n} \mathbb{E} |\mathrm{tr} \mathbf{D} - \mathrm{tr} \mathbf{D}_j|^4 + \mathbb{E} \left(\sum_{j=1}^{n} \mathbb{E}_k |\mathrm{tr} \mathbf{D} - \mathrm{tr} \mathbf{D}_j|^2 \right)^2 \right\} \mathbb{1}_{U_n}$$
$$\leqslant \frac{K}{n^2}, \tag{S2.32}$$

where the last "$\leqslant$" follows from the fact that $|\text{tr}\mathbf{D} - \text{tr}\mathbf{D}_j|\mathbb{1}_{U_n}$ is bounded.

Combining Lemma S1.1, (S2.31) and (S2.32), we obtain

$$\mathbb{E}|\mu_k|^4\mathbb{1}_{U_n} \leqslant K\left(\frac{1}{n} + \frac{1}{n^2} + \frac{1}{np}\right).$$

The proof of lemma is complete. □

S2.13. *Proof of Lemma S1.20.*

PROOF. Let

$$\varepsilon_k = \frac{1}{z + (npb_p)^{-1}\text{tr}\mathbf{M}_k^{(1)}},$$

$$\mu_k = \frac{1}{\sqrt{npb_p}}(\mathbf{x}_k'\mathbf{\Sigma}_p\mathbf{x}_k - pa_p) - \gamma_{k1} - \left(\frac{1}{npb_p}\text{tr}\mathbf{M}_k^{(1)} - \frac{1}{npb_p}\mathbb{E}\text{tr}\mathbf{M}_k^{(1)}\right).$$

We have the decomposition:

$$M_n^{(1)}(z) = \text{tr}\mathbf{D} - \mathbb{E}\text{tr}\mathbf{D} = \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\text{tr}(\mathbf{D} - \mathbf{D}_k)$$

$$= \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\big(-\beta_k(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k)\big)$$

$$= \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\Big\{-(\beta_k - \varepsilon_k)(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k) - \varepsilon_k\gamma_{k2}\Big\} \tag{S2.33}$$

$$= \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\Big\{-\big(\varepsilon_k^2\mu_k + \beta_k\varepsilon_k^2\mu_k^2\big)(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k) - \varepsilon_k\gamma_{k2}\Big\} \tag{S2.34}$$

$$= \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\Big\{-\varepsilon_k^2\mu_k\Big(1 + \frac{1}{npb_p}\text{tr}\mathbf{M}_k^{(2)}\Big) - \varepsilon_k^2\mu_k\gamma_{k2}$$

$$- \beta_k\varepsilon_k^2\mu_k^2(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k) - \varepsilon_k\gamma_{k2}\Big\}$$

$$=: \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})(u_{k1} + u_{k2} + u_{k3} + u_{k4}) \tag{S2.35}$$

Below we explain (S2.33) and (S2.34) in more details:

- (S2.33) follows from

$$(\mathbb{E}_k - \mathbb{E}_{k-1})\varepsilon_k(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k) = (\mathbb{E}_k - \mathbb{E}_{k-1})\varepsilon_k(\mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k)$$

$$= (\mathbb{E}_k - \mathbb{E}_{k-1})\varepsilon_k\Big(\mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k - \frac{1}{npb_p}\text{tr}\mathbf{M}_k^{(2)}\Big) = (\mathbb{E}_k - \mathbb{E}_{k-1})\varepsilon_k\gamma_{k2}.$$

- (S2.34) follows from the identity $\beta_k = \varepsilon_k + \beta_k\varepsilon_k\mu_k = \varepsilon_k + \varepsilon_k^2\mu_k + \beta_k\varepsilon_k^2\mu_k^2$.

By Lemma S1.19, we have

$$\mathbb{E}|u_{ki}|^4\mathbb{1}_{U_n} = O(n^{-1}), \qquad i = 1, 2, 3, 4,$$

which, together with the decomposition (S2.35) and Burkholder's inequality (Lemma S1.3), implies that

$$\mathbb{E}\big|M_n^{(1)}(z)\mathbb{1}_{U_n}\big|^2 \leqslant K.$$

This completes the proof. □

S2.14. *Proof of Lemma S1.21.*

PROOF. Using Lemma S1.1, we have

$$\mathbb{E}(\mathbf{x}_i'\mathbf{\Sigma}_p^2\mathbf{x}_i - pb_p)^2 \leqslant K\nu_4 \mathsf{tr}(\mathbf{\Sigma}_p^4) \leqslant K\cdot p\|\mathbf{\Sigma}_p^4\| \leqslant Kp. \tag{S2.36}$$

Note that $\mathsf{tr}(\mathbf{A}^*\mathbf{B})$ is the inner product of $\mathsf{vec}(\mathbf{A})$ and $\mathsf{vec}(\mathbf{B})$ for any $n\times m$ matrices $\mathbf{A}$ and $\mathbf{B}$. It follows from the Cauchy–Schwarz inequality that

$$\big|\mathsf{tr}(\mathbf{A}^*\mathbf{B})\big|^2 \leqslant \mathsf{tr}(\mathbf{A}^*\mathbf{A})\cdot\mathsf{tr}(\mathbf{B}^*\mathbf{B}). \tag{S2.37}$$

By using (S2.37), we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{npb_p}\mathsf{tr}\mathbf{M}_k^{(1)}(z) - \frac{1}{n}\mathsf{tr}\mathbf{D}_k(z)\right|^2\\
&= \frac{1}{(npb_p)^2}\mathbb{E}\left|\mathsf{tr}\Big\{\mathbf{D}_k(z)\big(\mathbf{X}_k'\mathbf{\Sigma}_p^2\mathbf{X}_k - pb_p\mathbf{I}_{n-1}\big)\Big\}\right|^2\\
&\leqslant \frac{1}{(npb_p)^2}\mathbb{E}\left\{\mathsf{tr}\big(\mathbf{D}_k(\bar z)\mathbf{D}_k(z)\big)\cdot\mathsf{tr}\big(\mathbf{X}_k'\mathbf{\Sigma}_p^2\mathbf{X}_k - pb_p\mathbf{I}_{n-1}\big)^2\right\}\\
&\leqslant \frac{1}{(npb_p)^2}\mathbb{E}\left\{n\big\|\mathbf{D}_k(\bar z)\mathbf{D}_k(z)\big\|\cdot\mathsf{tr}\big(\mathbf{X}_k'\mathbf{\Sigma}_p^2\mathbf{X}_k - pb_p\mathbf{I}_{n-1}\big)^2\right\}\\
&\leqslant \frac{1}{n(pb_pv_1)^2}\mathbb{E}\left\{\mathsf{tr}\big(\mathbf{X}_k'\mathbf{\Sigma}_p^2\mathbf{X}_k - pb_p\mathbf{I}_{n-1}\big)^2\right\}.
\end{aligned}$$

Indeed, by using (S2.36) and the fact $\mathbb{E}(\mathbf{x}_i'\mathbf{\Sigma}_p^2\mathbf{x}_j)^2 = \mathsf{tr}(\mathbf{\Sigma}_p^4)$ for $i\neq j$, we have

$$\begin{aligned}
\mathbb{E}\Big\{\mathsf{tr}\big(\mathbf{X}_k'\mathbf{\Sigma}_p^2\mathbf{X}_k - pb_p\mathbf{I}_{n-1}\big)^2\Big\} &= \sum_{i\neq k}\mathbb{E}(\mathbf{x}_i'\mathbf{\Sigma}_p^2\mathbf{x}_i - pb_p)^2 + \sum_{i\neq j, i\neq k, j\neq k}\mathbb{E}(\mathbf{x}_i'\mathbf{\Sigma}_p^2\mathbf{x}_j)^2\\
&\leqslant (n-1)\cdot pK + (n-1)(n-2)\cdot pK.
\end{aligned}$$

Thus we have

$$\mathbb{E}\left|\frac{1}{npb_p}\mathsf{tr}\mathbf{M}_k^{(1)}(z) - \frac{1}{n}\mathsf{tr}\mathbf{D}_k(z)\right|^2 \leqslant \frac{Kn}{p}. \tag{S2.38}$$

Moreover, by (S2.48) and (S1.4), we have

$$\left|\frac{1}{n}\mathsf{tr}\mathbf{D}(z) - \frac{1}{n}\mathsf{tr}\mathbf{D}_k(z)\right| \overset{(S2.48)}{=} \frac{1}{n}\left|\beta_k(1+\mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k)\right| \overset{(S1.4)}{\leqslant} \frac{1}{nv_1}, \tag{S2.39}$$

which, together with (S2.38) and the fact that $m_n(z) \xrightarrow{a.s.} m$, completes the proof. □

S2.15. *Proof of Lemma 6.1.*

PROOF. In general, this proof extends the result of Chen and Pan (2015). We denote the non-diagonal part of $\mathbf{A}_n = (\mathbf{X}'\mathbf{\Sigma}_p\mathbf{X} - pa_p\mathbf{I}_n)/\sqrt{npb_p}$ as

$$\mathbf{B}_n = \mathbf{A}_n - \mathsf{diag}(\mathbf{A}_n) = \frac{1}{\sqrt{npb_p}}\big(\mathbf{X}'\mathbf{\Sigma}_p\mathbf{X} - \mathsf{diag}(\mathbf{X}'\mathbf{\Sigma}_p\mathbf{X})\big),$$

where we use the notation $\mathsf{diag}(\mathbf{A})$ to denote the diagonal matrix of $\mathbf{A}$ (replacing all off-diagonal entries with zero). Then we have

$$\max_{1\leqslant j\leqslant n}|\lambda_j^{\mathbf{A}_n}| = \|\mathbf{A}_n\| \leqslant \|\mathbf{B}_n\| + \|\mathsf{diag}(\mathbf{A}_n)\| = \|\mathbf{B}_n\| + \max_{1\leqslant i\leqslant n}|A_{ii}|, \tag{S2.40}$$

where $\|\cdot\|$ denotes the spectral norm and A_{ii} is the (i,i)-th entry of $\mathbf{A}_n$.

Note that

$$\begin{aligned}
\max_{1\leqslant i\leqslant n}|A_{ii}| &= \frac{1}{\sqrt{npb_p}}\max_{1\leqslant i\leqslant n}\left|\sum_{s=1}^{p}\sum_{t=1}^{p}(\mathbf{\Sigma}_p)_{st}X_{si}X_{ti} - pa_p\right| \\
&= \frac{1}{\sqrt{npb_p}}\max_{1\leqslant i\leqslant n}\left|\sum_{s=1}^{p}(\mathbf{\Sigma}_p)_{ss}(X_{si}^2-1) + \sum_{s\neq t}(\mathbf{\Sigma}_p)_{st}X_{si}X_{ti}\right| \\
&\leqslant \frac{1}{\sqrt{npb_p}}\max_{1\leqslant i\leqslant n}\left|\sum_{s=1}^{p}(\mathbf{\Sigma}_p)_{ss}(X_{si}^2-1)\right| + \frac{1}{\sqrt{npb_p}}\max_{1\leqslant i\leqslant n}\left|\sum_{s\neq t}(\mathbf{\Sigma}_p)_{st}X_{si}X_{ti}\right|.
\end{aligned}\tag{S2.41}$$

From (S2.40) and (S2.41), it suffices to prove that, for any $\varepsilon > 0$,

$$\Pr\big(\|\mathbf{B}_n\| \geqslant \eta + \varepsilon\big) = o(n^{-1}), \tag{S2.42}$$

$$\Pr\left(\frac{1}{\sqrt{npb_p}}\max_{1\leqslant i\leqslant n}\left|\sum_{s=1}^{p}(\mathbf{\Sigma}_p)_{ss}(X_{si}^2-1)\right| \geqslant \varepsilon\right) = o(n^{-1}), \tag{S2.43}$$

and

$$\Pr\left(\frac{1}{\sqrt{npb_p}}\max_{1\leqslant i\leqslant n}\left|\sum_{s\neq t}(\mathbf{\Sigma}_p)_{st}X_{si}X_{ti}\right| \geqslant \varepsilon\right) = o(n^{-1}). \tag{S2.44}$$

Since $(\mathbf{\Sigma}_p)_{ss} \leqslant \|\mathbf{\Sigma}_p\|$, (S2.43) follows from inequality (9) in Chen and Pan (2012).

Next, we consider (S2.42). From the well-known Courant-Fischer theorem, we have

$$\begin{aligned}
\|\mathbf{B}_n\|^2 = \max_{\|\mathbf{z}\|=1}\|\mathbf{B}_n\mathbf{z}\|^2 &= \frac{1}{npb_p}\max_{\|\mathbf{z}\|=1}\left\|\Big[\mathbf{X}'\mathbf{\Sigma}_p\mathbf{X} - \mathsf{diag}(\mathbf{X}'\mathbf{\Sigma}_p\mathbf{X})\Big]\mathbf{z}\right\|^2 \\
&= \frac{1}{npb_p}\max_{\|\mathbf{z}\|=1}\sum_{i=1}^{n}\left(\sum_{j\neq i}(\mathbf{x}_i'\mathbf{\Sigma}_p\mathbf{x}_j)z_j\right)^2 \\
&\leqslant \frac{\|\mathbf{\Sigma}_p\|^2}{b_p}\max_{\|\mathbf{z}\|=1}\frac{1}{np}\sum_{i=1}^{n}\left(\sum_{j\neq i}(\mathbf{x}_i'\mathbf{x}_j)z_j\right)^2 \leqslant \left(\frac{\eta}{2}\right)^2\|\widehat{\mathbf{B}}_n\|^2,
\end{aligned}\tag{S2.45}$$

where $\widehat{\mathbf{B}}_n = \frac{1}{\sqrt{np}}\big(\mathbf{X}'\mathbf{X} - \mathsf{diag}(\mathbf{X}'\mathbf{X})\big)$. For any sequence of positive numbers $k = k_p \to \infty$ and $\varepsilon > 0$, we have

$$\mathbb{P}\big(\|\widehat{\mathbf{B}}_n\| \geqslant 2+\varepsilon\big) \leqslant \frac{\mathbb{E}\,\|\widehat{\mathbf{B}}_n\|^{2k}}{(2+\varepsilon)^{2k}} \leqslant \frac{\mathbb{E}\,\sum_{j=1}^{n}\big(\lambda_j^{\widehat{\mathbf{B}}_n}\big)^{2k}}{(2+\varepsilon)^{2k}} = \frac{\mathbb{E}\,\mathsf{tr}(\widehat{\mathbf{B}}_n^{2k})}{(2+\varepsilon)^{2k}}. \tag{S2.46}$$

With an appropriate choice of the sequence $\{k = k_p\}$ and some sophisticated combinational techniques, (Chen and Pan, 2012, p. 1413 - 1418) proved that for any $\varepsilon > 0$,

$$(2+\varepsilon)^{-2k}\mathbb{E}\,\mathrm{tr}(\widehat{\mathbf{B}}_n^{2k}) = o(n^{-1}),$$

which, together with (S2.45) (S2.46), implies (S2.42).

Finally, we consider (S2.44). For and $\varepsilon > 0$ and $\delta > 0$, we have

$$\begin{aligned}
&\Pr\left(\frac{1}{\sqrt{npb_p}}\max_{1\leqslant i\leqslant n}\left|\sum_{s\neq t}(\boldsymbol{\Sigma}_p)_{st}X_{si}X_{ti}\right| > \varepsilon\right)\\
&\leqslant n\Pr\left(\left|\sum_{s\neq t}(\boldsymbol{\Sigma}_p)_{st}X_{s1}X_{t1}\right| > \varepsilon\sqrt{npb_p}\right)\\
&\leqslant n\cdot(\varepsilon\sqrt{npb_p})^{-(4+\delta)}\cdot\mathbb{E}\left|\sum_{s\neq t}(\boldsymbol{\Sigma}_p)_{st}X_{s1}X_{t1}\right|^{4+\delta}\\
&\leqslant n\cdot(\varepsilon\sqrt{npb_p})^{-(4+\delta)}\cdot K\cdot[\mathrm{tr}(\boldsymbol{\Sigma}_p^2)]^{2+\delta/2} \qquad \text{(S2.47)}\\
&= o(n^{-1}),
\end{aligned}$$

the estimation (S2.47) follows from Lemma S1.2 with $k = 4+\delta$. This completes the proof. □

S2.16. *Proof of Lemma 6.2.* As explained in the main text, we first decompose the random part $M_n^{(1)}(z)$ as a sum of martingale difference, which is given in (S2.54). Then, we apply the martingale CLT (Lemma S1.4) to obtain the asymptotic distribution of $M_n^{(1)}(z)$.

Step 1: Martingale difference decomposition of $M_n^{(1)}(z)$.

First, we introduce some notations. Define

$$\mathbf{X}_k = (\mathbf{x}_1,\ldots,\mathbf{x}_{k-1},\mathbf{x}_{k+1},\ldots,\mathbf{x}_n), \qquad \mathbf{A}_k = \tfrac{1}{\sqrt{npb_p}}\left(\mathbf{X}_k'\boldsymbol{\Sigma}_p\mathbf{X}_k - pa_p\mathbf{I}_{n-1}\right),$$

$$\mathbf{D} = (\mathbf{A}_n - z\mathbf{I}_n)^{-1}, \qquad \mathbf{D}_k = (\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1}, \qquad \mathbf{M}_k^{(s)} = \boldsymbol{\Sigma}_p\mathbf{X}_k\mathbf{D}_k^s\mathbf{X}_k'\boldsymbol{\Sigma}_p,\ s=1,2,$$

$$a_{kk}^{\mathsf{diag}} = A_{kk} - z = \tfrac{1}{\sqrt{npb_p}}\left(\mathbf{x}_k'\boldsymbol{\Sigma}_p\mathbf{x}_k - pa_p\right) - z, \qquad \mathbf{q}_k' = \tfrac{1}{\sqrt{npb_p}}\left(\mathbf{x}_k'\boldsymbol{\Sigma}_p\mathbf{X}_k\right)$$

$$\beta_k = \frac{1}{-a_{kk}^{\mathsf{diag}} + \mathbf{q}_k'\mathbf{D}_k\mathbf{q}_k}, \qquad \beta_k^{\mathsf{tr}} = \frac{1}{z + (npb_p)^{-1}\mathrm{tr}\mathbf{M}_k^{(1)}},$$

$$\gamma_{ks} = -\frac{1}{npb_p}\mathrm{tr}\mathbf{M}_k^{(s)} + \mathbf{q}_k'\mathbf{D}_k^s\mathbf{q}_k,\ s=1,2, \qquad \eta_k = \tfrac{1}{\sqrt{npb_p}}\left(\mathbf{x}_k'\boldsymbol{\Sigma}_p\mathbf{x}_k - pa_p\right) - \gamma_{k1},$$

$$\ell_k = -\beta_k\beta_k^{\mathsf{tr}}\eta_k\left(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k\right).$$

Note that a_{kk}^{diag} is the k-th diagonal element of $\mathbf{D}^{-1}$ and $\mathbf{q}_k'$ is the vector from the k-th row of $\mathbf{D}^{-1}$ by deleting the k-th element. By applying Theorem A.5 in Bai and Silverstein (2010), we obtain the equality

$$\mathrm{tr}\mathbf{D} - \mathrm{tr}\mathbf{D}_k = -\frac{1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k}{-a_{kk}^{\mathsf{diag}} + \mathbf{q}_k'\mathbf{D}_k\mathbf{q}_k} = -\beta_k(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k). \qquad \text{(S2.48)}$$

Straightforward calculation gives:

$$\beta_k - \beta_k^{\mathsf{tr}} = \beta_k \beta_k^{\mathsf{tr}} \eta_k, \tag{S2.49}$$

and

$$(\mathbb{E}_k - \mathbb{E}_{k-1})\beta_k^{\mathsf{tr}}(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k) = \mathbb{E}_k\big(\beta_k^{\mathsf{tr}}\gamma_{k2}\big), \qquad \mathbb{E}_{k-1}\big(\beta_k^{\mathsf{tr}}\gamma_{k2}\big) = 0, \tag{S2.50}$$

where $\mathbb{E}_k(\cdot)$ is the expectation with respect to the σ-field generated by the first k columns of $\mathbf{X}$.

By the definition of $\mathbf{D}_k$, we obtain a basic identity:

$$\mathbf{D}_k\mathbf{X}_k'\mathbf{\Sigma}_p\mathbf{X}_k = pa_p\mathbf{D}_k + \sqrt{npb_p}(\mathbf{I}_{n-1} + z\mathbf{D}_k). \tag{S2.51}$$

If $\mathbf{\Sigma}_p = \mathbf{I}_p$, it is straightforward to derive that the limit of $\mathsf{tr}\big(\mathbf{M}_k^{(1)}(z)\big)/(npb_p)$ is $m(z)$ by using (S2.51). However, when $\mathbf{\Sigma}_p \neq \mathbf{I}_p$, we need a more detailed estimate (see Lemma S1.21).

Applying (S2.48) – (S2.50), we have the following decomposition:

$$M_n^{(1)}(z) = \mathsf{tr}\mathbf{D} - \mathbb{E}\mathsf{tr}\mathbf{D} = \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\big(\mathsf{tr}\mathbf{D} - \mathsf{tr}\mathbf{D}_k\big)$$

$$= -\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\beta_k\Big(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k\Big)$$

$$\overset{\text{(S2.49)}}{=} \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})(-\beta_k\beta_k^{\mathsf{tr}}\eta_k)\Big(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k\Big) - \sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\beta_k^{\mathsf{tr}}\Big(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k\Big)$$

$$= \sum_{k=1}^{n}\Big[(\mathbb{E}_k - \mathbb{E}_{k-1})\ell_k - \mathbb{E}_k\big(\beta_k^{\mathsf{tr}}\gamma_{k2}\big)\Big]. \tag{S2.52}$$

By using (S2.49), we can split ℓ_k as

$$\ell_k = -(\beta_k^{\mathsf{tr}} + \beta_k\beta_k^{\mathsf{tr}}\eta_k)\beta_k^{\mathsf{tr}}\eta_k\Big(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k\Big)$$

$$= -\big[(\beta_k^{\mathsf{tr}})^2\eta_k + \beta_k(\beta_k^{\mathsf{tr}})^2\eta_k^2\big]\Big(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k\Big)$$

$$= -(\beta_k^{\mathsf{tr}})^2\eta_k\Big(1 + \frac{1}{npb_p}\mathsf{tr}\mathbf{M}_k^{(2)}\Big) - (\beta_k^{\mathsf{tr}})^2\eta_k\gamma_{k2} - \beta_k(\beta_k^{\mathsf{tr}})^2\eta_k^2\Big(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k\Big)$$

$$= \ell_{k1} + \ell_{k2} + \ell_{k3}, \tag{S2.53}$$

where $\ell_{k1} = -(\beta_k^{\mathsf{tr}})^2\eta_k\big(1 + \mathsf{tr}\mathbf{M}_k^{(2)}/(npb_p)\big), \ell_{k2} = -(\beta_k^{\mathsf{tr}})^2\eta_k\gamma_{k2}, \ell_{k3} = -\beta_k(\beta_k^{\mathsf{tr}})^2\eta_k^2\big(1 + \mathbf{q}_k'\mathbf{D}_k^2\mathbf{q}_k\big)$.

By Lemma S1.10 and Lemma S1.11, it is not difficult to verify that

$$\mathbb{E}\Big|\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\ell_{k2}\Big|^2 = o(1), \qquad \mathbb{E}\Big|\sum_{k=1}^{n}(\mathbb{E}_k - \mathbb{E}_{k-1})\ell_{k3}\Big|^2 = o(1).$$

These estimates, together with (S2.52) and (S2.53), imply that

$$M_n^{(1)}(z) = \sum_{k=1}^{n}\mathbb{E}_k\Big\{-(\beta_k^{\mathsf{tr}})^2\eta_k\Big(1 + \frac{1}{npb_p}\mathsf{tr}\mathbf{M}_k^{(2)}\Big) - \beta_k^{\mathsf{tr}}\gamma_{k2}\Big\} + o_P(1)$$

$$=: \sum_{k=1}^{n}Y_k(z) + o_P(1), \tag{S2.54}$$

where $Y_k(z)$ is a sequence of martingale differences.

Step 2: Application of martingales CLT to (S2.54).

To prove finite-dimensional convergence of $M_n^{(1)}(z), z \in \mathbb{C}_1$, we need only to consider the limit of the following martingale difference decomposition:

$$\sum_{j=1}^{r} a_j M_n^{(1)}(z_j) = \sum_{j=1}^{r} a_j \sum_{k=1}^{n} Y_k(z_j) + o(1) = \sum_{k=1}^{n} \left(\sum_{j=1}^{r} a_j Y_k(z_j) \right) + o(1),$$

where $\{a_j\}$ is any complex sequence and r is any positive integer. We apply the martingale CLT (Lemma S1.4) to this martingale difference decomposition of $\sum_{j=1}^{r} a_j M_n^{(1)}(z_j)$. To this end, we need to check two conditions:

Condition 1. For any $\varepsilon > 0$,

$$\sum_{k=1}^{n} \mathbb{E}\left(\left| \sum_{j=1}^{r} a_j Y_k(z_j) \right|^2 \mathbb{1}_{\{\sum_{j=1}^{r} a_j Y_k(z_j) \geqslant \varepsilon\}} \right) = o(1).$$

Condition 2. For $z_1, z_2 \in \mathbb{C}_1$, the sum

$$\Lambda_n(z_1, z_2) := \sum_{k=1}^{n} \mathbb{E}_{k-1}\big(Y_k(z_1) Y_k(z_2)\big) \tag{S2.55}$$

converges in probability to $\Lambda(z_1, z_2)$ defined in (33).

First, we verify Condition 1. By Lemma S1.10 and Lemma S1.11, we have

$$\mathbb{E}|Y_j(z)|^4 \leqslant K \frac{\delta_n^4}{n} + K\left(\frac{1}{n^2} + \frac{n}{p^2} \right),$$

which implies that, for each $\varepsilon > 0$,

$$\sum_{k=1}^{n} \mathbb{E}\left(\left| \sum_{j=1}^{r} a_j Y_k(z_j) \right|^2 \mathbb{1}_{\{\sum_{j=1}^{r} a_j Y_k(z_j) \geqslant \varepsilon\}} \right) \leqslant \frac{1}{\varepsilon^2} \sum_{k=1}^{n} \mathbb{E}\left| \sum_{j=1}^{r} a_j Y_k(z_j) \right|^4 = o(1),$$

thus, Condition 1 is satisfied.

Then, we verify Condition 2. Note that

$$-\big(\beta_k^{\mathsf{tr}}\big)^2 \eta_k \left(1 + \frac{1}{npb_p} \mathrm{tr} \mathbf{M}_k^{(2)} \right) - \beta_k^{\mathsf{tr}} \gamma_{k2} = \frac{\partial}{\partial z} \Big\{ \beta_k^{\mathsf{tr}}(z) \eta_k(z) \Big\},$$

thus, we can rewrite $\Lambda_n(z_1, z_2)$ as

$$\Lambda_n(z_1, z_2) = \frac{\partial^2}{\partial z_2 \partial z_1} \sum_{k=1}^{n} \mathbb{E}_{k-1}\Big[\mathbb{E}_k\big\{ \beta_k^{\mathsf{tr}}(z_1) \eta_k(z_1) \big\} \cdot \mathbb{E}_k\big\{ \beta_k^{\mathsf{tr}}(z_2) \eta_k(z_2) \big\} \Big]. \tag{S2.56}$$

It is enough to consider the limit of

$$\sum_{k=1}^{n} \mathbb{E}_{k-1}\Big[\mathbb{E}_k\big\{ \beta_k^{\mathsf{tr}}(z_1) \eta_k(z_1) \big\} \cdot \mathbb{E}_k\big\{ \beta_k^{\mathsf{tr}}(z_2) \eta_k(z_2) \big\} \Big]. \tag{S2.57}$$

By equation (4), Lemma S1.21, and the dominated convergence theorem, we conclude that

$$\mathbb{E}\Big| \beta_k^{\mathsf{tr}}(z) + m(z) \Big|^2 = o(1). \tag{S2.58}$$

Combining (S2.57) and (S2.58) yields that

$$\sum_{k=1}^{n}\mathbb{E}_{k-1}\Big[\mathbb{E}_k\big\{\beta_k^{\mathsf{tr}}(z_1)\eta_k(z_1)\big\}\cdot\mathbb{E}_k\big\{\beta_k^{\mathsf{tr}}(z_2)\eta_k(z_2)\big\}\Big]$$

$$= m(z_1)m(z_2)\sum_{k=1}^{n}\mathbb{E}_{k-1}\big\{\mathbb{E}_k\eta_k(z_1)\cdot\mathbb{E}_k\eta_k(z_2)\big\}+o_P(1)$$

$$=: m(z_1)m(z_2)\widetilde{\Lambda}_n(z_1,z_2)+o_P(1). \tag{S2.59}$$

In view of (S2.55) – (S2.59), it suffices to derive the limit of $\widetilde{\Lambda}_n(z_1,z_2)$, which further gives the limit of (S2.55).

Since $\mathbb{E}_k\big\{\eta_k(z)\big\}=(1/\sqrt{npb_p})\big(\mathbf{x}_k'\mathbf{\Sigma}_p\mathbf{x}_k-pa_p\big)-\mathbb{E}_k\big[\gamma_{k1}(z)\big]$, we have

$$\mathbb{E}_{k-1}\big\{\mathbb{E}_k\eta_k(z_1)\cdot\mathbb{E}_k\eta_k(z_2)\big\}=\frac{1}{n}\bigg\{\frac{\tilde{b}_p}{b_p}(\nu_4-3)+2\bigg\}+A_1^{(k)}+A_2^{(k)}+A_3^{(k)}, \tag{S2.60}$$

where

$$A_1^{(k)}=\mathbb{E}_{k-1}\big\{\mathbb{E}_k\gamma_{k1}(z_1)\cdot\mathbb{E}_k\gamma_{k1}(z_2)\big\},$$

$$A_2^{(k)}=-\mathbb{E}_{k-1}\bigg\{\frac{1}{\sqrt{npb_p}}\big(\mathbf{x}_k'\mathbf{\Sigma}_p\mathbf{x}_k-pa_p\big)\cdot\mathbb{E}_k\gamma_{k1}(z_1)\bigg\},$$

$$A_3^{(k)}=-\mathbb{E}_{k-1}\bigg\{\frac{1}{\sqrt{npb_p}}\big(\mathbf{x}_k'\mathbf{\Sigma}_p\mathbf{x}_k-pa_p\big)\cdot\mathbb{E}_k\gamma_{k1}(z_2)\bigg\}.$$

First, we show that $A_2^{(k)}$ and $A_3^{(k)}$ are negligible. Denote $\mathbf{M}_k^{(1)}(z)=\big(m_{ij}^{(1)}(z)\big)_{p\times p}$, using the independence between $\mathbf{x}_k$ and $\mathbf{M}_k^{(1)}$, we have

$$A_2^{(k)}=\frac{-1}{(npb_p)^{3/2}}\mathbb{E}_{k-1}\Bigg[\bigg(\sum_{i,j}(\mathbf{\Sigma}_p)_{ij}X_{ik}X_{jk}-pa_p\bigg)\times\bigg\{\sum_{i\neq j}X_{ik}X_{jk}\mathbb{E}_k m_{ij}^{(1)}+\sum_{i=1}^{p}(X_{ik}^2-1)\mathbb{E}_k m_{ii}^{(1)}\bigg\}\Bigg]$$

$$=\frac{-1}{(npb_p)^{3/2}}\mathbb{E}_{k-1}\bigg\{\sum_{i\neq j}(\mathbf{\Sigma}_p)_{ij}X_{ik}^2X_{jk}^2\mathbb{E}_k m_{ij}^{(1)}+\sum_{i=1}^{p}(\mathbf{\Sigma}_p)_{ii}X_{ik}^2(X_{ik}^2-1)\mathbb{E}_k m_{ii}^{(1)}\bigg\}$$

$$=\frac{-1}{(npb_p)^{3/2}}\bigg\{\sum_{i\neq j}(\mathbf{\Sigma}_p)_{ij}\mathbb{E}_k m_{ij}^{(1)}+(\nu_4-1)\sum_{i=1}^{p}(\mathbf{\Sigma}_p)_{ii}\mathbb{E}_k m_{ii}^{(1)}\bigg\}$$

$$=\frac{-1}{\sqrt{npb_p}}\mathbb{E}_k\bigg\{\frac{1}{npb_p}\mathsf{tr}\Big(\mathbf{\Sigma}_p\mathbf{M}_k^{(1)}\Big)-\frac{\nu_4-2}{npb_p}\sum_{i=1}^{p}(\mathbf{\Sigma}_p)_{ii}m_{ii}^{(1)}\bigg\}. \tag{S2.61}$$

As for the first term in the bracket of (S2.61), we can estimate it by using a similar argument as in the proof of Lemma S1.21. Replacing pb_p and $\mathbf{M}_k^{(1)}$ in the proof of Lemma S1.21 with $\mathsf{tr}(\mathbf{\Sigma}_p^3)$ and $\mathbf{\Sigma}_p\mathbf{M}_k^{(1)}$, we can prove that

$$\mathbb{E}\bigg|\frac{1}{n\mathsf{tr}(\mathbf{\Sigma}_p^3)}\mathsf{tr}\Big(\mathbf{\Sigma}_p\mathbf{M}_k^{(1)}\Big)-\frac{1}{n}\mathsf{tr}\mathbf{D}_k\bigg|^2\leqslant\frac{Kn}{p}.$$

Moreover, by the fact $\frac{b_p^2}{a_p} \leqslant \mathsf{tr}(\mathbf{\Sigma}_p^3) \leqslant Kp$, the first inequality of which follows from Cauchy–Schwarz inequality, we conclude that

$$\frac{1}{npb_p}\mathsf{tr}\Big(\mathbf{\Sigma}_p\mathbf{M}_k^{(1)}\Big) = \frac{\mathsf{tr}(\mathbf{\Sigma}_p^3)}{pb_p}\cdot\frac{1}{n\mathsf{tr}(\mathbf{\Sigma}_p^3)}\mathsf{tr}\Big(\mathbf{\Sigma}_p\mathbf{M}_k^{(1)}\Big) = O_P(1).$$

As for the second term in the bracket of (S2.61), we have

$$\frac{1}{npb_p}\sum_{i=1}^p(\mathbf{\Sigma}_p)_{ii}a_{ii}^{(1)} \leqslant \frac{\|\mathbf{\Sigma}_p\|}{npb_p}\sum_{i=1}^p a_{ii}^{(1)} = \frac{\|\mathbf{\Sigma}_p\|}{npb_p}\mathsf{tr}\mathbf{M}_k^{(1)} = O_P(1).$$

Thus, the term in the square bracket of (S2.61) is bounded in probability. Therefore, we conclude that $\Big|\sum_{k=1}^n A_2^{(k)}\Big| \to 0$. Similarly, we can show that $\Big|\sum_{k=1}^n A_3^{(k)}\Big| \to 0$.

Now we consider $A_1^{(k)}$ with the notation $\mathbf{M}_k^{(1)}(z) = \big(m_{ij}^{(1)}(z)\big)_{p\times p}$,

$$\begin{aligned}
A_1^{(k)} &= \frac{1}{(npb_p)^2}\mathbb{E}_{k-1}\Bigg[\Bigg\{\sum_{i\neq j}X_{ik}X_{jk}\mathbb{E}_k m_{ij}^{(1)}(z_1) + \sum_{i=1}^p(X_{ik}^2-1)\mathbb{E}_k m_{ii}^{(1)}(z_1)\Bigg\} \\
&\qquad\times\Bigg\{\sum_{i\neq j}X_{ik}X_{jk}\mathbb{E}_k m_{ij}^{(1)}(z_2) + \sum_{i=1}^p(X_{ik}^2-1)\mathbb{E}_k m_{ii}^{(1)}(z_2)\Bigg\}\Bigg] \\
&= \frac{1}{(npb_p)^2}\mathbb{E}_{k-1}\Bigg\{2\sum_{i\neq j}X_{ik}^2X_{jk}^2\mathbb{E}_k m_{ij}^{(1)}(z_1)\mathbb{E}_k m_{ij}^{(1)}(z_2) + \sum_{i=1}^p(X_{ik}^2-1)^2\mathbb{E}_k m_{ii}^{(1)}(z_1)\mathbb{E}_k m_{ii}^{(1)}(z_2)\Bigg\} \\
&= \frac{1}{(npb_p)^2}\Bigg\{2\sum_{i,j}\mathbb{E}_k m_{ij}^{(1)}(z_1)\mathbb{E}_k m_{ij}^{(1)}(z_2) + (\nu_4-3)\sum_{i=1}^p\mathbb{E}_k m_{ii}^{(1)}(z_1)\mathbb{E}_k m_{ii}^{(1)}(z_2)\Bigg\} \\
&= \frac{2}{(npb_p)^2}\mathsf{tr}\Big(\mathbb{E}_k\mathbf{M}_k^{(1)}(z_1)\cdot\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\Big) + o_P(1),
\end{aligned}$$

where the last step follows from

$$\begin{aligned}
&\mathbb{E}\Bigg|\sum_{i=1}^p\mathbb{E}_k m_{ii}^{(1)}(z_1)\cdot\mathbb{E}_k m_{ii}^{(1)}(z_2)\Bigg|^2 \leqslant p\cdot\sum_{i=1}^p\mathbb{E}\Big|\mathbb{E}_k m_{ii}^{(1)}(z_1)\cdot\mathbb{E}_k m_{ii}^{(1)}(z_2)\Big|^2 \\
&\leqslant p\cdot\sum_{i=1}^p\Big(\mathbb{E}\Big|\mathbb{E}_k m_{ii}^{(1)}(z_1)\Big|^4\Big)^{1/2}\cdot\Big(\mathbb{E}\Big|\mathbb{E}_k m_{ii}^{(1)}(z_2)\Big|^4\Big)^{1/2} \\
&\leqslant p\cdot\sum_{i=1}^p\Big(\mathbb{E}\Big|m_{ii}^{(1)}(z_1)\Big|^4\Big)^{1/2}\cdot\Big(\mathbb{E}\Big|m_{ii}^{(1)}(z_2)\Big|^4\Big)^{1/2} \leqslant K(n^4p^2+n^2p^3).
\end{aligned}$$

By the above estimates, we obtain

$$\begin{aligned}
\widetilde{\Lambda}_n(z_1,z_2) &= \frac{2}{(npb_p)^2}\sum_{k=1}^n\mathsf{tr}\Big(\mathbb{E}_k\mathbf{M}_k^{(1)}(z_1)\cdot\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\Big) + \Big\{\frac{\tilde{b}_p}{b_p}(\nu_4-3)+2\Big\} + o_P(1), \\
&= \frac{2}{n}\sum_{k=1}^n\mathbb{Z}_k + \frac{\tilde{b}_p}{b_p}(\nu_4-3) + 2 + o_P(1), \qquad \text{(S2.62)}
\end{aligned}$$

where

$$\mathbb{Z}_k = \frac{1}{n(pb_p)^2}\mathsf{tr}\Big(\mathbb{E}_k\mathbf{M}_k^{(1)}(z_1)\cdot\mathbb{E}_k\mathbf{M}_k^{(1)}(z_2)\Big).$$

In Lemma S1.7, we derive the asymptotic expression of $\mathbb{Z}_k$. This asymptotic expression ensures that

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{Z}_k \to \int_0^1 \frac{tm(z_1)m(z_2)}{1-tm(z_1)m(z_2)}\mathrm{d}t = -1 - \frac{\log\big(1-m(z_1)m(z_2)\big)}{m(z_1)m(z_2)}. \tag{S2.63}$$

By (S2.56), (S2.59), (S2.62) and (S2.63), we have

$$\widetilde{\Lambda}_n(z_1,z_2) \xrightarrow{p} \frac{\omega}{\theta}(\nu_4-3) - \frac{2\log\big(1-m(z_1)m(z_2)\big)}{m(z_1)m(z_2)}.$$

Therefore,

$$\begin{aligned}\Lambda_n(z_1,z_2) &\xrightarrow{p} \frac{\partial^2}{\partial z_1\partial z_2}\left\{\frac{\omega}{\theta}(\nu_4-3)m(z_1)m(z_2) - 2\log\big(1-m(z_1)m(z_2)\big)\right\}\\ &= m'(z_1)m'(z_2)\left\{\frac{\omega}{\theta}(\nu_4-3) + 2\big(1-m(z_1)m(z_2)\big)^{-2}\right\}.\end{aligned}$$

The verification of Condition 2 is then complete.

S3. Proofs of equations (35) and (36). This section contains proofs of equations (35) and (36).

PROOF. We only consider the case $j=\ell,0$. (The case $j=r$ is similar to $j=\ell$.) From Proposition 6.1, we have $\mathbb{E}|M(z)|^2=\Lambda(z,\bar z)=O(1)$. Figure S.1 shows the decomposition of the contour $\mathcal{C}=\mathcal{C}_\ell\cup\mathcal{C}_r\cup\mathcal{C}_u\cup\mathcal{C}_0$. Let $\|\mathcal{C}_j\|$ denote the length of $\mathcal{C}_j$, $j=\ell,0$, then

$$\int_{\mathcal{C}_0}\mathbb{E}|M(z)|^2\,\mathrm{d}z \leqslant |\Lambda(z,\bar z)|\cdot\|\mathcal{C}_0\| = |\Lambda(z,\bar z)|\cdot(2\xi_n/n)\to 0$$

and

$$\lim_{v_1\downarrow 0}\int_{\mathcal{C}_\ell}\mathbb{E}|M(z)|^2\,\mathrm{d}z \leqslant \lim_{v_1\downarrow 0}|\Lambda(z,\bar z)|\cdot\|\mathcal{C}_\ell\| = \lim_{v_1\downarrow 0}|\Lambda(z,\bar z)|\cdot 2(v_1-\xi_n/n)=0.$$

Thus, the estimate (36) holds for $z\in\mathcal{C}_j,\ j=\ell,0$.

We choose the event $U_n=\{\max_{j\leqslant n}|\lambda_j^{\mathbf{A}_n}|<\eta+\varepsilon\}$ with $\varepsilon=(u_1-\eta)/2$. By Lemma 6.1, the probability of U_n^c decays to zero faster than n^{-1}, that is,

$$\mathbb{P}(U_n^c)=o(n^{-1}).$$

When the event U_n happens, for any $z\in\mathcal{C}_0$, we have $|m_n(z)|\leqslant 2/(u_1-\eta)$ and $|m(z)|\leqslant 1$. Thus, we have

$$\begin{aligned}\int_{\mathcal{C}_0}\mathbb{E}\big|M_n(z)\mathbb{1}_{U_n}\big|^2\,\mathrm{d}z &= \int_{\mathcal{C}_0}\mathbb{E}\big|n[m_n(z)-m(z)-\mathcal{X}_n(m)]\mathbb{1}_{U_n}\big|^2\,\mathrm{d}z\\ &\leqslant n\left(\frac{2}{u_1-\eta}+1+o(1)\right)^2\|\mathcal{C}_0\|\\ &= 2\left(\frac{2}{u_1-\eta}+1+o(1)\right)^2\xi_n,\end{aligned}$$

since $\xi_n\to 0$ as $n\to\infty$, we conclude that (35) is true for $z\in\mathcal{C}_0$.

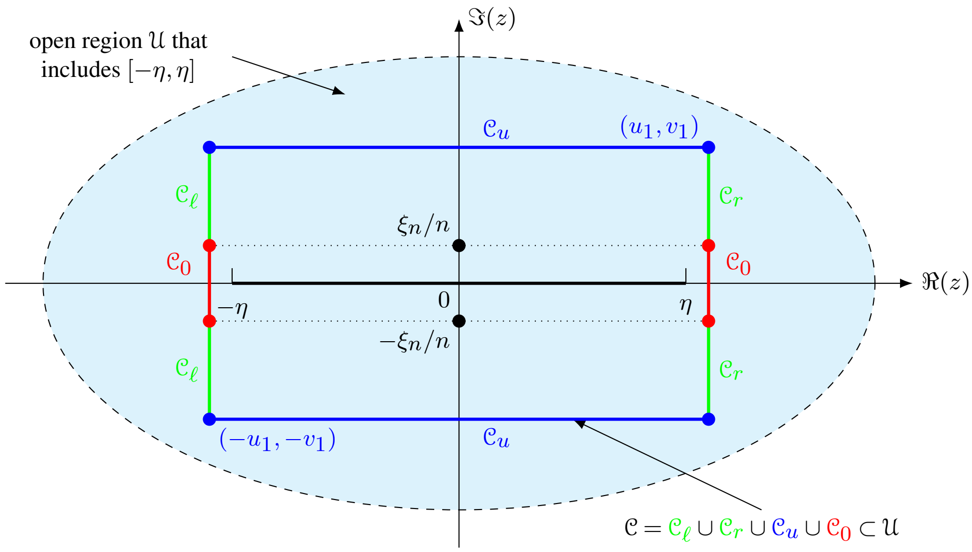


Fig S.1: Open region $\mathcal{U}$ and decomposition of the closed contour $\mathcal{C}$

Recall that we decompose $M_n(z)$ into a random part $M_n^{(1)}(z)$ and a deterministic part $M_n^{(2)}(z)$ for $z \in \mathcal{C}$, where

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

thus

$$\text{(S3.1)} \qquad \int_{\mathcal{C}_\ell} \mathbb{E}\Big|M_n(z)\mathbb{1}_{U_n}\Big|^2 \mathrm{d}z \leqslant K\int_{\mathcal{C}_\ell} \mathbb{E}\Big|M_n^{(1)}(z)\mathbb{1}_{U_n}\Big|^2 \mathrm{d}z + K\int_{\mathcal{C}_\ell} \Big|M_n^{(2)}(z)\mathbb{1}_{U_n}\Big|^2 \mathrm{d}z.$$

By Lemma S1.20, we have

$$0 \leqslant \int_{\mathcal{C}_\ell} \mathbb{E}\Big|M_n^{(1)}(z)\mathbb{1}_{U_n}\Big|^2 \mathrm{d}z \leqslant K\|\mathcal{C}_\ell\| \to 0, \qquad \text{as } n\to\infty,\ v_1 \downarrow 0.$$

Similarly, we have

$$\int_{\mathcal{C}_\ell} \Big|M_n^{(2)}(z)\mathbb{1}_{U_n}\Big|^2 \mathrm{d}z \to 0, \qquad \text{as } n\to\infty,\ v_1 \downarrow 0.$$

Plugging these estimation into (S3.1) implies that (35) is true for $z \in \mathcal{C}_j$, $j = \ell$. □

S4. Proofs in applications. This section contains proofs of equation (15), Theorem 4.2, and Proposition 4.1.

S4.1. *Proof of equation* (15).

PROOF. By Lemma 2.2 in Wang and Yao (2013), under the high-dimensional setting $c_n = p/n \to c$ as $p \to \infty$, we have

$$n\begin{pmatrix} \frac{1}{p}\mathrm{tr}(\mathbf{S}_n^2) - \left(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}\right) \\ \frac{1}{p}\mathrm{tr}(\mathbf{S}_n) - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \frac{1}{c^2}\mathbf{H}\right),$$

where

$$\mathbf{H} = \begin{pmatrix} 4c^2 + 4(\nu_4 - 1)(1+c)^2 c & 2(\nu_4 - 1)(1+c)c \\ 2(\nu_4 - 1)(1+c)c & (\nu_4 - 1)c \end{pmatrix}.$$

Define the function $f(x,y) = x - 2y + 1 - py^2/n + p/n$, then $W = f\big(\mathrm{tr}(\mathbf{S}_n^2)/p, \mathrm{tr}(\mathbf{S}_n)/p\big)$, and

$$\frac{\partial f}{\partial x}\Big(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}, 1\Big) = 1$$

$$\frac{\partial f}{\partial y}\Big(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}, 1\Big) = -2\Big(1 + \frac{p}{n}\Big)$$

$$f\Big(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}, 1\Big) = \frac{p}{n} + \frac{c(\nu_4-2)}{p}.$$

By the delta method, we obtain

$$n\bigg(W - f\Big(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}, 1\Big)\bigg) \xrightarrow{d} \mathcal{N}(0, \lim D),$$

where

$$D = \begin{pmatrix} \frac{\partial f}{\partial x}(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}, 1) \\ \frac{\partial f}{\partial y}(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}, 1) \end{pmatrix}' \Big(\frac{1}{c^2}\mathbf{H}\Big) \begin{pmatrix} \frac{\partial f}{\partial x}(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}, 1) \\ \frac{\partial f}{\partial y}(1 + \frac{p}{n} + \frac{c(\nu_4-2)}{p}, 1) \end{pmatrix} \to 4$$

as $n \to \infty$. Thus,

$$nW - p - (\nu_4 - 2) \xrightarrow{d} \mathcal{N}(0, 4),$$

and the proof of (15) is complete. □

S4.2. *Proof of Theorem 4.2.*

PROOF. Recall that $\mathbf{S}_n = \mathbf{Y}\mathbf{Y}'/n$, $a_p = \mathrm{tr}(\mathbf{\Sigma}_p)/p$ and $b_p = \mathrm{tr}(\mathbf{\Sigma}_p^2)/p$. Let $\mathbf{A}_n = \frac{1}{\sqrt{npb_p}}(\mathbf{Y}'\mathbf{Y} - pa_p\mathbf{I}_n) = \frac{1}{\sqrt{npb_p}}(\mathbf{X}'\mathbf{\Sigma}_p\mathbf{X} - pa_p\mathbf{I}_n)$. By some elementary calculations, we obtain two identities:

$$\mathrm{tr}(\mathbf{S}_n) = \sqrt{\frac{pb_p}{n}}\mathrm{tr}(\mathbf{A}_n) + pa_p, \qquad \mathrm{tr}(\mathbf{S}_n^2) = \frac{pb_p}{n}\mathrm{tr}(\mathbf{A}_n^2) + \frac{2pa_p}{n}\sqrt{\frac{pb_p}{n}}\mathrm{tr}(\mathbf{A}_n) + \frac{(pa_p)^2}{n}.$$

Then W can be written as

$$W = \frac{b_p}{n}\mathrm{tr}(\mathbf{A}_n^2) - \frac{2}{p}\sqrt{\frac{pb_p}{n}}\mathrm{tr}(\mathbf{A}_n) - \frac{b_p}{n^2}\big[\mathrm{tr}(\mathbf{A}_n)\big]^2 + \frac{p}{n} - 2a_p + 1.$$

Li and Yao (2016) derived the limiting joint distribution of $\big(\mathrm{tr}(\mathbf{A}_n^2)/n, \mathrm{tr}(\mathbf{A}_n)/n\big)$ (see their Lemma 3.1) as follows:

$$n\begin{pmatrix} \frac{1}{n}\mathrm{tr}(\mathbf{A}_n^2) - 1 - \frac{1}{n}\big(\frac{\omega}{\theta}(\nu_4 - 3) + 1\big) \\ \frac{1}{n}\mathrm{tr}(\mathbf{A}_n) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & \frac{\omega}{\theta}(\nu_4 - 3) + 2 \end{pmatrix}\right). \tag{S4.1}$$

Define the function

$$g(x,y) = b_p x - \frac{2n}{p}\sqrt{\frac{pb_p}{n}}y - b_p y^2 + \frac{p}{n} - 2a_p + 1,$$

then $W = g\big(\mathrm{tr}(\mathbf{A}_n^2)/n, \mathrm{tr}(\mathbf{A}_n)/n\big)$, we have

$$\frac{\partial g}{\partial x}\Big(1+\frac{1}{n}\big(\frac{\omega}{\theta}(\nu_4-3)+1\big),0\Big) = b_p,$$

$$\frac{\partial g}{\partial y}\Big(1+\frac{1}{n}\big(\frac{\omega}{\theta}(\nu_4-3)+1\big),0\Big) = -\frac{2n}{p}\sqrt{\frac{pb_p}{n}},$$

$$g\Big(1+\frac{1}{n}\big(\frac{\omega}{\theta}(\nu_4-3)+1\big),0\Big) = b_p + \frac{b_p}{n}\Big(\frac{\omega}{\theta}(\nu_4-3)+1\Big) + \frac{p}{n} - 2a_p + 1.$$

By (S4.1), we have

$$n\bigg(W - g\Big(1+\frac{1}{n}\big(\frac{\omega}{\theta}(\nu_4-3)+1\big),0\Big)\bigg) \xrightarrow{d} \mathcal{N}(0, \lim A),$$

where

$$A = \begin{pmatrix} \frac{\partial g}{\partial x}(1+\frac{1}{n}(\frac{\omega}{\theta}(\nu_4-3)+1),0) \\ \frac{\partial g}{\partial x}(1+\frac{1}{n}(\frac{\omega}{\theta}(\nu_4-3)+1),0) \end{pmatrix}' \begin{pmatrix} 4 & 0 \\ 0 & \frac{\omega}{\theta}(\nu_4-3)+2 \end{pmatrix} \begin{pmatrix} \frac{\partial g}{\partial x}(1+\frac{1}{n}(\frac{\omega}{\theta}(\nu_4-3)+1),0) \\ \frac{\partial g}{\partial x}(1+\frac{1}{n}(\frac{\omega}{\theta}(\nu_4-3)+1),0) \end{pmatrix} \to 4\theta^2.$$

Thus,

$$n\Big(W - b_p - \frac{b_p}{n}\Big(\frac{\omega}{\theta}(\nu_4-3)+1\Big) + 2a_p - 1 - \frac{p}{n}\Big) \xrightarrow{d} \mathcal{N}(0, 4\theta^2),$$

that is,

$$nW - p - \theta\Big(\frac{\omega}{\theta}(\nu_4-3)+1\Big) + n(2\gamma-1-\theta) \xrightarrow{d} \mathcal{N}(0, 4\theta^2).$$

The proof of Theorem 4.2 is complete. □

S4.3. *Proof of Proposition 4.1.*

PROOF. For the test based on statistic W, by Theorem 4.1 and 4.2, we have

$$\begin{aligned}\beta(H_1) &= \mathbb{P}\bigg(\frac{1}{2}\Big(nW - p - (\nu_4-2)\Big) \geqslant z_\alpha \,\bigg|\, H_1\bigg) \\ &= \mathbb{P}\bigg(nW - p - \theta\Big(\frac{\omega}{\theta}(\nu_4-3)+1\Big) + n(2\gamma-1-\theta) \\ &\qquad\qquad \geqslant 2z_\alpha - \theta\Big(\frac{\omega}{\theta}(\nu_4-3)+1\Big) + n(2\gamma-1-\theta) + (\nu_4-2) \,\bigg|\, H_1\bigg) \\ &= 1 - \Phi\bigg(\frac{1}{2\theta}\Big\{2z_\alpha - \omega(\nu_4-3) - \theta + n(2\gamma-1-\theta) + (\nu_4-2)\Big\}\bigg),\end{aligned}$$

since $2\gamma - 1 \leqslant \gamma^2 \leqslant \theta$, Proposition 4.1 follows. □

S5. Additional simulation results. This section contains some additional simulation results of the paper. The simulation settings are the same as those in Section 5.1 of the main paper except $p = n^{2.5}$.

TABLE S.2
Empirical mean and variance of $\overline{G}_n(f_i)$, $i = 1, 2, 3$ from 5000 replications. Theoretical mean and variance are 0 and 1, respectively. Dimension $p = n^{2.5}$.

	$\mathbf{\Sigma}_p = \mathbf{\Sigma}_A$		$\mathbf{\Sigma}_p = \mathbf{\Sigma}_B$		$\mathbf{\Sigma}_p = \mathbf{\Sigma}_C$		$\mathbf{\Sigma}_p = \mathbf{\Sigma}_D$	
n	mean	var	mean	var	mean	var	mean	var
50	-0.0024	1.0064	0.0087	0.9873	0.0008	1.0101	-0.0063	0.9999
100	0.0021	1.0039	0.0242	0.9834	-0.0185	0.9992	-0.023	0.9877
150	-0.0067	1.0312	0.0208	0.9798	0.0191	0.9923	0.0068	0.9977
200	0.0081	0.9752	-0.0271	0.9767	-0.0012	0.9817	0.0042	0.9924
$\overline{G}_n(f_1)$	**Gaussian**							
50	0.0064	0.9928	-0.0172	1.0451	0.0064	1.0145	0.0248	1.0085
100	0.0204	0.9853	-0.0105	0.9678	0.0201	1.0295	-0.0036	1.0107
150	0.0156	1.0115	-0.0024	0.9977	0.0143	0.9766	-0.0002	1.0046
200	0.0091	0.9842	-0.0201	0.9863	-0.0087	1.0251	0.0107	0.9621
$\overline{G}_n(f_1)$	**Non-Gaussian**							
50	0.0036	1.0309	-0.0089	1.0246	-0.0024	1.0002	-0.0032	1.0283
100	-0.0101	0.9941	-0.002	1.0386	-0.0238	0.9857	0.0002	1.023
150	-0.0131	1.0129	0.0031	0.9589	0.0012	0.9781	0.0106	1.0162
200	0.0199	0.998	-0.0177	1.0273	-0.0151	1.0115	0.0132	1.0278
$\overline{G}_n(f_2)$	**Gaussian**							
50	-0.0077	1.1137	-0.0114	1.1008	0.0085	1.1179	0.0116	1.0816
100	0.016	1.022	0.0128	1.0405	-0.0093	1.0207	-0.0118	1.0573
150	-0.0159	1.0203	-0.016	1.0384	0.0174	1.0538	0.0067	0.9585
200	-0.0017	1.0158	-0.0159	1.0257	-0.0038	1.0273	0.0192	1.0463
$\overline{G}_n(f_2)$	**Non-Gaussian**							
50	0.0049	1.0541	0.0208	1.0843	0.0443	1.0284	0.0107	1.0174
100	0.0115	1.06	0.0475	1.0313	-0.0048	1.0689	0.0051	1.0027
150	0.0059	1.0479	0.0317	1.0151	0.0309	1.0301	0.0184	1.0596
200	0.0209	0.9741	-0.0142	1.0178	-0.0019	0.9765	0.0304	1.0195
$\overline{G}_n(f_3)$	**Gaussian**							
50	0.0511	1.129	0.0423	1.1733	0.0584	1.1513	0.0861	1.1456
100	0.0485	1.0639	0.0198	1.0319	0.0488	1.0817	0.0252	1.0579
150	0.0288	1.0518	0.0189	1.0393	0.032	1.0263	0.0348	1.1
200	0.0128	1.0134	0.0055	1.0179	0.0215	1.0218	0.0386	0.9725
$\overline{G}_n(f_3)$	**Non-Gaussian**							

REFERENCES

BAI, Z. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability* **32**. https://doi.org/10.1214/aop/1078415845

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral analysis of large dimensional random matrices*, 2nd ed. Springer.

BAO, Z. (2015). On asymptotic expansion and central limit theorem of linear eigenvalue statistics for sample covariance matrices when $N/M \to 0$. *Theory of Probability & Its Applications* **59** 185–207. https://doi.org/10.1137/s0040585x97t987089

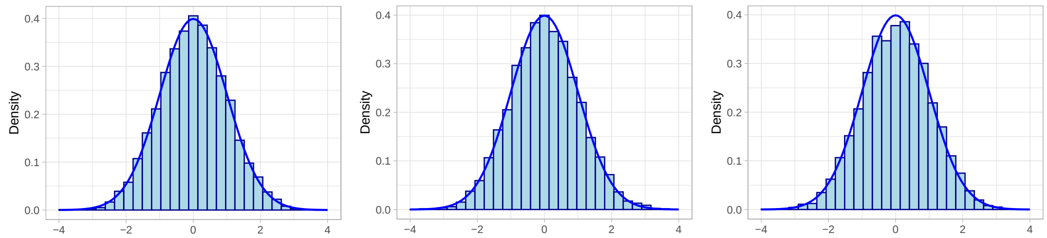
BILLINGSLEY, P. (1968). *Convergence of probability measures*. New York: Wiley.

BILLINGSLEY, P. (2008). *Probability and measure*. John Wiley & Sons.

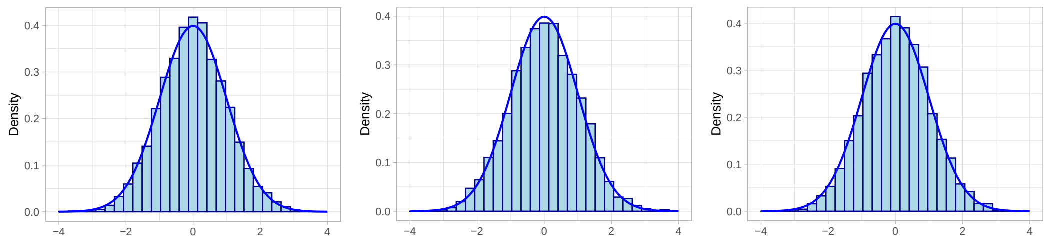
BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *The Annals of Probability* **1** 19–42. https://doi.org/10.1214/aop/1176997023

CHEN, B. and PAN, G. (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**. https://doi.org/10.3150/11-bej381

CHEN, B. and PAN, G. (2015). CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli* **21** 1089–1133. https://doi.org/10.3150/14-bej599



(a) $f_1 = x$, Gaussian data (b) $f_2 = x^2$, Gaussian data (c) $f_3 = x^3$, Gaussian data



(d) $f_1 = x$, non-Gaussian data (e) $f_2 = x^2$, non-Gaussian data (f) $f_3 = x^3$, non-Gaussian data

Fig S.2: Histograms of $\overline{G}_n(f_i), i = 1, 2, 3$ from 5000 replications under the case (D) with $(p, n) = (200^{2.5}, 200)$. The curves are density functions of standard normal distribution.

KHORUNZHY, A. M., KHORUZHENKO, B. A. and PASTUR, L. A. (1996). Asymptotic properties of large random matrices with independent entries. *Journal of Mathematical Physics* **37** 5033–5060. https://doi.org/10.1063/1.531589

LI, Z. and YAO, J. (2016). Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size. *Electronic Journal of Statistics* **10** 2973–3010. https://doi.org/10.1214/16-ejs1199

PAN, G. and ZHOU, W. (2011). Central limit theorem for Hotelling's T^2 statistic under large dimension. *The Annals of Applied Probability* **21** 1860–1910. https://doi.org/10.1214/10-aap742

WANG, Q. and YAO, J. (2013). On the sphericity test with large-dimensional observations. *Electronic Journal of Statistics* **7** 2164–2192. https://doi.org/10.1214/13-ejs842