The Annals of Statistics
2015, Vol. 43, No. 2, 546–591
DOI: 10.1214/14-AOS1292

SUBSTITUTION PRINCIPLE FOR CLT OF LINEAR SPECTRAL STATISTICS OF HIGH-DIMENSIONAL SAMPLE COVARIANCE MATRICES WITH APPLICATIONS TO HYPOTHESIS TESTING

BY SHURONG ZHENG[1], ZHIDONG BAI[2] AND JIANFENG YAO[3]

Northeast Normal University, Northeast Normal University and University of Hong Kong

Sample covariance matrices are widely used in multivariate statistical analysis. The central limit theorems (CLTs) for linear spectral statistics of high-dimensional noncentralized sample covariance matrices have received considerable attention in random matrix theory and have been applied to many high-dimensional statistical problems. However, known population mean vectors are assumed for noncentralized sample covariance matrices, some of which even assume Gaussian-like moment conditions. In fact, there are still another two most frequently used sample covariance matrices: the ME (moment estimator, constructed by subtracting the sample mean vector from each sample vector) and the unbiased sample covariance matrix (by changing the denominator n as $N = n - 1$ in the ME) without depending on unknown population mean vectors. In this paper, we not only establish the new CLTs for noncentralized sample covariance matrices when the Gaussian-like moment conditions do not hold but also characterize the nonnegligible differences among the CLTs for the three classes of high-dimensional sample covariance matrices by establishing a *substitution principle*: by substituting the *adjusted* sample size $N = n - 1$ for the actual sample size n in the centering term of the new CLTs, we obtain the CLT of the unbiased sample covariance matrices. Moreover, it is found that the difference between the CLTs for the ME and unbiased sample covariance matrix is nonnegligible in the centering term although the only difference between two sample covariance matrices is a normalization by n and $n - 1$, respectively. The new results are applied to two testing problems for high-dimensional covariance matrices.

1. Introduction. Consider a sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ of size n from a p-dimensional population $\mathbf{x}$ with unknown mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The

Received December 2013; revised July 2014.
[1]Supported in part by NSFC Grants NSFC-11171058 and NECT-11-0616.
[2]Supported in part by NSFC Grant NSFC-11171057, PCSIRT and Fundamental Research Funds for the Central Universities.
[3]Supported in part by a GRF Grant 705413.
MSC2010 subject classifications. 62H15, 15B52, 62H10.
Key words and phrases. CLT for linear spectral statistics, unbiased sample covariance matrix, substitution principle, testing on high-dimensional covariance matrix, high-dimensional sample covariance matrix, large Fisher matrix, high-dimensional data.

unbiased sample covariance matrix is defined by

$$\mathbf{S}_n = \frac{1}{N}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^*, \tag{1.1}$$

where $\overline{\mathbf{x}} = \frac{1}{n}\sum_j \mathbf{x}_j$ is the sample mean, $N = n - 1$ the *adjusted sample size* and $*$ denotes transpose and complex conjugate. Sample covariance matrices are widely applied in multivariate statistical analysis. For instance, in structure testing problems of population covariance matrices $\mathbf{\Sigma}$, many well-known test statistics are functionals of the eigenvalues $\{\lambda_j, 1 \le j \le p\}$ of $\mathbf{S}_n$ that have the form

$$T = \frac{1}{p}\sum_{j=1}^{p} g(\lambda_j) = \mu_{\mathbf{S}_n}(g), \tag{1.2}$$

for some given function g. Such statistics are referred hereafter as *linear spectral statistics* (LSS) of the unbiased sample covariance matrix $\mathbf{S}_n$. For example, the log-likelihood ratio statistic for testing the identity hypothesis for a Gaussian population covariance matrix is proportional to $\mu_{\mathbf{S}_n}(g)$ with $g(\lambda) = \lambda - 1 - \log\lambda$ (see Section 4 for more details). John's test for the sphericity hypothesis "$\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$" ($\sigma^2$ unspecified) uses the square of the coefficient of variation of the sample eigenvalues

$$U_n = \frac{p^{-1}\sum_{j=1}^{p}(\lambda_j - \overline{\lambda})^2}{\overline{\lambda}^2},$$

where $\overline{\lambda} = p^{-1}\sum_j \lambda_j = \mu_{\mathbf{S}_n}(\lambda)$. Clearly, U_n is a function of two linear spectral statistics $\mu_{\mathbf{S}_n}(\lambda^2)$ and $\mu_{\mathbf{S}_n}(\lambda)$ (see, e.g., [22] for more details on this test). Therefore, LSS $\mu_{\mathbf{S}_n}(g)$ of the sample covariance matrix $\mathbf{S}_n$ are important in multivariate analysis.

When the dimension p is much less than the sample size n, or equivalently, the RDS (*ratio of dimension-to-sample size*) p/n is close to zero, classical large sample theory assesses that once $\mathbb{E}\|\mathbf{x}\|^4 < \infty$, the sample covariance matrix $\mathbf{S}_n$ is a consistent and asymptotic normal estimator of $\mathbf{\Sigma}$. Consequently, the same also holds for the sample eigenvalues $\{\lambda_j, j = 1, \ldots, p\}$ as an estimator of the population eigenvalues of $\mathbf{\Sigma}$. Therefore, for any smooth function g,

$$\sqrt{n}\{\mu_{\mathbf{S}_n}(g) - \mu_{\mathbf{\Sigma}}(g)\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2), \tag{1.3}$$

where the asymptotic variance s^2 is a function of $\mathbf{\Sigma}$ and g. Here and throughout the paper, $\mu_{\mathbf{A}} = p^{-1}\sum_i \delta_{\{\alpha_i\}}$ denotes the *empirical spectral distribution* (ESD) generated by the eigenvalues $\{\alpha_i, 1 \le i \le p\}$ of a matrix $\mathbf{A}$, so that for a given function g, $\mu_{\mathbf{A}}(g) = \frac{1}{p}\sum_i g(\alpha_i)$.

High-dimensional statistics have emerged in recent years as an important and active research area. Applications have been found in various fields such as genomic data analysis and wireless communications. Typically in these problems, the RDS p/n is no more close to zero and the above large sample theory (1.3) fails to provide meaningful inference procedures. Many efforts have been put in finding new procedures to deal with high-dimensional data. As an example, the inconsistency of $\mathbf{S}_n$ as an estimator of $\boldsymbol{\Sigma}$ has lead to an abundant literature on covariance matrix estimation (see, e.g., [7, 8, 10] and the references therein).

This paper is concerned with asymptotics of LSS $\mu_{\mathbf{S}_n}(g)$. An interesting question is what is the CLT replacing (1.3) in the high-dimensional context. Notice that it remains challenging to transform the above-mentioned results on covariance matrix estimation to limit theorem on LSS of interest. It turns out that when both the dimension p and the sample size n grow to infinity, limit theory for sample eigenvalues depend on how the RDS p/n behaves asymptotically. In this paper, we adopt the so-called *Marčenko–Pastur scheme* where it is assumed that

$$RDS = p/n \to y \in (0, \infty) \qquad \text{as } n \to \infty.$$

It has been demonstrated that such limiting scheme has a wide application scope for real-life high-dimensional data analysis [13].

The seminal paper [5] establishes such a CLT for the population with population mean $\boldsymbol{\mu} = 0$ (or equivalently, $\boldsymbol{\mu}$ is known and data can then be dealt with by subtracting $\boldsymbol{\mu}$) and Gaussian-like moment conditions (the population second-order and fourth-order moments are the same as those of real or complex Gaussian population), and the noncentralized sample covariance matrix is defined as

$$\mathbf{S}_n^0 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^0 \mathbf{x}_i^{0*}. \tag{1.4}$$

(The superscript 0 is here to remind the fact that the population $\mathbf{x}^0$ has a zero mean.) Let $y_n = p/n$ and $H_p = \mu_{\boldsymbol{\Sigma}}$ be the population ESD of $\boldsymbol{\Sigma}$. As $p, n \to \infty$, it is assumed that the ratio $y_n \to y \in (0, \infty)$ and $H_p \to H$ (weakly) for some probability distribution H. Then the ESD $\mu_{\mathbf{S}_n^0}$ converges to a nonrandom distribution $F^{y,H}$, called *limiting spectral distribution* (LSD), which depends on y and the population limiting distribution H. This LSD is referred as the *generalized Marčenko–Pastur distribution* with index (y, H) (for background on Marčenko–Pastur distributions, the reader is referred to [3], Chapter 3). Therefore, in the simplest form, the CLT in [5] states that

$$p\{\mu_{\mathbf{S}_n^0}(g) - F^{y_n,H_p}(g)\} \xrightarrow{\mathcal{D}} \mathcal{N}(m(g), v(g)), \tag{1.5}$$

a Gaussian distribution whose parameters $m(g)$ and $v(g)$ depend only on the LSD $F^{y,H}$ and g. The crucial issue here is that the centering term

$$p \cdot F^{y_n,H_p}(g) = p \int g(x)\, dF^{y_n,H_p}(x),$$

uses a finite-horizon proxy F^{y_n,H_p} of the LSD $F^{y,H}$ obtained by substituting, in the LSD, the current RDS $y_n = p/n$ and the population ESD H_p for their limits y

and H, respectively. Since p and n have a same order, any mis-estimation of order n^{-1} in $F^{y_n,H_p}(g)$ will affect the asymptotic mean $m(g)$.

This scenario of populations with a known mean $\boldsymbol{\mu}$, is however a bit too ideal and real-life data analyses rely on the unbiased sample covariance matrix $\mathbf{S}_n$ (1.1) after subtraction of the sample mean. It has been believed for a while in the literature in high-dimensional statistics that both sample covariance matrices $\mathbf{S}_n$ and $\mathbf{S}_n^0$ share a same CLT for their LSS, that is, the CLT (1.5) might apply equally to the matrix $\mathbf{S}_n$. Unfortunately, this is indeed untrue. The problem can be best explained by observing the Gaussian case. Actually, for a Gaussian population,

$$N\mathbf{S}_n := \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^*$$

has a Wishart distribution $\mathbf{W}_N(\boldsymbol{\Sigma})$ with $N = n - 1$ degrees of freedom. Since for a Gaussian population with known population mean the matrix $N\mathbf{S}_N^0 = \sum_{i=1}^{N} \mathbf{x}_i^0 \mathbf{x}^{0*}$ has the same Wishart distribution, we conclude that the fluctuations of the eigenvalues $\{\lambda_j\}$ of $\mathbf{S}_n$ are the same as the matrix $\mathbf{S}_N^0$ so that by (1.5), it holds

$$p\{\mu_{\mathbf{S}_n}(g) - F^{y_N,H_p}(g)\} \xrightarrow{\mathcal{D}} \mathcal{N}\big(m(g), v(g)\big). \tag{1.6}$$

In words, in the Gaussian case, the CLT for populations with unknown means is the same as the CLT for populations with known means provided that in the centering term $pF^{y_n,H_p}(g)$, one substitutes the adjusted sample size $N = n - 1$ for the sample size n. This result will be named hereafter as the *substitution principle*. Notice that typically the difference between $F^{y_N,H_p}(g)$ and $F^{y_n,H_p}(g)$ is of order $O(n^{-1})$ and as explained above, such a difference is nonnegligible because of the multiplication by p in the CLT. As an example, when $\boldsymbol{\Sigma} = \mathbf{I}_p$ we have $H_p = \delta_1$ and for $g(\lambda) = \lambda^2$, it is well known that $F^{y_n,\delta_1}(g) = 1 + y_n$. Therefore, the difference

$$p\{F^{y_n,H_p}(g) - F^{y_N,H_p}(g)\} = p(y_n - y_N)$$

tends to $-y^2$, a nonnegligible negative constant.

This substitution principle is indeed a remarkable result and provides an elegant solution to the question of CLT for LSS of the unbiased covariance matrix $\mathbf{S}_n$ from a Gaussian population. It then raises the question whether the principle is universal, that is, valid for general populations other than Gaussian. One of the main results from the paper establishes this universality for arbitrary populations provided the existence of a fourth-order moment. Meanwhile, most of the existing methods in hypothesis testing or regression analysis with high-dimensional data assume either Gaussian-like moment conditions or populations with known means (see, e.g., [1, 2, 9, 14, 20, 21]), so that LSS of the sample covariance matrices are approximated using either the CLT (1.6) or the CLT (1.5). The universality of the substitution principle established in this paper for these CLTs will then help the existing methods to cover more general high-dimensional data. Consider the ME of $\boldsymbol{\Sigma}$

$$\hat{\boldsymbol{\Sigma}}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^*. \tag{1.7}$$

By the decomposition

$$\mu_{\hat{\boldsymbol{\Sigma}}_n}(g) - F^{y_N,H_p}(g) = \big(\mu_{\mathbf{S}_n}(g) - F^{y_N,H_p}(g)\big) + \big(\mu_{\hat{\boldsymbol{\Sigma}}_n}(g) - \mu_{\mathbf{S}_n}(g)\big),$$

the CLT (1.6) established in this paper and the fact $\hat{\boldsymbol{\Sigma}}_n = (1-1/n)\mathbf{S}_n$, it readily follows that

$$\begin{aligned} m_1(g) &= p\big(\mu_{\hat{\boldsymbol{\Sigma}}_n}(g) - \mu_{\mathbf{S}_n}(g)\big) \\ &= \sum_{i=1}^{p}\big\{g\big((1-/n)\lambda_i\big) - g(\lambda_i)\big\} \to -yF^{y,H}\big(\lambda g'(\lambda)\big). \end{aligned}$$

That is,

$$p\big\{\mu_{\hat{\boldsymbol{\Sigma}}_n}(g) - F^{y_N,H_p}(g)\big\} + m_1(g) \xrightarrow{\mathcal{D}} \mathcal{N}\big(m(g), v(g)\big), \tag{1.8}$$

which shows that the difference of CLTs between the ME and the unbiased sample covariance is nonnegligible, and that the CLT for LSS $\mu_{\hat{\boldsymbol{\Sigma}}_n}(g)$ for the ME $\hat{\boldsymbol{\Sigma}}_n$ can be seen as a direct consequence of the substitution principle (1.6) established in this paper. Another major contribution of the paper is to establish a new CLT for LSS for $\mathbf{S}_n^0$ when the Gaussian-like moment conditions are not met. In a related work, [17] removes the Gaussian-like fourth-order moment condition, but their assumptions of replacement, made on both the population covariance matrices $\boldsymbol{\Sigma}_x$ and the Stieltjes transform of the LSD $F^{y,H}$, are not easy to verify in applications. The new CLT of this paper removes the Gaussian-like second-order and fourth-order moment condition restrictions and the given conditions are not only easy to satisfy but also are unremovable as demonstrated by three counterexamples given in the Appendix.

We next address the same problems for the class of Fisher matrices. From now on, for the sample $\mathbf{x}_i$'s we will use the notation $\boldsymbol{\Sigma}_x$ and $\mathbf{S}_x$ for its population and sample covariance matrices, respectively. Consider another sample $\mathbf{y}_1, \dots, \mathbf{y}_m$ of size m from a p-dimensional population $\mathbf{y}$ with mean $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Sigma}_y$. The corresponding unbiased sample covariance matrix is

$$\mathbf{S}_y = \frac{1}{M}\sum_{j=1}^{m}(\mathbf{y}_j - \overline{\mathbf{y}})(\mathbf{y}_j - \overline{\mathbf{y}})^*, \tag{1.9}$$

where $\overline{\mathbf{y}} = \frac{1}{m}\sum_j \mathbf{y}_j$ is the sample mean and $M = m-1$ the adjusted sample size. The so-called Fisher matrix

$$\mathbf{F} = \mathbf{S}_x\mathbf{S}_y^{-1}$$

is a natural statistic for the two-sample test of the hypothesis "$\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$" that the populations have the same covariance matrix. The CLT for LSS $\mu_{\mathbf{F}}(g)$ of $\mathbf{F}$ has been established in [23] assuming that both populations have zero means, that is, $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$ and standardized, that is, $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = \mathbf{I}_p$. While keeping the standardization assumption but dropping the condition $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$, we prove a similar substitution principle: the CLT fr LSS of $\mathbf{F} = \mathbf{S}_x\mathbf{S}_y^{-1}$ with arbitrary population means and population distribution (provided that a fourth-moment exists) is the same as the CLT in [23] for populations with known means provided that one substitutes the adjusted sample sizes

$$(N, M) = (n-1, m-1)$$

for the sample sizes (n, m) in the centering term of the CLT in [23]. This second substitution principle can be viewed as a consequence of the first substitution principle for sample covariance matrices.

They have been other proposals in the literature for testing hypotheses about high-dimensional covariance matrices. In particular, procedures are proposed in [12, 15] using a family of well-chosen U-statistics and the asymptotic theory of these procedures does not require that p/n tends to a positive limit. In another perspective, a minimax analysis for the one-sample identity test $\mathbf{\Sigma}_x = \mathbf{I}_p$ has been recently proposed in [11]. All these proposals are however not directly linked to the substitutions principles discussed in this paper since they do not rely on LSS $\mu_{\mathbf{S}_n}(g)$ or $\mu_{\mathbf{F}}(g)$ studied in this paper.

The main results of the paper, the two substitution principles and the new CLT are presented in Sections 2 and 3. To demonstrate the importance of these principles, we develop in Section 4 new procedures for hypothesis testing about high-dimensional covariance matrices extending previous results to cover general Gaussian or non-Gaussian populations with unknown populations means. Technical proofs are relegated to Section 5.

2. Substitution principle for the unbiased sample covariance matrix $\mathbf{S}_x$. Before introducing the first substitution principle, we give a new CLT of LSS for noncentralized sample covariance matrix whenever the Gaussian-like moment conditions hold or not.

Assumption (a). Samples are $\{\mathbf{x}_j = \boldsymbol{\mu} + \mathbf{\Gamma}\mathbf{X}_j, j = 1, \dots, n\}$ where $\mathbf{X}_j = (X_{1j}, \dots, X_{pj})^T$. For each p, $\{X_{ij}, i \le p, j \le n\}$ are independent random variables with common moments

$$EX_{ij} = 0, \qquad E|X_{ij}|^2 = 1,$$
$$\beta_x = E|X_{ij}|^4 - |EX_{11}^2|^2 - 2, \qquad \alpha_x = |EX_{11}^2|^2$$

and satisfying the following Lindeberg condition:

$$\frac{1}{np}\sum_{j=1}^{p}\sum_{k=1}^{n} E\{|X_{jk}|^4 \mathbf{1}_{\{|X_{jk}| \ge \eta\sqrt{n}\}}\} \to 0 \qquad \text{for any fixed } \eta > 0,$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function.

Assumption (b). The RDS $y_n = p/n$ tends to a positive $y > 0$ as $n, p \to \infty$.

Assumption (c). The sequence of $\{\boldsymbol{\Sigma}_x = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^*\}_{p\geq 1}$ is bounded in spectral norm and the ESD $H_p = \mu_{\boldsymbol{\Sigma}_x}$ of $\boldsymbol{\Sigma}_x$ converges weakly to a LSD H as $p \to \infty$.

Assumption (d1). $\boldsymbol{\Gamma}$ is real or the variables X_{ij} are complex satisfying $\alpha_x = 0$.

Assumption (d2). $\boldsymbol{\Gamma}^*\boldsymbol{\Gamma}$ is diagonal or $\beta_x = 0$.

In fact, assumption (d1) is for the second-order moment condition of X_{ij} and assumption (d2) is for the fourth-order moment condition of X_{ij}. Assumption (d2) can be interpreted as follows: suppose the singular decomposition of $\boldsymbol{\Gamma}$ is $\boldsymbol{\Gamma} = \mathbf{U}^*\mathbf{L}^{1/2}\mathbf{V}$, then $\boldsymbol{\Sigma}_x = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^* = \mathbf{U}^*\mathbf{L}\mathbf{U}$ where $\mathbf{L}$ is the diagonal matrix formed by eigenvalues and $\mathbf{U}^*$ by eigenvectors of $\boldsymbol{\Sigma}_x$. Then we can see that $\boldsymbol{\Gamma}^*\boldsymbol{\Gamma} = \mathbf{V}^*\mathbf{L}\mathbf{V}$ is diagonal if the unitary matrix $\mathbf{V}$ is an identity. It is the case especially that $\mathbf{V}\mathbf{Y}$ has the same distribution as $\mathbf{Y}$. So it shows that assumption (d2) is easy to satisfy.

Write $\mathbf{x}_j = \boldsymbol{\mu} + \mathbf{x}_j^0$ with $\mathbf{x}_j^0 = \boldsymbol{\Gamma}\mathbf{X}_j$ and define the corresponding noncentralized sample covariance matrix as

$$\mathbf{S}_x^0 = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j^0 \mathbf{x}_j^{0*}. \tag{2.1}$$

Under assumptions (a)–(b)–(c), it is well known that both the unbiased sample covariance matrix $\mathbf{S}_x$ and noncentralized sample covariance matrix $\mathbf{S}_x^0$ have the same LSD $F^{y,H}$, namely the Marčenko–Pastur distribution of index (y, H). We recall some useful facts about these distributions (see [3] for details). The LSD has support

$$[a, b] = \Big[(1-\sqrt{y})^2 I_{(0<y<1)} \liminf_n \lambda_{\min}^{\boldsymbol{\Sigma}_x}, (1+\sqrt{y})^2 \limsup_n \lambda_{\max}^{\boldsymbol{\Sigma}_x}\Big], \tag{2.2}$$

where it has a density function. Moreover, the LSD $F^{y,H}$ has a Dirac mass $1 - 1/y$ at the origin when $y > 1$. Define $\underline{m}_y$ to be the Stieltjes transform of the companion LSD

$$\underline{F}^{y,H} = (1-y)\delta_0 + yF^{y,H},$$

where δ_0 is the point distribution at zero. Then $\underline{m}_y$ is the unique solution in

$$\frac{1-y}{z} + \underline{m}_y \in \mathbb{C}^+ = \{z : \Im(z) > 0\}$$

of the equation

$$z = -\frac{1}{\underline{m}_y(z)} + y\int \frac{t\,dH(t)}{1 + t\underline{m}_y(z)}, \qquad z \in \mathbb{C}^+ = \{z : \Im(z) > 0\}. \tag{2.3}$$

Notice that when a finite-horizon proxy F^{y_n,H_p} is substituted for the LSD $F^{y,H}$, these properties and relationships hold with the parameters (y, H) replaced by (y_n, H_p).

For Gaussian-like moment conditions $\beta_x = 0$ or $\alpha_x = 0$ for complex population, the CLT (1.5) for LSS of the noncentralized sample covariance matrix $\mathbf{S}_x^0$ has been established first in [5] where the explicit limiting mean and covariance functions are given. However, this result has a limitation in that it requires Gaussian-like moment conditions, that is, $\beta_x = 0$, $\alpha_x = 0$ for complex population. There have been many efforts in the literature for removing this restriction; see [16] and [18]. The CLT in [18] removes the Gaussian-like fourth-order condition $\beta_x = 0$. However, their assumptions of replacement, made on both the population covariance matrices $\mathbf{\Sigma}_x$ and the Stieltjes transform of the LSD $F^{y,H}$, are not easy to verify in applications. Moreover, it is of practical importance to remove the Gaussian-like second condition but rare literature has mentioned it. In this section, we propose a new CLT under assumptions (d2) and (d1) without assuming these Gaussian-like moment conditions made in [5]. Three counterexamples are provided in the Appendix to show that these assumptions (d1) and (d2) cannot be removed for a general CLT for LSS of the sample covariance matrix $\mathbf{S}_x^0$.

THEOREM 2.1. *Assume that all assumptions* (a)–(b)–(c)–(d1)–(d2) *hold. Let* $f_1, \dots, f_k$ *be functions analytic on an open domain of the complex plan containing the support of the LSD* $F^{y,H}$ *and define*

$$X_p(f_\ell) = p\{\mu_{\mathbf{S}_x^0}(f_\ell) - pF^{y_n,H_p}(f_\ell)\} = \sum_{i=1}^{p} f_\ell(\lambda_i^0) - pF^{y_n,H_p}(f_\ell), \tag{2.4}$$

where $\{\lambda_i^0\}_{i=1}^p$ *are eigenvalues of* $\mathbf{S}_x^0$ *and*

$$F^{y_n,H_p}(f_\ell) = \int f_\ell(x)\, dF^{y_n,H_p}(x).$$

Then the random vector $(X_p(f_1), \dots, X_p(f_k))$ *converges to a* k*-dimensional Gaussian random vector* $(X_{f_1}, \dots, X_{f_k})$ *with mean function*

$$\begin{aligned}
\mathrm{E}X_{f_\ell} = -\frac{\alpha_x}{2\pi\mathbf{i}} \oint_{\mathcal{C}} f_\ell(z) & \left(y \int \underline{m}_y^3(z) t^2 \left(1 + t\underline{m}_y(z)\right)^{-3} dH(t) \right) \\
& \Big/ \left(\left(1 - y \int \frac{\underline{m}_y^2(z) t^2}{(1 + t\underline{m}_y(z))^2}\, dH(t) \right) \right. \\
& \quad \left. \times \left(1 - \alpha_x y \int \frac{\underline{m}_y^2(z) t^2}{(1 + t\underline{m}_y(z))^2}\, dH(t) \right) \right) dz \\
& - \frac{\beta_x}{2\pi\mathbf{i}} \cdot \oint_{\mathcal{C}} f_\ell(z) \frac{y \int \underline{m}_y^3(z) t^2 (1 + t\underline{m}_y(z))^{-3}\, dH(t)}{1 - y \int \underline{m}_y^2(z) t^2 (1 + t\underline{m}_y(z))^{-2}\, dH(t)}\, dz,
\end{aligned}$$

and variance–covariance function

$$
\begin{aligned}
&\operatorname{Cov}(X_{f_j}, X_{f_\ell}) \\
&\quad = -\frac{1}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{f_j(z_1) f_\ell(z_2)}{(\underline{m}_y(z_1) - \underline{m}_y(z_2))^2} \, d\underline{m}_y(z_1) \, d\underline{m}_y(z_2) \\
&\qquad - \frac{y\beta_x}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} f_j(z_1) f_\ell(z_2) \biggl[\int \frac{t}{(\underline{m}_y(z_1)t + 1)^2} \\
&\qquad\qquad \times \frac{t}{(\underline{m}_y(z_2)t + 1)^2} \, dH(t) \biggr] d\underline{m}_y(z_1) \, d\underline{m}_y(z_2) \\
&\qquad - \frac{1}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} f_j(z_1) f_\ell(z_2) \biggl[\frac{\partial^2}{\partial z_1 \, \partial z_2} \log\bigl(1 - a(z_1, z_2)\bigr) \biggr] dz_1 \, dz_2,
\end{aligned} \tag{2.5}
$$

where $\mathcal{C}$, $\mathcal{C}_1$ and $\mathcal{C}_2$ are closed contours in the complex plan enclosing the support of the LSD $F^{y,H}$, and $\mathcal{C}_1$ and $\mathcal{C}_2$ being nonoverlapping. Finally, the function $a(z_1, z_2)$ is

$$
a(z_1, z_2) = \alpha_x \biggl(1 + \frac{\underline{m}_y(z_1) \underline{m}_y(z_2)(z_1 - z_2)}{\underline{m}_y(z_2) - \underline{m}_y(z_1)} \biggr).
$$

The proof of this refinement is given in Section 5.3.3. Moreover, as said earlier, new assumptions (d1) and (d2) are used as a replacement and they will be proven to be necessary by examples shown in the Appendix of the paper. The major advantage of this CLT is that the fourth-order and second-order population moments can be arbitrary instead of matching Gaussian-like populations, that is, the parameters β_x and α_x may be nonzero.

When the Gaussian-like second-order moment condition ($\alpha_x = 1$ for real $\{X_{ij}\}$ and $\alpha_x = 0$ for complex $\{X_{ij}\}$) holds, it can be easily checked that the previous limiting mean and variance–covariance functions reduce to

$$
\begin{aligned}
\mathrm{E} X_{f_\ell} &= -\frac{\alpha_x}{2\pi \mathbf{i}} \oint_{\mathcal{C}} f_\ell(z) \frac{y \int \underline{m}_y^3(z) t^2 (1 + t\underline{m}_y(z))^{-3} \, dH(t)}{[1 - y \int \underline{m}_y^2(z) t^2 (1 + t\underline{m}_y(z))^{-2} \, dH(t)]^2} \, dz \\
&\quad - \frac{\beta_x}{2\pi \mathbf{i}} \cdot \oint_{\mathcal{C}} f_\ell(z) \frac{y \int \underline{m}_y^3(z) t^2 (1 + t\underline{m}_y(z))^{-3} \, dH(t)}{1 - y \int \underline{m}_y^2(z) t^2 (1 + t\underline{m}_y(z))^{-2} \, dH(t)} \, dz,
\end{aligned} \tag{2.6}
$$

and variance function

$$
\begin{aligned}
&\operatorname{Cov}(X_{f_j}, X_{f_\ell}) \\
&\quad = -\frac{\alpha_x + 1}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{f_j(z_1) f_\ell(z_2)}{(\underline{m}_y(z_1) - \underline{m}_y(z_2))^2} \, d\underline{m}_y(z_1) \, d\underline{m}_y(z_2) \\
&\qquad - \frac{y\beta_x}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} f_j(z_1) f_\ell(z_2)
\end{aligned} \tag{2.7}
$$

$$\times \left[\int \frac{t}{(\underline{m}_y(z_1)t+1)^2} \times \frac{t}{(\underline{m}_y(z_2)t+1)^2}\, dH(t) \right] d\underline{m}_y(z_1)\, d\underline{m}_y(z_2).$$

In particular, under Gaussian-like second-order and fourth-order moment conditions, we recover the CLT (1.5) of [5].

Coming to the unbiased sample covariance matrix $\mathbf{S}_x$ with unknown population means, as a second main result of the paper, we establish the following substitution principle. Recall that $N = n - 1$ denotes the adjusted sample size.

THEOREM 2.2 (One sample substitution principle). *Under the same conditions as in Theorem 2.1, define*

$$Y_p(f_\ell) = p\{\mu_{\mathbf{S}_x}(f_\ell) - F^{y_N, H_p}(f_\ell)\} = \sum_{i=1}^{p} f_\ell(\lambda_i) - pF^{y_N, H_p}(f_\ell), \tag{2.8}$$

where $\{\lambda_i\}_{i=1}^p$ are the eigenvalues of the unbiased sample covariance matrix $\mathbf{S}_x$ and $N = n - 1$. Then the random vector $(Y_p(f_1), \dots, Y_p(f_k))$ converges in distribution to the same Gaussian vector $(X_{f_1}, \dots, X_{f_k})$ given in Theorem 2.1.

The proof of Theorem 2.2 is postponed to Section 5.3.

3. Substitution principle for the two-sample Fisher matrix. In this section, we investigate the effect in the CLT for LSS of $\mathbf{F} = \mathbf{S}_x\mathbf{S}_y^{-1}$ when the unbiased covariance matrices $\mathbf{S}_x$ and $\mathbf{S}_y$ are used. The following assumptions for the second sample $\mathbf{y}_1, \dots, \mathbf{y}_m$ mimic assumptions (a)–(b)–(c) set for the first sample $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Assumption (a′). Samples are $\{\mathbf{y}_j = \boldsymbol{\nu} + \boldsymbol{\Gamma}_y\mathbf{Y}_j, j = 1, \dots, m\}$ where $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{pj})^T$. For each p, the elements of the data matrix $\{Y_{ij}, i \le p, j \le m\}$ are independent random variables with common moments

$$EY_{ij} = 0, \qquad E|Y_{ij}|^2 = 1, \qquad \beta_y = E|Y_{ij}|^4 - |EY_{11}^2|^2 - 2,$$

especially $\mathbb{E}Y_{ij}^2 = 0$ in complex case and satisfying the following Lindeberg condition:

$$\frac{1}{mp}\sum_{j=1}^{p}\sum_{k=1}^{m} E\{|Y_{jk}|^4 \mathbf{1}_{\{|Y_{jk}| \ge \eta\sqrt{m}\}}\} \to 0 \qquad \text{for any fixed } \eta > 0.$$

Assumption (b′). The RDS ratio $y_m = p/m \to y_2 \in (0, 1)$ as $m, p \to \infty$.

Assumption (c′). The sequence $\{\mathbf{\Sigma}_y = \mathbf{\Gamma}_y\mathbf{\Gamma}_y^*\}_{p\ge 1}$ is bounded in spectral norm and the ESD $H_{2,p}$ of $\mathbf{\Sigma}_y$ converges to a LSD H_2 as $p\to\infty$.

Regarding the distinction between real-valued and complex-valued populations, an indicator κ is used for both populations $\mathbf{x}$ and $\mathbf{y}$ since the mixed situation where one population is real-valued while the other is complex-valued is rarely realistic in applications. When $\{\mathbf{X}_j\}_{j=1}^n$ and $\{\mathbf{Y}_j\}_{j=1}^m$ are real, let $\kappa=2$. When $\{\mathbf{X}_j\}_{j=1}^n$ and $\{\mathbf{Y}_j\}_{j=1}^m$ are complex, let $\kappa=1$.

Consider first the noncentralized Fisher and sample covariance matrices

$$\mathbf{S}_x^0=\frac{1}{n}\sum_{j=1}^n \mathbf{X}_j\mathbf{X}_j^*,\qquad \mathbf{S}_y^0=\frac{1}{m}\sum_{j=1}^m \mathbf{Y}_j\mathbf{Y}_j^*,\qquad \mathbf{F}^0=\mathbf{S}_x^0\mathbf{S}_y^{0^{-1}},\tag{3.1}$$

where $\boldsymbol{\nu}=\mathbf{0}$. Assume that assumptions (a)–(b)–(c) and (a′)–(b′)–(c′) are fulfilled. In this section, if both populations are complex, we assume that the second moments are null, that is, $\mathbb{E}X_{ij}^2=\mathbb{E}Y_{ij}^2=0$. From now onward, for notation convenience, the limiting ratio $y=\lim p/n$ of the $\mathbf{x}$-sample is denoted by y_1. It is well known from random matrix theory that the ESD of $\mu_{\mathbf{F}^0}$ converges to a LSD $G_{(y_1,y_2)}$ with compact support [6, 19]. Moreover, let $f_1,\dots,f_k$ be analytic functions on an open set of the complex plan enclosing the support of $G_{(y_1,y_2)}$. Consider linear spectral statistics

$$Z_p(f_\ell)=p\{\mu_{\mathbf{F}^0}(f_\ell)-G_{(y_n,y_m)}(f_\ell)\},\tag{3.2}$$

where, similar to CLTs for sample covariance matrices, $G_{(y_n,y_m)}$ is a finite-horizon proxy for the LSD $G_{(y_1,y_2)}$ obtained by substituting the current ratios $(y_n,y_m)=(p/n,p/m)$ for their limits $(y_1,y_2)=\lim(p/n,p/m)$. Let $h=(y_1^2+y_2^2+y_1y_2)^{1/2}$. Then the CLT in [23] establishes that the random vector $(Z_p(f_1),\dots,Z_p(f_k))$ converges to a k-dimensional Gaussian vector $(Z_{f_1},\dots,Z_{f_k})$ with mean function

$$\begin{aligned}\mathrm{E}Z_{f_\ell}&=\lim_{r\downarrow 1}\frac{\kappa-1}{4\pi\mathbf{i}}\oint_{|\xi|=1} f_\ell\left(\frac{1+h^2+2h\Re(\xi)}{(1-y_2)^2}\right)\\&\qquad\times\left[\frac{1}{\xi-r^{-1}}+\frac{1}{\xi+r^{-1}}-\frac{2}{\xi+y_2/h}\right]d\xi\\&\quad+\frac{\beta_x y_1(1-y_2)^2}{2\pi\mathbf{i}\cdot h^2}\oint_{|\xi|=1} f_\ell\left(\frac{1+h^2+2h\Re(\xi)}{(1-y_2)^2}\right)\frac{1}{(\xi+y_2/h)^3}\,d\xi\\&\quad+\frac{\beta_y(1-y_2)}{4\pi\mathbf{i}}\oint_{|\xi|=1} f_\ell\left(\frac{1+h^2+2h\Re(\xi)}{(1-y_2)^2}\right)\frac{\xi^2-y_2/h^2}{(\xi+y_2/h)^2}\\&\qquad\times\left[\frac{1}{\xi-\sqrt{y_2}/h}+\frac{1}{\xi+\sqrt{y_2}/h}-\frac{2}{\xi+y_2/h}\right]d\xi,\end{aligned}\tag{3.3}$$

and covariance function

$$
\begin{aligned}
&\operatorname{Cov}(Z_{f_j}, Z_{f_\ell}) \\
&\quad = -\lim_{r\downarrow 1} \frac{\kappa}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \biggl(\biggl(f_j\biggl(\frac{1+h^2+2h\Re(\xi_1)}{(1-y_2)^2}\biggr) \\
&\qquad\qquad \times f_\ell\biggl(\frac{1+h^2+2h\Re(\xi_2)}{(1-y_2)^2}\biggr)\biggr) \\
&\qquad\qquad \Big/ (\xi_1 - r\xi_2)^2 \biggr)\, d\xi_1\, d\xi_2 \qquad (3.4) \\
&\qquad - \frac{(\beta_x y_1 + \beta_y y_2)(1-y_2)^2}{4\pi^2 h^2} \\
&\qquad \times \oint_{|\xi_1|=1} f_j\biggl(\frac{1+h^2+2h\Re(\xi_1)}{(1-y_2)^2}\biggr) \Big/ \biggl(\xi_1 + \frac{y_2}{h}\biggr)^2 d\xi_1 \\
&\qquad \times \oint_{|\xi_2|=1} f_\ell\biggl(\frac{1+h^2+2h\Re(\xi_2)}{(1-y_2)^2}\biggr) \Big/ \biggl(\xi_2 + \frac{y_2}{h}\biggr)^2 d\xi_2.
\end{aligned}
$$

For the Fisher matrix of interest $\mathbf{F} = \mathbf{S}_x\mathbf{S}_y^{-1}$ from populations with unknown population means and as the second main result of the paper, we establish the following substitution principle under an additional condition of equal covariance matrices.

THEOREM 3.1 (Two-sample substitution principle). *Assume that the assumptions* (a)–(b)–(c) *and* (a′)–(b′)–(c′) *are fulfilled with* $y_2 \in (0,1)$ *and that* $\mathbf{\Gamma}_x = \mathbf{\Gamma}_y$. *Let* $f_1, \dots, f_k$ *be functions analytic on an open domain of the complex plan enclosing the support of the LSD* $G_{(y_1,y_2)}$ *and define linear spectral statistics*

$$
W_p(f_\ell) = p\{\mu_{\mathbf{F}}(f_\ell) - G_{(y_N,y_M)}(f_\ell)\}, \qquad (3.5)
$$

where $N = n-1$ *and* $M = m-1$ *are the adjusted sample sizes*, $y_N = p/N$ *and* $y_M = p/M$. *Then the random vector* $(W_p(f_1), \dots, W_p(f_k))$ *converges to the same limiting* k*-dimensional Gaussian vector* $(Z_{f_1}, \dots, Z_{f_k})$ *defined in* [23] *with the mean and covariance functions* (3.3)–(3.4).

The proof of this theorem is given in Section 5.4.

4. Applications to hypothesis testing on large covariance matrices. This section is devoted to illustrate the importance of the substitution principles proposed in this paper. We consider the problem of testing hypotheses about large covariance matrices based on the unbiased sample covariance matrices when population means are to be estimated. In this manner, Sections 4.1 and 4.2 generalize the main results of [1] on the one-sample and two-sample likelihood ratio tests on large covariance matrices. The generalized test procedures apply for non-Gaussian populations with unknown population means. To our best knowledge, few procedures exist for such testing problems on large sample covariance matrices, two exceptions being [12, 15]; see also [11] on a minimax study for the identity test.

4.1. *Testing the hypothesis that $\mathbf{\Sigma}_x$ is equal to a given matrix.* Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a sample from a p-dimensional population with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}_x$. Consider first a one-sample test for the hypothesis $H_0 : \mathbf{\Sigma}_x = \mathbf{I}_p$ that a p-dimensional covariance matrix $\mathbf{\Sigma}_x$ equals the identity matrix. The *corrected likelihood ratio* test in [1] is developed by assuming that the population is Gaussian and $\boldsymbol{\mu} = 0$ (or equivalently, $\boldsymbol{\mu}$ is given). The test statistic equals

$$L^0 = \operatorname{tr} \mathbf{S}_x^0 - \log|\mathbf{S}_x^0| - p, \tag{4.1}$$

where $\mathbf{S}_x^0$ is the noncentralized sample covariance matrix given in (1.4). The following theorem is established in [1].

PROPOSITION 4.1 (Theorem 3.1 of [1]). *Assume that the population is real Gaussian with mean $\boldsymbol{\mu} = 0$ and covariance matrix $\mathbf{\Sigma}_x$, and the dimension p and the sample size tend to infinity such that $y_n = p/n \to y \in (0, 1)$. Then under H_0,*

$$\upsilon_n(g)^{-1/2}\bigl[L^0 - p \cdot F^{y_n}(g) - m_n(g)\bigr] \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \tag{4.2}$$

where F^{y_n} is the Marčenko–Pastur law of index y_n, $g(x) = x - \log x - 1$ and

$$\begin{aligned} F^{y_n}(g) &= 1 - \frac{y_n - 1}{y_n} \log(1 - y_n), \\ m_n(g) &= -\frac{\log(1 - y_n)}{2}, \\ \upsilon_n(g) &= -2\log(1 - y_n) - 2y_n. \end{aligned}$$

At the significance level α, the test will reject the null hypothesis if the statistic in (4.2) exceeds z_α, the upper $\alpha\%$ quantile of the standard Gaussian distribution. The test has been proved to have good powers against the inflation of the dimension p. To extend this result to general populations with the unknown population mean vector, we start by assuming that the population $\mathbf{x}$ fulfills assumptions (a)–(b)–(c) of Section 2. The corrected likelihood ratio test statistic (CLRT) is defined to be

$$L^* = \operatorname{tr} \mathbf{S}_x - \log|\mathbf{S}_x| - p, \tag{4.3}$$

where $\mathbf{S}_x$ is the unbiased sample covariance matrix given in (1.1).

THEOREM 4.1. *Assume that the population $\mathbf{x}$ fulfills assumptions* (a)–(b)–(c) *where $y_n = p/n \to y \in (0, 1)$. Then under the null hypothesis $H_0 : \mathbf{\Sigma}_x = \mathbf{I}_p$, for the* *unbiased sample covariance matrix* $\mathbf{S}_x$ *in* (1.1) *and the LRT statistic* L^* *in* (4.3), *we have*

$$\upsilon_N^*(g)^{-1/2}\big[L^* - p \cdot F^{y_N}(g) - m_N^*(g)\big] \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{4.4}$$

where

$$m_N^*(g) = (\kappa - 1) m_N(g) + \frac{\beta_x}{2} y_N,$$

$$\upsilon_N^*(g) = \frac{\kappa}{2} \upsilon_N(g),$$

the function g, *the values* $F^{y_N}(g)$, $m_N(g)$ *and* $\upsilon_N(g)$ *are the same as in Proposition* 4.1 *and* $\kappa = 2$ *or* $\kappa = 1$ *if* $\{\mathbf{X}_j\}_{j=1}^n$ *is real or complex*, *respectively* (*notice however the substitution of* N *for* n *in these quantities*).

Let us explain how this result extends considerably the previous Proposition 4.1 proposed in [1]. For real Gaussian observations, we have $\kappa = 2$ and $\beta_x = 0$, then $m_N^*(g) = m_N(g)$ and $\upsilon_N^*(g) = \upsilon_N(g)$, so that the new CLT gives an extension of Proposition 4.1 to Gaussian populations. If the observations are complex Gaussian, $\kappa = 1$ and $\beta_x = 0$, we have $m_N^*(g) = 0$ and the variance $\upsilon_N^*(g) = \frac{1}{2}\upsilon_N(g)$, which is half the variance for the real Gaussian case. For general non-Gaussian and centered populations, the new CLT provides a novel procedure for the one-sample test on large covariance matrix. In this case, the variance $\upsilon_N^*(g)$ stays the same as for Gaussian observations, but there is an additional term $\frac{1}{2}\beta_x y_N$ in the asymptotic mean.

We conclude the section by reporting a small Monte-Carlo experiment that demonstrates the importance of the sample size substitution proposed in Theorem 4.1. We simulate a p-dimensional standard Gaussian population $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$ but we do not assume to know anything about the mean and the covariance matrix so that the test will be based on the statistic L^* of (4.3). Simulation results are listed in Table 1. For the distribution of the CLRT statistic L^*, the experiment shows that the formula for its asymptotic mean and variance with adjusted RDS $y_N = p/N$ always outperforms the formula without the adjustment using $y_n = p/n$. The difference is quite significant for $p/n = 0.8$. This is an interesting improvement since when p/n is getting close to 1, the sample covariance matrix has more small eigenvalues near 0 and the presence of the logarithm function in the LRT statistic makes it more sensible with a larger variance. So a more accurate approximation for its asymptotic distribution is particularly valuable in such situations.

4.2. *Testing the equality of two large covariance matrices*. The second test problem we consider is about the equality between two large covariance matrices. As in Section 3, let $\mathbf{x}_1, \dots, \mathbf{x}_n$ and $\mathbf{y}_1, \dots, \mathbf{y}_m$ be samples from two p-dimensional populations with mean and covariance matrix $(\boldsymbol{\mu}, \boldsymbol{\Sigma}_x)$ and $(\boldsymbol{\nu}, \boldsymbol{\Sigma}_y)$, respectively.

TABLE 1
Effects of the sample size substitution for the corrected one-sample LRT with unbiased covariance matrix. Standard normal population with 10,000 *independent replications*

	$(pF^{y_N}(g)+m_N^*(g), v_N^*(g))$	$(pF^{y_n}(g)+m_n^*(g), v_n^*(g))$	**Empirical mean and variance of L^***
		$p/n = 0.5$	
$(p,n) = (25, 50)$	(8.226, 0.407)	(8.017, 0.386)	(8.234, 0.452)
$(p,n) = (50, 100)$	(15.889, 0.396)	(15.689, 0.386)	(15.886, 0.405)
$(p,n) = (100, 200)$	(31.228, 0.391)	(31.031, 0.386)	(31.231, 0.410)
$(p,n) = (150, 300)$	(46.570, 0.390)	(46.374, 0.386)	(46.569, 0.404)
		$p/n = 0.8$	
$(p,n) = (32, 40)$	(20.835, 1.794)	(19.929, 1.618)	(20.895, 2.158)
$(p,n) = (64, 80)$	(39.909, 1.702)	(39.053, 1.618)	(39.931, 1.851)
$(p,n) = (96, 120)$	(59.018, 1.673)	(58.178, 1.618)	(59.051, 1.739)
$(p,n) = (128, 160)$	(78.135, 1.659)	(77.302, 1.618)	(78.132, 1.714)

To test the hypothesis $H_0 : \mathbf{\Sigma}_x = \mathbf{\Sigma}_y$, a *corrected likelihood ratio* test is developed in [1] by assuming that both populations are Gaussian and $\boldsymbol{\mu} = \boldsymbol{\nu} = 0$ (or equivalently, they are given). Under the null hypothesis and because of the Gaussian assumption, one can assume without loss of generality that $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = \mathbf{I}$. Therefore the sample covariance matrices $\mathbf{S}_X^0$ and $\mathbf{S}_Y^0$ are as defined in (3.1) and the normalized Fisher matrix is $\mathbf{F}^0 = \mathbf{S}_x^0 \mathbf{S}_y^{0\,-1}$. The LRT statistic is

$$T^0 = \frac{|\mathbf{S}_x^0|^{n/2} \cdot |\mathbf{S}_y^0|^{m/2}}{|c_1 \mathbf{S}_x^0 + c_2 \mathbf{S}_y^0|^{(n+m)/2}}, \tag{4.5}$$

where $c_1 = n/(n+m)$ and $c_2 = m/(n+m)$. Recall the ratios $y_n = \frac{p}{n}$, $y_m = \frac{p}{m}$ and set

$$h_{n,m} = (y_n + y_m - y_n y_m)^{1/2}.$$

The following result is established in [1].

PROPOSITION 4.2 (Theorem 4.1 of [1]). *Assume that both populations are real Gaussian with respective mean* 0 *and covariance matrices* $\mathbf{\Sigma}_x, \mathbf{\Sigma}_y$ *and that* $p \wedge n \wedge m \to \infty$ *such that* $y_n \to y_1 > 0$, $y_m \to y_2 \in (0,1)$. *Then under* H_0,

$$\upsilon_{n,m}(f)^{-1/2}\left[-\frac{2\log T^0}{n+m} - p \cdot G_{y_n,y_m}(f) - a_{n,m}(f)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{4.6}$$

where

$$f(x) = \log(y_n + y_m x) - \frac{y_m}{y_n + y_m}\log x - \log(y_n + y_m), \tag{4.7}$$

$$G_{y_n,y_m}(f) = \frac{h_{n,m}^2}{y_n y_m} \log \frac{y_n + y_m}{h_{n,m}^2} + \frac{y_n(1-y_m)}{y_m(y_n+y_m)} \log(1-y_m) + \frac{y_m(1-y_n)}{y_n(y_n+y_m)} \log(1-y_n), \tag{4.8}$$

$$a_{n,m}(f) = \frac{1}{2}\left[\log\left(\frac{h_{n,m}^2}{y_n+y_m}\right) - \frac{y_n}{y_n+y_m}\log(1-y_m) - \frac{y_m}{y_n+y_m}\log(1-y_n)\right], \tag{4.9}$$

$$\upsilon_{n,m}(f) = -\frac{2y_m^2}{(y_n+y_m)^2}\log(1-y_n) - \frac{2y_n^2}{(y_n+y_m)^2}\log(1-y_m) + 2\log\frac{h_{n,m}^2}{y_n+y_m}. \tag{4.10}$$

Again, a corrected LRT is obtained based on this limiting distribution and has been proved to have good powers for large dimensions p. To extend this result to general non-Gaussian populations with unknown population means, we start by assuming that the population $\mathbf{x}$ fulfills assumptions (a)–(b)–(c) of Section 2 and the population $\mathbf{y}$ fulfills assumptions (a′)–(b′)–(c′) of Section 3. The corrected likelihood ratio test statistic (CLRT) is defined to be

$$T^* = \frac{|\mathbf{S}_x|^{N/2} \cdot |\mathbf{S}_y|^{M/2}}{|(N/(N+M))\mathbf{S}_x + (M/(N+M))\mathbf{S}_y|^{(N+M)/2}}. \tag{4.11}$$

Here, the unbiased sample covariance matrices $\mathbf{S}_x$ and $\mathbf{S}_y$ are defined in (1.1) and (1.9), respectively.

THEOREM 4.2. *Assume that the populations* $\mathbf{x}$ *and* $\mathbf{y}$ *satisfy assumptions* (a)–(b)–(c) *and* (a′)–(b′)–(c′), *respectively. Then under the null hypothesis* $H_0 : \mathbf{\Sigma}_x = \mathbf{\Sigma}_y$,

$$\upsilon_{N,M}^*(f^*)^{-1/2}\left[-\frac{2\log T^*}{N+M} - p\cdot G_{y_N,y_M}(f^*) - a_{N,M}^*(f^*)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{4.12}$$

where

$$f^*(x) = \log(y_N + y_M x) - \frac{y_M}{y_N+y_M}\log x - \log(y_N+y_M),$$

$$a_{N,M}^*(f^*) = (\kappa-1)a_{N,M}(f^*) + \frac{y_N y_M}{2(y_N+y_M)^2}(\beta_x y_N + \beta_y y_M),$$

$$\upsilon_{N,M}^*(f^*) = \frac{\kappa}{2}\upsilon_{N,M}(f^*),$$

TABLE 2

Effects of the sample size substitution for the corrected two-sample LRT with unbiased covariance matrices. Standard normal populations with 10,000 *independent replications*

	$(pF^{y_N,y_M}(f^*)+a^*_{N,M}(f^*)$, $v^*_{N,M}(f^*))$	$(pF^{y_n,y_m}(f^*)+a^*_{n,m}(f^*)$, $v^*_{n,m}(f^*))$	**Empirical mean and variance of T^***
		$p/n=0.5$	
$(p,n)=(20,40)$	(3.731, 0.127)	(3.601, 0.118)	(3.729, 0.134)
$(p,n)=(50,100)$	(8.820, 0.121)	(8.698, 0.118)	(8.819, 0.122)
$(p,n)=(80,160)$	(13.916, 0.120)	(13.795, 0.118)	(13.918, 0.125)
		$p/n=0.8$	
$(p,n)=(20,25)$	(8.551, 0.714)	(7.827, 0.588)	(8.520, 0.776)
$(p,n)=(60,75)$	(23.011, 0.625)	(22.382, 0.588)	(23.012, 0.661)
$(p,n)=(100,125)$	(37.550, 0.610)	(36.937, 0.588)	(37.552, 0.616)

where the values $G_{y_N,y_M}(f^*)$, $a_{N,M}(f^*)$ *and* $\upsilon_{N,M}(f^*)$ *are similar to those in Proposition* 4.2 [*notice however the substitution of* $(N,M)=(n-1,m-1)$ *for* (n,m) *in these formula*].

Again it is interesting to compare this CLT to the previous one in Proposition 4.2. When both populations are real Gaussian, $\kappa=2$ and $\beta_x=\beta_y=0$, we have $a^*_{N,M}(f^*)=a_{N,M}(f^*)$ and $\upsilon^*_{N,M}(f^*)=\upsilon_{N,M}(f^*)$, the new CLT is an extension of Proposition 4.2 to Gaussian populations. When there are both complex Gaussian, $\kappa=1$ and $\beta_x=\beta_y=0$, $a^*_{N,M}(f^*)=0$ and the variance $\upsilon^*_{N,M}(f^*)$ is reduced by half. For general non-Gaussian populations with unknown population means, there will be always a shift in the mean, but the variance again remains *unchanged* compared to the Gaussian situation. In summary, the substitution principle allows a full generalization of the corrected likelihood ratio two-sample test for large covariance matrices from non-Gaussian populations with unknown population means.

We conclude the section by reporting two Monte-Carlo experiments. The first is similar to the one in the previous section and designed to examine the effect of the sample size substitution proposed in Theorem 4.2. We adopt standard Gaussian population for both populations $\mathbf{x},\mathbf{y}\sim N(\mathbf{0}_p,\mathbf{I}_p)$ but we do not assume to know anything about these parameters so that the test will be based on the statistic T^* of (4.11). Simulation results are listed in Table 2. For the distribution of the CLRT statistic T^*, the limiting parameters with adjusted RDS y_N and y_M are much more accurate than using the original ones y_n and y_m.

The second Monte-Carlo experiment is designed to compare the new CLRT proposed in this paper with that proposed in Li and Chen test proposed in [15]. First of all, as reported in [1] and confirmed in [15], the classical likelihood ratio test (LRT) using a χ^2 approximation for $-2\log T^*$ fails completely even in moderate dimensions. The reason is that the χ^2 approximation is highly biased in such situ-

TABLE 3
Empirical sizes and powers of the CLRT and Li–Chen tests with 1000 *replications. Gaussian innovations for the upper half and Gamma innovations for the lower half. Moving average parameters are* $\theta_1 = 2$ *and* θ_2 *varying from* 0 *to* 0.4 (*see* [15] *for other details of the experimental design*)

			Power			
$(\boldsymbol{p}, \boldsymbol{n}, \boldsymbol{m})$	**Method**	**Size**	$\boldsymbol{\theta_2 = 0.1}$	$\boldsymbol{\theta_2 = 0.2}$	$\boldsymbol{\theta_2 = 0.3}$	$\boldsymbol{\theta_2 = 0.4}$
(40, 120, 120)	CLRT	0.05	0.092	0.285	0.793	0.999
	Li–Chen	0.051	0.062	0.109	0.214	0.347
(60, 180, 180)	CLRT	0.049	0.103	0.507	0.986	1.000
	Li–Chen	0.043	0.059	0.126	0.321	0.560
(40, 120, 120)	CLRT	0.044	0.087	0.253	0.776	0.998
	Li–Chen	0.077	0.072	0.134	0.216	0.378
(60, 180, 180)	CLRT	0.045	0.111	0.474	0.965	1.000
	Li–Chen	0.056	0.087	0.135	0.291	0.580

ations so that the actual test size becomes much higher than the nominal level (say 5%). In contrary, both the CLRT discussed in [1] and this paper and the Li–Chen test are able to deal with the high-dimensional effect. Second, comparing these two tests is not an obvious task since the power function for the CLRT is currently unavailable (the one for Li–Chen test has been derived in [15]). One thus resorts to Monte-Carlo experiments for a comparison, but the results would depend on the experimental design used. For example, Table 1 in [15] reports a design where the Li–Chen test is uniformly better than the CLRT. We have tested several other designs and the general feeling is that the powers of these two procedures are in general comparable. It is, however, reminded here that by nature, the CLRT are only available when the RDSs p/n_j are smaller than 1 while the Li–Chen test has no such a priori limitations and has then a wider application scope.

We now report a design under which the CLRT is uniformly better than the Li–Chen test. This design is indeed proposed in [15]: the **x** population is a first-order moving average with parameter θ_1 and the **y** population is a second-order moving average with parameters (θ_1, θ_2); see equations (4.1) and (4.2) in [15]. The innovations can be standard Gaussian or standardized Gamma distributed as $\{\text{Gamma}(\frac{5}{2}, \frac{1}{2}) - 5\}/\sqrt{10}$. Notice that when $\theta_2 = 0$, the two populations are coincident and in particular, $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$. As in [15], the empirical sizes and powers are evaluated with 1000 independent replications. The Li–Chen test uses the code provided on the web-page of the authors. Results are shown in Table 3. In particular, when the CLRT has reached the maximum power of 1, the powers of the Li–Chen test are 0.347, 0.56, 0.378 and 0.58 in the considered cases. It is also worth noticing that by construction, the Li–Chen procedure is very time-consuming. In the tested scenarios, 1000 replications of the Li–Chen procedure takes 33.0 minutes on a laptop while the CLRT uses only 3.5 minutes when $(p, n, m) = (40, 120, 120)$. The

Li–Chen procedure and CLRT take 180 minutes and 6.5 minutes, respectively, when $(p, n, m) = (60, 180, 180)$.

5. Proofs. Some of the proofs below use several technical lemmas which are collected and proved in Section 5.5.

5.1. *Proof of Theorem* 4.1. Under the null hypothesis $\mathbf{\Sigma}_x = \mathbf{I}_p$ and by the substitution principle of Theorem 2.2, it is enough to consider the sample covariance matrix

$$\mathbf{S}_N^0 = \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i^*,$$

where the $\mathbf{X}_i$'s have i.i.d. (0, 1) components. Applying the formula in Proposition A.1, we only need to evaluate the mean and variance parameter in equations (A.1)–(A.2) with the function $g(x) = x - \log x - 1$, that is,

$$m_N^*(g) = (\kappa - 1) I_1(g) + \beta_x I_2(g)$$

and

$$\upsilon_N^*(g) = \kappa J_1(g, g) + \beta_x J_2(g, g).$$

Note that the forms $\{I_\ell\}$ and $\{J_\ell\}$ are linear and bi-linear, respectively, and null on constants. Using their values on the functions x and $\log x$ calculated in equation (6.1) to equation (6.10) in [22], we readily find the claimed formula for $m_N^*(g)$ and $\upsilon_N^*(g)$.

5.2. *Proof of Theorem* 4.2. Under the null hypothesis, according to [1], the likelihood ratio statistic $-2(n+m)^{-1} \log T^*$ is a LSS of a Fisher matrix. Moreover, by the substitution principle of Theorem 3.1, it is enough to consider a Fisher matrix with the RDS $y_N = p/N$ and $y_M = p/M$ (instead of y_n and y_m). We thus use the CLT of [23] with these ratios and the test function f defined in (4.7), namely,

$$f^*(x) = \log(y_N + y_M x) - \frac{y_M}{y_N + y_M} \log(x) - \log(y_N + y_M).$$

Define

$$f_1(x) = \log(y_N + y_M x), \qquad f_2(x) = \log(x),$$

so that

$$f^* = f_1 - \frac{y_M}{y_N + y_M} f_2 - \log(y_N + y_M).$$

The asymptotic mean $E(X_{f_k})$ and the variance–covariance functions $\mathrm{Cov}(X_{f_k}, X_{f_\ell})$, $k, \ell = 1, 2$ are found using the calculations done in Example 4.1 of [23] with the following values of the parameters c, d, c' and d':

$$c' = \frac{1}{1 - y_M}, \qquad d' = \frac{\sqrt{y_N + y_M - y_N y_M}}{1 - y_M},$$
$$c = \frac{\sqrt{y_N + y_M - y_N y_M}}{1 - y_M}, \qquad d = \frac{y_M}{1 - y_M}.$$

That is, the mean function is

$$\begin{aligned} EX_{f_1} &= \frac{\kappa - 1}{2} \log\left(\frac{(c^2 - d^2)(h_{N,M})^2}{(ch_{N,M} - y_M d)^2}\right) - \frac{\beta_x y_N (1 - y_M)^2 d^2}{2(ch_{N,M} - dy_M)^2} \\ &\quad + \frac{\beta_y (1 - y_M)}{2}\left[\frac{2dy_M}{ch_{N,M} - dy_M} + \frac{d^2(y_M^2 - y_M)}{(ch_{N,M} - dy_M)^2}\right] \\ &= \frac{\kappa - 1}{2} \log \frac{y_N + y_M - y_N y_M}{(y_N + y_M)(1 - y_M)} - \frac{\beta_x}{2} \frac{y_N y_M^2}{(y_N + y_M)^2} + \frac{\beta_y}{2} \frac{y_M^2(2y_N + y_M)}{(y_N + y_M)^2} \end{aligned}$$

and

$$\begin{aligned} EX_{f_2} &= \frac{\kappa - 1}{2} \log\left(\frac{((c')^2 - (d')^2)h^2}{(c'h_{N,M} - y_M d')^2}\right) - \frac{\beta_x y_N (1 - y_M)^2 (d')^2}{2(c'h_{N,M} - d'y_M)^2} \\ &\quad + \frac{\beta_y (1 - y_M)}{2}\left[\frac{2d'y_M}{c'h_{N,M} - d'y_M} + \frac{(d')^2(y_M^2 - y_M)}{(c'h_{N,M} - d'y_M)^2}\right] \\ &= \frac{\kappa - 1}{2} \log \frac{1 - y_N}{1 - y_M} - \frac{\beta_x}{2} y_N + \frac{\beta_y}{2} y_M, \end{aligned}$$

where $h_{N,M} = \sqrt{y_N + y_M - y_N y_M}$. And the variance–covariance function is

$$\begin{aligned} \mathrm{Var}(X_{f_1}) &= \kappa \log\left(\frac{c^2}{c^2 - d^2}\right) + \frac{(\beta_x y_N + \beta_y y_M)(1 - y_M)^2 d^2}{(ch_{N,M} - dy_M)^2} \\ &= \kappa \log \frac{(h_{N,M})^2}{(y_{N,M} + y_M)(1 - y_M)} + (\beta_x y_N + \beta_y y_M) \frac{y_M^2}{(y_N + y_M)^2} \\ &= \kappa \log \frac{y_N + y_M - y_N y_M}{(y_N + y_M)(1 - y_M)} + (\beta_x y_N + \beta_y y_M) \frac{y_M^2}{(y_N + y_M)^2}, \\ \mathrm{Var}(X_{f_2}) &= \kappa \log\left(\frac{(c')^2}{(c')^2 - (d')^2}\right) + \frac{(\beta_x y_N + \beta_y y_M)(1 - y_M)^2 (d')^2}{(c'h_{N,M} - d'y_M)^2} \\ &= \kappa \log \frac{1}{(1 - y_N)(1 - y_M)} + (\beta_x y_N + \beta_y y_M), \end{aligned}$$

$$\begin{aligned}\operatorname{Cov}(X_{f_1}, X_{f_2}) &= \kappa \log\left(\frac{cc'}{cc' - dd'}\right) + \frac{(\beta_x y_N + \beta_y y_M)(1 - y_M)^2 dd'}{(ch_{N,M} - dy_M)(c'h_{N,M} - d'y_M)} \\ &= \kappa \log \frac{1}{1 - y_M} + (\beta_x y_N + \beta_y y_M)\frac{y_M}{y_N + y_M}.\end{aligned}$$

As by the definition of f,

$$a_{N,M}^*(f^*) = EX_{f_1} - \frac{y_M}{y_N + y_M} \cdot EX_{f_2}$$

and

$$\upsilon_{N,M}^*(f^*) = \operatorname{Var}(X_{f_1}) + \frac{y_M^2}{(y_N + y_M)^2} \cdot \operatorname{Var}(X_{f_2}) - \frac{2y_M}{y_N + y_M} \cdot \operatorname{Cov}(X_{f_1}, X_{f_2}),$$

by plugging in the calculations above, we readily find the announced formula for $a_{N,M}^*(f^*)$ and $\upsilon_{N,M}^*(f^*)$.

5.3. *Proof of Theorems* 2.1 *and* 2.2. Some notation is introduced as follows: $N = n - 1$, $\frac{1}{p}\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1}$ is the Stieltjes transform of the empirical spectral distribution of $\mathbf{S}_x = \frac{1}{N}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^*$,

$$\mathbf{A}(z) = \mathbf{B}_x - z\mathbf{I}_p, \qquad \mathbf{B}_x = \sum_{i=1}^n \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^*, \qquad \boldsymbol{\Delta} = \frac{1}{N}\sum_{j \neq k} \boldsymbol{\gamma}_j \boldsymbol{\gamma}_k^*,$$

$$\boldsymbol{\gamma}_j = \frac{\mathbf{x}_j}{\sqrt{n}} = \frac{\boldsymbol{\Gamma}\mathbf{X}_j}{\sqrt{n}}, \qquad \underline{m}_n^{(0)}(z) = -\frac{1 - y_n}{z} + y_n m_n^{(0)}(z),$$

$$z = -\frac{1}{\underline{m}_n^{(0)}(z)} + \frac{p}{n}\int \frac{t}{1 + t\underline{m}_n^{(0)}(z)}\, dH_p(t), \tag{5.1}$$

$$z = -\frac{1}{\underline{m}_N^{(0)}(z)} + \frac{p}{N}\int \frac{t}{1 + t\underline{m}_N^{(0)}(z)}\, dH_p(t), \tag{5.2}$$

$$z = -\frac{1}{\underline{m}_y(z)} + y\int \frac{t}{1 + t\underline{m}_y(z)}\, dH(t) \tag{5.3}$$

and H_p and H are the empirical spectral distribution and the limiting spectral distribution of $\boldsymbol{\Sigma}_x$. The support set of the LSD of $\mathbf{S}_x$ is

$$\left[(1 - \sqrt{y})^2 I_{(0<y<1)} \liminf_n \lambda_{\min}^{\boldsymbol{\Sigma}_x}, (1 + \sqrt{y})^2 \limsup_n \lambda_{\max}^{\boldsymbol{\Sigma}_x}\right].$$

Let x_r be a number greater than $(1 + \sqrt{y})^2 \limsup_n \lambda_{\max}^{\boldsymbol{\Sigma}_x}$. If $y < 1$, then let x_l be a number between 0 and

$$(1 - \sqrt{y})^2 \liminf_n \lambda_{\min}^{\boldsymbol{\Sigma}_x}.$$

If $y \geq 1$, let x_l be a negative number. Let η_l and η_r satisfy

$$x_l < \eta_l < (1-\sqrt{y})^2 I_{(0<y<1)} \liminf_n \lambda_{\min}^{\boldsymbol{\Sigma}_x} < (1+\sqrt{y})^2 \limsup_n \lambda_{\max}^{\boldsymbol{\Sigma}_x} < \eta_r < x_r.$$

Define a contour $\mathcal{C} = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_b \cup \mathcal{C}_r$ where

$$\mathcal{C}_u = \{x + i v_0 : x \in [x_l, x_r]\}, \qquad \mathcal{C}_l = \{x_l + i v : |v| \leq v_0\},$$
$$\mathcal{C}_b = \{x - i v_0 : x \in [x_l, x_r]\}, \qquad \mathcal{C}_r = \{x_r + i v : |v| \leq v_0\},$$

and $\mathcal{C}_n = \mathcal{C} \cap \{z : \Im(z) > n^{-2}\}$. As f is analytic, we have by Cauchy integral theorem

$$\begin{aligned} Y_p(f) &= \sum_{i=1}^{p} f(\lambda_i) - p \int f(x)\, dF^{y_N, H_p}(x) \\ &= -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z)\{\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - p m_N^{(0)}(z)\}\, dz, \end{aligned} \tag{5.4}$$

where $\oint_{\mathcal{C}}$ is the contour integration in the anti-clockwise direction and $\{\lambda_j\}_{j=1}^{p}$ are eigenvalues of $\mathbf{S}_x$. It remains to find the asymptotic distribution of

$$\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - p m_N^{(0)}(z)$$

in order to obtain the asymptotic distribution of $Y_p(f_j)$.

5.3.1. *Outline of the main steps of the proof and of the key differences between this proof and the proof of the CLT in Bai and Silverstein* [5]. When the population mean is $\mathbf{0}$, because $\mathbf{S}_x$ is close to $\mathbf{B}_x$, we will express their relationship as

$$\mathbf{S}_x = \mathbf{B}_x - \boldsymbol{\Delta}.$$

Therefore, our goal turns out to show that

$$\begin{aligned} &\{\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - p m_N^{(0)}(z)\} - \{\operatorname{tr} \mathbf{A}^{-1}(z) - p m_n^{(0)}(z)\} \\ &\quad = p m_n^{(0)}(z) - p m_N^{(0)}(z) + \operatorname{tr}(\mathbf{A}(z) - \boldsymbol{\Delta})^{-1} - \operatorname{tr} \mathbf{A}^{-1}(z) \to 0 \\ &\quad \text{in probability.} \end{aligned} \tag{5.5}$$

Then the main results under the Gaussian-like moment conditions follow from Bai and Silverstein [5].

When the Gaussian-like moment conditions are not met, as pointed out in Section 2, the CLT of LSS of $\mathbf{S}_x$ or $\mathbf{B}_x$ are not true generally and we have to pose some alternative conditions. Because Bai and Silverstein [5] have proved that

$$\left\{ \oint_{\mathcal{C}} f(z)(\operatorname{tr} \mathbf{A}^{-1}(z) - p m_n^{(0)}(z))\, dz \right\}$$

is tight, even without the Gaussian like assumptions, then our main work turns out to derive the asymptotic means and covariances of

$$\left\{\oint_{\mathcal{C}} f(z)\big(\operatorname{tr}\mathbf{A}^{-1}(z) - pm_n^{(0)}(z)\big)\,dz\right\}$$

under the alternative and unremovable conditions.

5.3.2. *The proof of* (5.5) *uniformly.* We have

$$\mathbf{S}_x = \frac{n}{N}\sum_{i=1}^{n}(\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}})(\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}})^* = \sum_{i=1}^{n}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^* - \frac{1}{N}\sum_{i\neq j}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_j^* = \mathbf{B}_x - \Delta,$$

where $\bar{\boldsymbol{\gamma}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\gamma}_i$. Moreover, we have

$$\begin{aligned}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} &= \big(\mathbf{A}(z) - \boldsymbol{\Delta}\big)^{-1} = \mathbf{A}^{-1}(z) + \big(\mathbf{A}(z) - \boldsymbol{\Delta}\big)^{-1}\boldsymbol{\Delta}\mathbf{A}^{-1}(z)\\ &= \mathbf{A}^{-1}(z) + \mathbf{A}^{-1}(z)\boldsymbol{\Delta}\mathbf{A}^{-1}(z) + \mathbf{A}^{-1}(z)\big(\boldsymbol{\Delta}\mathbf{A}^{-1}(z)\big)^2\\ &\quad + \big(\mathbf{A}(z) - \boldsymbol{\Delta}\big)^{-1}\big(\boldsymbol{\Delta}\mathbf{A}^{-1}(z)\big)^3.\end{aligned}$$

Therefore,

$$\begin{aligned}&\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - pm_N^{(0)}(z)\\ &\quad = \operatorname{tr}\big(\mathbf{A}(z) - \boldsymbol{\Delta}\big)^{-1} - pm_n^{(0)}(z) + pm_n^{(0)}(z) - pm_N^{(0)}(z)\\ &\quad = \operatorname{tr}\mathbf{A}^{-1}(z) - pm_n^{(0)}(z) + p\big(m_n^{(0)}(z) - m_N^{(0)}(z)\big) + \operatorname{tr}\mathbf{A}^{-2}(z)\boldsymbol{\Delta}\\ &\qquad + \operatorname{tr}\mathbf{A}^{-1}(z)\big(\boldsymbol{\Delta}\mathbf{A}^{-1}(z)\big)^2 + \operatorname{tr}\big(\mathbf{A}(z) - \boldsymbol{\Delta}\big)^{-1}\big(\boldsymbol{\Delta}\mathbf{A}^{-1}(z)\big)^3.\end{aligned}\tag{5.6}$$

By (5.6), Lemmas 5.1 and 5.6, we have

$$\begin{aligned}\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - pm_N^{(0)}(z) &= \operatorname{tr}\mathbf{A}^{-1}(z) - pm_n^{(0)}(z) + o_P(1)\\ &= \operatorname{tr}(\mathbf{B}_x - z\mathbf{I}_p)^{-1} - pm_n^{(0)}(z) + o_P(1),\end{aligned}\tag{5.7}$$

where the error term $o_P(1)$ is uniformly in $z \in \mathcal{C}$. We also need to check the tightness of $\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - pm_N^{(0)}(z)$. Because

$$\begin{aligned}&\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - pm_N^{(0)}(z)\\ &\quad = \operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - \operatorname{tr}(\mathbf{B}_x - z\mathbf{I}_p)^{-1} + \operatorname{tr}(\mathbf{B}_x - z\mathbf{I}_p)^{-1}\\ &\qquad - pm_n^{(0)}(z) + pm_n^{(0)}(z) - pm_N^{(0)}(z)\end{aligned}$$

and the tightness of

$$\big\{\operatorname{tr}(\mathbf{B}_x - z\mathbf{I}_p)^{-1} - pm_n^{(0)}(z)\big\}$$

is proved in [5], then we only prove the tightness of

$$\{\operatorname{tr}(\mathbf{S}_x - z\mathbf{I}_p)^{-1} - \operatorname{tr}(\mathbf{B}_x - z\mathbf{I}_p)^{-1}\}.$$

Let $\{\tilde{\lambda}_i\}$ and $\{\lambda_i\}$ be the eigenvalues of $\mathbf{B}_x$ and $\mathbf{S}_x$, respectively, and be arranged in descending order. Let the event $\mathcal{B}_n$ is defined as

$$\eta_l < \lambda_p < \frac{n}{N}\tilde{\lambda}_1 < \eta_r.$$

Then it is well known from random matrix theory that for any positive number t, it holds for large enough n that $\mathrm{P}(\mathcal{B}_n^c) = o(n^{-t})$; see, for example, [4]. Notice that

$$\mathbf{S}_x = \frac{n}{N}\mathbf{B}_x - \frac{n}{N}\boldsymbol{\Gamma}\bar{\mathbf{X}}\bar{\mathbf{X}}^*\boldsymbol{\Gamma}^*,$$

where $\bar{\mathbf{X}} = \sum_{j=1}^n \mathbf{X}_j/n$. Similar to arguments in [5], we only need to prove that there is an absolute constant M such that for any $z_1, z_2 \in \mathcal{C}_n$

$$\begin{aligned}
&\frac{E|\operatorname{tr}(\mathbf{B}_x - z_1\mathbf{I}_p)^{-1} - \operatorname{tr}(\mathbf{B}_x - z_2\mathbf{I}_p)^{-1} - \operatorname{tr}(\mathbf{S}_x - z_1\mathbf{I}_p)^{-1} + \operatorname{tr}(\mathbf{S}_x - z_2\mathbf{I}_p)^{-1}|^2}{|z_1 - z_2|^2}\\
&\quad = \mathrm{E}\big|\operatorname{tr}(\mathbf{B}_x - z_1\mathbf{I}_p)^{-1}(\mathbf{B}_x - z_2\mathbf{I}_p)^{-1} - \operatorname{tr}(\mathbf{S}_x - z_1\mathbf{I}_p)^{-1}(\mathbf{S}_x - z_2\mathbf{I}_p)^{-1}\big|^2\\
&\quad = \mathrm{E}\left|\sum_{i=1}^n \frac{(\tilde{\lambda}_i - \lambda_i)(\lambda_i + \tilde{\lambda}_i - z_1 - z_2)}{(\lambda_i - z_1)(\lambda_i - z_2)(\tilde{\lambda}_i - z_1)(\tilde{\lambda}_i - z_2)}\right|^2 \qquad (5.8)\\
&\quad \le K\mathrm{E}\left\{\left|\sum_{i=1}^n |\lambda_i - \tilde{\lambda}_i|\right|^2 I_{\mathcal{B}_n}\right\} + o(1) \le M,
\end{aligned}$$

where the last step of (5.8) follows from the fact that

$$\begin{aligned}
\sum_{i=1} |\lambda_i - \tilde{\lambda}_i| I_{\mathcal{B}_n} &= \sum_{i=1}\left|\lambda_i - \frac{n}{N}\tilde{\lambda}_i\right| + \frac{1}{N}\sum_{i=1}^n \tilde{\lambda}_i\\
&= \sum_{i=1}\frac{n}{N}\tilde{\lambda}_i - \lambda_i + \frac{1}{N}\sum_{i=1}^n \tilde{\lambda}_i\\
&\le \frac{n}{N}\tilde{\lambda}_1 - \lambda_p + \frac{1}{N}\sum_{i=1}^n \tilde{\lambda}_i \le 2\eta_r - \eta_l,
\end{aligned}$$

by the interlacing theorem. Finally, the proof of (5.5) uniformly is completed.

5.3.3. *Proof of Theorem 2.1 under the alternative conditions.* For brevity, we introduce several notations. Let E_j be the conditional expectation on $\mathbf{r}_1, \ldots, \mathbf{r}_{j-1}$,

$$\mathbf{A}_j = \sum_{i=1}^n \mathbf{r}_i\mathbf{r}_i^* - z\mathbf{I}_p - \mathbf{r}_j\mathbf{r}_j^*, \qquad \mathbf{A}_{kj} = \mathbf{A}_j - \mathbf{r}_k\mathbf{r}_k^*,$$

$$\beta_k = \frac{1}{1 + \mathbf{r}_k^* \mathbf{A}_k^{-1} \mathbf{r}_k}, \qquad \beta_{k(j)} = \frac{1}{1 + \mathbf{r}_k^* \mathbf{A}_{kj}^{-1} \mathbf{r}_k}.$$

$\breve{\mathbf{A}}_j$, $\breve{\mathbf{A}}_{jk}$, $\breve{\beta}_j$ and $\breve{\beta}_{k(j)}$ are similarly defined as $\mathbf{A}_j$, $\mathbf{A}_{jk}$, β_j and $\beta_{k(j)}$ with variables $\mathbf{r}_{j+1}, \dots, \mathbf{r}_n$ replaced by $\breve{\mathbf{r}}_{j+1}, \dots, \breve{\mathbf{r}}_n$, an independent copy of $\mathbf{r}_{j+1}, \dots, \mathbf{r}_n$, and $b(z) = -z\underline{m}_y(z)$.

Now we assume that the matrix $\mathbf{\Gamma}$ is real. Then we consider (2.7) of Bai and Silverstein [5]. According to Bai and Silverstein [5], it is easy to obtain

$$\begin{aligned} &\frac{1}{n}\mathrm{E}_j\{z_1 \operatorname{tr} \mathbf{\Sigma}_x \mathbf{A}_j^{-1}(z_1) - z_2 \operatorname{tr} \mathbf{\Sigma}_x (\breve{\mathbf{A}}_j')^{-1}(z_2)\} \\ &\qquad \to z_1(b^{-1}(z_1) - 1) - z_2(b^{-1}(z_2) - 1), \qquad \text{a.s.} \end{aligned} \tag{5.9}$$

In fact, we have that $\{\operatorname{tr}(\mathbf{S}_x^0 - z\mathbf{I}_p)^{-1} - pm_n^{(0)}(z)\}$ converges to a Gaussian process with the mean function as the limit of

$$\underline{m}_y(z) \cdot y_n \sum_{j=1}^{n} \mathrm{E}\beta_j d_j \Big/ \left(1 - y \int \frac{(\underline{m}_y^2(z) t^2 \, dH(t))}{(1 + t\underline{m}_y(z))^2}\right) \tag{5.10}$$

and the covariance function as the limit of

$$\frac{b_n(z_1) b_n(z_2)}{n^2} \sum_{j=1}^{n} \operatorname{tr} \mathrm{E}_j \mathbf{\Sigma}_x \mathbf{A}_j^{-1}(z_1) \mathrm{E}_j(\mathbf{\Sigma}_x \mathbf{A}_j^{-1}(z_2)) \tag{5.11}$$

$$+ \frac{\alpha_x b_n(z_1) b_n(z_2)}{n^2} \sum_{j=1}^{n} \operatorname{tr} \mathrm{E}_j \mathbf{\Sigma}_x \mathbf{A}_j^{-1}(z_1) \mathrm{E}_j(\mathbf{\Sigma}_x (\mathbf{A}_j')^{-1}(z_2)) \tag{5.12}$$

$$+ \frac{\beta_x b_n(z_1) b_n(z_2)}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{p} \mathbf{e}_i' \mathbf{\Gamma}_x^* \mathbf{A}_j^{-1}(z_1) \mathbf{\Gamma}_x \mathbf{e}_i \cdot \mathbf{e}_i' \mathbf{\Gamma}_x^* \mathbf{A}_j^{-1}(z_2) \mathbf{\Gamma}_x \mathbf{e}_i, \tag{5.13}$$

where

$$d_j = y^{-1} \boldsymbol{\gamma}_j^* \mathbf{A}_j^{-1}(z) \{\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\}^{-1} \boldsymbol{\gamma}_j - p^{-1} \operatorname{tr} \mathbf{A}^{-1}(z) \{\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\}^{-1} \mathbf{\Sigma}_x,$$

$\mathbf{e}_i$ is a p-dimensional vector with the ith element equal to 1 and other elements equal to 0.

First, in [5] it is proved that the limit of (5.11) is

$$\left(\frac{\partial}{\partial z_1} \underline{m}_y(z_1) \frac{\partial}{\partial z_2} \underline{m}_y(z_2)\right) \Big/ (\underline{m}_y(z_1) - \underline{m}_y(z_2))^2.$$

In [23]'s (40)–(41) or [17], when Γ_x is diagonal, it is proved that the limit of (5.13) is

$$y\beta_x \int \frac{1}{(\underline{m}_y(z_1)t + 1)^2} \frac{1}{(\underline{m}_y(z_1)t + 1)^2} \, dH(t).$$

It remains to find the limits of (5.12) and (5.10). When Γ_x is real, we have

$$
\begin{aligned}
&\frac{1}{n}\{z_1 \mathrm{E}_j \operatorname{tr} \boldsymbol{\Sigma}_x \mathbf{A}_j^{-1}(z_1) - z_2 \operatorname{tr} \Sigma_x (\breve{\mathbf{A}}_j')^{-1}(z_2)\} \\
&\quad = \frac{1}{n}\mathrm{E}_j \operatorname{tr} \Sigma_x \mathbf{A}_j^{-1}(z_1)\Bigg[\sum_{i=1}^{j-1}(z_1 \bar{\mathbf{r}}_i \mathbf{r}_i' - z_2 \mathbf{r}_i \mathbf{r}_i^*) \\
&\qquad\qquad + \sum_{i=j+1}^{n}(z_1 \bar{\breve{\mathbf{r}}}_i \breve{\mathbf{r}}_i' - z_2 \mathbf{r}_i \mathbf{r}_i^*)\Bigg](\breve{\mathbf{A}}_j')^{-1}(z_2) \\
&\quad = \frac{1}{n}\sum_{i=1}^{j-1} \mathrm{E}_j \{z_1 \breve{\beta}_{ji}(z_2) \mathbf{r}_i' (\breve{\mathbf{A}}_{ji}')^{-1}(z_2) \\
&\qquad\qquad \times \Sigma_x [\mathbf{A}_{ji}^{-1}(z_1) - \mathbf{A}_{ji}^{-1}(z_1)\mathbf{r}_i \mathbf{r}_i^* \mathbf{A}_{ji}^{-1}(z_1)\beta_{ji}(z_1)]\bar{\mathbf{r}}_i \\
&\qquad\qquad - z_2 \mathbf{r}_i^* [(\breve{\mathbf{A}}_{ji}')^{-1}(z_2) - \breve{\beta}_{ji}(z_2)(\breve{\mathbf{A}}_{ji}')^{-1}(z_2)\bar{\mathbf{r}}_i \mathbf{r}_i' (\breve{\mathbf{A}}_{ji}')^{-1}(z_2)] \\
&\qquad\qquad\qquad \times \Sigma_x \mathbf{A}_{ji}^{-1}(z_1)\beta_{ji}(z_1)\mathbf{r}_i\} \\
&\qquad + \frac{1}{n}\sum_{i=j+1}^{n} \mathrm{E}_j [z_1 \breve{\beta}_{ji}(z_2) \breve{\mathbf{r}}_i' (\breve{\mathbf{A}}_{ji}')^{-1}(z_2) \Sigma_x \mathbf{A}_j^{-1}(z_1) \bar{\breve{\mathbf{r}}}_i \\
&\qquad\qquad - z_2 \mathbf{r}_i^* (\breve{\mathbf{A}}_j')^{-1}(z_2) \Sigma_x \mathbf{A}_{ji}^{-1}(z_1) \mathbf{r}_i \beta_{ji}(z_1)] \\
&\quad = \frac{1}{n}\sum_{i=1}^{j-1} \mathrm{E}_j \Bigg\{ z_1 \breve{\beta}_{ji}(z_2) \bigg(\frac{1}{n} \operatorname{tr}(\breve{\mathbf{A}}_{ji}')^{-1}(z_2) \Sigma_x \mathbf{A}_{ji}^{-1}(z_1) \Sigma_x \\
&\qquad\qquad - \frac{\alpha_x}{n^2} \operatorname{tr}(\breve{\mathbf{A}}_{ji}')^{-1}(z_2) \Sigma_x \mathbf{A}_{ji}^{-1}(z_1) \Sigma_x \\
&\qquad\qquad\qquad \times \operatorname{tr} \mathbf{A}_{ji}^{-1}(z_1) \Sigma_x \beta_{ij}(z_1)\bigg) \\
&\qquad\qquad - z_2 \beta_{ij}(z_1) \bigg(\frac{1}{n} \operatorname{tr}(\breve{\mathbf{A}}_{ji}')^{-1}(z_2) \Sigma_x \mathbf{A}_{ji}^{-1}(z_1) \Sigma_x \\
&\qquad\qquad - \frac{\alpha_x}{n^2} \operatorname{tr}(\breve{\mathbf{A}}_{ji}')^{-1}(z_2) \Sigma_x \\
&\qquad\qquad\qquad \times \operatorname{tr}(\breve{\mathbf{A}}_{ji}')^{-1}(z_2) \Sigma_x \mathbf{A}_{ji}^{-1}(z_1) \Sigma_x \breve{\beta}_{ji}(z_2)\bigg)\Bigg\} \\
&\qquad + \frac{1}{n}\sum_{i=j+1}^{n} \mathrm{E}_j \{z_1 \breve{\beta}_{ji}(z_2) \breve{\mathbf{r}}_i' (\breve{\mathbf{A}}_{ji}')^{-1}(z_2) \Sigma_x \mathbf{A}_j^{-1}(z_1) \bar{\breve{\mathbf{r}}}_i \\
&\qquad\qquad - z_2 \mathbf{r}_i^* (\breve{\mathbf{A}}_j')^{-1}(z_2) \Sigma_x \mathbf{A}_{ji}^{-1}(z_1) \mathbf{r}_i \beta_{ji}(z_1)\} + o_{\mathrm{a.s.}}(1)
\end{aligned}
$$

$$= \left\{ -\frac{j-1}{n} \alpha_x \left[\left(z_1 b(z_2) - z_2 b(z_1) \right) - b(z_1) b(z_2)(z_1 - z_2) \right] \right.$$
$$\left. + \left(z_1 b(z_2) - z_2 b(z_1) \right) \right\}$$
$$\times \mathrm{E}_j \frac{1}{n} \operatorname{tr} \Sigma_x \mathbf{A}_j^{-1}(z_1) \Sigma_x \left(\breve{\mathbf{A}}_j' \right)^{-1}(z_2) + o_{\text{a.s.}}(1),$$

where $\alpha_x = |EX_{11}^2|^2$.

Comparing the estimate above with (5.9), we obtain

$$\mathrm{E}_j \frac{1}{n} \operatorname{tr} \Sigma_x \breve{\mathbf{A}}_j^{-1}(z_2) \Sigma_x \mathbf{A}_j^{-1}(z_1)$$
$$= \left(z_1 \left(b^{-1}(z_1) - 1 \right) - \left(z_2 b^{-1}(z_2) - 1 \right) + o_{\text{a.s.}}(1) \right)$$
$$\Big/ \left(-\frac{j-1}{n} \alpha_x \left[\left(z_1 b(z_2) - z_2 b(z_1) \right) - b(z_1) b(z_2)(z_1 - z_2) \right] \right.$$
$$\left. + \left(z_1 b(z_2) - z_2 b(z_1) \right) \right).$$

Consequently, we obtain the limit of (5.12) as follows:

$$\frac{1}{n^2} \sum_{j=1}^{n} \alpha_x b_n(z_1) b_n(z_2) \operatorname{tr} \mathrm{E}_j \Sigma_x \mathbf{A}_j^{-1}(z_1) \mathrm{E}_j \left(\Sigma_x \left(\mathbf{A}_j' \right)^{-1}(z_2) \right)$$
$$\to a(z_1, z_2) \int_0^1 \frac{1}{1 - t a(z_1, z_2)} \, dt = \int_0^{a(z_1, z_2)} \frac{1}{1 - z} \, dz,$$

where

$$a(z_1, z_2) = \frac{\alpha_x b(z_1) b(z_2) (z_1 (b^{-1}(z_1) - 1) - z_2 (b^{-1}(z_2) - 1))}{z_1 b(z_2) - z_2 b(z_1)}$$
$$= \alpha_x \left(1 + \frac{b(z_1) b(z_2)(z_2 - z_1)}{z_1 b(z_2) - z_2 b(z_1)} \right)$$
$$= \alpha_x \left(1 + \frac{\underline{m}_y(z_1) \underline{m}_y(z_2)(z_1 - z_2)}{\underline{m}_y(z_2) - \underline{m}_y(z_1)} \right).$$

Moreover, we have

$$\frac{\partial^2}{\partial z_2 \, \partial z_1} \int_0^{a(z_1, z_2)} \frac{1}{1 - z} \, dz = \frac{\partial^2}{\partial z_1 \, \partial z_2} \log\left(1 - a(z_1, z_2) \right).$$

So the covariance function $\operatorname{Cov}(X_{f_j}, X_{f_\ell})$ will then have the following form:

$$\operatorname{Cov}(X_{f_j}, X_{f_\ell})$$
$$= -\frac{1}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} \frac{f_j(z_1) f_\ell(z_2)}{(\underline{m}_y(z_1) - \underline{m}_y(z_2))^2} \, d\underline{m}_y(z_1) \, d\underline{m}_y(z_2) \tag{5.14}$$

$$
\begin{aligned}
&- \frac{y\beta_x}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} f_j(z_1) f_\ell(z_2) \\
&\qquad\qquad \times \biggl[\int \frac{t}{(\underline{m}_y(z_1)t+1)^2} \frac{t}{(\underline{m}_y(z_1)t+1)^2}\, dH(t)\biggr] dz_1\, dz_2 \\
&- \frac{1}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} f_j(z_1) f_\ell(z_2) \biggl[\frac{\partial^2}{\partial z_1\, \partial z_2} \log\bigl(1 - a(z_1, z_2)\bigr)\biggr] dz_1\, dz_2.
\end{aligned}
$$

Next for the limiting mean function, we have by (9.11.12) of [3]

$$
\begin{aligned}
&\underline{m}_y(z) \cdot y_n \sum_{j=1}^n \mathrm{E}\beta_j d_j \\
&\quad = \beta_x y_n z^2 \underline{m}_y^3(z) \frac{1}{pn} \sum_{j=1}^n \sum_{i=1}^p \mathbf{e}_i' \mathbf{\Gamma}_x^* \mathbf{A}_j^{-1}(z) \{\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\}^{-1} \\
&\qquad\qquad \times \mathbf{\Gamma}_x \mathbf{e}_i \cdot \mathbf{e}_i' \mathbf{\Gamma}_x^* \mathbf{A}_j^{-1}(z) \mathbf{\Gamma}_x \mathbf{e}_i \\
&\qquad + \underline{m}_y(z) \cdot \alpha_x \frac{z^2 \underline{m}_y^2(z)}{n^2} \sum_{j=1}^n \mathrm{E}\,\mathrm{tr}\bigl[\mathbf{A}_j^{-1}(z) \{\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\}^{-1} \\
&\qquad\qquad \times \mathbf{\Sigma}_x \bigl(\mathbf{A}_j'\bigr)^{-1}(z) \mathbf{\Sigma}_x\bigr] + o(1).
\end{aligned}
\tag{5.15}
$$

First, we have

$$
\begin{aligned}
&\alpha_x \frac{z^2 \underline{m}_y^2(z)}{n^2} \sum_{j=1}^n \mathrm{E}\,\mathrm{tr}\bigl[\mathbf{A}_j^{-1}(z) \{\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\}^{-1} \mathbf{\Sigma}_x \bigl(\mathbf{A}_j'\bigr)^{-1}(z) \mathbf{\Sigma}_x\bigr] \\
&\quad = \alpha_x \frac{\underline{m}_y^2(z)}{n} \mathrm{E}\,\mathrm{tr}\bigl(\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\bigr)^{-3} \mathbf{\Sigma}_x^2 \\
&\qquad + \alpha_x \frac{z^2 \underline{m}_y^4(z)}{n^2} \\
&\qquad \times \sum_{i \neq j} \mathrm{E}\,\mathrm{tr}\biggl\{\bigl(\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\bigr)^{-1} \mathbf{\Sigma}_x \bigl(\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\bigr)^{-1} \\
&\qquad\qquad \times \biggl(\mathbf{r}_i \mathbf{r}_i^* - \frac{1}{n} \mathbf{\Sigma}_x\biggr) \mathbf{A}_{ij}^{-1}(z) \\
&\qquad\qquad \times \bigl(\underline{m}_y(z) \mathbf{\Sigma}_x + \mathbf{I}_p\bigr)^{-1} \mathbf{\Sigma}_x \bigl(\mathbf{A}_{ij}'\bigr)^{-1}(z) \biggl(\bar{\mathbf{r}}_i \mathbf{r}_i' - \frac{1}{n} \mathbf{\Sigma}_x\biggr)\biggr\} + o(1) \\
&\quad = \alpha_x \frac{\underline{m}_y^2(z)}{n} \mathrm{tr}\bigl(\underline{m}_y(z) \Sigma_x + \mathbf{I}_p\bigr)^{-3} \Sigma_x^2
\end{aligned}
$$

$$+\frac{\alpha_x^2 z^2 \underline{m}_y^4(z)}{n^3}\sum_{j=1}^{n}\{\operatorname{tr}(\underline{m}_y(z)\boldsymbol{\Sigma}_x+\mathbf{I}_p)^{-1}\boldsymbol{\Sigma}_x\}^2$$

$$\times \operatorname{tr}\mathbf{A}_j^{-1}(z)(\underline{m}_y(z)\boldsymbol{\Sigma}_x+\mathbf{I}_p)^{-1}\boldsymbol{\Sigma}_x(\mathbf{A}_j')^{-1}(z)\boldsymbol{\Sigma}_x+o(1).$$

Then we have

$$\begin{aligned}
&\alpha_x\frac{z^2\underline{m}_y^2(z)}{n^2}\sum_{j=1}^{n}\mathrm{E}\operatorname{tr}[\mathbf{A}_j^{-1}(z)\{\underline{m}_y(z)\boldsymbol{\Sigma}_x+\mathbf{I}_p\}^{-1}\boldsymbol{\Sigma}_x(\mathbf{A}_j')^{-1}(z)\boldsymbol{\Sigma}_x]\\
&\quad=\frac{\alpha_x(\underline{m}_y^2(z)/n)\operatorname{tr}\{(\underline{m}_y(z)\Sigma_x+\mathbf{I}_p)^{-3}\Sigma_x^2\}}{1-(\alpha_x\underline{m}_y^2(z)/n)\operatorname{tr}\{(\underline{m}_y(z)\Sigma_x+\mathbf{I}_p)^{-1}\Sigma_x\}^2}+o(1)\\
&\quad=\alpha_x y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^3}\Big/\left(1-\alpha_x y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)+o(1).
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
&\left(\alpha_x\frac{z^2\underline{m}_y^2(z)}{n^2}\sum_{j=1}^{n}\mathrm{E}\operatorname{tr}[\mathbf{A}_j^{-1}(z)\{\underline{m}_y(z)\boldsymbol{\Sigma}_x+\mathbf{I}_p\}^{-1}\boldsymbol{\Sigma}_x(\mathbf{A}_j')^{-1}(z)\boldsymbol{\Sigma}_x]\right)\\
&\quad\Big/\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)\\
&\qquad=\left(\alpha_x y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^3}\right)\\
&\qquad\quad\Big/\left(\left\{1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right\}\left\{1-\alpha_x y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right\}\right)+o(1).
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&\left(\beta_x y_n z^2\underline{m}_y^3(z)\frac{1}{pn}\sum_{j=1}^{n}\sum_{i=1}^{p}\mathrm{E}\mathbf{e}_i'\boldsymbol{\Gamma}_x^*\mathbf{A}_j^{-1}(z)\{\underline{m}_y(z)\boldsymbol{\Sigma}_x+\mathbf{I}_p\}^{-1}\right.\\
&\qquad\qquad\left.\times\boldsymbol{\Gamma}_x\mathbf{e}_i\cdot\mathbf{e}_i'\boldsymbol{\Gamma}_x^*\mathbf{A}_j^{-1}(z)\boldsymbol{\Gamma}_x\mathbf{e}_i\right)\\
&\quad\Big/\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)\\
&\qquad=\beta_x y\int\frac{\underline{m}_y^3(z)t^2}{(\underline{m}_y(z)t+1)^3}\,dH(t)\Big/\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)+o(1).
\end{aligned}$$

That is,

$$\begin{aligned}
&\underline{m}_y(z)\cdot y_n\sum_{j=1}^n \mathrm{E}\beta_j d_j\Big/\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)\\
&\quad=\alpha_x y\int\frac{\underline{m}_y^3(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^3}\\
&\qquad\Big/\left(\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)\left(1-\alpha_x y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)\right)\\
&\qquad+y\int\frac{\underline{m}_y^3(z)t^2}{(\underline{m}_y(z)t+1)^3}\,dH(t)\Big/\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right).
\end{aligned}\tag{5.16}$$

Then the mean function $\mathrm{E}X_f$ of Bai and Silverstein [5] will be

$$\begin{aligned}
&-\frac{\alpha_x}{2\pi\mathbf{i}}\oint_{\mathcal{C}} f(z)\cdot y\int\frac{\underline{m}_y^3(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^3}\\
&\quad\Big/\left(\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)\left(1-\alpha_x y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)\right)dz\\
&\quad-\frac{\beta_x}{2\pi\mathbf{i}}\oint_{\mathcal{C}} f(z)y\int\frac{\underline{m}_y^3(z)t^2}{(\underline{m}_y(z)t+1)^3}\,dH(t)\Big/\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)dz.
\end{aligned}$$

Therefore, we obtain that

$$\begin{aligned}
X_p(f_j)&=\sum_{i=1}^p f_j(\lambda_i)-p\int f_j(x)\,dF^{y_n,H_p}(x)\\
&=\oint f_j(z)\{\mathrm{tr}(\mathbf{S}_x^0-z\mathbf{I}_p)^{-1}-p\cdot m_n^{(0)}(z)\}\,dz,
\end{aligned}$$

converges to a Gaussian vector with mean EX_{f_ℓ} and covariance function $\mathrm{Cov}(X_{f_j},X_{f_\ell})$ as follows:

$$\begin{aligned}
\mathrm{E}X_{f_\ell}&=-\frac{\alpha_x}{2\pi\mathbf{i}}\oint f_\ell(z)y\int\frac{\underline{m}_y^3(z)t^2}{(1+t\underline{m}_y(z))^3}\,dH(t)\\
&\quad\Big/\left(\left(1-y\int\frac{\underline{m}_y^2(z)t^2}{(1+t\underline{m}_y(z))^2}\,dH(t)\right)\right.\\
&\qquad\left.\times\left(1-\alpha_x y\int\frac{\underline{m}_y^2(z)t^2}{(1+t\underline{m}_y(z))^2}\,dH(t)\right)\right)dz\\
&\quad-\frac{\beta_x}{2\pi\mathbf{i}}\oint f_\ell(z)y\int\frac{\underline{m}_y^3(z)t^2}{(\underline{m}_y(z)t+1)^3}\,dH(t)\Big/\left(1-y\int\frac{\underline{m}_y^2(z)t^2\,dH(t)}{(1+t\underline{m}_y(z))^2}\right)dz
\end{aligned}$$

and

$$
\begin{aligned}
&\operatorname{Cov}(X_{f_j}, X_{f_\ell}) \\
&\quad = -\frac{1}{4\pi^2}\oint\oint \frac{f_j(z_1) f_\ell(z_2)}{(\underline{m}_y(z_1) - \underline{m}_y(z_2))^2}\, d\underline{m}_y(z_1)\, d\underline{m}_y(z_2) \\
&\qquad - \frac{y\beta_x}{4\pi^2}\oint\oint f_j(z_1) f_\ell(z_2)\biggl[\int \frac{t}{(\underline{m}_y(z_1)t+1)^2} \\
&\qquad\qquad \times \frac{t}{(\underline{m}_y(z_1)t+1)^2}\, dH(t)\biggr] dz_1\, dz_2 \\
&\qquad - \frac{1}{4\pi^2}\oint\oint f_j(z_1) f_\ell(z_2)\biggl[\frac{\partial^2}{\partial z_1\,\partial z_2}\log\bigl(1 - a(z_1, z_2)\bigr)\biggr] dz_1\, dz_2.
\end{aligned}
$$

The proof of Theorem 2.1 is complete.

5.3.4. *Proof of Theorem* 2.2. By (5.7) and Theorem 2.1, we obtain Theorem 2.2.

5.4. *Proof of Theorem* 3.1. Recall that $N = n - 1$ and $M = m - 1$ are the adjusted sample sizes. The proof has two steps following the decomposition:

$$
\begin{aligned}
\operatorname{tr}(\mathbf{F} - z\mathbf{I}_p)^{-1} - p m_{(y_N, y_M)}(z) &= \bigl[\operatorname{tr}\bigl(\mathbf{S}_x\mathbf{S}_y^{-1} - z\mathbf{I}_p\bigr)^{-1} - p m^{(y_N, F^{\mathbf{S}_y^{-1}})}(z)\bigr] \\
&\quad + p\bigl[m^{(y_N, F^{\mathbf{S}_y^{-1}})}(z) - m_{(y_N, y_M)}(z)\bigr],
\end{aligned}
$$

where:

- $F^{\mathbf{S}_y^{-1}}(t)$ and $F^{\mathbf{S}_y}(t)$ are the ESDs of $\mathbf{S}_y^{-1}$ and $\mathbf{S}_y$;
- $m_{(y_1,y_2)}(z)$ is the Stieltjes transform of the LSD $G_{(y_1,y_2)}$ of $\mathbf{F}$, and to simplify the notation we simply write $m(z) = m_{(y_1,y_2)}(z)$ if no confusion is possible, and $\underline{m}(z) = -\frac{1-y_1}{z} + y_1 m(z)$;
- $m_{(y_N,y_M)}(z)$ is obtained by replacing (y_1, y_2) by (y_N, y_M) in $m(z) = m_{(y_1,y_2)}(z)$; and
- $\underline{m}^{\{y_N, F^{\mathbf{S}_y^{-1}}\}} = -\frac{1-y_N}{z} + y_N m^{\{y_N, F^{\mathbf{S}_y^{-1}}\}}(z)$ and

$$
\begin{aligned}
z &= -\frac{1}{\underline{m}^{\{y_N, F^{\mathbf{S}_y^{-1}}\}}} + y_N \int \frac{t}{1 + t\underline{m}^{\{y_N, F^{\mathbf{S}_y^{-1}}\}}}\, dF^{\mathbf{S}_y^{-1}}(t) \\
&= -\frac{1}{\underline{m}^{\{y_N, F^{\mathbf{S}_y^{-1}}\}}} + y_N \int \frac{1}{t + \underline{m}^{\{y_N, F^{\mathbf{S}_y^{-1}}\}}}\, dF^{\mathbf{S}_y}(t).
\end{aligned}
\tag{5.17}
$$

Step 1. Given $\mathbf{S}_y$, in the proof of Theorem 2.2, we have proved that the process

$$
\bigl\{\operatorname{tr}\bigl(\mathbf{S}_x\mathbf{S}_y^{-1} - z\mathbf{I}_p\bigr)^{-1} - p m^{\{y_N, F^{\mathbf{S}_y^{-1}}\}}(z)\bigr\}
$$

weakly tends to a Gaussian process $M_1(z)$ on the contour $\mathcal{C}$ with mean function

$$\begin{aligned}\mathrm{E}\big(M_1(z)|\mathbf{S}_2\big) &= (\kappa-1)\cdot\frac{y_1\int \underline{m}(z)^3 x[x+\underline{m}(z)]^{-3}\,dF_{y_2}(x)}{[1-y_1\int \underline{m}^2(z)(x+\underline{m}(z))^{-2}\,dF_{y_2}(x)]^2}\\ &\quad+\beta_x\cdot\left(\left(y_1\cdot\underline{m}^3(z)\cdot\int\frac{dF_{y_2}(x)}{x+\underline{m}(z)}\int\frac{x\cdot dF_{y_2}(x)}{(x+\underline{m}(z))^2}\right)\right.\\ &\qquad\left./\left(1-y_1\int\underline{m}^2(z)\big(x+\underline{m}(z)\big)^{-2}\,dF_{y_2}(x)\right)\right),\end{aligned}$$

for $z\in\mathcal{C}$ and covariance function

$$\begin{aligned}&\operatorname{Cov}\big(M_1(z_1),M_1(z_2)|\mathbf{S}_2\big)\\ &\quad=\kappa\cdot\left(\frac{\underline{m}'(z_1)\cdot\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}-\frac{1}{(z_1-z_2)^2}\right)\\ &\qquad+\beta_x\cdot y_1\cdot\int\frac{\underline{m}'(z_1)\cdot x\cdot dF_{y_2}(x)}{(x+\underline{m}(z_1))^2}\int\frac{\underline{m}'(z_2)\cdot x\cdot dF_{y_2}(x)}{(x+\underline{m}(z_2))^2},\end{aligned}$$

for $z_1,z_2\in\mathcal{C}$ where F_{y_2} is the Marčenko–Pastur law with the ratio y_2.

Step 2. By (5.17) and the truth of

$$z=-\frac{1}{\underline{m}_{\{y_N,y_M\}}}+y_N\int\frac{1}{t+\underline{m}_{\{y_N,y_M\}}}\,dF_{y_M}(t), \tag{5.18}$$

where F_{y_M} is the Marčenko–Pastur law with the ratio $y_M=p/M$. Subtracting both sides of (5.17) from those of (5.18) and by Theorem 2.2, we obtain

$$\begin{aligned}&p\cdot\big[m^{\{y_N,F^{\mathbf{S}_y^{-1}}\}}(z)-m_{\{y_N,y_M\}}(z)\big]\\ &\quad=-y_N\underline{m}_{\{y_{n-},y_M\}}\underline{m}^{\{y_N,F^{\mathbf{S}_y^{-1}}\}}\\ &\qquad\times\left(\big(\operatorname{tr}\big(\mathbf{S}_y+\underline{m}_{\{y_N,y_M\}}\mathbf{I}_p\big)^{-1}-pm_{y_M}(-\underline{m}_{\{y_N,y_M\}})\big)\right.\\ &\qquad\left./\left(1-y_N\cdot\int\frac{\underline{m}_{\{y_N,y_M\}}\cdot\underline{m}^{\{y_N,F^{\mathbf{S}_y^{-1}}\}}\,dF_M(t)}{(t+\underline{m}_{\{y_N,y_M\}})\cdot(t+\underline{m}^{\{y_N,F^{\mathbf{S}_x^{-1}}\}})}\right)\right)\end{aligned} \tag{5.19}$$

which converges weakly to a Gaussian process $M_2(\cdot)$ on $z\in\mathcal{C}$ with mean function

$$\begin{aligned}\mathrm{E}\big(M_2(z)\big) &= -(\kappa-1)\cdot\frac{y_2\underline{m}'(z)\cdot[\underline{m}_{y_2}(-\underline{m}(z))]^3\cdot[1+\underline{m}_{y_2}(-\underline{m}(z))]^{-3}}{[1-y_2\cdot(\underline{m}_{y_2}(-\underline{m}(z))/(1+\underline{m}_{y_2}(-\underline{m}(z))))^2]^2}\\ &\quad-\beta_y\cdot\underline{m}'(z)\frac{y_2\cdot m_0^3(z)\cdot(1+m_0(z))^{-3}}{1-y_2\cdot m_0^2(z)\cdot(1+m_0(z))^{-2}},\end{aligned}$$

and covariance function

$$\begin{aligned}
&\operatorname{Cov}\big(M_2(z_1), M_2(z_2)\big)\\
&\quad = \kappa \underline{m}'(z_1)\underline{m}'(z_2)\bigg(\frac{\underline{m}'_{y_2}(-\underline{m}(z_1)) \cdot \underline{m}'_{y_2}(-\underline{m}(z_2))}{[\underline{m}_{y_2}(-\underline{m}(z_1)) - \underline{m}_{y_2}(-\underline{m}(z_2))]^2} - \frac{1}{(\underline{m}(z_1) - \underline{m}(z_2))^2}\bigg)\\
&\qquad + \beta_y \cdot y_2 \cdot \frac{\underline{m}'(z_1)\underline{m}'_{y_2}(-\underline{m}(z_1))}{(1+\underline{m}_{y_2}(-\underline{m}(z_1)))^2} \cdot \frac{\underline{m}'(z_2)\underline{m}'_{y_2}(-\underline{m}(z_2))}{(1+\underline{m}_{y_2}(-\underline{m}(z_2)))^2}
\end{aligned}$$

for $z_1, z_2 \in \mathcal{C}$, where we have used the relationships

$$z = -\frac{1}{\underline{m}_{y_M}} + \frac{y_M}{1+\underline{m}_{y_M}}, \qquad \underline{m}_{y_M}(z) = -\frac{1-y_M}{z} + y m_{y_M}(z)$$

$$\text{with } m_0(z) = \underline{m}_{y_2}\big(-\underline{m}(z_2)\big).$$

Thus

$$\operatorname{tr}(\mathbf{F} - z\mathbf{I}_p)^{-1} - p m_{(y_N, y_M)}(z)$$

converges to a Gaussian process $\{M_1(z) + M_2(z)\}$ with the following mean and covariance functions:

$$\begin{aligned}
&\mathrm{E}\big(M_1(z) + M_2(z)\big)\\
&\quad = (\kappa - 1) \cdot \frac{y_1 \int \underline{m}^3(z) x [x + \underline{m}(z)]^{-3}\, dF_{y_2}(x)}{[1 - y_1 \int \underline{m}^2(z)(x + \underline{m}(z))^{-2}\, dF_{y_2}(x)]^2}\\
&\qquad + \beta_x \cdot \frac{y_1 \cdot \underline{m}^3(z) \cdot \int (dF_{y_2}(x)/(x + \underline{m}(z))) \int (x \cdot dF_{y_2}(x)/(x + \underline{m}(z))^2)}{1 - y_1 \int \underline{m}^2(z)(x + \underline{m}(z))^{-2}\, dF_{y_2}(x)}\\
&\qquad - (\kappa - 1) \cdot \underline{m}'(z) \frac{y_2 \cdot [m_0(z)]^3 \cdot [1 + m_0(z)]^{-3}}{[1 - y_2 \cdot (m_0(z)/(1 + m_0(z)))^2]^2}\\
&\qquad - \beta_y \cdot \underline{m}'(z) \frac{y_2 \cdot m_0^3(z) \cdot (1 + m_0(z))^{-3}}{1 - y_2 \cdot m_0^2(z) \cdot (1 + m_0(z))^{-2}},
\end{aligned}$$

and

$$\begin{aligned}
&\operatorname{Cov}\big(M_1(z_1) + M_2(z_1), M_1(z_2) + M_2(z_2)\big)\\
&\quad = \beta_x \cdot y_1 \cdot \int \frac{\underline{m}'(z_1) \cdot x \cdot dF_{y_2}(x)}{(x + \underline{m}(z_1))^2} \int \frac{\underline{m}'(z_2) \cdot x \cdot dF_{y_2}(x)}{(x + \underline{m}(z_2))^2} - \frac{\kappa}{(z_1 - z_2)^2}\\
&\qquad + \kappa \cdot \frac{m_0'(z_1) \cdot m_0'(z_2)}{[m_0(z_1) - m_0(z_2)]^2}\\
&\qquad + \beta_y \cdot y_2 \frac{m_0'(z_1)}{(1 + m_0(z_1))^2} \cdot \frac{(m_0'(z_2))}{(1 + m_0(z_2))^2}.
\end{aligned}$$

Then by Corollary 3.2 of Zheng [23], we obtain that the random vector $(W_p(f_1), \dots, W_p(f_k))$ where

$$W_p(f_j) = \sum_{i=1}^{p} f_j(\lambda_i) - p \int f_j(x)\, dF^{\{y_N, y_M\}}(x)$$

with eigenvalues $\{\lambda_i\}_{i=1}^{p}$ of $\mathbf{F}$ converges weakly to a Gaussian vector (Z_{f_j}) with mean and covariance functions

$$\begin{aligned}
\mathrm{E}Z_{f_j} = {} & \lim_{r\downarrow 1} \frac{\kappa - 1}{4\pi \mathbf{i}} \oint_{|\xi|=1} f_j\left(\frac{1+h^2+2h\Re(\xi)}{(1-y_2)^2}\right) \\
& \qquad \times \left[\frac{1}{\xi - r^{-1}} + \frac{1}{\xi + r^{-1}} - \frac{2}{\xi + y_2/h}\right] d\xi \\
& + \frac{\beta_x y_1 (1-y_2)^2}{2\pi \mathbf{i} \cdot h^2} \oint_{|\xi|=1} f_j\left(\frac{1+h^2+2h\Re(\xi)}{(1-y_2)^2}\right) \frac{1}{(\xi + y_2/h)^3}\, d\xi \\
& + \frac{\beta_y (1-y_2)}{4\pi \mathbf{i}} \oint_{|\xi|=1} f_j\left(\frac{1+h^2+2h\Re(\xi)}{(1-y_2)^2}\right) \frac{\xi^2 - y_2/h^2}{(\xi + y_2/h)^2} \\
& \qquad \times \left[\frac{1}{\xi - \sqrt{y_2}/h} + \frac{1}{\xi + \sqrt{y_2}/h} - \frac{2}{\xi + y_2/h}\right] d\xi,
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{Cov}(Z_{f_j}, Z_{f_\ell}) \\
& \quad = -\lim_{r\downarrow 1} \frac{\kappa}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \left(f_j\left(\frac{1+h^2+2h\Re(\xi_1)}{(1-y_2)^2}\right) \right. \\
& \qquad\qquad \times f_\ell\left(\frac{1+h^2+2h\Re(\xi_2)}{(1-y_2)^2}\right) \\
& \qquad\qquad \left. \Big/ (\xi_1 - r\xi_2)^2 \right) d\xi_1\, d\xi_2 \\
& \qquad - \frac{(\beta_x y_1 + \beta_y y_2)(1-y_2)^2}{4\pi^2 h^2} \\
& \qquad \times \oint_{|\xi_1|=1} \left(f_j\left(\frac{1+h^2+2h\Re(\xi_1)}{(1-y_2)^2}\right) \Big/ \left(\xi_1 + \frac{y_2}{h}\right)^2 \right) d\xi_1 \\
& \qquad \times \oint_{|\xi_2|=1} \left(f_\ell\left(\frac{1+h^2+2h\Re(\xi_2)}{(1-y_2)^2}\right) \Big/ \left(\xi_2 + \frac{y_2}{h}\right)^2 \right) d\xi_2.
\end{aligned}$$

The proof of Theorem 3.1 is complete.

5.5. *Technical lemmas*. For brevity, we first introduce several notation as follows:

$$\beta_j = b_j - \beta_j b_j \varepsilon_j = b_j - b_j^2 \varepsilon_j + \beta_j b_j^2 \varepsilon_j^2, \tag{5.20}$$
$$\beta_{j(k)} = b_{j(k)} - \beta_{j(k)} b_{j(k)} \varepsilon_{j(k)} = b_{j(k)} - b_{j(k)}^2 \varepsilon_{j(k)} + \beta_{j(k)} b_{j(k)}^2 \varepsilon_{j(k)}^2,$$

with

$$b_j = \frac{1}{1 + \mathrm{E}\boldsymbol{\gamma}_j^* \mathbf{A}_j^{-1}(z)\boldsymbol{\gamma}_j}, \qquad \varepsilon_j = \boldsymbol{\gamma}_j^* \mathbf{A}_j^{-1}(z)\boldsymbol{\gamma}_j - E\boldsymbol{\gamma}_j^* \mathbf{A}_j^{-1}(z)\boldsymbol{\gamma}_j$$

and $b_{j(k)}$ and $\varepsilon_{j(k)}$ are similarly defined by replacing $\mathbf{A}_j^{-1}(z)$ by $\mathbf{A}_{jk}^{-1}(z)$. All the lemmas in this section assume that the conditions of Theorem 2.1 are satisfied. Under the assumed moment conditions, we can truncate the variables X_{ij} as $X_{ij} I_{(|X_{ij}| \le \eta_n \sqrt{n})}$ where $\eta_n = o(1)$ without affecting the final results; see, for example, page 183 of [3] for a justification. For simplicity, we still denote by X_{ij} the normalization of the truncated variable $X_{ij} I_{(|X_{ij}| \le \eta_n \sqrt{n})}$.

LEMMA 5.1. *After truncation and normalization, for every* $z \in \mathbb{C}^+ = \{z : \Im(z) > 0\}$, *we have*

$$p\big(m_n^{(0)} - m_N^{(0)}\big) \xrightarrow{a.s.} (1 + z\underline{m}_y) \cdot \frac{\underline{m}_y + z\underline{m}_y'}{z\underline{m}_y}.$$

PROOF. We have

$$\underline{m}_n^{(0)}(z) = -\left(1 - \frac{p}{n}\right) \cdot \frac{1}{z} + \frac{p}{n} m_n^0(z), \tag{5.21}$$
$$\underline{m}_N^{(0)}(z) = -\left(1 - \frac{p}{N}\right) \cdot \frac{1}{z} + \frac{p}{N} m_N^0(z),$$

where $p/n \to y_1 > 0$. By (5.3), we obtain

$$\underline{m}_y'(z) = \frac{1}{1/\underline{m}_y^2(z) - y\int (t^2/(1 + t\underline{m}_y(z))^2)\, dH(t)}, \tag{5.22}$$
$$y_1 \int \frac{t}{1 + t\underline{m}_y(z)}\, dH(t) = \frac{1 + z\underline{m}_y(z)}{\underline{m}_y(z)}.$$

For brevity, $\underline{m}_n^{(0)}(z)$, $\underline{m}_N^{(0)}(z)$, $\underline{m}_y(z)$ are simplified as $\underline{m}_n^{(0)}$, $\underline{m}_N^{(0)}$ and $\underline{m}_y$. Using (5.1)–(5.2), we obtain

$$0 = \frac{\underline{m}_n^{(0)} - \underline{m}_N^{(0)}}{\underline{m}_n^{(0)} \underline{m}_N^{(0)}} - \big(\underline{m}_n^{(0)} - \underline{m}_N^{(0)}\big) \frac{p}{n} \int \frac{t^2}{(1 + t\underline{m}_n^{(0)})(1 + t\underline{m}_N^{(0)})}\, dH_p(t)$$
$$- \frac{p}{n(n-1)} \int \frac{t}{1 + t\underline{m}_N^{(0)}}\, dH_p(t),$$

that is,

$$
\begin{aligned}
&n\big(\underline{m}_n^{(0)} - \underline{m}_N^{(0)}\big) \\
&\qquad = \frac{p}{N}\int \frac{t}{1+t\underline{m}_N^{(0)}}\,dH_p(t) \\
&\qquad\qquad \Big/\bigg(\frac{1}{\underline{m}_n^{(0)}\underline{m}_N^{(0)}} - \frac{p}{n}\int \frac{t^2}{(1+t\underline{m}_n^{(0)})(1+t\underline{m}_N^{(0)})}\,dH_p(t)\bigg) \qquad (5.23)\\
&\qquad \to y\int \frac{t}{1+t\underline{m}_y}\,dH(t)\Big/\bigg(\frac{1}{\underline{m}_y^2} - y\int \frac{t^2}{(1+t\underline{m}_y)^2}\,dH(t)\bigg).
\end{aligned}
$$

By (5.21), (5.22) and (5.23), we have

$$
\begin{aligned}
p\big(m_n^{(0)} - m_N^{(0)}\big) &= n\underline{m}_n^{(0)} + \frac{n-p}{z} - \bigg((n-1)\underline{m}_N^{(0)} + \frac{n-1-p}{z}\bigg) \\
&= n\big(\underline{m}_n^{(0)} - \underline{m}_N^{(0)}\big) + \underline{m}_N^{(0)}(z) + \frac{1}{z} \qquad (5.24)\\
&\to \frac{y\int (t/(1+t\underline{m}_y))\,dH(t)}{1/\underline{m}_y^2 - y\int (t^2/(1+t\underline{m}_y)^2)\,dH(t)} + \frac{1+z\underline{m}_{y_1}(z)}{z} \\
&= \underline{m}_y'\cdot\frac{1+z\underline{m}_y}{\underline{m}_y} + \frac{1+z\underline{m}_y(z)}{z} = (1+z\underline{m}_y)\cdot\frac{\underline{m}_y + z\underline{m}_y'}{z\underline{m}_y}.
\end{aligned}
$$

Thus, Lemma 5.1 is proved. □

In the sequel, we shall use Vitali lemma frequently. Let

$$\boldsymbol{\Delta} = \frac{1}{n}\sum_{j\neq k}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_k^*.$$

The normalization is by $1/n$ here instead of the previously used $1/N$ but this difference does not affect the limits calculated here. We will derive the limit of $\operatorname{tr}(\mathbf{A}(z) - \boldsymbol{\Delta})^{-1} - \operatorname{tr}(\mathbf{A}^{-1}(z))$.

LEMMA 5.2. *After truncation and normalization*, *we have*

$$\mathrm{E}\big|\boldsymbol{\gamma}_k^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_k - \big(1+z\underline{m}_y(z)\big)\big|^2 \le Kn^{-1},$$

for every $z\in\mathbb{C}^+$ with a constant K.

PROOF. We have

$$\boldsymbol{\gamma}_k^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_k = \boldsymbol{\gamma}_k^*\mathbf{A}_k^{-1}(z)\boldsymbol{\gamma}_k\beta_k = 1-\beta_k,$$

where

$$\mathbf{A}_k(z) = \mathbf{A}(z) - \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^*, \qquad \beta_k = \left(1 + \boldsymbol{\gamma}_k^* \mathbf{A}_k^{-1} \boldsymbol{\gamma}_k\right)^{-1}.$$

Therefore, by (1.15) and (2.17) of Bai and Silverstein [5], we have

$$\mathrm{E}\left|\boldsymbol{\gamma}_k^* \mathbf{A}^{-1}(z) \boldsymbol{\gamma}_k - \left(1 + z\underline{m}_y(z)\right)\right|^2 = \mathrm{E}\left|\beta_k + z\underline{m}_y(z)\right|^2 \le K n^{-1}.$$

Then the lemma is proved. $\square$

COROLLARY 5.1. *After truncation and normalization, we have*

$$\mathrm{E}\left|\boldsymbol{\gamma}_k^* \mathbf{A}^{-2}(z) \boldsymbol{\gamma}_k - \frac{d}{dz}\left(1 + z\underline{m}_y(z)\right)\right|^2 \le K n^{-1}$$

for every $z \in \mathbb{C}^+$.

PROOF. By the Cauchy integral formula, we have

$$\boldsymbol{\gamma}_k^* \mathbf{A}^{-2}(z) \boldsymbol{\gamma}_k = \frac{1}{2\pi i} \oint_{|\zeta - z| = \Im(z)/2} \frac{\boldsymbol{\gamma}_k^* \mathbf{A}^{-1}(\zeta) \boldsymbol{\gamma}_k}{(\zeta - z)^2}\, d\zeta$$

and

$$\frac{d}{dz}\left(1 + z\underline{m}_y(z)\right) = \frac{1}{2\pi i} \oint_{|\zeta - z| = \Im(z)/2} \frac{1 + \zeta \underline{m}_y(\zeta)}{(\zeta - z)^2}\, d\zeta.$$

Then $\mathrm{E}|\boldsymbol{\gamma}_k^* \mathbf{A}^{-2} \boldsymbol{\gamma}_k - \frac{d}{dz}(1 + z\underline{m}_y(z))|^2 \le K n^{-1}$ follows from Lemma 5.2. $\square$

LEMMA 5.3. *After truncation and normalization, we have*

$$\mathrm{E}\left|\operatorname{tr} \mathbf{A}^{-1}(z) \boldsymbol{\Delta}\right|^2 \le K n^{-1}$$

for every $z \in \mathbb{C}^+$. Especially for every $z \in \mathbb{C}^+$,

$$\mathrm{E}\left|\operatorname{tr}\left(\mathbf{A}^{-2}(z) \boldsymbol{\Delta}\right)\right|^2 = \mathrm{E}\left|\frac{1}{n} \sum_{j \ne k \in \mathcal{U}} \boldsymbol{\gamma}_j^* \mathbf{A}^{-2}(z) \boldsymbol{\gamma}_k\right|^2 = O\left(n^{-1}\right), \qquad \mathcal{U} = \{1, 2, \dots, n\}.$$

PROOF. We have

$$\operatorname{tr} \mathbf{A}^{-1}(z) \boldsymbol{\Delta} = \frac{1}{n} \sum_{j \ne k \in \mathcal{U}} \boldsymbol{\gamma}_j^* \mathbf{A}^{-1}(z) \boldsymbol{\gamma}_k = \frac{1}{n} \sum_{j \ne k \in \mathcal{U}} \boldsymbol{\gamma}_j^* \mathbf{A}_{jk}^{-1}(z) \boldsymbol{\gamma}_k \beta_j \beta_{k(j)},$$

where

$$\mathbf{A}_{jk}(z) = \mathbf{A}_k(z) - \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^*, \qquad \beta_{k(j)} = \left(1 + \boldsymbol{\gamma}_k^* \mathbf{A}_{jk}^{-1}(z) \boldsymbol{\gamma}_k\right)^{-1}.$$

We will similarly define $\mathbf{A}_{ijk}(z)$ and $\beta_{k(ij)}$ for later use. Then we obtain

$$
\begin{aligned}
&\mathrm{E}\big|\mathrm{tr}\big(\mathbf{A}^{-1}(z)\boldsymbol{\Delta}\big)\big|^2 \\
&\quad = \mathrm{E}\Big\{\frac{1}{n}\sum_{j_1\neq k_1\in\mathcal{U}} \boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1k_1}^{-1}\boldsymbol{\gamma}_{k_1}\beta_{j_1}\beta_{k_1(j_1)}\frac{1}{n}\sum_{j_2\neq k_2\in\mathcal{U}}\overline{\boldsymbol{\gamma}_{j_2}^*\mathbf{A}_{j_2k_2}^{-1}\boldsymbol{\gamma}_{k_2}\beta_{j_2}\beta_{k_2(j_2)}}\Big\} \\
&\quad = \sum_{(2)}+\sum_{(3)}+\sum_{(4)},
\end{aligned}
$$

where the index $(\cdot)$ denotes the number of distinct integers in the set $\{j_1, k_1, j_2, k_2\}$. By the facts that $|\beta_j| \le \frac{|z|}{v}$ and $v = \Im(z)$, we have

$$
\begin{aligned}
\sum_{(2)} &\le \frac{2|z|^4}{n^2v^4}\sum_{j\neq k\in\mathcal{U}} \mathrm{E}\big|\boldsymbol{\gamma}_j^*\mathbf{A}_{jk}^{-1}\boldsymbol{\gamma}_k\big|^2 \\
&\le \frac{|z|^4}{v^4n^4}\sum_{j\neq k\in\mathcal{U}} \mathrm{E}\,\mathrm{tr}\big(\boldsymbol{\Sigma}_x\mathbf{A}_{jk}^{-1}\boldsymbol{\Sigma}_x\overline{\mathbf{A}_{jk}^{-1}}\big) \le \frac{p}{n^2}\frac{|z|^4\|\mathbf{T}\|^2}{v^6} \le Kn^{-1},
\end{aligned}
$$

where K is a constant. Moreover, we have

$$
\sum_{(4)} = \frac{1}{n^2}\sum_{j_1\neq k_1\neq j_2\neq k_2\in\mathcal{U}} \mathrm{E}\big\{\boldsymbol{\gamma}_{j_1}^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}_{j_2}^*\overline{\mathbf{A}^{-1}(z)}\boldsymbol{\gamma}_{k_2}\big\}.
$$

To evaluate the sum above, we expand $\boldsymbol{\gamma}_{j_1}^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_1}$ as

$$
\begin{aligned}
&\boldsymbol{\gamma}_{j_1}^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_1} \\
&\quad = \beta_{j_1}\beta_{k_1(j_1)}\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1k_1}^{-1}\boldsymbol{\gamma}_{k_1} \\
&\quad = \beta_{j_1}\beta_{k_1(j_1)}\big[\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1k_1k_2}^{-1}\boldsymbol{\gamma}_{k_1} - \beta_{k_2(j_1k_1)}\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1k_1k_2}^{-1}\boldsymbol{\gamma}_{k_2}\boldsymbol{\gamma}_{k_2}^*\mathbf{A}_{j_1k_1k_2}^{-1}\boldsymbol{\gamma}_{k_1}\big] \\
&\quad = \beta_{j_1}\beta_{k_1(j_1)}\big[\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_1} - \beta_{j_2(j_1k_1k_2)}\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_1} \\
&\qquad\qquad - \beta_{k_2(j_1k_1)}\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_2}\boldsymbol{\gamma}_{k_2}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_1} \\
&\qquad\qquad + \beta_{k_2(j_1k_1)}\beta_{j_2(j_1k_1k_2)}\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^* \\
&\qquad\qquad\quad \times \mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_2}\boldsymbol{\gamma}_{k_2}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_1} \\
&\qquad\qquad + \beta_{k_2(j_1k_1)}\beta_{j_2(j_1k_1k_2)}\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_2}\boldsymbol{\gamma}_{k_2}^* \\
&\qquad\qquad\quad \times \mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_1} \\
&\qquad\qquad - \beta_{k_2(j_1k_1)}\beta_{j_2(j_1k_1k_2)}^2\boldsymbol{\gamma}_{j_1}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_2} \\
&\qquad\qquad\qquad\qquad \times \boldsymbol{\gamma}_{k_2}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^*\mathbf{A}_{j_1j_2k_1k_2}^{-1}\boldsymbol{\gamma}_{k_1}\big].
\end{aligned}
$$

And analogously, expand $\boldsymbol{\gamma}_{j_2}^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_2}$ into similar 6 terms and then we will estimate the expectations of the 36 products in the expansion of

$$\boldsymbol{\gamma}_{j_1}^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_1}(\boldsymbol{\gamma}_{j_2}^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_2})^*.$$

Case 1. *Terms containing at least five* $\mathbf{A}_{j_1j_2k_1k_2}^{-1}$*'s in* $\sum_{(4)}$. We will prove that these terms are bounded by $O(n^{-3})$. We shall use the fact that all β-factors $\beta_j, \beta_{j(k)}, \beta_{k_2(j_1k_1)}, \beta_{j_2(j_1k_1k_2Z)}$ are bounded by $|z|/v \le K$. Let $\mathbf{B} = \mathbf{A}_{j_1j_2k_1k_2}^{-1}$. As an example, consider the product of the last terms of the two expansions, its expectation is bounded by

$$\begin{aligned}
&E|(\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{k_2}\boldsymbol{\gamma}_{k_2}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{k_1})(\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}_{k_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_2})^*| \\
&\quad\le (\mathrm{E}|\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}_{k_2}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{k_1}|^2 \\
&\qquad\times \mathrm{E}|\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{k_2}\boldsymbol{\gamma}_{k_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{k_2}|^2)^{1/2},
\end{aligned}$$

where, when applying Cauchy–Schwarz, we have exchanged the two factors $\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1}$ and $\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{k_2}$ in the two groups. We have

$$\begin{aligned}
&\mathrm{E}|(\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}_{k_2}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{k_1})|^2 \\
&\quad= \frac{1}{n}\mathrm{E}|(\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}_{k_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_2})|^2\boldsymbol{\gamma}_{j_2}^*\mathbf{B}^*\boldsymbol{\Sigma}_x\mathbf{B}\boldsymbol{\gamma}_{j_2} \\
&\quad\le \frac{K}{n^5}\mathrm{E}|\boldsymbol{\gamma}_{j_2}^*\mathbf{B}\boldsymbol{\Sigma}_x\mathbf{B}\boldsymbol{\Sigma}_x\mathbf{B}\boldsymbol{\gamma}_{j_2}|^2\boldsymbol{\gamma}_{j_2}^*\mathbf{B}^*\boldsymbol{\Sigma}_x\mathbf{B}\boldsymbol{\gamma}_{j_2} \le \frac{K}{n^5}E(\boldsymbol{\gamma}_{j_2}^*\mathbf{B}^*\boldsymbol{\Sigma}_x\mathbf{B}\boldsymbol{\gamma}_{j_2})^3 \\
&\quad\le \frac{K}{n^5}\Big[\|\boldsymbol{\Gamma}^*\mathbf{B}\boldsymbol{\Sigma}_x\mathbf{B}^*\boldsymbol{\Gamma}\|^3\big(1+\max_{jk}E|x_{ij}^6|\big)\Big] = o(n^{-4}),
\end{aligned} \tag{5.25}$$

where we have used the facts that

$$\max_{ij}\mathrm{E}|X_{ij}^6| \le \eta_n^2 n \quad \text{and} \quad \mathbf{e}_i'\boldsymbol{\Gamma}^*\mathbf{B}\boldsymbol{\Sigma}_x\mathbf{B}^*\boldsymbol{\Gamma}\mathbf{e}_i \le \|\boldsymbol{\Sigma}_x\|^2/v^2.$$

Similarly, we can show that the other factor is also bounded by $o(n^{-4})$. As another example, we consider the product of the first term of $\boldsymbol{\gamma}_{j_1}^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_1}$ and the last term of $\boldsymbol{\gamma}_{j_2}^*\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_2}$. It is bounded by

$$\begin{aligned}
&\mathrm{E}|(\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1})^3\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_2}| \\
&\quad\le (\mathrm{E}|(\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1})^2\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}|^2\mathrm{E}|\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_2}|^2)^{1/2} \\
&\quad= o(n^{-2}),
\end{aligned}$$

where, similar to the proof of (5.25), one can show that

$$\mathrm{E}|\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_2}|^2 = \frac{1}{n^2}\mathrm{E}|\boldsymbol{\gamma}_{j_1}^*\mathbf{B}^*\boldsymbol{\Sigma}_x\mathbf{B}\boldsymbol{\gamma}_{j_1}|^2 = O(n^{-2}),$$

$$\mathrm{E}|(\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{k_1})^2\boldsymbol{\gamma}_{j_1}^*\mathbf{B}\boldsymbol{\gamma}_{j_2}|^2 \le \frac{K}{n^2}\mathrm{E}|\boldsymbol{\gamma}_{j_1}^*\mathbf{B}^*\boldsymbol{\Sigma}_x\mathbf{B}\boldsymbol{\gamma}_{j_1}|^3 = o(n^{-2}).$$

By using a similar approach, one can prove that the expectation of the other products with the number of $\mathbf{B}$ less than or equal to 5 are bounded by $o(n^{-2})$.

Case 2. *Terms with four* $\mathbf{A}^{-1}_{j_1 j_2 k_1 k_2}(z)$*'s in* $\sum_{(4)}$. We shall apply the first term expansions of β_{j_1} or β_{j_2} as it is needed and then use the bound $|z|/v \le K$ for β's. Then we can show that such terms are also bounded by $o(n^{-2})$. As an example, consider the product of the first term of $\boldsymbol{\gamma}^*_{j_1}\mathbf{A}^{-1}(z)\boldsymbol{\gamma}_{k_1}$ and the 5th term of $\boldsymbol{\gamma}^*_{j_2}\overline{\mathbf{A}^{-1}(z)}\boldsymbol{\gamma}_{k_2}$. Its expectation is bounded by

$$
\begin{aligned}
&\big|\mathrm{E}\big(\beta_{j_1}\beta_{k_1(j_1)}\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_1}\big)\big(\beta_{j_2}\beta_{k_2(j_2)}\beta_{k_1(j_2k_2)}\beta_{j_1(j_2k_1k_2)}\boldsymbol{\gamma}^*_{j_2}\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}^*_{k_1}\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_2}\big)^*\big|\\
&\quad= \big|\mathrm{E}\big((\beta_{j_1}\beta_{k_1(j_1)}\beta_{j_2} - b_{j_1}b_{k_1(j_1)}b_{j_2})\beta_{k_2(j_2)}\beta_{k_1(j_2k_2)}\beta_{j_1(j_2k_1k_2)}\big)\\
&\qquad\qquad \times \boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_1}\big(\boldsymbol{\gamma}^*_{j_2}\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}^*_{k_1}\mathbf{B}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_2}\big)^*\big|\\
&\quad\le K\big(\mathrm{E}\big|(\beta_{j_1}\beta_{k_1(j_1)}\beta_{j_2} - b_{j_1}b_{k_1(j_1)}b_{j_2})\big(\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}^*_{j_2}\mathbf{B}\boldsymbol{\gamma}_{j_1}\big)\big|^2\\
&\qquad\qquad\qquad \times \mathrm{E}\big|\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_1}\big|^2\big|\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_2}\big|^2\big)^{1/2}\\
&\quad\le o\big(n^{-2}\big).
\end{aligned}
$$

Here, we have used the fact that by expanding of the β-functions, we have

$$
\big|\beta_{j_1}\beta_{k_1(j_1)}\beta_{j_2} - b_{j_1}b_{k_1(j_1)}b_{j_2}\big| \le K\big(|\varepsilon_{j_1}| + |\varepsilon_{j_2}| + |\varepsilon_{k_1(j_1)}|\big).
$$

Then by the same approach employed in case 1, one can show that the bounds for

$$
\mathrm{E}\big|\varepsilon_{j_2}\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}^*_{j_2}\mathbf{B}\boldsymbol{\gamma}_{j_1}\big|^2 = O\big(n^{-3}\big), \qquad \mathrm{E}\big|\varepsilon_{k_1(j_1)}\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}^*_{j_2}\mathbf{B}\boldsymbol{\gamma}_{j_1}\big|^2 = O\big(n^{-3}\big)
$$

and

$$
\mathrm{E}\big|\varepsilon_{j_j}\boldsymbol{\gamma}^*_{j_1}\mathbf{B}\boldsymbol{\gamma}_{k_1}\boldsymbol{\gamma}^*_{j_2}\mathbf{B}\boldsymbol{\gamma}_{j_1}\big|^2 = o\big(n^{-2}\big).
$$

Thus, the bound for the first factor is $o(n^{-2})$ and one can easily show that the bound for the second factor is $O(n^{-2})$.

Case 3. *Terms with less than four* $\mathbf{A}^{-1}_{j_1 j_2 k_1 k_2}(z)$*'s in* $\sum_{(4)}$. If there are three $\mathbf{A}^{-1}_{j_1 j_2 k_1 k_2}(z)$'s in $\sum_{(4)}$, we need to use the first term expansion of β_{j_1} and further expand the matrix $\mathbf{A}_{j_1}$ in ε_{j_1} as

$$
\mathbf{A}^{-1}_{j_1 j_2} - \mathbf{A}^{-1}_{j_1 j_2}\boldsymbol{\gamma}_{j_2}\boldsymbol{\gamma}^*_{j_2}\mathbf{A}^{-1}_{j_1 j_2}\beta_{j_2(j_1)},
$$

and expand

$$
\mathbf{A}^{-1}_{j_2} = \mathbf{A}^{-1}_{j_1 j_2} - \mathbf{A}^{-1}_{j_1 j_2}\boldsymbol{\gamma}_{j_1}\boldsymbol{\gamma}^*_{j_1}\mathbf{A}^{-1}_{j_1 j_2}\beta_{j_1(j_2)}
$$

in ε_{j_2}, and then use the approach employed in case 2 to obtain the desired bound. Then we can also show that the expectation of the product is controlled by $o(n^{-2})$. If there are two $\mathbf{A}^{-1}_{j_1 j_2 k_1 k_2}(z)$, we need to further expand the inverses of $\mathbf{A}$-matrices.

The details are omitted. Finally, we obtain that $\sum_{(4)} = o(1)$. Similarly, we have $\sum_{(3)} = o(1)$. Because

$$\operatorname{tr}(\mathbf{A}^{-2}\boldsymbol{\Delta}) = \frac{d}{dz}\operatorname{tr}(\mathbf{A}^{-1}\boldsymbol{\Delta}),$$

then applying Cauchy integral, we can prove that

$$\mathrm{E}\big|\operatorname{tr}(\mathbf{A}^{-2}\boldsymbol{\Delta})\big|^2 = \mathrm{E}\left|\frac{1}{n}\sum_{j\neq k}\boldsymbol{\gamma}_j^*\mathbf{A}^{-2}\boldsymbol{\gamma}_k\right|^2 = o(1).$$

The lemma is proved. $\square$

LEMMA 5.4. *After truncation and normalization, for $z \in \mathbb{C}^+$, $\operatorname{tr}(\mathbf{A}^{-2}\boldsymbol{\Delta}\mathbf{A}^{-1}\boldsymbol{\Delta})$ converges to*

$$\big(\underline{m}_y(z) + z\underline{m}_y'(z)\big)\big(1 + z\underline{m}_y(z)\big) \qquad \text{in } L_2.$$

PROOF. Set

$$\operatorname{tr}\mathbf{A}^{-1}(z_1)\boldsymbol{\Delta}\mathbf{A}^{-1}(z_2)\boldsymbol{\Delta} = \frac{1}{n^2}\sum_{i\neq k, j\neq t}\boldsymbol{\gamma}_i^*\mathbf{A}^{-1}(z_1)\boldsymbol{\gamma}_k\boldsymbol{\gamma}_j^*\mathbf{A}^{-1}(z_2)\boldsymbol{\gamma}_t = Q_1 + Q_2,$$

where

$$Q_1 = \frac{1}{n^2}\sum_{j\neq k}^{n}\boldsymbol{\gamma}_j^*\mathbf{A}^{-1}(z_1)\boldsymbol{\gamma}_j\boldsymbol{\gamma}_k^*\mathbf{A}^{-1}(z_2)\boldsymbol{\gamma}_k \quad \text{and}$$

$$Q_2 = \frac{1}{n^2}\sum_{\substack{i\neq k, j\neq t\\ i\neq j, \text{ or } k\neq t}}\boldsymbol{\gamma}_i^*\mathbf{A}^{-1}(z_1)\boldsymbol{\gamma}_k\boldsymbol{\gamma}_j^*\mathbf{A}^{-1}(z_2)\boldsymbol{\gamma}_t.$$

By Lemmas 5.2 and 5.3, we obtain

$$\mathrm{E}\big|Q_1 - \big(1 + z\underline{m}_y(z_1)\big)\big(1 + z\underline{m}_y(z_2)\big)\big|^2 \le Kn^{-1}$$

and $\mathrm{E}|Q_2|^2 = o(1)$. We thus have

$$\mathrm{E}\big|\operatorname{tr}\mathbf{A}^{-1}(z_1)\boldsymbol{\Delta}\mathbf{A}^{-1}(z_2)\boldsymbol{\Delta} - \big(1 + z\underline{m}_y(z_1)\big)\big(1 + z\underline{m}_y(z_2)\big)\big|^2 = o(1).$$

Consequently, using

$$\operatorname{tr}\mathbf{A}^{-2}(z_1)\boldsymbol{\Delta}\mathbf{A}^{-1}(z_2)\boldsymbol{\Delta} = \frac{\partial\operatorname{tr}\mathbf{A}^{-1}(z_1)\boldsymbol{\Delta}\mathbf{A}^{-1}(z_2)\boldsymbol{\Delta}}{\partial z_1},$$

we have

$$\mathrm{E}\left|\operatorname{tr}\mathbf{A}^{-2}(z_1)\boldsymbol{\Delta}\mathbf{A}^{-1}(z_2)\boldsymbol{\Delta} - \frac{\partial}{\partial z_1}g(z_1)g(z_2)\right|^2 = o(1).$$

That is, $\operatorname{tr}\mathbf{A}^{-2}(z_1)\boldsymbol{\Delta}\mathbf{A}^{-1}(z_2)\boldsymbol{\Delta}$ converges to $g(z_2)g'(z_1)$ in L_2 where $g(z) = 1 + z\underline{m}_y(z)$. By setting $z_1 = z_2 = z$, we obtain $\operatorname{tr}(\mathbf{A}^{-2}\boldsymbol{\Delta}\mathbf{A}^{-1}\boldsymbol{\Delta})$ converges to $g(z)g'(z)$ in L_2. $\square$

LEMMA 5.5. *After truncation and normalization, we have for* $z \in \mathbb{C}^+$,

$$\operatorname{tr}(\mathbf{A}^{-1}\boldsymbol{\Delta})^3(\mathbf{A}-\boldsymbol{\Delta})^{-1} = g(z)\operatorname{tr}\bigl((\mathbf{A}^{-1}\boldsymbol{\Delta})^2(\mathbf{A}-\boldsymbol{\Delta})^{-1}\bigr) + o_p(1).$$

PROOF. We have

$$\begin{aligned}
\operatorname{tr}(\mathbf{A}^{-1}\boldsymbol{\Delta})^3(\mathbf{A}-\boldsymbol{\Delta})^{-1} &= E\frac{1}{n^3}\sum_{\substack{i\neq t, j\neq g\\ h\neq s}} \boldsymbol{\gamma}_i^*\mathbf{A}^{-1}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_g^*\mathbf{A}^{-1}\boldsymbol{\gamma}_h\boldsymbol{\gamma}_s^*(\mathbf{A}-\boldsymbol{\Delta})^{-1}\mathbf{A}^{-1}\boldsymbol{\gamma}_t \\
&= \frac{1}{n^3}\sum_{\substack{i\neq t, j\neq g\\ i=j, h\neq s}} \boldsymbol{\gamma}_i^*\mathbf{A}^{-1}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_g^*\mathbf{A}^{-1}\boldsymbol{\gamma}_h\boldsymbol{\gamma}_s^*(\mathbf{A}-\boldsymbol{\Delta})^{-1}\mathbf{A}^{-1}\boldsymbol{\gamma}_t \\
&\quad + \frac{1}{n^3}\sum_{\substack{i\neq t, j\neq g\\ i\neq j, h\neq s}} \boldsymbol{\gamma}_i^*\mathbf{A}^{-1}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_g^*\mathbf{A}^{-1}\boldsymbol{\gamma}_h\boldsymbol{\gamma}_s^*(\mathbf{A}-\boldsymbol{\Delta})^{-1}\mathbf{A}^{-1}\boldsymbol{\gamma}_t \\
&= g(z)\frac{1}{n^2}\sum_{h\neq s}\boldsymbol{\gamma}_g^*\mathbf{A}^{-1}\boldsymbol{\gamma}_h\boldsymbol{\gamma}_s^*(\mathbf{A}-\boldsymbol{\Delta})^{-1}\mathbf{A}^{-1}\boldsymbol{\gamma}_t + o_p(1) \\
&= g(z)\operatorname{tr}\bigl((\mathbf{A}^{-1}\boldsymbol{\Delta})^2(\mathbf{A}-\boldsymbol{\Delta})^{-1}\bigr) \\
&\quad + g(z)\frac{1}{n^2}\sum_{\substack{g=t\\ h\neq s}}\boldsymbol{\gamma}_g^*\mathbf{A}^{-1}\boldsymbol{\gamma}_h\boldsymbol{\gamma}_s^*(\mathbf{A}-\boldsymbol{\Delta})^{-1}\mathbf{A}^{-1}\boldsymbol{\gamma}_t + o_p(1) \\
&= g(z)\operatorname{tr}\bigl((\mathbf{A}^{-1}\boldsymbol{\Delta})^2(\mathbf{A}-\boldsymbol{\Delta})^{-1}\bigr) + o_p(1).
\end{aligned}$$

Then the lemma is proved. $\square$

Finally, by Lemma 5.4, we have

$$\begin{aligned}
\operatorname{tr}(\mathbf{A}^{-1}\boldsymbol{\Delta})^2(\mathbf{A}-\boldsymbol{\Delta})^{-1} &= \operatorname{tr}(\mathbf{A}^{-1}\boldsymbol{\Delta})^2(\mathbf{A})^{-1} + \operatorname{tr}(\mathbf{A}^{-1}\boldsymbol{\Delta})^3(\mathbf{A}-\boldsymbol{\Delta})^{-1} \\
&= \operatorname{tr}(\mathbf{A}^{-1}\boldsymbol{\Delta})^2(\mathbf{A})^{-1} + g(z)\operatorname{tr}(\mathbf{A}^{-1}\boldsymbol{\Delta})^2(\mathbf{A}-\boldsymbol{\Delta})^{-1} + o_p(1) \\
&= \frac{(1+z\underline{m}_y(z))(\underline{m}_y(z)+z\underline{m}'_y(z))}{1-g(z)} + o_p(1).
\end{aligned}$$

Hence, we obtain the following lemma.

LEMMA 5.6. *After truncation and normalization, for* $z \in \mathbb{C}^+$, *we have*

$$\begin{aligned}
&\operatorname{tr}\mathbf{A}^{-2}(z)\boldsymbol{\Delta} + \operatorname{tr}\mathbf{A}^{-1}(z)\bigl(\boldsymbol{\Delta}\mathbf{A}^{-1}(z)\bigr)^2 + \operatorname{tr}\bigl(\mathbf{A}(z)-\boldsymbol{\Delta}\bigr)^{-1}\bigl(\boldsymbol{\Delta}\mathbf{A}^{-1}(z)\bigr)^3 \\
&\quad = \frac{(\underline{m}_y(z)+z\underline{m}'_y(z))(1+z\underline{m}_y(z))}{-z\underline{m}_y(z)} + o_p(1).
\end{aligned}$$

APPENDIX: COMPLEMENTS ON THE CLT THEOREM 2.1

This appendix is intended to give more discussions on the CLT in Theorem 2.1.

A.1. The special case where $\boldsymbol{\Sigma}_x \equiv \boldsymbol{I}_p$. In this special case, the CLT for linear spectral statistics is well known since [5] and the limiting mean and covariance functions can be simplified significantly. Here, we report a recent version proposed in [22]. Then $H_p \equiv \delta_{\{1\}} = H$ and the LSD $F^{y_1,H}$ becomes the standard Marčenko–Pastur distribution F^{y_1} of index y_1.

PROPOSITION A.1. *Under the conditions of Theorem 2.1 and assume moreover that $\boldsymbol{\Sigma}_x \equiv \mathbf{I}_p$. Then the mean and covariance function of the Gaussian limit $(X_{f_1}, \dots, X_{f_k})$ equal to*

$$\mathbb{E}[X_f] = (\kappa - 1)I_1(f) + \beta_x I_2(f), \tag{A.1}$$

$$\operatorname{Cov}(X_f, X_g) = \kappa J_1(f, g) + \beta_x J_2(f, g), \tag{A.2}$$

where with $h_0 = \sqrt{y}_1$,

$$I_1(f) = \lim_{r \downarrow 1} \frac{1}{2\pi \mathbf{i}} \oint_{|\xi|=1} f\left(|1 + h_0 \xi|^2\right) \left[\frac{\xi}{\xi^2 - r^{-2}} - \frac{1}{\xi} \right] d\xi, \tag{A.3}$$

$$I_2(f) = \frac{1}{2\pi \mathbf{i}} \oint_{|\xi|=1} f\left(|1 + h_0 \xi|^2\right) \frac{1}{\xi^3} \, d\xi, \tag{A.4}$$

$$J_1(f, g) = \lim_{r \downarrow 1} \frac{-1}{4\pi^2} \oint_{|\xi_1|=1} \oint_{|\xi_2|=1} \frac{f(|1 + h_0 \xi_1|^2) g(|1 + h_0 \xi_2|^2)}{(\xi_1 - r \xi_2)^2} \, d\xi_1 \, d\xi_2, \tag{A.5}$$

$$J_2(f, g) = -\frac{1}{4\pi^2} \oint_{|\xi_1|=1} \frac{f(|1 + h_0 \xi_1|^2)}{\xi_1^2} \, d\xi_1 \oint_{|\xi_2|=1} \frac{g(|1 + h_0 \xi_2|^2)}{\xi_2^2} \, d\xi_2. \tag{A.6}$$

A.2. Comparison with the CLTs in [5] and [18]. Compared to the CLT in [5], Theorem 2.1 removes Gaussian-like second-order and fourth-order moment conditions and can then be applied to a broader range of populations, for example, non-Gaussian populations. The new CLT relies on a new condition that $\boldsymbol{\Gamma}^* \boldsymbol{\Gamma}$ *is diagonal*. This condition can hardly be relaxed as shown by the following examples. The first example is for complex population. The example shows that although $\boldsymbol{\Sigma}_x = \operatorname{diag}[1, 2, \dots, 1, 2]$ is diagonal and Gaussian-like fourth moment condition exists, there still is a counterexample that shows that the convergence of LSS of sample covariance matrices does not happen when $EX_{ij}^2 \neq 0$, $\boldsymbol{\Gamma}^* \boldsymbol{\Gamma}$ is not diagonal and $\boldsymbol{\Gamma}$ is complex. The counterexample shows that the conditions of real $\boldsymbol{\Gamma}$ and diagonal $\boldsymbol{\Gamma}^* \boldsymbol{\Gamma}$ are unremovable for Theorem 2.1 when the Gaussian-like second moment condition is not satisfied.

EXAMPLE A.1. Let $p = 2m$ and $\widetilde{\mathbf{T}} = \mathbf{\Gamma}^*\mathbf{\Gamma} = \mathbf{U}^*\mathbf{L}\mathbf{U}$ (i.e., $\mathbf{\Gamma} = \mathbf{L}^{1/2}\mathbf{U}$), where $\mathbf{\Sigma}_x = \mathbf{L} = \operatorname{diag}[1, 2, \dots, 1, 2]$,

$$\mathbf{U}^* = \frac{1}{\sqrt{2}} \operatorname{diag}\left[\begin{pmatrix} 1 & e^{i\theta_{1m}} \\ e^{i\theta_{2m}} & -e^{i(\theta_{1m}+\theta_{2m})} \end{pmatrix}, \dots, \begin{pmatrix} 1 & e^{i\theta_{1m}} \\ e^{i\theta_{2m}} & -e^{i(\theta_{1m}+\theta_{2m})} \end{pmatrix}\right].$$

The X_{ij}'s are i.i.d. and have a mixture distribution: with probability τ, $X_{ij} \stackrel{\mathcal{D}}{=} \frac{\sqrt{3}}{2} Y + \frac{\mathbf{i}}{2} Z$ and with probability $1 - \tau$, $X_{ij} \stackrel{\mathcal{D}}{=} \frac{\sqrt{3}}{2} W + \frac{\mathbf{i}}{2} V$ where Y, Z are i.i.d. standard normal and W, V are i.i.d. and take values ± 1 with probability $\frac{1}{2}$. Then, it is easy to verify that

$$\mathrm{E}X_{ij} = 0, \qquad \mathrm{E}|X_{ij}^2| = 1, \qquad \mathrm{E}X_{ij}^2 = \frac{1}{2}, \qquad \mathrm{E}|X_{ij}^4| = \frac{5\tau}{4} + 1.$$

Taking $\tau = \frac{4}{5}$, we will have $\mathrm{E}|X_{ij}^4| = 2$. For $f(x) = x$, the random part of the corresponding LSS is

$$A_n(f) = \operatorname{tr} \mathbf{S}_x^0 - \operatorname{tr} \widetilde{\mathbf{T}} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{x}_j^* \mathbf{x}_j - \operatorname{tr} \widetilde{\mathbf{T}}),$$

where $\mathbf{X}_j = (X_{1j}, \dots, X_{pj})'$ and $\mathbf{x}_j = \mathbf{\Gamma}\mathbf{X}_j$. The variance of $A_n(f)$ is

$$\frac{\beta_x}{n} \sum_{j=1}^{p} \tilde{t}_{jj}^2 + \frac{1}{n} \operatorname{tr} \widetilde{\mathbf{T}}^2 + \frac{|\mathrm{E}X_{11}^2|^2}{n} \operatorname{tr} \widetilde{\mathbf{T}}\widetilde{\mathbf{T}}' = \frac{\beta_x 3m + 5m}{n} + \frac{m(18 + 2\cos 2\theta_{2m})}{4n},$$

where $\beta_x = E|X_{11}^4| - |EX_{11}^2|^2 - 2$. Noting that $m/n \to y/2$, hence the normalized LSS does not converge in distribution if we choose θ_{2m} such that $\cos^2 \theta_{2m}$ does not have a limit.

The following example with a complex population shows that although $\mathbf{\Sigma}_x = \operatorname{diag}[1, 2, \dots, 1, 2]$ is diagonal and Gaussian-like second moment condition is satisfied, there still is a counterexample that shows that the convergence of the LSS of sample covariance matrices does not happen when $E|X_{ij}^4| \neq 2$ and $\mathbf{\Gamma}^*\mathbf{\Gamma}$ is not diagonal. Therefore, the condition that $\mathbf{\Gamma}^*\mathbf{\Gamma}$ is diagonal is unremovable for Theorem 2.1 when the Gaussian-like fourth moment condition is not satisfied.

EXAMPLE A.2. Let $p = 2m$, $\mathbf{\Sigma}_x = \mathbf{\Gamma}\mathbf{\Gamma}^*$ with $\mathbf{\Gamma} = \mathbf{L}^{1/2}\mathbf{U}$ and $\widetilde{\mathbf{T}} = \mathbf{\Gamma}^*\mathbf{\Gamma} = \mathbf{U}^*\mathbf{L}\mathbf{U}$, where $\mathbf{L} = \operatorname{diag}[1, 2, \dots, 1, 2]$ and

$$\mathbf{U}^* = \operatorname{diag}\left[\begin{pmatrix} \cos\theta_m & \sin\theta_m \\ -\sin\theta_m & \cos\theta_m \end{pmatrix}, \dots, \begin{pmatrix} \cos\theta_m & \sin\theta_m \\ -\sin\theta_m & \cos\theta_m \end{pmatrix}\right].$$

The X_{ij}'s are i.i.d. and have a mixture distribution: with probability τ, $X_{ij} \stackrel{\mathcal{D}}{=} \frac{1}{\sqrt{2}}(Y + \mathbf{i}Z)$ and with probability $1 - \tau$, $X_{ij} \stackrel{\mathcal{D}}{=} \frac{1}{\sqrt{2}}(W + \mathbf{i}V)$ where Y, Z, W and V have the same distribution as given in Example A.1. Then it is easy to verify that

$$\mathrm{E}X_{ij} = 0, \qquad \mathrm{E}|X_{ij}^2| = 1, \qquad \mathrm{E}X_{ij}^2 = 0, \qquad \mathrm{E}|X_{ij}^4| = 1 + \tau.$$

Choose $f(x) = x$, then the random part of the corresponding LSS is

$$A_n(f) = \operatorname{tr} \mathbf{S}_x^0 - \operatorname{tr} \widetilde{\mathbf{T}} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{x}_j^* \mathbf{x}_j - \operatorname{tr} \widetilde{\mathbf{T}}),$$

where $\mathbf{X}_j = (X_{1j}, \ldots, X_{pj})'$ and $\mathbf{x}_j = \mathbf{\Gamma}\mathbf{X}_j$. The variance of $A_n(f)$ is

$$\frac{1}{n} \operatorname{tr} \widetilde{\mathbf{T}}^2 + \frac{\tau - 1}{n} \sum_{i=1}^{p} \tilde{t}_{ii}^2 = \frac{5m}{n} + \frac{m(\tau - 1)(5 - 2\cos^2\theta_m \sin^2\theta_m)}{n}.$$

Again, the normalized LSS does not converge in distribution if we choose θ_m such that $\cos^2\theta_m$ does not have a limit.

The following example with a real population shows that although $\mathbf{\Sigma}_x = \operatorname{diag}[1, 2, \ldots, 1, 2]$ is diagonal and Gaussian-like second moment condition is satisfied, there still is a counterexample where the convergence of LSS of sample covariance matrices does not happen when $EX_{ij}^4 \neq 3$ and $\mathbf{\Gamma}^*\mathbf{\Gamma}$ is not diagonal. Therefore, the condition that $\mathbf{\Gamma}^*\mathbf{\Gamma}$ is diagonal is unremovable for Theorem 2.1 when the Gaussian-like fourth moment condition is not satisfied.

EXAMPLE A.3. Choose $\mathbf{\Gamma}$ as same as in Example A.2 and let X_{ij}'s be i.i.d. and distributed as $\sqrt{3/5}$ times a t-distribution with degrees of freedom 5. Then it is easy to verify that

$$\mathrm{E}X_{ij} = 0, \qquad \mathrm{E}X_{ij}^2 = 1, \qquad \mathrm{E}X_{ij}^4 \neq 3.$$

Again the variance of $A_n(f)$ is

$$\frac{2}{n} \operatorname{tr} \widetilde{\mathbf{T}}^2 + \frac{\mathrm{E}X_{ij}^4 - 3}{n} \sum_{j=1}^{p} \tilde{t}_{jj}^2 = \frac{10m}{n} + \frac{(\mathrm{E}X_{ij}^4 - 3)m(5 - 2\cos^2\theta_m \sin^2\theta_m)}{n}.$$

Hence, the normalized LSS does not converge in distribution if we choose θ_m such that $\cos^2\theta_m$ does not have a limit.

Acknowledgments. The authors thank the anonymous reviewers including an Associate Editor, as well as the Editors, for their helpful comments and suggestions that led to the improvement of this paper.

REFERENCES

[1] 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

[2] BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2013). Testing linear hypotheses in high-dimensional regressions. *Statistics* **47** 1207–1223. MR3175743

[3] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

[4] BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051

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

[6] BAI, Z. D., YIN, Y. Q. and KRISHNAIAH, P. R. (1987). On limiting empirical distribution function of the eigenvalues of a multivariate F matrix. *Theory Probab. Appl.* **32** 537–548.

[7] BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

[8] BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

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

[10] CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. MR2847949

[11] CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557

[12] CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

[13] JOHNSTONE, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I* 307–333. Eur. Math. Soc., Zürich. MR2334195

[14] 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

[15] LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938

[16] LYTOVA, A. and PASTUR, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840. MR2561434

[17] PAN, G. (2014). Comparison between two types of large sample covariance matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 655–677. MR3189088

[18] PAN, G. M. and ZHOU, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244

[19] SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408

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

[21] SRIVASTAVA, M. S., KOLLO, T. and VON ROSEN, D. (2011). Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *J. Multivariate Anal.* **102** 1090–1103. MR2793878

[22] WANG, Q. and YAO, J. (2013). On the sphericity test with large-dimensional observations. *Electron. J. Stat.* **7** 2164–2192. MR3104916

[23] ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263

S. ZHENG
Z. BAI
SCHOOL OF MATHEMATICS AND STATISTICS AND KLAS
NORTHEAST NORMAL UNIVERSITY
CHANGCHUN CITY, 130024
P. R. CHINA
E-MAIL: zhengsr@nenu.edu.cn
baizd@nenu.edu.cn

J. YAO
DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF HONG KONG
POKFULAM ROAD, HONG KONG
P. R. CHINA
E-MAIL: jeffyao@hku.hk