Bernoulli **28**(1), 2022, 606–637
https://doi.org/10.3150/21-BEJ1360

On eigenvalues of a high-dimensional spatial-sign covariance matrix

WEIMING LI[1], QINWEN WANG[2], JIANFENG YAO[3] and WANG ZHOU[4]

[1]*School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China. E-mail: li.weiming@shufe.edu.cn*
[2]*School of Data Science, Fudan University, Shanghai, China. E-mail: wqw@fudan.edu.cn*
[3]*Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China. E-mail: jeffyao@hku.hk*
[4]*Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117546.. E-mail: stazw@nus.edu.sg*

This paper investigates limiting spectral properties of a high-dimensional sample spatial-sign covariance matrix when both the dimension of the observations and the sample size grow to infinity. The underlying population is general enough to include the popular independent components model and the family of elliptical distributions. The first result of the paper shows that the empirical spectral distribution of a high dimensional sample spatial-sign covariance matrix converges to a generalized Marčenko-Pastur distribution. Secondly, a new central limit theorem for a class of related linear spectral statistics is established.

Keywords: Central limit theorem; eigenvalue distribution; linear spectral statistics; spatial-sign covariance matrix

1. Introduction

The so-called *spatial-sign covariance matrix* (SSCM) originates from the field of robust statistics. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ be a sequence of *independent and identically distributed* (i.i.d.) observations from a common population $\mathbf{x} \in \mathbb{R}^p$ structured as

$$\mathbf{x} = \mathbf{m} + w\mathbf{A}^{\frac{1}{2}}\mathbf{z}, \tag{1.1}$$

where $\mathbf{m} \in \mathbb{R}^p$ is a known location vector, $\mathbf{A}$ is a $p \times p$ deterministic and positive definite matrix, $w \in \mathbb{R}$ and $\mathbf{z} \in \mathbb{R}^p$ are two (possibly dependent) random quantities. Such structure encompasses the popular *independent components model* (ICM, when the random variable w reduces to some fixed constant) and the family of *elliptical distributions* (when the random vector $\mathbf{z}$ is restricted to be standard normal $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$) as special cases (detailed discussions are referred to Remarks 2.1 and 2.2). The sample SSCM formed by the sample $\{\mathbf{x}_j\}$, referred as $\mathbf{B}_n$, throughout this paper is defined as

$$\mathbf{B}_n = \frac{p}{n}\sum_{j=1}^{n}\mathbf{s}(\mathbf{x}_j - \mathbf{m})\mathbf{s}(\mathbf{x}_j - \mathbf{m})', \tag{1.2}$$

where $\mathbf{s}(\mathbf{y}) = I_{(\mathbf{y}\neq 0)}\mathbf{y}/\|\mathbf{y}\|$ is the spatial-sign transform projecting the vector $\mathbf{y}$ onto the unit sphere in $\mathbb{R}^p$. In Locantore et al. [20] and Visuri, Koivunen and Oja [28], the authors demonstrated that SSCM is able to mitigate the impact of extreme outliers for the purpose of robust principal components analysis. Since then, SSCM has been widely adopted for statistical inference especially when the sample data exhibit heavy tails or tail dependence as in the case of elliptical distributions. Recent works concerning the properties of SSCM and its applications include Magyar and Tyler [22], Dürre, Vogel and

1350-7265

Fried [11], Li, Wang and Zou [19], Feng and Sun [12], Feng, Zou and Wang [13] and Chakraborty and Chaudhuri [9]. Despite the popularity of SSCM, asymptotic behaviors of its eigenvalues are not fully developed in high dimensional regimes, which then motivates the current topic studied in this manuscript.

This paper investigates both the first and second order spectrum limits of the sample SSCM $\mathbf{B}_n$ with general data structure (1.1) under the Marčenko-Pastur asymptotic regime [23], where the dimension of the population p diverges to infinity along with the sample size n, that is,

$$n \to \infty, \quad p = p(n) \to \infty, \quad p/n = c_n \to c \in (0, \infty).$$

This asymptotic regime is commonly adopted in the literature of *random matrix theory* (RMT). The first result of the paper is a new generalized *Marčenko-Pastur* (MP) law for the *empirical spectral distribution* (ESD) of the sample SSCM $\mathbf{B}_n$. The MP law was originally introduced in [23] for the limiting spectrum of a high dimensional *sample covariance matrix* (SCM), which was then refined and extended in several works, say [26,30] and [3]. With this knowledge, by a comparison between $\mathbf{B}_n$ and its associated SCM, our result is derived under certain moment condition. The second contribution of this paper is a new *central limit theorem* (CLT) for general *linear spectral statistics* (LSSs) of $\mathbf{B}_n$. CLT for LSSs of certain random matrix ensembles has been actively studied in recent decades in RMT. Most of early works in this area concern Hermitian (symmetric) Wigner matrices. [17] presented a CLT for LSSs given the eigenvalues' joint density for Gaussian-type random Hermitian matrices. Using the moment method, [27] derived a CLT for polynomial functions of Wigner-type matrices and [1] obtained a CLT for a class of band random matrices. CLT for general Wigner matrices with arbitrary entries was first derived in [6] via Stieltjes transforms which provided explicit formulas for the mean and covariance functions of the limiting Gaussian distribution of the LSSs. A related approach using Gaussian interpolation for both Wigner matrices and Wishart matrices was proposed in [21]. As for SCMs, the earliest work dates back to [18] for Wishart matrices. The seminal paper [5] established the CLT under the ICM, which was later extended in [25] and [31]. Other extensions on CLT for SCMs are recently proposed in [15] and [16] for the class of elliptical distributions.

From the technical point of view, the structure of the sample SSCM $\mathbf{B}_n$ is quite different from the commonly studied SCM. Although $\mathbf{B}_n$ can be considered as SCM type by treating the spatial-sign transform of the original data $\{\mathbf{s}(\mathbf{x}_j - \mathbf{m})\}$ as a new data sample, the spatial-sign transform does introduce at the same time, complex non-linear correlations between the p-coordinates of $\mathbf{s}(\mathbf{x}_j - \mathbf{m})$ through the normalization by the Euclidean norm $\|\mathbf{x}_j - \mathbf{m}\|$ in its denominator. Such new correlations make the analysis more intricate in high dimensions. Specifically, let us compare the situation with a sample covariance matrix $\mathbf{S}_n = n^{-1}\sum_{j=1}^{n}(\mathbf{x}_j - \mathbf{m})(\mathbf{x}_j - \mathbf{m})'$ from the ICM (letting the random variable w in (1.1) be some constant). Here the correlations among the coordinates of a sample vector $\mathbf{x}_j - \mathbf{m}$ have only one source, coming from the shape matrix $\mathbf{A}$. However, in the case of sample SSCM $\mathbf{B}_n = pn^{-1}\sum_{j=1}^{n}\mathbf{s}(\mathbf{x}_j - \mathbf{m})\mathbf{s}(\mathbf{x}_j - \mathbf{m})'$, the correlations among the coordinates of $\mathbf{s}(\mathbf{x}_j - \mathbf{m})$ can originate from both the shape matrix $\mathbf{A}$ and the normalization factor $\|\mathbf{x}_j - \mathbf{m}\|$ in the denominator of $\mathbf{s}(\mathbf{x}_j - \mathbf{m})$. Therefore, a main task in our analysis is to find new approaches for decoupling these two sources of correlation in $\mathbf{s}(\mathbf{x}_j - \mathbf{m})$. To this end, by giving an asymptotic expansion of $\mathbf{s}(\mathbf{x}_j - \mathbf{m})$ to certain order, we develop new lemmas concerning the covariance and stochastic order of certain quadratic forms, which turns out to be one of the cornerstones for establishing our new CLT (see Section A.1). Another technical innovation of the paper, compared to the classical approach in [5], is that we introduce a new and more straightforward method to find the limiting mean function of LSSs, see Step 3 in the proof given in Section 3.2.

The rest of the paper is organized as follows. Section 2 presents our main theoretical results including both the convergence of the ESD of $\mathbf{B}_n$ and the CLT for its linear spectral statistics. Proofs of these asymptotic conclusions are presented in Sections 3.1 and 3.2. Some supporting lemmas and their proofs are relegated into the Appendix.

2. High-dimensional theory for eigenvalues of sample SSCMs

2.1. Preliminary

Let $\mathbf{M}_p$ be a $p \times p$ symmetric or Hermitian matrix with eigenvalues $(\lambda_j)_{1 \le j \le p}$. Its ESD is by definition the probability measure

$$F^{\mathbf{M}_p} = \frac{1}{p} \sum_{j=1}^{p} \delta_{\lambda_j},$$

where δ_b denotes the Dirac mass at b. If the ESD sequence $\{F^{\mathbf{M}_p}\}$ has a limit when $p \to \infty$, this limit is referred as the *limiting spectral distribution* (LSD). For a probability measure G, its Stieltjes transform is defined as

$$m_G(z) = \int \frac{1}{x - z} dG(x), \quad z \in \mathbb{C}^+,$$

where $\mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im(z) > 0\}$. This definition can be extended to the whole complex plane except the support set of G. Inversion formula from the Stieltjes transform $m_G(z)$ to its corresponding probability measure G can be found in [2].

2.2. Model assumptions

We consider a sequence of i.i.d. observations $\mathbf{x}_1, \dots, \mathbf{x}_n$ admitting the following stochastic representation

$$\mathbf{x}_j = \mathbf{m} + w_j \mathbf{A}^{\frac{1}{2}} \mathbf{z}_j, \quad j = 1, \dots, n, \tag{2.1}$$

where

(i) the location vector $\mathbf{m} \in \mathbb{R}^p$;
(ii) the scalar random variable w_j is real-valued satisfying $\mathrm{P}(w_j \ne 0) = 1$;
(iii) the matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$, referred as the *shape matrix* or *scatter matrix* of the population, is deterministic, positive definite, and normalized by $\mathrm{tr}(\mathbf{A}) = p$ for the identification in the triple product $w_j \mathbf{A}^{1/2} \mathbf{z}_j$, since we can always move any scalar factor related to $\mathbf{A}$ into the scalar random variable w_j;
(iv) the vector $\mathbf{z}_j = (z_{1j}, \dots, z_{pj})' \in \mathbb{R}^p$ is an array of i.i.d. standardized random variables and possibly dependent of the scalar random variable w_j.

Our main assumptions are as follows.

Assumption (a). Both the sample size n and population dimension p tend to infinity in such a way that $n \to \infty$, $p = p(n) \to \infty$ and $p/n = c_n \to c \in (0, \infty)$.

Assumption (b). The ESD H_p of the shape matrix $\mathbf{A}$ has a bounded support, that is, $\mathrm{Supp}(H_p) \subset [a, b]$ for some $a, b \in (0, \infty)$, and converges weakly to a probability distribution H as $p \to \infty$.

Assumption (c). The random variables (z_{ij}) are i.i.d. and satisfy

$$\mathbb{E}(z_{ij}) = 0, \qquad \mathbb{E}(z_{ij}^2) = 1, \qquad \mathbb{E}(z_{ij}^4) = \tau, \qquad \mathbb{E}|z_{ij}|^{4+\delta} < \infty,$$

for some $\delta > 0$.

Assumptions (a) and (b) are standard in RMT while Assumption (c) poses slightly higher moment restriction than that for the sample covariance matrices studied in [5]. Such stronger moment condition is imposed here for controlling the fluctuation of the normalization factor $\|\mathbf{x}_j - \mathbf{m}\|$ in the denominator of $\mathbf{s}(\mathbf{x}_j - \mathbf{m})$ in the definition of SSCM $\mathbf{B}_n$ in (1.2).

Remark 2.1. Recall that in the literature on high-dimensional SCMs, the following ICM is routinely considered [5,25,29,31]

$$\mathbf{x}_j = \mathbf{m} + \mathbf{A}^{\frac{1}{2}}\mathbf{z}_j, \tag{2.2}$$

where $\mathbf{m}$ and $\mathbf{z}_j$ are the same as in model (2.1), $\mathbf{A}$ is a $p \times p$ positive definite population covariance matrix. Clearly the model (2.2) is a particular case of the model (2.1) where $\{w_j\}$ degenerate to a constant parameter.

Remark 2.2. The model (2.1) contains also the family of elliptical distributions. Indeed, a generalized elliptically distributed sample $\mathbf{x}_j \in \mathbb{R}^p$ has the form

$$\mathbf{x}_j = \mathbf{m} + v_j\mathbf{A}^{\frac{1}{2}}\mathbf{u}_j, \tag{2.3}$$

where v_j is a scalar random variable and $\mathbf{u}_j$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$. Let $\mathbf{u}_j = \mathbf{z}_j/\|\mathbf{z}_j\|$, $w_j = v_j/\|\mathbf{z}_j\|$ and $\mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ in (2.3), we have

$$\mathbf{x}_j = \mathbf{m} + v_j\mathbf{A}^{\frac{1}{2}}\mathbf{u}_j = \mathbf{m} + w_j\mathbf{A}^{\frac{1}{2}}\mathbf{z}_j.$$

Certainly the moment conditions in Assumption (c) are satisfied with $\tau = 3$ for such standard Gaussian random vectors $\{\mathbf{z}_j\}$. Thus the generalized elliptical distributions described by (2.3) are also special cases of our model (2.1).

2.3. Limiting spectral distribution of $\mathbf{B}_n$

Our first result establishes the convergence of the ESD $F^{\mathbf{B}_n}$ of the sample SSCM $\mathbf{B}_n$ defined in (1.2).

Theorem 2.1. *Suppose that Assumptions (a)–(c) hold. Then, almost surely, the empirical spectral distribution $F^{\mathbf{B}_n}$ converges weakly to a probability distribution $F^{c,H}$, whose Stieltjes transform $m = m(z)$ is the unique solution to the equation*

$$m = \int \frac{1}{t(1 - c - czm) - z} dH(t), \qquad z \in \mathbb{C}^+, \tag{2.4}$$

in the set $\{m \in \mathbb{C} : -(1-c)/z + cm \in \mathbb{C}^+\}$.

Theorem 2.1 demonstrates that the ESD $F^{\mathbf{B}_n}$ converges to the generalized MP law $F^{c,H}$ defined through the equation (2.4), see [23]. Let $\underline{F}^{c,H} = cF^{c,H} + (1-c)\delta_0$ be the companion distribution of $F^{c,H}$ and $\underline{m} = \underline{m}(z)$ be the Stieltjes transform of $\underline{F}^{c,H}$. Then (2.4) can be rewritten as

$$z = -\frac{1}{\underline{m}} + c\int \frac{t}{1+t\underline{m}}\, dH(t), \quad z \in \mathbb{C}^+, \tag{2.5}$$

see [26]. For procedures on finding the density function of $F^{c,H}$ and its support set from (2.4) or (2.5), one is referred to [2]. The proof of this theorem is presented in Section 3.1.

2.4. CLT for linear spectral statistics of $\mathbf{B}_n$

In this section, we study the fluctuation of LSSs. Given a measurable function f, the LSS of $\mathbf{B}_n$ associated with f is defined to be the statistic

$$\int f(x)\, dF^{\mathbf{B}_n}(x). \tag{2.6}$$

To centralize this statistic, we introduce a matrix $\mathbf{T}$ that is closely related to the shape matrix $\mathbf{A}$,

$$\mathbf{T} = \mathbf{A} - \frac{2}{p}\mathbf{A}^2 - \frac{\tau-3}{p}\mathbf{A}^{\frac{1}{2}}\operatorname{diag}(\mathbf{A})\mathbf{A}^{\frac{1}{2}} + \left(\frac{2}{p^2}\operatorname{tr}\mathbf{A}^2 + \frac{\tau-3}{p^2}\operatorname{tr}(\mathbf{A}\circ\mathbf{A})\right)\mathbf{A}, \tag{2.7}$$

where "$\circ$" denotes the Hadamard product of two matrices (more detailed discussion on the matrix $\mathbf{T}$ is referred to Remark 2.3). Let $\tilde{H}_p$ denote the ESD of $\mathbf{T}$ and $\underline{m}_0(z)$ be the finite-horizon proxy for $\underline{m}(z)$ by replacing the two limits (c, H) with their finite counterparts $(c_n, \tilde{H}_p)$ in (2.5), that is, the solution to

$$z = -\frac{1}{\underline{m}_0(z)} + c_n\int \frac{t}{1+t\underline{m}_0(z)}\, d\tilde{H}_p(t), \quad z \in \mathbb{C}^+. \tag{2.8}$$

Such $\underline{m}_0(z)$ uniquely defines a probability distribution, denoted by $F^{c_n,\tilde{H}_p}$, through

$$\underline{m}_0(z) = -\frac{1-c_n}{z} + c_n\int \frac{1}{x-z}\, dF^{c_n,\tilde{H}_p}(x). \tag{2.9}$$

By means of this distribution, we center the LSS in (2.6) as

$$G_n(f) \triangleq \int f(x)\, dG_n(x) = \int f(x)\, d\big[F^{\mathbf{B}_n}(x) - F^{c_n,\tilde{H}_p}(x)\big]. \tag{2.10}$$

In addition, we assume the limits of the following three auxiliary quantities exist, that is,

$$\begin{aligned}
\zeta_p &= \frac{1}{p}\operatorname{tr}[\mathbf{A}\circ\mathbf{A}] \to \zeta, \\
h_p(u) &= \frac{1}{p}\operatorname{tr}\big[\mathbf{A}^{\frac{1}{2}}(\mathbf{A}-u\mathbf{I})^{-1}\mathbf{A}^{\frac{1}{2}}\circ\mathbf{A}\big] \to h(u), \\
g_p(u,v) &= \frac{1}{p}\operatorname{tr}\big[\big(\mathbf{A}^{\frac{1}{2}}(\mathbf{A}-u\mathbf{I})^{-1}\mathbf{A}^{\frac{1}{2}}\big)\circ\big(\mathbf{A}^{\frac{1}{2}}(\mathbf{A}-v\mathbf{I})^{-1}\mathbf{A}^{\frac{1}{2}}\big)\big] \to g(u,v),
\end{aligned} \tag{2.11}$$

where u and v are two complex variables in $\mathbb{C}^+$. Such limits will contribute to the CLT for $G_n(f)$ when the fourth moment $\tau \neq 3$.

Theorem 2.2. *Suppose that Assumptions (a)–(c) hold with $\delta = 1$. Let $f_1, \ldots, f_k$ be k functions analytic on an open set that includes the interval*

$$I_c = \Big[\liminf_{p\to\infty} \lambda_{\min}^{\mathbf{A}} \delta_{(0,1)}(c)(1-\sqrt{c})^2, \limsup_{p\to\infty} \lambda_{\max}^{\mathbf{A}}(1+\sqrt{c})^2\Big].$$

Also let

$$\mathbf{Y}_n = p\big\{G_n(f_1), \ldots, G_n(f_k)\big\}$$

be the vector of k normalized LSSs with respect to $f_1, \ldots, f_k$. Then $\mathbf{Y}_n$ converges in distribution to a k-dimensional Gaussian random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ with mean function

$$\mathbb{E}(\xi_j) = -\frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}_1} f_j(z)\big[\mu_1(z) + (\tau-3)\mu_2(z)\big]\,dz,$$

where

$$\begin{aligned}
\mu_1(z) &= \int \frac{c(\underline{m}'t)^2\,dH(t)}{\underline{m}(1+\underline{m}t)^3} - \int \frac{2\underline{m}'(1+z\underline{m})t^2\,dH(t)}{(1+\underline{m}t)^2} \\
&\quad + \int \frac{(\alpha_2 t - t^2)\,dH(t)}{1+\underline{m}t} \int \frac{2c\underline{m}\underline{m}'t\,dH(t)}{(1+\underline{m}t)^2}, \\
\mu_2(z) &= \frac{c\underline{m}'}{\underline{m}^2} g_u'(u,v)\bigg|_{u=v=\frac{-1}{\underline{m}}} + \zeta \int \frac{(1+z\underline{m})t\underline{m}'\,dH(t)}{(1+\underline{m}t)^2} \\
&\quad - \frac{(1+z\underline{m})\underline{m}'}{\underline{m}^2} h'(u)\bigg|_{u=\frac{-1}{\underline{m}}} - \int \frac{c\underline{m}'t\,dH(t)}{(1+\underline{m}t)^2} h\bigg(\frac{-1}{\underline{m}}\bigg),
\end{aligned}$$

and covariance function

$$\operatorname{Cov}(\xi_j, \xi_\ell) = -\frac{1}{4\pi^2} \oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f_j(z) f_\ell(\tilde{z})\big[\sigma_1(z,\tilde{z}) + (\tau-3)\sigma_2(z,\tilde{z})\big]\,dz\,d\tilde{z},$$

where

$$\begin{aligned}
\sigma_1(z,\tilde{z}) &= \frac{2\partial^2}{\partial z \partial \tilde{z}}\bigg[\log \frac{\underline{m}(z) - \underline{m}(\tilde{z})}{\underline{m}(z)\underline{m}(\tilde{z})(z-\tilde{z})} + \bigg(\frac{\alpha_2}{c} + \frac{1}{c\underline{m}(z)} + \frac{1}{c\underline{m}(\tilde{z})}\bigg)\big(1+z\underline{m}(z)\big)\big(1+\tilde{z}\underline{m}(\tilde{z})\big) \\
&\quad - z\underline{m}(z) - \tilde{z}\underline{m}(\tilde{z}) - 2\bigg], \\
\sigma_2(z,\tilde{z}) &= \frac{\partial^2}{\partial z \partial \tilde{z}}\bigg[cg\bigg(\frac{-1}{\underline{m}(z)}, \frac{-1}{\underline{m}(\tilde{z})}\bigg) + \frac{\zeta}{c}\big(1+z\underline{m}(z)\big)\big(1+\tilde{z}\underline{m}(\tilde{z})\big) \\
&\quad - \big(1+z\underline{m}(z)\big)h\bigg(\frac{-1}{\underline{m}(\tilde{z})}\bigg) - \big(1+\tilde{z}\underline{m}(\tilde{z})\big)h\bigg(\frac{-1}{\underline{m}(z)}\bigg)\bigg],
\end{aligned}$$

in which $\alpha_2 = \int t^2 dH(t)$. The contours $\mathcal{C}_1$ and $\mathcal{C}_2$ are non-overlapping, closed, counter-clockwise orientated in the complex plane and enclosing the interval I_c.

The proof of this theorem is presented in Section 3.2.

Remark 2.3. The matrix $\mathbf{T}$ defined in (2.7) is actually an approximation of the population SSCM $\mathbf{\Sigma} \triangleq \mathbb{E}\mathbf{B}_n$. With the condition $\mathbb{E}|z_{ij}|^5 < \infty$, that is, $\delta = 1$ as assumed in Theorem 2.2, we can later prove that in terms of spectral norm, $\|\mathbf{\Sigma} - \mathbf{T}\| = o(p^{-1})$, see Lemma A.3. This ensures that we can centralize the LSS of $\mathbf{B}_n$ by $\int f(x)\,dF^{c_n,\tilde{H}_p}(x)$ (see (2.10)), which relies on the spectrum of the matrix $\mathbf{T}$. Note that under our general model settings in (2.1), the spectrum of the matrix $\mathbf{T}$ depends not only on the eigenvalues of the shape matrix $\mathbf{A}$ but also on its eigenvectors. However, for those elliptical distributions where we could assume $\mathbf{z}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ as discussed in Remark 2.2 and then the fourth moment $\tau = 3$, the spectrum of the matrix $\mathbf{T}$ depends only on the eigenvalues of $\mathbf{A}$ (up to the order p^{-1}). Indeed, Dürre, Tyler and Vogel [10] has already show that for those elliptical distributions, the shape matrix $\mathbf{A}$ shares the same eigenvectors as the population SSCM $\mathbf{\Sigma}$ and their eigenvalues have a one-to-one correspondence which can be represented through certain integrals. Our approximation, that is the one given in (2.7) and Lemma A.3, is however explicit and is not restricted to elliptical distributions only.

Remark 2.4. When the shape matrix $\mathbf{A} = \mathbf{I}_p$, we note that from (2.7), $\mathbf{T} \equiv \mathbf{A} = \mathbf{I}_p$ this time. Then the three auxiliary quantities defined in (2.11) are equal to their limits, that is, $\zeta_p = \zeta = 1$, $h_p(u) = h(u) = (1-u)^{-1}$ and $g_p(u,v) = g(u,v) = (1-u)^{-1}(1-v)^{-1}$. After a little bit calculation, we have all the quantities in Theorem 2.2 involving the factor $\tau - 3$ equal zero, which gives the fact that the CLT is actually independent of the fourth moment when the underlying shape matrix is identity.

Remark 2.5. Theorem 2.2 contains the CLT for LSSs of high dimensional correlation matrices [14]. To see this, consider the simplest case that $\mathbf{m} = \mathbf{0}$, $w_j \equiv 1$ and $\mathbf{A} = \mathbf{I}_p$ in (2.1), then the sample SSCM under study can be written as

$$\mathbf{B}_n = \frac{p}{n}\sum_{j=1}^{n} \frac{\mathbf{z}_j}{\|\mathbf{z}_j\|}\frac{\mathbf{z}_j'}{\|\mathbf{z}_j\|} = \frac{p}{n}\left(\frac{\mathbf{z}_1}{\|\mathbf{z}_1\|}, \dots, \frac{\mathbf{z}_n}{\|\mathbf{z}_n\|}\right)\left(\frac{\mathbf{z}_1}{\|\mathbf{z}_1\|}, \dots, \frac{\mathbf{z}_n}{\|\mathbf{z}_n\|}\right)'.$$

Denote its companion matrix as

$$\underline{\mathbf{B}}_n = \frac{p}{n}\left(\frac{\mathbf{z}_1}{\|\mathbf{z}_1\|}, \dots, \frac{\mathbf{z}_n}{\|\mathbf{z}_n\|}\right)'\left(\frac{\mathbf{z}_1}{\|\mathbf{z}_1\|}, \dots, \frac{\mathbf{z}_n}{\|\mathbf{z}_n\|}\right), \tag{2.12}$$

which shares the same non-zero eigenvalues as $\mathbf{B}_n$. Thus the result in Theorem 2.2 gives the CLT for LSSs of $\underline{\mathbf{B}}_n$. Now let's denote the data matrix as $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n) = (\mathbf{v}_1, \dots, \mathbf{v}_p)'$, where $\mathbf{z}_j$ is the j-th column (j-th observation) and $\mathbf{v}_j'$ is the j-th row (j-th coordinate) of $\mathbf{Z}$. Moreover, the table $\mathbf{Z}$ consists of independent and identically distributed entries across both the rows and columns so permuting the entries in $\mathbf{Z}$ will not change its distribution. On the other hand, the correlation matrix $\mathbf{R}_n$ associated with the data set $\mathbf{Z}$ can be expressed as

$$\mathbf{R}_n = \left(\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \dots, \frac{\mathbf{v}_p}{\|\mathbf{v}_p\|}\right)'\left(\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \dots, \frac{\mathbf{v}_p}{\|\mathbf{v}_p\|}\right), \tag{2.13}$$

which has the same structure (up to a constant factor) as $\underline{\mathbf{B}}_n$ in (2.12) by interchanging the roles of p and n. Therefore in the case of $\mathbf{A} = \mathbf{I}_p$, the CLT for LSSs of $\mathbf{R}_n$ is readily derived from an application of Theorem 2.2 to the matrix $\underline{\mathbf{B}}_n$.

2.5. Example

As an illustration, we exhibit the CLT for a widely used LSS which is the second moment of the eigenvalues of $\mathbf{B}_n$, denoted by

$$\hat{\beta}_2 = \frac{1}{p}\operatorname{tr}\bigl(\mathbf{B}_n^2\bigr).$$

We consider the case where the population shape matrix $\mathbf{A}$ is diagonal. Thus the matrix $\mathbf{T}$ given in (2.7) can be simplified as

$$\mathbf{T} = \mathbf{A} - \frac{\tau - 1}{p}\mathbf{A}^2 + \frac{\tau - 1}{p^2}\operatorname{tr}\mathbf{A}^2 \cdot \mathbf{A}, \tag{2.14}$$

whose spectrum depends on the eigenvalues of $\mathbf{A}$ only. Let $\alpha_{k,p} = \frac{1}{p}\operatorname{tr}(\mathbf{T}^k) = \int t^k \, d\tilde{H}_p(t)$, we have according to (2.14), $\alpha_{k,p} \to \alpha_k \triangleq \int t^k \, dH(t)$. Moreover, the three auxiliary quantities in (2.11) have limits

$$\zeta = \int t^2 \, dH(t), \qquad h(u) = \int \frac{t^2}{t-u}\, dH(t), \qquad g(u, v) = \int \frac{t^2}{(t-u)(t-v)}\, dH(t),$$

which are also functions of the eigenvalues of $\mathbf{A}$ only.

By the relations in (2.8) and (2.9), the centering term for the statistic $\hat{\beta}_2$ is

$$\beta_{2,p} \triangleq \int x^2 \, dF^{c_n, \tilde{H}_p}(x) = \alpha_{2,p} + c_n.$$

The limiting mean and variance of $p[\hat{\beta}_2 - \beta_{2,p}]$ can be figured out through the residue theorem. For illustration, we calculate the integral corresponding to the first term in $\mu_1(z)$, that is,

$$-\frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}_1} z^2 \int \frac{c(\underline{m}'(z)t)^2 \, dH(t)}{\underline{m}(z)(1 + \underline{m}(z)t)^3}\, dz. \tag{2.15}$$

Taking derivatives with respect to z on both sides of (2.8), we obtain

$$\underline{m}'(z) = \left(\frac{1}{\underline{m}^2(z)} - c\int \frac{t^2}{(1 + t\underline{m}(z))^2}\, dH(t)\right)^{-1}.$$

It then follows that

$$\begin{aligned}
(2.15) &= \int -\frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}_1} z^2 \frac{c\underline{m}'(z)t^2}{\underline{m}(z)(1 + \underline{m}(z)t)^3}\, d\underline{m}(z)\, dH(t) \\
&= \int -\frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}_1} \frac{ct^2(z\underline{m}(z))^2}{\underline{m}(z)(1 + \underline{m}(z)t)^3}\left(1 - c\int \frac{u^2\underline{m}^2(z)}{(1 + u\underline{m}(z))^2}\, dH(u)\right)^{-1} d\underline{m}(z)\, dH(t) \\
&= -\int \left\{ \frac{ct^2(z\underline{m}(z))^2}{(1 + \underline{m}(z)t)^3}\left(1 - c\int \frac{u^2\underline{m}^2(z)}{(1 + u\underline{m}(z))^2}\, dH(u)\right)^{-1} \Bigg|_{\underline{m}(z)=0} \right\} dH(t) \\
&= -c\alpha_2.
\end{aligned}$$

Similar procedure can be repeated to find the values of the remaining contour integrals. As a result and by Theorem 2.2, the distribution of $p[\hat{\beta}_2 - \beta_{2,p}]$ converges to a Gaussian distribution $N(\mu, \sigma^2)$, where the mean and variance parameters are given by

$$\mu = -c\alpha_2,$$
$$\sigma^2 = 8c\big(\alpha_2^3 - 2\alpha_2\alpha_3 + \alpha_4\big) + 4c^2\alpha_2^2 + 4c(\tau - 3)\big(\alpha_2^3 - 2\alpha_2\alpha_3 + \alpha_4\big).$$

3. Proofs of the main results

This section presents the proofs of Theorem 2.1 and Theorem 2.2. In all the proofs, we assume the location vector $\mathbf{m} = \mathbf{0}$, otherwise, it can be directly subtracted from the sample $\{\mathbf{x}_j\}$. We will denote by K some constants appearing in inequalities that can vary from place to place.

3.1. Proof of Theorem 2.1

Let $g_j = p/(\mathbf{z}_j'\mathbf{A}\mathbf{z}_j)$ for $j = 1, \dots, n$, and denote

$$\mathbf{Z} = (z_{ij}), \qquad \mathbf{G} = \operatorname{diag}(g_1, \dots, g_n), \qquad \mathbf{B}_n = \frac{1}{n}\mathbf{A}^{\frac{1}{2}}\mathbf{Z}\mathbf{G}\mathbf{Z}'\mathbf{A}^{\frac{1}{2}}, \qquad \mathbf{S}_n = \frac{1}{n}\mathbf{A}^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}'\mathbf{A}^{\frac{1}{2}}. \tag{3.1}$$

Under Assumptions (a)–(c), the generalized MP law holds true for the sample covariance matrix $\mathbf{S}_n$ [26]. Thus it is sufficient to show

$$\|\mathbf{B}_n - \mathbf{S}_n\| \xrightarrow{a.s.} 0. \tag{3.2}$$

To this end, with the moment conditions in Assumption (c), we shall truncate the variables (z_{ij}) at $n^{2/\gamma}$ for some $\gamma \in (4, 4+\delta]$. Some relevant quantities are denoted as below. For $i = 1, \dots, p$ and $j = 1, \dots, n$,

$$\begin{aligned} &\hat{z}_{ij} = z_{ij} I\big(|z_{ij}|^\gamma \le n^2\big), \qquad \hat{\mathbf{z}}_j = (\hat{z}_{1j}, \dots, \hat{z}_{pj})', \qquad \hat{g}_j = p/\big(\hat{\mathbf{z}}_j'\mathbf{A}\hat{\mathbf{z}}_j\big), \\ &\widehat{\mathbf{Z}} = (\hat{z}_{ij}), \qquad \widehat{\mathbf{G}} = \operatorname{diag}(\hat{g}_1, \dots, \hat{g}_n), \\ &\widehat{\mathbf{B}}_n = \frac{1}{n}\mathbf{A}^{\frac{1}{2}}\widehat{\mathbf{Z}}\widehat{\mathbf{G}}\widehat{\mathbf{Z}}'\mathbf{A}^{\frac{1}{2}}, \qquad \widehat{\mathbf{S}}_n = \frac{1}{n}\mathbf{A}^{\frac{1}{2}}\widehat{\mathbf{Z}}\widehat{\mathbf{Z}}'\mathbf{A}^{\frac{1}{2}}. \end{aligned} \tag{3.3}$$

Note that for the truncated variables $(\hat{z}_{ij})$, the following results hold automatically

$$\begin{aligned} &|\mathbb{E}\hat{z}_{ij}| = o\big(n^{-2+2/\gamma}\big), \qquad \mathbb{E}\big(\hat{z}_{ij}^2\big) = 1 + o\big(n^{-2+4/\gamma}\big), \\ &\mathbb{E}\big(\hat{z}_{ij}^4\big) = \tau + o(1), \qquad \mathbb{E}|\hat{z}_{ij}|^\gamma < \infty, \quad |\hat{z}_{ij}|^\gamma < n^2, \end{aligned} \tag{3.4}$$

and

$$\sum_k 2^k \mathbb{E}|z_{ij}|^{\gamma/2} I\big(|z_{ij}| > 2^{2k/\gamma}\big) < \infty. \tag{3.5}$$

From (3.5) and similar arguments as in the proof of Lemma 5.12 in [2], we have

$$\mathrm{P}(\widehat{\mathbf{B}}_n \ne \mathbf{B}_n, \text{ i.o.}) = \mathrm{P}(\widehat{\mathbf{S}}_n \ne \mathbf{S}_n, \text{ i.o.}) = 0. \tag{3.6}$$

Next, we will prove that for any $\varepsilon > 0$ and $k \geq 2$,

$$\mathrm{P}\big(\|\widehat{\mathbf{B}}_n - \widehat{\mathbf{S}}_n\| > \varepsilon\big) \leq K\varepsilon^{-k}\big(n^{-\frac{k}{2}+1} + n^{-\frac{k(\gamma-4)}{\gamma}}\big). \tag{3.7}$$

Notice that the spectral norm of the difference between $\widehat{\mathbf{B}}_n$ and $\widehat{\mathbf{S}}_n$ can be bounded by

$$\|\widehat{\mathbf{B}}_n - \widehat{\mathbf{S}}_n\| \leq \|\mathbf{A}\| \frac{\|\widehat{\mathbf{Z}}\widehat{\mathbf{Z}}'\|}{n} \max_{1\leq j\leq n} |\hat{g}_j - 1|. \tag{3.8}$$

From [4], almost surely, the spectral norm $\|\widehat{\mathbf{Z}}\widehat{\mathbf{Z}}'\|/n$ is bounded for all large n. Thus, we only need to control the convergence rate of $\max_j |\hat{g}_j - 1|$ or $\max_j |1/\hat{g}_j - 1|$. By Markov's inequality, for any $\varepsilon > 0$ and $k \geq 2$, we have

$$\mathrm{P}\left(\max_j\left|\frac{1}{\hat{g}_j} - 1\right| > \varepsilon\right) \leq np^{-k}\varepsilon^{-k}\mathbb{E}\big|\hat{\mathbf{z}}_1'\mathbf{A}\hat{\mathbf{z}}_1 - p\big|^k. \tag{3.9}$$

To bound the expectation in (3.9), we divide it into three parts as follows

$$\mathbb{E}\big|\hat{\mathbf{z}}_1'\mathbf{A}\hat{\mathbf{z}}_1 - p\big|^k \leq K\mathbb{E}\big|\hat{\mathbf{z}}_1'\mathbf{A}\hat{\mathbf{z}}_1 - \tilde{\mathbf{z}}_1'\mathbf{A}\tilde{\mathbf{z}}_1\big|^k + K\mathbb{E}\big|\tilde{\mathbf{z}}_1'\mathbf{A}\tilde{\mathbf{z}}_1 - \mathbb{E}\tilde{\mathbf{z}}_1'\mathbf{A}\tilde{\mathbf{z}}_1\big|^k + K\big|\mathbb{E}\tilde{\mathbf{z}}_1'\mathbf{A}\tilde{\mathbf{z}}_1 - p\big|^k,$$

where $\tilde{\mathbf{z}}_1 \triangleq \hat{\mathbf{z}}_1 - \mathbb{E}\hat{\mathbf{z}}_1$. From (3.4), the boundedness of $\|\mathbf{A}\|$ and Lemma A.1, the first term can be controlled by

$$\begin{aligned}
\mathbb{E}\big|\hat{\mathbf{z}}_1'\mathbf{A}\hat{\mathbf{z}}_1 - \tilde{\mathbf{z}}_1'\mathbf{A}\tilde{\mathbf{z}}_1\big|^k &\leq K\mathbb{E}\big|\tilde{\mathbf{z}}_1'\mathbf{A}\mathbb{E}\hat{\mathbf{z}}_1\big|^k + K\big|\mathbb{E}\hat{\mathbf{z}}_1'\mathbf{A}\mathbb{E}\hat{\mathbf{z}}_1\big|^k \\
&\leq K\mathbb{E}^{\frac{1}{2}}\big|\tilde{\mathbf{z}}_1'\tilde{\mathbf{z}}_1\big|^k\big|\mathbb{E}\hat{\mathbf{z}}_1'\mathbb{E}\hat{\mathbf{z}}_1\big|^{\frac{k}{2}} + Kn^{(-3+4/\gamma)k} \\
&\leq K\big[\mathbb{E}\big|\tilde{\mathbf{z}}_1'\tilde{\mathbf{z}}_1 - \mathbb{E}\tilde{\mathbf{z}}_1'\tilde{\mathbf{z}}_1\big|^k + \big|\mathbb{E}\tilde{\mathbf{z}}_1'\tilde{\mathbf{z}}_1\big|^k\big]^{\frac{1}{2}} n^{(-3/2+2/\gamma)k} + Kn^{(-3+4/\gamma)k} \\
&\leq K\mathbb{E}^{\frac{1}{2}}\big|\tilde{\mathbf{z}}_1'\tilde{\mathbf{z}}_1 - \mathbb{E}\tilde{\mathbf{z}}_1'\tilde{\mathbf{z}}_1\big|^k n^{(-3/2+2/\gamma)k} + Kn^{(-3+4/\gamma)k} \\
&\leq Kn^{-1/2+(-3/2+4/\gamma)k} + Kn^{(-1+2/\gamma)k},
\end{aligned} \tag{3.10}$$

where we use the abbreviation $\mathbb{E}^{\frac{1}{2}}(\cdot) \triangleq [\mathbb{E}(\cdot)]^{1/2}$ above and also throughout all the remaining proofs. Again from Lemma A.1, the second term is bounded by

$$\mathbb{E}\big|\tilde{\mathbf{z}}_1'\mathbf{A}\tilde{\mathbf{z}}_1 - \mathbb{E}\tilde{\mathbf{z}}_1'\mathbf{A}\tilde{\mathbf{z}}_1\big|^k \leq K\big(n^{k/2} + n\mathbb{E}|\tilde{z}_{11}|^{2k}\big) \leq K\big(n^{k/2} + n^{-1+4k/\gamma}\big). \tag{3.11}$$

For the third one, we have from (3.4)

$$\big|\mathbb{E}\tilde{\mathbf{z}}_1'\mathbf{A}\tilde{\mathbf{z}}_1 - p\big|^k = p^k\big|\mathrm{Var}(\hat{\mathbf{z}}_{11}) - 1\big|^k \leq Kn^{(-1+4/\gamma)k}. \tag{3.12}$$

Collecting the results in (3.10)–(3.12) yields

$$\mathbb{E}\big|\hat{\mathbf{z}}_1'\mathbf{A}\hat{\mathbf{z}}_1 - p\big|^k \leq K\big(n^{k/2} + n^{-1+4k/\gamma}\big), \tag{3.13}$$

which together with (3.8) and (3.9) give the result in (3.7). Hence, the conclusion of (3.2) follows from (3.6) and (3.7) by taking some large k. The proof is then complete.

3.2. Proof of Theorem 2.2

3.2.1. *Sketch of the proof*

Following the truncation step in the proof of Theorem 2.1, we now centralize the truncated variables. In addition to the notations in (3.1) and (3.3), some quantities are denoted as below.

$$\tilde{z}_{ij} = \hat{z}_{ij} - \mathbb{E}(\hat{z}_{ij}), \qquad \tilde{\mathbf{z}}_j = (\tilde{z}_{1j}, \ldots, \tilde{z}_{pj})', \qquad \tilde{g}_j = p/\big(\tilde{\mathbf{z}}_j'\mathbf{A}\tilde{\mathbf{z}}_j\big), \qquad \widetilde{\mathbf{G}} = \operatorname{diag}(\tilde{g}_1, \ldots, \tilde{g}_n),$$
$$\widetilde{\mathbf{Z}} = (\tilde{z}_{ij}), \qquad \widetilde{\mathbf{B}}_n = \frac{1}{n}\mathbf{A}^{\frac{1}{2}}\widetilde{\mathbf{Z}}\widetilde{\mathbf{G}}\widetilde{\mathbf{Z}}'\mathbf{A}^{\frac{1}{2}}, \qquad \overline{\mathbf{B}}_n = \frac{1}{n}\mathbf{A}^{\frac{1}{2}}\widetilde{\mathbf{Z}}\widehat{\mathbf{G}}\widetilde{\mathbf{Z}}'\mathbf{A}^{\frac{1}{2}}, \qquad \widetilde{\mathbf{S}}_n = \frac{1}{n}\mathbf{A}^{\frac{1}{2}}\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}'\mathbf{A}^{\frac{1}{2}}.$$

Similar to the derivation of (3.7), one may show that

$$\max\big\{\mathrm{P}\big(\|\widetilde{\mathbf{B}}_n - \widetilde{\mathbf{S}}_n\| > \varepsilon\big), \mathrm{P}\big(\|\overline{\mathbf{B}}_n - \widetilde{\mathbf{S}}_n\| > \varepsilon\big)\big\} \le K\varepsilon^{-k}\big(n^{-\frac{k}{2}+1} + n^{-\frac{k(\gamma-4)}{\gamma}}\big). \tag{3.14}$$

It thus follows from [4] that, almost surely, $\limsup_n \|\widehat{\mathbf{B}}_n\|$, $\limsup_n \|\widetilde{\mathbf{B}}_n\|$ and $\limsup_n \|\overline{\mathbf{B}}_n\|$ are all bounded.

Let $F^{\mathbf{B}_n}$, $F^{\widehat{\mathbf{B}}_n}$, $F^{\overline{\mathbf{B}}_n}$, and $F^{\widetilde{\mathbf{B}}_n}$ be the ESDs of the matrices $\mathbf{B}_n$, $\widehat{\mathbf{B}}_n$, $\overline{\mathbf{B}}_n$, and $\widetilde{\mathbf{B}}_n$, respectively. Then, for each function $f_j(x)$, we have from (3.6)

$$p\left|\int f_j(x)\,dF^{\mathbf{B}_n} - \int f_j(x)\,dF^{\widehat{\mathbf{B}}_n}\right| \xrightarrow{a.s.} 0. \tag{3.15}$$

By Corollary A.37 in [2], it holds that

$$\begin{aligned} p&\left|\int f_j(x)\,dF^{\widehat{\mathbf{B}}_n} - \int f_j(x)\,dF^{\overline{\mathbf{B}}_n}\right| \\ &\le K_j\sum_{k=1}^{p}\big|\lambda_k^{\widehat{\mathbf{B}}_n} - \lambda_k^{\overline{\mathbf{B}}_n}\big| \\ &\le 2K_j\big[c_n \operatorname{tr}\mathbf{A}^{\frac{1}{2}}(\widehat{\mathbf{Z}} - \widetilde{\mathbf{Z}})\widehat{\mathbf{G}}(\widehat{\mathbf{Z}} - \widetilde{\mathbf{Z}})'\mathbf{A}^{\frac{1}{2}}\big(\|\widehat{\mathbf{B}}_n\| + \|\overline{\mathbf{B}}_n\|\big)\big]^{1/2}. \end{aligned} \tag{3.16}$$

where K_j is an upper bound on $|f_j'(x)|$ and $\lambda_k^{\mathbf{B}}$ denotes the k-th largest eigenvalue of the matrix $\mathbf{B}$. By (3.4) and (3.9), one may get

$$\big|\operatorname{tr}\mathbf{A}^{\frac{1}{2}}(\widehat{\mathbf{Z}} - \widetilde{\mathbf{Z}})\widehat{\mathbf{G}}(\widehat{\mathbf{Z}} - \widetilde{\mathbf{Z}})'\mathbf{A}^{\frac{1}{2}}\big| \le \|\mathbf{A}\| \max_j |\hat{g}_j| \operatorname{tr}\big(\mathbb{E}\widehat{\mathbf{Z}}\mathbb{E}\widehat{\mathbf{Z}}'\big) \xrightarrow{a.s.} 0,$$

and thus (3.16) is $o_{a.s.}(1)$. Moreover, from (3.9) and (3.10), applying Markov's inequality, we have also

$$\begin{aligned} p\left|\int f_j(x)\,dF^{\overline{\mathbf{B}}_n} - \int f_j(x)\,dF^{\widetilde{\mathbf{B}}_n}\right| &\le K_j p\|\overline{\mathbf{B}}_n - \widetilde{\mathbf{B}}_n\| \\ &\le K_j p\|\mathbf{A}\| \cdot \left\|\frac{1}{n}\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}'\right\| \max_j |\hat{g}_j - \tilde{g}_j| \xrightarrow{a.s.} 0. \end{aligned} \tag{3.17}$$

Collecting (3.15), (3.16), and (3.17), we get

$$p\left|\int f_j(x)\,dF^{\mathbf{B}_n} - \int f_j(x)\,dF^{\widetilde{\mathbf{B}}_n}\right| \xrightarrow{a.s.} 0. \tag{3.18}$$

Therefore, it is sufficient to prove the theorem by replacing the matrix $\mathbf{B}_n$ with its truncated and centralized version $\widetilde{\mathbf{B}}_n$, or equivalently, we assume

$$\begin{aligned} &\mathbb{E}(z_{11}) = 0, \qquad \mathbb{E}\big(z_{11}^2\big) = 1, \qquad \mathbb{E}\big(z_{11}^4\big) = \tau + o(1), \\ &\mathbb{E}\big(|z_{11}|^\gamma\big) < \infty, \quad \max_{i,j} |z_{ij}|^\gamma < n^2 \end{aligned} \tag{3.19}$$

for the proof of the theorem. Note that for those $(\tilde{z}_{ij})$ after truncation and centralization, its variance might not equal to 1, however since the sample spatial-sign vectors $\{\mathbf{A}^{1/2}\tilde{\mathbf{z}}_j/\|\mathbf{A}^{1/2}\tilde{\mathbf{z}}_j\|\}$ are all self-normalized, we could then assume $\mathbb{E}(\tilde{z}_{11}^2) = 1$ as in (3.19).

Next we define a rectangular contour enclosing the interval $I_c = [s_l, s_r]$,

$$s_l = \liminf_{p\to\infty} \lambda_{\min}^{\mathbf{A}}(1-\sqrt{c})^2 I_{(0,1)}(c) \quad \text{and} \quad s_r = \limsup_{p\to\infty} \lambda_{\max}^{\mathbf{A}}(1+\sqrt{c})^2, \tag{3.20}$$

and thus enclosing all supports of the LSDs $\{F^{c_n,\tilde{H}_p}\}$. Choosing two numbers $x_l < x_r$ such that $[s_l, s_r] \subset (x_l, x_r)$ and letting $v_0 > 0$ be arbitrary, then the contour can be described as

$$\mathcal{C} = \big\{x \pm i v_0 : x \in [x_l, x_r]\big\} \cup \big\{x + iv : x \in \{x_r, x_l\}, v \in [-v_0, v_0]\big\}.$$

Denote

$$\begin{aligned} m_n(z) &= \int \frac{1}{x-z}\, dF^{\mathbf{B}_n}(x), \qquad \underline{m}_n(z) = -\frac{1-c_n}{z} + c_n m_n(z), \\ m_0(z) &= \int \frac{1}{x-z}\, dF^{c_n,\tilde{H}_p}(x), \qquad \underline{m}_0(z) = -\frac{1-c_n}{z} + c_n m_0(z). \end{aligned}$$

We then define a random process on $\mathcal{C}$ as

$$M_n(z) = p\big[m_n(z) - m_0(z)\big] = n\big[\underline{m}_n(z) - \underline{m}_0(z)\big], \quad z \in \mathcal{C}.$$

From Cauchy's integral formula, for any k analytic functions (f_ℓ) and complex numbers (a_ℓ), we have

$$\sum_{\ell=1}^k p a_\ell \int f_\ell(x)\, dG_n(x) = -\sum_{\ell=1}^k \frac{a_\ell}{2\pi \mathrm{i}} \oint_{\mathcal{C}} f_\ell(z) M_n(z)\, dz$$

when all sample eigenvalues fall in the interval (x_l, x_r), which holds with probability $1 - o(n^{-s})$ for any $s > 0$. That is,

$$\mathrm{P}\big(\|\mathbf{B}_n\| > x_r\big) = o\big(n^{-s}\big) \quad \text{and} \quad \mathrm{P}\big(\lambda_{\min}^{\mathbf{B}_n} < x_l\big) = o\big(n^{-s}\big), \quad \forall s > 0, \tag{3.21}$$

which follows from (3.14) and a similar conclusion for $\mathbf{S}_n$, see [5]. In order to deal with the small probability event where some eigenvalues are outside the interval (x_l, x_r) in finite dimensional situations, [5] suggested truncating $M_n(z)$ as, for $z = x + \mathrm{i}v \in \mathcal{C}$,

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

where $\mathcal{C}_n = \{x \pm \mathrm{i}v_0 : x \in [x_l, x_r]\} \cup \{x \pm \mathrm{i}v : x \in \{x_l, x_r\}, v \in [n^{-1}\varepsilon_n, v_0]\}$, a regularized version of $\mathcal{C}$ excluding a small segment near the real line, and the positive sequence (ε_n) decreases to zero satisfying $\varepsilon_n > n^{-a}$ for some $a \in (0, 1)$. From this and (3.21), one may thus find

$$\oint_{\mathcal{C}} f_\ell(z) M_n(z)\, dz = \oint_{\mathcal{C}} f_\ell(z) \widehat{M}_n(z)\, dz + o_p(1),$$

for every $\ell \in \{1, \ldots, k\}$. Hence, the proof of Theorem 2.2 can be completed by verifying the convergence of $\widehat{M}_n(z)$ on $\mathcal{C}$ as stated in the following lemma.

Lemma 3.1. *In addition to Assumptions (a)–(c), suppose that the conditions in (3.19) hold with $\gamma = 5$. We have*

$$\widehat{M}_n(z) \stackrel{d}{=} M(z) + o_p(1), \quad z \in \mathcal{C},$$

where the random process $M(z)$ is a two-dimensional Gaussian process. The mean function is

$$\mathbb{E}M(z) = \mu_1(z) + (\tau - 3)\mu_2(z), \tag{3.22}$$

and the covariance function is

$$\operatorname{Cov}\big(M(z), M(\tilde{z})\big) = \sigma_1(z, \tilde{z}) + (\tau - 3)\sigma_2(z, \tilde{z}),$$

where $\mu_1(z)$, $\mu_2(z)$, $\sigma_1(z, \tilde{z})$, $\sigma_2(z, \tilde{z})$ are defined in Theorem 2.2.

3.2.2. *Proof of Lemma 3.1*

At the beginning of the proof of Lemma 3.1, we list some quantities below which will be used frequently throughout this proof.

$$\mathbf{s}_j = \mathbf{s}(\mathbf{x}_j), \qquad \mathbf{r}_j = \sqrt{p/n}\,\mathbf{s}_j, \qquad \mathbf{B}_n = \sum_{j=1}^n \mathbf{r}_j \mathbf{r}_j', \qquad \boldsymbol{\Sigma} = \mathbb{E}\mathbf{B}_n = n\mathbb{E}\mathbf{r}_1\mathbf{r}_1',$$

$$\mathbf{D}(z) = \mathbf{B}_n - zI, \qquad \mathbf{D}_j(z) = \mathbf{D}(z) - \mathbf{r}_j\mathbf{r}_j', \qquad \mathbf{D}_{ij}(z) = \mathbf{D}(z) - \mathbf{r}_i\mathbf{r}_i' - \mathbf{r}_j\mathbf{r}_j', \quad (i \neq j),$$

$$\varepsilon_j(z) = \mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{n}\operatorname{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z), \qquad \gamma_j(z) = \mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{n}\mathbb{E}\operatorname{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z),$$

$$\delta_j(z) = \mathbf{r}_j'\mathbf{D}_j^{-2}(z)\mathbf{r}_j - \frac{1}{n}\operatorname{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-2}(z),$$

$$\beta_j(z) = \frac{1}{1 + \mathbf{r}_j'\mathbf{D}_j^{-1}(z)\mathbf{r}_j}, \qquad \bar{\beta}_j(z) = \frac{1}{1 + n^{-1}\operatorname{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z)}, \qquad b_n(z) = \frac{1}{1 + n^{-1}\mathbb{E}\operatorname{tr}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}(z)},$$

$$\beta_{jk}(z) = \frac{1}{1 + \mathbf{r}_j'\mathbf{D}_{kj}^{-1}(z)\mathbf{r}_j}, \qquad \bar{\beta}_{jk}(z) = \frac{1}{1 + n^{-1}\operatorname{tr}\boldsymbol{\Sigma}\mathbf{D}_{kj}^{-1}(z)}, \qquad \bar{b}_n(z) = \frac{1}{1 + n^{-1}\mathbb{E}\operatorname{tr}\boldsymbol{\Sigma}\mathbf{D}_{kj}^{-1}(z)}.$$

Note that the last six quantities are all bounded in absolute value by $|z|/v$ for any $z = u + iv \in \mathbb{C}^+$ (see, e.g., Page 568 in [5]). Now we split $\widehat{M}_n(z)$ into two parts as

$$\begin{aligned}\widehat{M}_n(z) =& p\big[m_n(z) - \mathbb{E}m_n(z)\big] + p\big[\mathbb{E}m_n(z) - m_0(z)\big]\\ :=& M_n^{(1)}(z) + M_n^{(2)}(z).\end{aligned}$$

Hence, the convergence of $\widehat{M}_n(z)$ can be obtained through the following three steps.

Step 1: Finite dimensional convergence of $M_n^{(1)}(z)$. Let $z_1, \dots, z_q$ be any q complex numbers on $\mathcal{C}_n$, this step approximates joint distribution of

$$\left[M_n^{(1)}(z_1), \dots, M_n^{(1)}(z_q)\right] \tag{3.23}$$

through martingale CLT [8]. Beyond the techniques used in [5], a particularly important problem is to find new approaches to deal with the non-linear correlation structure among the entries of $\mathbf{s}(\mathbf{x}_j)$. And such non-linear correlation is actually introduced by the spatial-sign transform of the data, to be precise, the norm $\|\mathbf{x}_j\|$ that appears in the denominator of $\mathbf{s}(\mathbf{x}_j)$. To this end, by giving an asymptotic expansion of $\mathbf{s}(\mathbf{x}_j)$, we develop Lemma A.2 concerning the covariance of certain quadratic forms, which turns out to be one of the cornerstones for establishing our new CLT. Moreover, we develop Lemma A.3 for bounding the spectral norm of the difference between the population SSCM $\mathbf{\Sigma}$ and the matrix $\mathbf{T}$ introduced in (2.7). In this way, we link the population SSCM $\mathbf{\Sigma}$ to the shape matrix $\mathbf{A}$ of our model and once coming across the first order asymptotic concerning the matrix $\mathbf{\Sigma}$, we can always replace it with the shape matrix $\mathbf{A}$. That is, for any analytical function f, we have the following convergence

$$\lim_{p\to\infty}\int f(t)\,d\tilde{H}_p(t) = \lim_{p\to\infty}\int f(t)\,dH_p(t) = \int f(t)\,dH(t).$$

Step 2: Tightness of $M_n^{(1)}(z)$ on $\mathcal{C}_n$. We illustrate in this step the basic idea for proving the tightness. The result in (3.21) controls the probability of extreme eigenvalues falling outside the contour $\mathcal{C}$. By virtue of this and Lemma A.4, the tightness can be obtained following similar arguments in [5].

Step 3: Convergence of $M_n^{(2)}(z)$. In this final step, we approximate the quantity $M_n^{(2)}(z)$. In parallel with Step 1, dealing with the nonlinear effects as shown in Lemma A.2 is the main focus in this part. As will be seen, such nonlinear effects will contribute several new terms to the mean of $\boldsymbol{\xi}$.

Step 1: Finite dimensional convergence of $M_n^{(1)}(z)$ in distribution.

Let $\mathbb{E}_0(\cdot)$ denote expectation and $\mathbb{E}_j(\cdot)$ denote conditional expectation with respect to the σ-field generated by $\mathbf{r}_1, \dots, \mathbf{r}_j$, $j = 1, \dots, n$. From the martingale decomposition and the identity

$$\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z) = -\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j'\mathbf{D}_j^{-1}(z)\beta_j(z), \tag{3.24}$$

we get

$$\begin{aligned}
M_n^{(1)}(z) &= \sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\operatorname{tr}\left[\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z)\right] \\
&= -\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\frac{\mathbf{r}_j'\mathbf{D}_j^{-2}\mathbf{r}_j}{1+\mathbf{r}_j'\mathbf{D}_j^{-1}\mathbf{r}_j} \\
&= -\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\left[\frac{\mathbf{r}_j'\mathbf{D}_j^{-2}\mathbf{r}_j}{1+\mathbf{r}_j'\mathbf{D}_j^{-1}\mathbf{r}_j} - \frac{n^{-1}\operatorname{tr}\mathbf{D}_j^{-2}\mathbf{\Sigma}}{1+n^{-1}\operatorname{tr}\mathbf{D}_j^{-1}\mathbf{\Sigma}}\right]
\end{aligned}$$

$$= -\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\left[\frac{d}{dz}\log\beta_j^{-1}(z) - \frac{d}{dz}\log\bar{\beta}_j^{-1}(z)\right]$$

$$= \sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\frac{d}{dz}\log\bigl(\beta_j(z)/\bar{\beta}_j(z)\bigr)$$

$$= \frac{d}{dz}\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\log\bigl[1 - \bar{\beta}_j(z)\varepsilon_j(z) + \bar{\beta}_j(z)\beta_j(z)\varepsilon_j^2(z)\bigr], \tag{3.25}$$

where the last equality is from the identity $\beta_j(z) = \bar{\beta}_j(z) - \bar{\beta}_j^2(z)\varepsilon_j(z) + \bar{\beta}_j^2(z)\beta_j(z)\varepsilon_j^2(z)$. Note that $\beta_j(z)$ and $\bar{\beta}_j(z)$ are bounded because the imaginary part of their respective denominator is lower bounded. Thus from Lemma A.4 we have

$$\mathbb{E}\Bigl|\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\bar{\beta}_j(z)\beta_j(z)\varepsilon_j^2(z)\Bigr|^2 \le Kn\mathbb{E}\bigl|\varepsilon_j(z)\bigr|^4 \to 0.$$

Thus applying Taylor's expansion to the log function in (3.25), one may conclude

$$M_n^{(1)}(z) = -\frac{d}{dz}\sum_{j=1}^{n}(\mathbb{E}_j - \mathbb{E}_{j-1})\bar{\beta}_j(z)\varepsilon_j(z) + o_p(1)$$

$$= -\frac{d}{dz}\sum_{j=1}^{n}\mathbb{E}_j\bar{\beta}_j(z)\varepsilon_j(z) + o_p(1).$$

Therefore, we turn to consider the martingale difference sequence

$$Y_{nj}(z) := \frac{d}{dz}\mathbb{E}_j\bar{\beta}_j(z)\varepsilon_j(z), \quad j = 1, \ldots, n.$$

The Lyapunov condition for this sequence is guaranteed by the fact that

$$\sum_{j=1}^{n}\mathbb{E}\bigl|Y_{nj}(z)\bigr|^4 = \sum_{j=1}^{n}\mathbb{E}\left|\mathbb{E}_j\left(\delta_j(z)\bar{\beta}_j(z) - \varepsilon_j(z)\bar{\beta}_j^2(z)\frac{1}{n}\operatorname{tr}\mathbf{\Sigma}\mathbf{D}_j^{-2}(z)\right)\right|^4$$

$$\le K\sum_{j=1}^{n}\left(\frac{|z|^4\mathbb{E}|\delta_j(z)|^4}{v^4} + \frac{|z|^8 p^4\mathbb{E}|\varepsilon_j(z)|^4}{v^{16}n^4}\right) \to 0,$$

where the convergence is from Lemma A.4.

We next consider the sum $\sigma_n(z,\tilde{z}) \triangleq \sum_{j=1}^{n}\mathbb{E}_{j-1}[Y_{nj}(z)Y_{nj}(\tilde{z})]$, for $z \ne \tilde{z} \in \{z_1, \ldots, z_w\}$. From similar arguments on Pages 571 and 576 of [5], we have

$$\mathbb{E}\bigl|\bar{\beta}_j(z) - b_n(z)\bigr| \le \frac{K}{n}\mathbb{E}\bigl|\operatorname{tr}\mathbf{\Sigma}\mathbf{D}_j^{-1}(z) - \mathbb{E}\operatorname{tr}\mathbf{\Sigma}\mathbf{D}_j^{-1}\bigr| \to 0 \quad \text{and} \quad b_n(z) + z\underline{m}(z) \to 0, \tag{3.26}$$

which implies

$$\sigma_n(z,\tilde{z}) = \frac{\partial^2}{\partial z\partial\tilde{z}}z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})\sum_{j=1}^{n}\mathbb{E}_{j-1}\bigl(\mathbb{E}_j\varepsilon_j(z)\mathbb{E}_j\varepsilon_j(\tilde{z})\bigr) + o_p(1).$$

Moreover, applying Lemma A.2 to the above conditional expectations, one may get

$$z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})\sum_{j=1}^{n}\mathbb{E}_{j-1}\big(\mathbb{E}_j\varepsilon_j(z)\mathbb{E}_j\varepsilon_j(\tilde{z})\big)$$
$$=2T_1+\frac{2}{p}\operatorname{tr}\big(\mathbf{A}^2\big)T_2-2T_3-2T_4+(\tau-3)(T_5+T_6-T_7-T_8)+o(1),$$

where

$$T_1=\frac{z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})}{n^2}\sum_{j=1}^{n}\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(z)\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(\tilde{z})\big],$$

$$T_2=\frac{z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})}{pn^2}\sum_{j=1}^{n}\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(z)\big]\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(\tilde{z})\big],$$

$$T_3=\frac{z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})}{pn^2}\sum_{j=1}^{n}\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}^2\mathbf{D}_j^{-1}(z)\big]\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(\tilde{z})\big],$$

$$T_4=\frac{z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})}{pn^2}\sum_{j=1}^{n}\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(z)\big]\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}^2\mathbf{D}_j^{-1}(\tilde{z})\big],$$

$$T_5=\frac{z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})}{n^2}\sum_{j=1}^{n}\operatorname{tr}\big[\mathbb{E}_j\big(\mathbf{A}^{\frac12}\mathbf{D}_j^{-1}(z)\mathbf{A}^{\frac12}\big)\circ\mathbb{E}_j\big(\mathbf{A}^{\frac12}\mathbf{D}_j^{-1}(\tilde{z})\mathbf{A}^{\frac12}\big)\big],$$

$$T_6=\frac{z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})}{p^2n^2}\sum_{j=1}^{n}\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(z)\big]\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(\tilde{z})\big]\operatorname{tr}[\mathbf{A}\circ\mathbf{A}],$$

$$T_7=\frac{z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})}{pn^2}\sum_{j=1}^{n}\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(z)\big]\operatorname{tr}\big[\mathbb{E}_j\big(\mathbf{A}^{\frac12}\mathbf{D}_j^{-1}(\tilde{z})\mathbf{A}^{\frac12}\big)\circ\mathbf{A}\big],$$

$$T_8=\frac{z\tilde{z}\underline{m}(z)\underline{m}(\tilde{z})}{pn^2}\sum_{j=1}^{n}\operatorname{tr}\big[\mathbb{E}_j\big(\mathbf{A}^{\frac12}\mathbf{D}_j^{-1}(z)\mathbf{A}^{\frac12}\big)\circ\mathbf{A}\big]\operatorname{tr}\big[\mathbb{E}_j\mathbf{A}\mathbf{D}_j^{-1}(\tilde{z})\big].$$

Following similar steps as in [5] and [15], applying Lemma A.3 and Lemma A.4, we obtain

$$T_1=\log\frac{\underline{m}(z)-\underline{m}(\tilde{z})}{\underline{m}(z)\underline{m}(\tilde{z})(z-\tilde{z})}+o_p(1),$$
$$T_2=\frac{T_6}{\zeta}=c\int\frac{t\underline{m}(z)\,dH(t)}{1+t\underline{m}(z)}\int\frac{t\underline{m}(\tilde{z})\,dH(t)}{1+t\underline{m}(\tilde{z})}+o_p(1)=\frac{[1+z\underline{m}(z)][1+\tilde{z}\underline{m}(\tilde{z})]}{c}+o_p(1).$$

Notice that T_3 and T_4 will reduce to T_2 if $\mathbf{A}^2$ is replaced with $\mathbf{A}$. By this, we have

$$T_3=c\int\frac{t^2\underline{m}(z)\,dH(t)}{1+t\underline{m}(z)}\int\frac{t\underline{m}(\tilde{z})\,dH(t)}{1+t\underline{m}(\tilde{z})}+o_p(1)=\bigg[1-\frac{1+z\underline{m}(z)}{c\underline{m}(z)}\bigg]\big[1+\tilde{z}\underline{m}(\tilde{z})\big]+o_p(1),$$

$$T_4 = c\int \frac{t\underline{m}(z)\,dH(t)}{1+t\underline{m}(z)} \int \frac{t^2\underline{m}(\tilde{z})\,dH(t)}{1+t\underline{m}(\tilde{z})} + o_p(1) = \left[1 - \frac{1+\tilde{z}\underline{m}(\tilde{z})}{c\underline{m}(\tilde{z})}\right]\left[1+z\underline{m}(z)\right] + o_p(1).$$

For the terms T_5, T_7, and T_8, following similar procedure as in [25] for proving their Theorem 1.4, using Lemma A.3, Lemma A.4 and Theorem 2.1, one may get

$$\begin{aligned}
T_5 &= \frac{1}{n}\operatorname{tr}\left[\left(\mathbf{A}^{\frac{1}{2}}\left(\underline{m}^{-1}(z)\mathbf{I}+\mathbf{\Sigma}\right)^{-1}\mathbf{A}^{\frac{1}{2}}\right)\circ\left(\mathbf{A}^{\frac{1}{2}}\left(\underline{m}^{-1}(\tilde{z})\mathbf{I}+\mathbf{\Sigma}\right)^{-1}\mathbf{A}^{\frac{1}{2}}\right)\right] + o_p(1) \\
&= cg\left(\frac{-1}{\underline{m}(z)}, \frac{-1}{\underline{m}(\tilde{z})}\right) + o_p(1), \\
T_7 &= \frac{1}{pn}\operatorname{tr}\left[\mathbf{A}\left(\underline{m}^{-1}(z)\mathbf{I}+\mathbf{\Sigma}\right)^{-1}\right]\operatorname{tr}\left[\left(\mathbf{A}^{\frac{1}{2}}\left(\underline{m}^{-1}(\tilde{z})\mathbf{I}+\mathbf{\Sigma}\right)^{-1}\mathbf{A}^{\frac{1}{2}}\right)\circ\mathbf{A}\right] + o_p(1) \\
&= h\left(\frac{-1}{\underline{m}(\tilde{z})}\right)\left[1+z\underline{m}(z)\right] + o_p(1), \\
T_8 &= \frac{1}{pn}\operatorname{tr}\left[\left(\mathbf{A}^{\frac{1}{2}}\left(\underline{m}^{-1}(z)\mathbf{I}+\mathbf{\Sigma}\right)^{-1}\mathbf{A}^{\frac{1}{2}}\right)\circ\mathbf{A}\right]\operatorname{tr}\left[\mathbf{A}\left(\underline{m}^{-1}(\tilde{z})\mathbf{I}+\mathbf{\Sigma}\right)^{-1}\right] + o_p(1) \\
&= h\left(\frac{-1}{\underline{m}(z)}\right)\left[1+\tilde{z}\underline{m}(\tilde{z})\right] + o_p(1).
\end{aligned}$$

Collecting the above results, we get

$$(3.23) \overset{d}{=} \left[M^{(1)}(z_1), \dots, M^{(1)}(z_q)\right] + o_p(1),$$

where $[M^{(1)}(z_1), \dots, M^{(1)}(z_q)]$ is a q-dimensional zero-mean Gaussian random vector with covariance function

$$\operatorname{Cov}\left[M^{(1)}(z), M^{(1)}(\tilde{z})\right] = \sigma_1(z,\tilde{z}) + (\tau-3)\sigma_2(z,\tilde{z}).$$

Step 2: Tightness of $M_n^{(1)}(z)$. The tightness can be established by verifying the moment condition (12.51) of [7]:

$$\sup_{n, z_1, z_2 \in \mathcal{C}_n} \frac{\mathbb{E}|M_n^{(1)}(z_1) - M_n^{(1)}(z_2)|^2}{|z_1 - z_2|^2} < \infty. \tag{3.27}$$

By (3.21) and arguments on Page 579 in [5], one may verify that moments of $\mathbf{D}^{-1}(z)$, $\mathbf{D}_j^{-1}(z)$ and $\mathbf{D}_{ij}^{-1}(z)$ are uniformly bounded in n and $z \in \mathcal{C}_n$, that is, for any positive q,

$$\max\left\{\mathbb{E}\left\|\mathbf{D}^{-1}(z)\right\|^q, \mathbb{E}\left\|\mathbf{D}_j^{-1}(z)\right\|^q, \mathbb{E}\left\|\mathbf{D}_{ij}^{-1}(z)\right\|^q\right\} \le K. \tag{3.28}$$

By such boundedness, the inequality in Lemma A.4 can be extended to

$$\left|\mathbb{E}\left[a(v)\prod_{l=1}^{k}\left(\mathbf{r}'\mathbf{B}_l(v)\mathbf{r} - \frac{1}{n}\operatorname{tr}\mathbf{\Sigma}\mathbf{B}_l(v)\right)\right]\right| \le Kn^{-1-k(\gamma-4)/\gamma}, \quad k \ge 2. \tag{3.29}$$

The matrices $\mathbf{B}_l(v)$ in (3.29) are independent of $\mathbf{r}$ and

$$\max\left\{\left|a(v)\right|, \left\|\mathbf{B}_l(v)\right\|\right\} \le K\left[1 + p^s I\left(\|\mathbf{B}_n\| \ge x_r \text{ or } \lambda_{\min}^{\tilde{\mathbf{B}}} \le x_l\right)\right]$$

for some positive s, where $\tilde{\mathbf{B}}$ denotes $\mathbf{B}_n = \sum \mathbf{r}_j\mathbf{r}_j'$, $\mathbf{B}_j = \sum_{k\neq j}\mathbf{r}_k\mathbf{r}_k'$, or $\mathbf{B}_{ij} = \sum_{k\neq i,j}\mathbf{r}_k\mathbf{r}_k'$. Finally, following similar procedure as in Section 3 of [5], and applying Lemma A.3, Lemma A.4 together with (3.21), (3.28), and (3.29), one may verify (3.27). The details are thus omitted.

Step 3: Convergence of $M_n^{(2)}(z)$. To finish the proof, it is enough to show that the sequence of $M_n^{(2)}(z)$ is bounded and equicontinuous, and is equal to the mean function (3.22) asymptotically. The boundedness and equicontinuity can be verified following the arguments on Pages 592-593 in [5]. We thus propose a novel method to approximate $M_n^{(2)}(z)$, which is quite different from the idea in [5]. This new procedure is more straightforward and easier to follow. Before the proof, we first list some results that will be used in this part:

$$\sup_{z\in\mathcal{C}_n}\mathbb{E}\big|\varepsilon_j(z)\big|^k \le Kn^{-k/2} + Kn^{-1-k(\gamma-4)/\gamma},$$
$$\sup_{z\in\mathcal{C}_n}\mathbb{E}\big|\gamma_j(z)\big|^k \le Kn^{-k/2} + Kn^{-1-k(\gamma-4)/\gamma}, \tag{3.30}$$

$$\sup_{n,z\in\mathcal{C}_n}\big|b_n(z)+z\underline{m}(z)\big|\to 0, \qquad \sup_{n,z\in\mathcal{C}_n}\big\|z\mathbf{I}-b_n(z)\mathbf{\Sigma}\big\|^{-1}<\infty, \tag{3.31}$$

$$\sup_{n,z\in\mathcal{C}_n}\mathbb{E}\big|\operatorname{tr}\mathbf{D}^{-1}(z)\mathbf{M}-\mathbb{E}\operatorname{tr}\mathbf{D}^{-1}(z)\mathbf{M}\big|^2\le K\|\mathbf{M}\|^2, \tag{3.32}$$

where $k\ge 2$ and $\mathbf{M}$ is a nonrandom $p\times p$ matrix. These results can be verified step by step following the discussions in [5] and we omit the details.

Writing $\mathbf{V}(z) = zI - b_n(z)\mathbf{\Sigma}$, we decompose $M_n^{(2)}(z)$ in two ways:

$$M_n^{(2)}(z) = \big[p\mathbb{E}m_n(z) + \operatorname{tr}\mathbf{V}^{-1}(z)\big] - \big[\operatorname{tr}\mathbf{V}^{-1}(z) + pm_0(z)\big] := S_n(z) - T_n(z),$$
$$M_n^{(2)}(z) = \big[n\mathbb{E}\underline{m}_n(z) + nb_n(z)/z\big] - \big[nb_n(z)/z + n\underline{m}_0(z)\big] := \underline{S}_n(z) - \underline{T}_n(z).$$

Notice that by Lemma A.3,

$$\begin{aligned}T_n(z) &= p\int\frac{d\tilde{H}_p(t)}{z-b_n(z)t} - p\int\frac{d\tilde{H}_p(t)}{z+z\underline{m}_0(z)t} + o(1)\\ &= p\big[b_n(z)+z\underline{m}_0(z)\big]\int\frac{t\,d\tilde{H}_p(t)}{(z-b_n(z)t)(z+z\underline{m}_0(z)t)} + o(1)\\ &= c\underline{T}_n(z)\int\frac{t\,dH(t)}{z(1+\underline{m}(z)t)^2} + o(1).\end{aligned}$$

From this and the convergence in (3.31), we have

$$\frac{M_n^{(2)}(z)-S_n(z)}{M_n^{(2)}(z)-\underline{S}_n(z)} = \frac{T_n(z)}{\underline{T}_n(z)} = \frac{c}{z}\int\frac{t\,dH(t)}{(1+\underline{m}(z)t)^2} + o(1). \tag{3.33}$$

Our next task is to study the convergence of $S_n(z)$ and $\underline{S}_n(z)$. For simplicity, we suppress the expression z in the sequel when it is served as independent variables of some functions. All expressions and convergence statements hold uniformly for $z\in\mathcal{C}_n$.

We first simplify the expression of S_n. Using the identity $\mathbf{r}_j'\mathbf{D}^{-1} = \mathbf{r}_j'\mathbf{D}_j^{-1}\beta_j$, we have

$$S_n = \mathbb{E}\,\mathrm{tr}\big(\mathbf{D}^{-1} + \mathbf{V}^{-1}\big) = \mathbb{E}\,\mathrm{tr}\left[\mathbf{V}^{-1}\left(\sum_{j=1}^{n}\mathbf{r}_j\mathbf{r}_j' - b_n\boldsymbol{\Sigma}\right)\mathbf{D}^{-1}\right]$$
$$= n\mathbb{E}\beta_1\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 - b_n\mathbb{E}\,\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{V}^{-1}. \tag{3.34}$$

From (3.24) and $\beta_1 = b_n - b_n\beta_1\gamma_1$,

$$\mathbb{E}\,\mathrm{tr}\,\mathbf{V}^{-1}\boldsymbol{\Sigma}\big(\mathbf{D}_1^{-1} - \mathbf{D}^{-1}\big) = \mathbb{E}\,\mathrm{tr}\,\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_1^{-1}\mathbf{r}_1\mathbf{r}_1'\mathbf{D}_1^{-1}\beta_1$$
$$= b_n\mathbb{E}(1-\beta_1\gamma_1)\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_1^{-1}\mathbf{r}_1,$$

where $|\mathbb{E}\beta_1\gamma_1\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_1^{-1}\mathbf{r}_1| \le Kn^{-1/2}$. From this and (3.34), we get

$$S_n = n\mathbb{E}\beta_1\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 - b_n\mathbb{E}\,\mathrm{tr}\,\boldsymbol{\Sigma}\mathbf{D}_1^{-1}\mathbf{V}^{-1} + \frac{1}{n}b_n^2\mathbb{E}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_1^{-1}\boldsymbol{\Sigma} + o(1).$$

Then plugging $\beta_1 = b_n - b_n^2\gamma_1 + b_n^3\gamma_1^2 - \beta_1 b_n^3\gamma_1^3$ into the first term in the above equation, we obtain

$$n\mathbb{E}\beta_1\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 = b_n\mathbb{E}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma} - nb_n^2\mathbb{E}\gamma_1\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1$$
$$+ nb_n^3\mathbb{E}\gamma_1^2\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 - nb_n^3\mathbb{E}\beta_1\gamma_1^3\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1.$$

Note that, from (3.29), (3.30) and (3.32),

$$\mathbb{E}\gamma_1\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 = \mathbb{E}\left[\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{r}_1 - \frac{1}{n}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\boldsymbol{\Sigma}\right]\left[\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 - \frac{1}{n}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right]$$
$$+ \frac{1}{n^2}\,\mathrm{Cov}\big(\mathrm{tr}\,\mathbf{D}_1^{-1}\boldsymbol{\Sigma}, \mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\big)$$
$$= \mathbb{E}\left[\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{r}_1 - \frac{1}{n}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\boldsymbol{\Sigma}\right]\left[\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 - \frac{1}{n}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right] + o\left(\frac{1}{n}\right),$$

$$\mathbb{E}\gamma_1^2\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 = \mathbb{E}\gamma_1^2\left[\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 - \frac{1}{n}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right]$$
$$+ \frac{1}{n}\,\mathrm{Cov}\big(\gamma_1^2, \mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\big) + \frac{1}{n}\mathbb{E}\gamma_1^2 E\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}$$
$$= \frac{1}{n}\mathbb{E}\gamma_1^2\mathbb{E}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma} + o\left(\frac{1}{n}\right),$$

$$\mathbb{E}\beta_1\gamma_1^3\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 = o\left(\frac{1}{n}\right).$$

We thus arrive at

$$S_n = -nb_n^2\mathbb{E}\left[\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{r}_1 - \frac{1}{n}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\boldsymbol{\Sigma}\right]\left[\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{V}^{-1}\mathbf{r}_1 - \frac{1}{n}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\right]$$
$$+ b_n^3\mathbb{E}\gamma_1^2\mathbb{E}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma} + \frac{1}{n}b_n^2\mathbb{E}\,\mathrm{tr}\,\mathbf{D}_1^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_1^{-1}\boldsymbol{\Sigma} + o(1).$$

On the other hand, by the identity $\mathbf{r}_j'\mathbf{D}^{-1} = \mathbf{r}_j'\mathbf{D}_j^{-1}\beta_j$, we have

$$p + z\,\mathrm{tr}\,\mathbf{D}^{-1} = \mathrm{tr}\big(\mathbf{B}_n\mathbf{D}^{-1}\big) = \sum_{j=1}^{n}\beta_j\mathbf{r}_j'\mathbf{D}_j^{-1}\mathbf{r}_j = n - \sum_{j=1}^{n}\beta_j,$$

which implies $nz\underline{m}_n = -\sum_{j=1}^{n}\beta_j$. From this, together with $\beta_1 = b_n - b_n^2\gamma_1 + b_n^3\gamma_1^2 - \beta_1 b_n^3\gamma_1^3$ and (3.29), we get

$$\underline{S}_n = -\frac{n}{z}\mathbb{E}(\beta_1 - b_n) = -\frac{n}{z}b_n^3\mathbb{E}\gamma_1^2 + o(1).$$

Applying Lemma A.2 to the simplified S_n and $\underline{S}_n$, and then replacing $\mathbf{D}_j$ with $\mathbf{D}$ in the derived results yield

$$\begin{aligned}
S_n = &-\frac{b_n^2}{n}\bigg[\mathbb{E}\,\mathrm{tr}\,\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A} + \frac{2}{p}\bigg(\frac{1}{n}\,\mathrm{tr}\,\mathbf{A}^2\mathbb{E}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1}\mathbf{V}^{-1} \\
&-\mathbb{E}\,\mathrm{tr}\,\mathbf{A}^2\mathbf{D}^{-1}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1}\mathbf{V}^{-1} - \mathbb{E}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1}\,\mathrm{tr}\,\mathbf{A}^2\mathbf{D}^{-1}\mathbf{V}^{-1}\bigg)\bigg] \\
&+\frac{2b_n^3}{n^2}\bigg[\mathbb{E}\,\mathrm{tr}\,\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A} + \frac{1}{p}\bigg(\frac{1}{n}\,\mathrm{tr}\,\mathbf{A}^2\mathbb{E}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1} - 2\mathbb{E}\,\mathrm{tr}\,\mathbf{A}^2\mathbf{D}^{-1}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1}\bigg)\bigg] \\
&\cdot\mathbb{E}\,\mathrm{tr}\,\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma} - \frac{(\tau-3)b_n^2}{n}\bigg[\mathbb{E}\,\mathrm{tr}\big[\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{A}^{\frac12}\big)\circ\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A}^{\frac12}\big)\big] \\
&+\frac{1}{p^2}\mathbb{E}\,\mathrm{tr}\big(\mathbf{D}^{-1}\mathbf{A}\big)\,\mathrm{tr}\big(\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A}\big)\,\mathrm{tr}[\mathbf{A}\circ\mathbf{A}] \\
&-\frac{1}{p}\mathbb{E}\,\mathrm{tr}\big(\mathbf{D}^{-1}\mathbf{A}\big)\,\mathrm{tr}\big[\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A}^{\frac12}\big)\circ\mathbf{A}\big] - \frac{1}{p}\mathbb{E}\,\mathrm{tr}\big(\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A}\big)\,\mathrm{tr}\big[\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{A}^{\frac12}\big)\circ\mathbf{A}\big]\bigg] \\
&+\frac{(\tau-3)b_n^3}{n^2}\bigg[\mathbb{E}\,\mathrm{tr}\big[\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{A}^{\frac12}\big)\circ\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{A}^{\frac12}\big)\big] + \frac{1}{p^2}\mathbb{E}\mathrm{tr}^2\big(\mathbf{D}^{-1}\mathbf{A}\big)\,\mathrm{tr}[\mathbf{A}\circ\mathbf{A}] \\
&-\frac{2}{p}\mathbb{E}\,\mathrm{tr}\big(\mathbf{D}^{-1}\mathbf{A}\big)\,\mathrm{tr}\big[\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{A}^{\frac12}\big)\circ\mathbf{A}\big]\bigg]\mathbb{E}\,\mathrm{tr}\,\mathbf{D}^{-1}\mathbf{V}^{-1}\boldsymbol{\Sigma} + o(1), \\
\underline{S}_n = &\frac{-2b_n^3}{zn}\bigg[\mathbb{E}\,\mathrm{tr}\,\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A} + \frac{1}{p}\bigg(\frac{1}{p}\,\mathrm{tr}\,\mathbf{A}^2\mathbb{E}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1} - 2\mathbb{E}\,\mathrm{tr}\,\mathbf{A}^2\mathbf{D}^{-1}\,\mathrm{tr}\,\mathbf{A}\mathbf{D}^{-1}\bigg)\bigg] \\
&-\frac{(\tau-3)b_n^3}{zn}\bigg[\mathbb{E}\,\mathrm{tr}\big[\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{A}^{\frac12}\big)\circ\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{A}^{\frac12}\big)\big] + \frac{1}{p^2}\mathbb{E}\mathrm{tr}^2\big(\mathbf{D}^{-1}\mathbf{A}\big)\,\mathrm{tr}[\mathbf{A}\circ\mathbf{A}] \\
&-\frac{2}{p}\mathbb{E}\,\mathrm{tr}\big(\mathbf{D}^{-1}\mathbf{A}\big)\,\mathrm{tr}\big[\big(\mathbf{A}^{\frac12}\mathbf{D}^{-1}\mathbf{A}^{\frac12}\big)\circ\mathbf{A}\big]\bigg] + o(1).
\end{aligned}$$

To study the convergence of S_n and $\underline{S}_n$, we need to figure out the difference between $\mathbf{D}^{-1}$ and $\mathbf{V}^{-1}$. Write

$$\mathbf{D}^{-1} + \mathbf{V}^{-1} = b_n\tilde{\mathbf{R}}_1 + \tilde{\mathbf{R}}_2 + \tilde{\mathbf{R}}_3, \tag{3.35}$$

where

$$\tilde{\mathbf{R}}_1 = \sum_{j=1}^{n} \mathbf{V}^{-1}\big(\mathbf{r}_j\mathbf{r}_j' - n^{-1}\boldsymbol{\Sigma}\big)\mathbf{D}_j^{-1}, \qquad \tilde{\mathbf{R}}_2 = \sum_{j=1}^{n} \mathbf{V}^{-1}\mathbf{r}_j\mathbf{r}_j'\mathbf{D}_j^{-1}(\beta_j - b_n),$$

$$\tilde{\mathbf{R}}_3 = \frac{1}{n}\sum_{j=1}^{n} b_n \mathbf{V}^{-1}\boldsymbol{\Sigma}\big(\mathbf{D}_j^{-1} - \mathbf{D}^{-1}\big).$$

From equations (4.15) and (4.16) in [5], we have for any $p \times p$ matrix $\mathbf{M}$,

$$|\mathbb{E}\operatorname{tr}\tilde{\mathbf{R}}_2\mathbf{M}| \le n^{1/2}K\big(\mathbb{E}\|\mathbf{M}\|^4\big)^{1/4} \quad \text{and} \quad |\operatorname{tr}\tilde{\mathbf{R}}_3\mathbf{M}| \le K\big(\mathbb{E}\|\mathbf{M}\|^2\big)^{1/2} \tag{3.36}$$

and, for nonrandom matrix $\mathbf{M}$,

$$|\mathbb{E}\operatorname{tr}\tilde{\mathbf{R}}_1\mathbf{M}| \le n^{1/2}K\|\mathbf{M}\|. \tag{3.37}$$

Taking a step further, for $\mathbf{M}$ nonrandom, we write

$$\operatorname{tr}\tilde{\mathbf{R}}_1\boldsymbol{\Sigma}\mathbf{D}^{-1}\mathbf{M} = \tilde{R}_{11} + \tilde{R}_{12} + \tilde{R}_{13}, \tag{3.38}$$

where

$$\tilde{R}_{11} = \operatorname{tr}\sum_{j=1}^{n} \mathbf{V}^{-1}\mathbf{r}_j\mathbf{r}_j'\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\big(\mathbf{D}^{-1} - \mathbf{D}_j^{-1}\big)\mathbf{M},$$

$$\tilde{R}_{12} = \operatorname{tr}\sum_{j=1}^{n} \mathbf{V}^{-1}\big(\mathbf{r}_j\mathbf{r}_j' - n^{-1}\boldsymbol{\Sigma}\big)\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\mathbf{M},$$

$$\tilde{R}_{13} = -\frac{1}{n}\operatorname{tr}\sum_{j=1}^{n} \mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{D}_j^{-1}\boldsymbol{\Sigma}\big(\mathbf{D}^{-1} - \mathbf{D}_j^{-1}\big)\mathbf{M}.$$

It's clear that $\mathbb{E}\tilde{R}_{12} = 0$ and moreover, using (3.28), (3.29) and (3.32), we get

$$|\mathbb{E}\tilde{R}_{13}| \le K\|\mathbf{M}\|, \tag{3.39}$$

$$\begin{aligned}
\mathbb{E}\tilde{R}_{11} &= -n\mathbb{E}\beta_1\mathbf{r}_1\mathbf{D}_1^{-1}\boldsymbol{\Sigma}\mathbf{D}_1^{-1}\mathbf{r}_1\mathbf{r}_1'\mathbf{D}_1^{-1}\mathbf{M}\mathbf{V}^{-1}\mathbf{r}_1 \\
&= -b_n n^{-1}\mathbb{E}\big(\operatorname{tr}\mathbf{D}_1^{-1}\boldsymbol{\Sigma}\mathbf{D}_1^{-1}\boldsymbol{\Sigma}\big)\big(\operatorname{tr}\mathbf{D}_1^{-1}\mathbf{M}\mathbf{V}^{-1}\boldsymbol{\Sigma}\big) + o(1) \\
&= -b_n n^{-1}\mathbb{E}\big(\operatorname{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma}\big)\big(\operatorname{tr}\mathbf{D}^{-1}\mathbf{M}\mathbf{V}^{-1}\boldsymbol{\Sigma}\big) + o(1) \\
&= -b_n n^{-1}\mathbb{E}\big(\operatorname{tr}\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}\boldsymbol{\Sigma}\big)\mathbb{E}\big(\operatorname{tr}\mathbf{D}^{-1}\mathbf{M}\mathbf{V}^{-1}\boldsymbol{\Sigma}\big) + o(1).
\end{aligned} \tag{3.40}$$

Applying (3.26), (3.35)–(3.40), and Lemma A.3, one may approximate each components of S_n and $\underline{S}_n$. Specifically, we have

$$\frac{1}{n}\mathbb{E}\operatorname{tr}\mathbf{D}^{-1}\mathbf{A}^k = -\int \frac{ct^k\,dH(t)}{z(1+\underline{m}t)} + o(1),$$

$$\frac{1}{n}\mathbb{E}\operatorname{tr}\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A}^k = -\int \frac{ct^k\,dH(t)}{z^2(1+\underline{m}t)^2} + o(1),$$

$$
\begin{aligned}
\frac{1}{n}\mathbb{E}\operatorname{tr}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A} &= -\frac{1}{n}\mathbb{E}\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A} - \frac{b_n^2}{n^2}\mathbb{E}\operatorname{tr}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\mathbb{E}\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A} + o(1) \\
&= -\frac{1}{n}\mathbb{E}\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\left[1 + \frac{b_n^2}{n}\mathbb{E}\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\right]^{-1} + o(1), \\
&= \int \frac{ct^2\,dH(t)}{z^2(1+\underline{m}t)^2}\left[1 - \int \frac{c\underline{m}^2t^2\,dH(t)}{(1+\underline{m}t)^2}\right]^{-1} + o(1), \\
\frac{1}{n}\mathbb{E}\operatorname{tr}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A} &= -\frac{1}{n}\mathbb{E}\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A}\left[1 + \frac{b_n^2}{n}\mathbb{E}\operatorname{tr}\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\right] + o(1) \\
&= -\frac{1}{n}\mathbb{E}\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{V}^{-1}\mathbf{A}\left[1 + \frac{b_n^2}{n}\mathbb{E}\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{A}\right]^{-1} + o(1), \\
&= \int \frac{ct^2\,dH(t)}{z^3(1+\underline{m}t)^3}\left[1 - \int \frac{c\underline{m}^2t^2\,dH(t)}{(1+\underline{m}t)^2}\right]^{-1} + o(1).
\end{aligned}
$$

Combining the above results, we obtain

$$
\begin{aligned}
T_n/\underline{T}_n &= \int \frac{ct\,dH(t)}{z(1+\underline{m}t)^2} + o(1), \\
S_n - \underline{S}_n T_n/\underline{T}_n &= -\int \frac{c\underline{m}^2t^2\,dH(t)}{z(1+\underline{m}t)^3}\left[1 - \int \frac{c\underline{m}^2t^2\,dH(t)}{(1+\underline{m}t)^2}\right]^{-1} \\
&\quad - \frac{2c\underline{m}^2}{z}\left[\int \frac{(\alpha_2 t - t^2)\,dH(t)}{1+\underline{m}t}\int \frac{t\,dH(t)}{(1+\underline{m}t)^2} - \int \frac{t\,dH(t)}{1+\underline{m}t}\int \frac{t^2\,dH(t)}{(1+\underline{m}t)^2}\right] \\
&\quad - c(\tau-3)\left\{\frac{1}{z\underline{m}}g_u'(u,v)|_{u=v=\frac{-1}{\underline{m}}} + \zeta\int \frac{t\underline{m}\,dH(t)}{1+\underline{m}t}\int \frac{t\underline{m}\,dH(t)}{z(1+\underline{m}t)^2}\right. \\
&\quad \left. - \left[\int \frac{t\,dH(t)}{z(1+\underline{m}t)}h'(u)|_{u=\frac{-1}{\underline{m}}} + \int \frac{t\underline{m}\,dH(t)}{z(1+\underline{m}t)^2}h\left(\frac{-1}{\underline{m}}\right)\right]\right\} + o(1).
\end{aligned}
$$

Therefore, from (3.33) and the identities

$$
\left[1 - \int \frac{ct\,dH(t)}{z(1+\underline{m}t)^2}\right]^{-1} = -z\underline{m}\left[1 - \int \frac{c\underline{m}^2t^2\,dH(t)}{(1+\underline{m}t)^2}\right]^{-1} = -\frac{z\underline{m}'}{\underline{m}},
$$

we obtain

$$
M_n^{(2)}(z) = \frac{S_n - \underline{S}_n T_n/\underline{T}_n}{1 - T_n/\underline{T}_n} = \mu_1(z) + (\tau-3)\mu_2(z) + o(1).
$$

The proof is complete.

Appendix: Additional lemmas and proofs

In this appendix, we list some supporting lemmas and their proofs. In order to ease the reading, we recall some notations that will be frequently encountered in this appendix. The population $\mathbf{x}$ has the structure $\mathbf{x} = w\mathbf{A}^{\frac{1}{2}}\mathbf{z}$, where $\mathbf{z} = (z_1, \dots, z_p)'$ is a random vector with i.i.d. standardized entries. Denote

$$\mathbf{s}(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}, \qquad \mathbf{r} = \sqrt{\frac{p}{n}}\mathbf{s}(\mathbf{x}) = \sqrt{\frac{p}{n}}\mathbf{s}\big(\mathbf{A}^{\frac{1}{2}}\mathbf{z}\big), \qquad \mathbf{\Sigma} = n\mathbb{E}(\mathbf{r}\mathbf{r}'),$$

$$\mathbf{T} = \mathbf{A} - \frac{2}{p}\mathbf{A}^2 - \frac{\tau-3}{p}\mathbf{A}^{\frac{1}{2}}\operatorname{diag}(\mathbf{A})\mathbf{A}^{\frac{1}{2}} + \left(\frac{2}{p^2}\operatorname{tr}\mathbf{A}^2 + \frac{\tau-3}{p^2}\operatorname{tr}(\mathbf{A}\circ\mathbf{A})\right)\mathbf{A}.$$

K is some constant that can vary from place to place and $\|\cdot\|$ denotes the Euclidean norm of a vector or the spectral norm of a matrix.

A.1. Lemmas

Lemma A.1 (Lemma 2.7 in [4]). *For $\mathbf{z} = (z_1, \dots, z_p)'$ i.i.d. standardized entries, $\mathbf{C}$ $p \times p$ matrix (complex) we have for any $k \geq 2$*

$$\mathbb{E}\big|\mathbf{z}'\mathbf{C}\mathbf{z} - \operatorname{tr}\mathbf{C}\big|^k \leq K\big[\big(\mathbb{E}|z_1|^4\operatorname{tr}\mathbf{C}\mathbf{C}^*\big)^{\frac{k}{2}} + \mathbb{E}|z_1|^{2k}\operatorname{tr}\big(\mathbf{C}\mathbf{C}^*\big)^{\frac{k}{2}}\big],$$

where K is a constant depending only on k.

Lemma A.2. *Suppose that Assumptions (a)–(c) and (3.19) hold with $\gamma \in (4, 4+\delta]$ for some $\delta > 0$. Then for any $p \times p$ complex matrices $\mathbf{C}$ and $\tilde{\mathbf{C}}$ with bounded spectral norms,*

$$\begin{aligned}
&\mathbb{E}\left(\mathbf{r}'\mathbf{C}\mathbf{r} - \frac{1}{n}\operatorname{tr}\mathbf{\Sigma}\mathbf{C}\right)\left(\mathbf{r}'\tilde{\mathbf{C}}\mathbf{r} - \frac{1}{n}\operatorname{tr}\mathbf{\Sigma}\tilde{\mathbf{C}}\right) \\
&\quad = \frac{1}{n^2}\operatorname{tr}\mathbf{A}\mathbf{C}\mathbf{A}\tilde{\mathbf{C}} + \frac{1}{n^2}\operatorname{tr}\mathbf{A}\mathbf{C}\mathbf{A}\tilde{\mathbf{C}}' + \frac{2}{p^2n^2}\operatorname{tr}\mathbf{A}^2\operatorname{tr}\mathbf{A}\mathbf{C}\operatorname{tr}\mathbf{A}\tilde{\mathbf{C}} - \frac{2}{pn^2}\operatorname{tr}\mathbf{A}^2\mathbf{C}\operatorname{tr}\mathbf{A}\tilde{\mathbf{C}} - \frac{2}{pn^2}\operatorname{tr}\mathbf{A}\mathbf{C}\operatorname{tr}\mathbf{A}^2\tilde{\mathbf{C}} \\
&\qquad + \frac{\tau-3}{n^2}\left\{\operatorname{tr}\big[\big(\mathbf{A}^{\frac{1}{2}}\mathbf{C}\mathbf{A}^{\frac{1}{2}}\big)\circ\big(\mathbf{A}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{A}^{\frac{1}{2}}\big)\big] + \frac{1}{p^2}\operatorname{tr}\mathbf{C}\mathbf{A}\operatorname{tr}\tilde{\mathbf{C}}\mathbf{A}\operatorname{tr}[\mathbf{A}\circ\mathbf{A}]\right. \\
&\qquad \left. - \frac{1}{p}\operatorname{tr}\mathbf{C}\mathbf{A}\operatorname{tr}\big[\big(\mathbf{A}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{A}^{\frac{1}{2}}\big)\circ\mathbf{A}\big] - \frac{1}{p}\operatorname{tr}\tilde{\mathbf{C}}\mathbf{A}\operatorname{tr}\big[\big(\mathbf{A}^{\frac{1}{2}}\mathbf{C}\mathbf{A}^{\frac{1}{2}}\big)\circ\mathbf{A}\big]\right\} + o\big(p^{-1}\big).
\end{aligned}$$

Lemma A.3. *Suppose that Assumptions (a)–(c) and (3.19) hold with $\gamma = 5$. We have*

$$\|\mathbf{\Sigma} - \mathbf{T}\| = o\big(p^{-1}\big).$$

Lemma A.4. *Suppose that Assumptions (a)–(c) and (3.19) hold with $\gamma \in (4, 4+\delta]$. For any $k \geq 2$ and $p \times p$ complex matrices $\mathbf{C}$ with bounded spectral norm,*

$$\begin{aligned}
\mathbb{E}\left|\mathbf{r}'\mathbf{C}\mathbf{r} - \frac{1}{n}\operatorname{tr}\mathbf{\Sigma}\mathbf{C}\right|^k &\leq Kn^{-k}\big[\mathbb{E}|z_1|^{2k}\operatorname{tr}(\mathbf{C}\mathbf{\Sigma})^k + \big(\mathbb{E}|z_1|^4\operatorname{tr}(\mathbf{C}\mathbf{\Sigma})^2\big)^{\frac{k}{2}} \\
&\qquad + \|\mathbf{C}\mathbf{\Sigma}\|^k\big(p^{\frac{k}{2}}\mathbb{E}^{\frac{k}{2}}|z_1|^4 + p\mathbb{E}|z_1|^{2k}\big)\big] \\
&\leq K\big(n^{-\frac{k}{2}} + n^{-1-\frac{k(\gamma-4)}{\gamma}}\big). \qquad \text{(A.1)}
\end{aligned}$$

where K is a constant depending only on k.

A.2. Proof of Lemma A.2

Denote $\mathbf{W} = \mathbf{A}^{\frac{1}{2}}\mathbf{C}\mathbf{A}^{\frac{1}{2}}$, $\mathbf{U} = \mathbf{A}^{\frac{1}{2}}\tilde{\mathbf{C}}\mathbf{A}^{\frac{1}{2}}$, and $s = \mathbf{z}'\mathbf{A}\mathbf{z}/p$. We consider the product of the quadratic form $n^2\mathbf{r}'\mathbf{C}\mathbf{r}\mathbf{r}'\tilde{\mathbf{C}}\mathbf{r} = \mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}/s^2$. From Lemma A.1 and the fact $\operatorname{tr}\mathbf{A} = p$, it holds that

$$\mathbb{E}|s-1|^k \le K\big(p^{-\frac{k}{2}} + p^{-1-\frac{k(\gamma-4)}{\gamma}}\big), \quad k \ge 2. \tag{A.2}$$

By the identity

$$\frac{1}{s^2} = 2 - s^2 + \big(1-s^2\big)^2 + s^{-2}\big(1-s^2\big)^3$$

and the inequality

$$\mathbb{E}\big(\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}\big)\big(s^{-2}\big(1-s^2\big)^3\big) \le Kp^2\mathbb{E}|1-s|^3 = o(p),$$

we have

$$n^2\mathbb{E}\mathbf{r}'\mathbf{C}\mathbf{r}\mathbf{r}'\tilde{\mathbf{C}}\mathbf{r} = \mathbb{E}\big(\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}\big)\big(6 - 8s + 3s^2\big) + o(p). \tag{A.3}$$

Therefore, the main task in the following is to derive the limits for the three terms $\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}$, $\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}s$ and $\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}s^2$ up to the order $O(p)$.

For the first term $\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}$, we have

$$\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z} = \mathbb{E}\sum_{i,j,k,\ell} z_i z_j z_k z_\ell \mathbf{W}_{ij}\mathbf{U}_{k\ell}.$$

Since all the p components z_i are independent and standardized, with mean zero, variance one and finite fourth moment, the terms that will contribute are the ones with their indexes either can be glued together or divided into two groups, that is, $i = j = k = \ell$, or $i = j \ne k = \ell$, or $i = k \ne j = \ell$ or $i = \ell \ne j = k$. All the four cases together gives

$$\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z} = \operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U} + \operatorname{tr}\mathbf{W}\mathbf{U} + \operatorname{tr}\mathbf{W}'\mathbf{U} + (\tau-3)\sum_i \mathbf{W}_{ii}\mathbf{U}_{ii} + o(p). \tag{A.4}$$

For the second term $\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}s$, we have

$$\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}s = \frac{1}{p}\mathbb{E}\sum_{i,j,k,\ell,s,u} z_i z_j z_k z_\ell z_s z_u \mathbf{W}_{ij}\mathbf{U}_{k\ell}\mathbf{A}_{su}. \tag{A.5}$$

The terms that will contribute up to order $O(p)$ are in $\sum_{(2)}$ and $\sum_{(3)}$, where the index $(\cdot)$ denotes the number of distinct integers in the set $\{i, j, k, \ell, s, u\}$. It can be checked that the following three cases should be counted in $\sum_{(2)}$ (all have the form of the product of two traces)

case 1: $i = j \ne k = \ell = s = u$,
case 2: $k = \ell \ne i = j = s = u$,
case 3: $s = u \ne i = j = k = \ell$,
while in $\sum_{(3)}$ the following four cases should be taken into account,
case 1: $k = s \ne \ell = u \ne i = j$ and $k = u \ne \ell = s \ne i = j$,
case 2: $i = s \ne j = u \ne k = \ell$ and $i = u \ne j = s \ne k = \ell$,
case 3: $i = \ell \ne j = k \ne s = u$ and $i = k \ne j = \ell \ne s = u$,
case 4: $i = j \ne k = \ell \ne s = u$.

Combining the contribution of each cases in $\sum_{(2)}$ and $\sum_{(3)}$, we have

$$\begin{aligned}
\text{case } 1 &= \frac{\tau+o(1)}{p}\sum_{i\neq k}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{kk} + \frac{2}{p}\sum_{i\neq k\neq \ell}\mathbf{W}_{ii}\mathbf{U}_{k\ell}\mathbf{A}_{\ell k}\\
&= \frac{\tau-2}{p}\sum_{i\neq k}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{kk} + \frac{2}{p}\sum_{i\neq k}\mathbf{W}_{ii}(\mathbf{UA})_{kk} + o(p)\\
&= \frac{\tau-2}{p}\operatorname{tr}\mathbf{W}\sum_{k}\mathbf{U}_{kk}\mathbf{A}_{kk} + \frac{2}{p}\operatorname{tr}\mathbf{W}\operatorname{tr}(\mathbf{UA}) + o(p),\\
\text{case } 2 &= \frac{\tau+o(1)}{p}\sum_{i\neq k}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ii} + \frac{2}{p}\sum_{i\neq j\neq k}\mathbf{W}_{ij}\mathbf{U}_{kk}\mathbf{A}_{ji}\\
&= \frac{\tau-2}{p}\sum_{i\neq k}\mathbf{W}_{ii}\mathbf{A}_{ii}\mathbf{U}_{kk} + \frac{2}{p}\sum_{i\neq k}\mathbf{U}_{kk}(\mathbf{WA})_{ii} + o(p)\\
&= \frac{\tau-2}{p}\operatorname{tr}\mathbf{U}\sum_{i}\mathbf{W}_{ii}\mathbf{A}_{ii} + \frac{2}{p}\operatorname{tr}\mathbf{U}\operatorname{tr}(\mathbf{WA}) + o(p),\\
\text{case } 3 &= \frac{\tau+o(1)}{p}\sum_{s\neq i}\mathbf{W}_{ii}\mathbf{U}_{ii}\mathbf{A}_{ss} + \frac{1}{p}\sum_{i\neq j\neq s}\mathbf{W}_{ij}\mathbf{U}_{ji}\mathbf{A}_{ss} + \frac{1}{p}\sum_{i\neq j\neq s}\mathbf{W}_{ij}\mathbf{U}^{*}_{ji}\mathbf{A}_{ss}\\
&= \frac{\tau-2}{p}\sum_{s\neq i}\mathbf{W}_{ii}\mathbf{U}_{ii}\mathbf{A}_{ss} + \frac{1}{p}\sum_{i\neq s}\mathbf{A}_{ss}(\mathbf{WU})_{ii} + \frac{1}{p}\sum_{i\neq s}\mathbf{A}_{ss}\big(\mathbf{WU}^{*}\big)_{ii} + o(p)\\
&= \frac{\tau-2}{p}\operatorname{tr}\mathbf{A}\sum_{i}\mathbf{W}_{ii}\mathbf{U}_{ii} + \frac{1}{p}\operatorname{tr}\mathbf{A}\operatorname{tr}(\mathbf{WU}) + \frac{1}{p}\operatorname{tr}\mathbf{A}\operatorname{tr}\big(\mathbf{WU}^{*}\big) + o(p),\\
\text{case } 4 &= \frac{1}{p}\sum_{i\neq k\neq s}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ss}\\
&= \frac{1}{p}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U}\operatorname{tr}\mathbf{A} - \frac{1}{p}\operatorname{tr}\mathbf{A}\sum_{i}\mathbf{W}_{ii}\mathbf{U}_{ii} - \frac{1}{p}\operatorname{tr}\mathbf{U}\sum_{i}\mathbf{W}_{ii}\mathbf{A}_{ii}\\
&\quad - \frac{1}{p}\operatorname{tr}\mathbf{W}\sum_{i}\mathbf{A}_{ii}\mathbf{U}_{ii} + o(p),
\end{aligned}$$

which further gives

$$\begin{aligned}
\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}s &= \text{case } 1 + \text{case } 2 + \text{case } 3 + \text{case } 4 + o(p)\\
&= \frac{1}{p}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U}\operatorname{tr}\mathbf{A} + \frac{2}{p}\operatorname{tr}\mathbf{W}\operatorname{tr}(\mathbf{UA}) + \frac{2}{p}\operatorname{tr}\mathbf{U}\operatorname{tr}(\mathbf{WA})\\
&\quad + \frac{1}{p}\operatorname{tr}\mathbf{A}\operatorname{tr}(\mathbf{WU}) + \frac{1}{p}\operatorname{tr}\mathbf{A}\operatorname{tr}\big(\mathbf{WU}^{*}\big) + \frac{\tau-3}{p}\operatorname{tr}\mathbf{W}\sum_{k}\mathbf{U}_{kk}\mathbf{A}_{kk}\\
&\quad + \frac{\tau-3}{p}\operatorname{tr}\mathbf{U}\sum_{i}\mathbf{W}_{ii}\mathbf{A}_{ii} + \frac{\tau-3}{p}\operatorname{tr}\mathbf{A}\sum_{i}\mathbf{W}_{ii}\mathbf{U}_{ii} + o(p). \qquad \text{(A.6)}
\end{aligned}$$

Finally, for the third term $\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}s^2$, we have

$$\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}s^2 = \frac{1}{p^2}\mathbb{E}\sum_{i,j,k,\ell,s,u,m,b} z_i z_j z_k z_\ell z_s z_u z_m z_b \mathbf{W}_{ij}\mathbf{U}_{k\ell}\mathbf{A}_{su}\mathbf{A}_{mb}.$$

The terms that will make the main contribution up to order $O(p)$ are in $\sum_{(3)}$ and $\sum_{(4)}$. For example, when considering $\sum_{(1)}$, we have

$$\sum_{(1)} = \mathbb{E}\sum_i \frac{1}{p^2} z_i^8 \mathbf{W}_{ii}\mathbf{U}_{ii}\mathbf{A}_{ii}^2 = O\big(p^{1-4(\gamma-4)/\gamma}\big) = o(p)$$

by using the assumptions in (3.19). Similar technique can be applied to dealing with the terms in $\sum_{(2)}$ and get their $o(p)$ bounds, which thus can be neglected. For terms in $\sum_{(3)}$ and $\sum_{(4)}$, we list in the following all the cases that should be counted, which are all up to order $O(p)$. For $\sum_{(3)}$, we have six cases

case 1: $i = j \neq k = \ell \neq s = u = m = b$,
case 2: $i = j = s = u \neq k = \ell \neq m = b$,
case 3: $i = j = m = b \neq k = \ell \neq s = u$,
case 4: $k = \ell = m = b \neq i = j \neq s = u$,
case 5: $i = j = k = \ell \neq s = u \neq m = b$,
case 6: $k = \ell = s = u \neq i = j \neq m = b$,
while in $\sum_{(4)}$, we have seven cases
case 1: $i = j \neq k = \ell \neq u = m \neq s = b$ and $i = j \neq k = \ell \neq s = m \neq u = b$,
case 2: $i = s \neq j = u \neq k = \ell \neq m = b$ and $i = u \neq j = s \neq k = \ell \neq m = b$,
case 3: $i = m \neq j = b \neq u = s \neq k = \ell$ ang $i = b \neq j = m \neq k = \ell \neq s = u$,
case 4: $k = m \neq \ell = b \neq i = j \neq s = u$ and $k = b \neq \ell = m \neq i = j \neq s = u$,
case 5: $i = k \neq j = \ell \neq s = u \neq m = b$ and $i = \ell \neq j = k \neq s = u \neq m = b$,
case 6: $k = s \neq \ell = u \neq i = j \neq m = b$ and $k = u \neq \ell = s \neq i = j \neq m = b$,
case 7: $i = j \neq k = \ell \neq s = u \neq m = b$.
Combining the above, we have

$$\begin{aligned}
\text{case } 1 &= \frac{2}{p^2}\sum_{i\neq k\neq m\neq s}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ms}\mathbf{A}_{ms} + \frac{\tau+o(1)}{p^2}\sum_{i\neq k\neq s}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ss}^2\\
&= \frac{2}{p^2}\sum_{i\neq k\neq s}\mathbf{W}_{ii}\mathbf{U}_{kk}(\mathbf{A}\mathbf{A})_{ss} + \frac{\tau-2}{p^2}\sum_{i\neq k\neq s}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ss}^2 + o(p)\\
&= \frac{2}{p^2}\operatorname{tr}\mathbf{A}^2\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U} + \frac{\tau-2}{p^2}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U}\sum_s\mathbf{A}_{ss}^2 + o(p),\\
\text{case } 2 &= \frac{2}{p^2}\sum_{i\neq j\neq k\neq m}\mathbf{W}_{ij}\mathbf{U}_{kk}\mathbf{A}_{ij}\mathbf{A}_{mm} + \frac{\tau+o(1)}{p^2}\sum_{i\neq k\neq m}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ii}\mathbf{A}_{mm}\\
&= \frac{2}{p^2}\sum_{i\neq k\neq m}(\mathbf{W}\mathbf{A})_{ii}\mathbf{U}_{kk}\mathbf{A}_{mm} + \frac{\tau-2}{p^2}\sum_{i\neq k\neq m}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ii}\mathbf{A}_{mm} + o(p)\\
&= \frac{2}{p^2}\operatorname{tr}(\mathbf{W}\mathbf{A})\operatorname{tr}\mathbf{U}\operatorname{tr}\mathbf{A} + \frac{\tau-2}{p^2}\operatorname{tr}\mathbf{U}\operatorname{tr}\mathbf{A}\sum_i\mathbf{W}_{ii}\mathbf{A}_{ii} + o(p),
\end{aligned}$$

$$
\begin{aligned}
\text{case } 3 &= \text{case } 2,\\
\text{case } 4 &= \frac{2}{p^2}\sum_{k\neq\ell\neq i\neq s}\mathbf{W}_{ii}\mathbf{U}_{k\ell}\mathbf{A}_{ss}\mathbf{A}_{k\ell}+\frac{\tau}{p^2}\sum_{k\neq i\neq s}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{kk}\mathbf{A}_{ss}\\
&= \frac{2}{p^2}\sum_{k\neq i\neq s}(\mathbf{UA})_{kk}\mathbf{W}_{ii}\mathbf{A}_{ss}+\frac{\tau-2}{p^2}\sum_{k\neq i\neq s}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ss}\mathbf{A}_{kk}+o(p)\\
&= \frac{2}{p^2}\operatorname{tr}(\mathbf{UA})\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{A}+\frac{\tau-2}{p^2}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{A}\sum_k\mathbf{U}_{kk}\mathbf{A}_{kk}+o(p),\\
\text{case } 5 &= \frac{1}{p^2}\sum_{i\neq j\neq s\neq m}\mathbf{W}_{ij}\mathbf{U}_{ij}\mathbf{A}_{ss}\mathbf{A}_{mm}+\frac{1}{p^2}\sum_{i\neq j\neq s\neq m}\mathbf{W}_{ij}\mathbf{U}_{ji}\mathbf{A}_{ss}\mathbf{A}_{mm}\\
&\quad+\frac{\tau}{p^2}\sum_{i\neq s\neq m}\mathbf{W}_{ii}\mathbf{U}_{ii}\mathbf{A}_{ss}\mathbf{A}_{mm}\\
&= \frac{1}{p^2}\sum_{i\neq s\neq m}(\mathbf{WU})_{ii}\mathbf{A}_{ss}\mathbf{A}_{mm}+\frac{1}{p^2}\sum_{i\neq s\neq m}\big(\mathbf{WU}^*\big)_{ii}\mathbf{A}_{ss}\mathbf{A}_{mm}\\
&\quad+\frac{\tau-2}{p^2}\sum_{i\neq s\neq m}\mathbf{W}_{ii}\mathbf{U}_{ii}\mathbf{A}_{ss}\mathbf{A}_{mm}+o(p)\\
&= \frac{1}{p^2}\operatorname{tr}(\mathbf{WU})(\operatorname{tr}\mathbf{A})^2+\frac{1}{p^2}\operatorname{tr}\big(\mathbf{WU}^*\big)(\operatorname{tr}\mathbf{A})^2+\frac{\tau-2}{p^2}(\operatorname{tr}\mathbf{A})^2\sum_i\mathbf{W}_{ii}\mathbf{U}_{ii}+o(p),\\
\text{case } 6 &= \text{case } 4,\\
\text{case } 7 &= \frac{1}{p^2}\sum_{i\neq k\neq s\neq m}\mathbf{W}_{ii}\mathbf{U}_{kk}\mathbf{A}_{ss}\mathbf{A}_{mm}\\
&= \frac{1}{p^2}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U}(\operatorname{tr}\mathbf{A})^2-\frac{1}{p^2}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U}\sum_s\mathbf{A}_{ss}^2-\frac{2}{p^2}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{A}\sum_s\mathbf{A}_{ss}\mathbf{U}_{ss}\\
&\quad-\frac{1}{p^2}(\operatorname{tr}\mathbf{A})^2\sum_i\mathbf{W}_{ii}\mathbf{U}_{ii}-\frac{2}{p^2}\operatorname{tr}\mathbf{U}\operatorname{tr}\mathbf{A}\sum_i\mathbf{W}_{ii}\mathbf{A}_{ii}+o(p),
\end{aligned}
$$

which finally leads to

$$
\begin{aligned}
&\mathbb{E}\big(\mathbf{z}'\mathbf{W}\mathbf{z}\mathbf{z}'\mathbf{U}\mathbf{z}\big)s^2\\
&\quad=\frac{1}{p^2}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U}(\operatorname{tr}\mathbf{A})^2+\frac{2}{p^2}\operatorname{tr}\mathbf{A}^2\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U}+\frac{4}{p^2}\operatorname{tr}(\mathbf{WA})\operatorname{tr}\mathbf{U}\operatorname{tr}\mathbf{A}\\
&\qquad+\frac{4}{p^2}\operatorname{tr}(\mathbf{UA})\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{A}+\frac{1}{p^2}\operatorname{tr}(\mathbf{WU})(\operatorname{tr}\mathbf{A})^2+\frac{1}{p^2}\operatorname{tr}\big(\mathbf{WU}^*\big)(\operatorname{tr}\mathbf{A})^2
\end{aligned}
$$

$$+ \frac{\tau-3}{p^2} \operatorname{tr} \mathbf{W} \operatorname{tr} \mathbf{U} \sum_s \mathbf{A}_{ss}^2 + \frac{2\tau-6}{p^2} \operatorname{tr} \mathbf{U} \operatorname{tr} \mathbf{A} \sum_i \mathbf{W}_{ii} \mathbf{A}_{ii}$$

$$+ \frac{2\tau-6}{p^2} \operatorname{tr} \mathbf{W} \operatorname{tr} \mathbf{A} \sum_k \mathbf{U}_{kk} \mathbf{A}_{kk} + \frac{\tau-3}{p^2} (\operatorname{tr} \mathbf{A})^2 \sum_i \mathbf{W}_{ii} \mathbf{U}_{ii} + o(p). \tag{A.7}$$

Collecting (A.3), (A.4), (A.6), (A.7), we have

$$\mathbb{E}\mathbf{r}'\mathbf{C}\mathbf{r}\mathbf{r}'\tilde{\mathbf{C}}\mathbf{r} = \frac{\tau-3}{n^2} \sum_i \mathbf{W}_{ii}\mathbf{U}_{ii} + \frac{1}{n^2} \operatorname{tr} \mathbf{W} \operatorname{tr} U + \frac{1}{n^2} \operatorname{tr}(\mathbf{W}\mathbf{U}) + \frac{1}{n^2} \operatorname{tr}\big(\mathbf{W}'\mathbf{U}\big) + \frac{6}{p^2 n^2} \operatorname{tr} \mathbf{A}^2 \operatorname{tr} \mathbf{W} \operatorname{tr} \mathbf{U}$$

$$- \frac{4}{pn^2} \operatorname{tr}(\mathbf{W}\mathbf{A}) \operatorname{tr} \mathbf{U} - \frac{4}{pn^2} \operatorname{tr}(\mathbf{U}\mathbf{A}) \operatorname{tr} \mathbf{W} + \frac{3(\tau-3)}{p^2 n^2} \operatorname{tr} \mathbf{W} \operatorname{tr} \mathbf{U} \sum_s \mathbf{A}_{ss}^2$$

$$- \frac{2(\tau-3)}{pn^2} \operatorname{tr} \mathbf{W} \sum_k \mathbf{U}_{kk}\mathbf{A}_{kk} - \frac{2(\tau-3)}{pn^2} \operatorname{tr} \mathbf{U} \sum_i \mathbf{W}_{ii}\mathbf{A}_{ii} + o\big(p^{-1}\big). \tag{A.8}$$

On the other hand, using the identity

$$\frac{1}{s} = 2 - s + (1-s)^2 + s^{-1}(1-s)^3$$

and the inequality (A.2), we can derive

$$n\mathbb{E}\mathbf{r}'\mathbf{C}\mathbf{r} = \mathbb{E}\frac{1}{s}\mathbf{z}'\mathbf{W}\mathbf{z} = \mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}\big(3 - 3s + s^2\big) + o(1). \tag{A.9}$$

It is trivial to have

$$\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z} = \operatorname{tr} \mathbf{W} \tag{A.10}$$

and by applying (A.4) and (A.6) again,

$$\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}s = \frac{\tau-3}{p} \sum_i \mathbf{W}_{ii}\mathbf{A}_{ii} + \operatorname{tr} \mathbf{W} + \frac{1}{p} \operatorname{tr}(\mathbf{W}\mathbf{A}) + \frac{1}{p} \operatorname{tr}\big(\mathbf{W}^*\mathbf{A}\big), \tag{A.11}$$

$$\mathbb{E}\mathbf{z}'\mathbf{W}\mathbf{z}s^2 = \operatorname{tr} \mathbf{W} + \frac{2}{p^2} \operatorname{tr} \mathbf{W} \operatorname{tr}\big(\mathbf{A}^2\big) + \frac{4}{p} \operatorname{tr}(\mathbf{W}\mathbf{A}) + \frac{2(\tau-3)}{p} \sum_i \mathbf{W}_{ii}\mathbf{A}_{ii}$$

$$+ \frac{\tau-3}{p^2} \operatorname{tr} \mathbf{W} \sum_i \mathbf{A}_{ii}^2 + o(1). \tag{A.12}$$

Collecting (A.9)–(A.12) leads to

$$n\mathbb{E}\mathbf{r}'\mathbf{C}\mathbf{r} = \operatorname{tr} \mathbf{W} + \frac{\tau-3}{p^2} \operatorname{tr} \mathbf{W} \sum_i \mathbf{A}_{ii}^2 + \frac{2}{p^2} \operatorname{tr} \mathbf{W} \operatorname{tr} \mathbf{A}^2$$

$$- \frac{\tau-3}{p} \sum_i \mathbf{W}_{ii}\mathbf{A}_{ii} - \frac{2}{p} \operatorname{tr}(\mathbf{W}\mathbf{A}) + o(1). \tag{A.13}$$

Therefore, combining (A.8)–(A.13), we have reached

$$
\begin{aligned}
&\mathbb{E}\left(\mathbf{r}'\mathbf{C}\mathbf{r} - \frac{1}{n}\operatorname{tr}\mathbf{\Sigma}\mathbf{C}\right)\left(\mathbf{r}'\tilde{\mathbf{C}}\mathbf{r} - \frac{1}{n}\operatorname{tr}\mathbf{\Sigma}\tilde{\mathbf{C}}\right)\\
&\quad = \mathbb{E}\mathbf{r}'\mathbf{C}\mathbf{r}\mathbf{r}'\tilde{\mathbf{C}}\mathbf{r} - \mathbb{E}\mathbf{r}'\mathbf{C}\mathbf{r}\mathbb{E}\mathbf{r}'\tilde{\mathbf{C}}\mathbf{r}\\
&\quad = \frac{1}{n^2}\operatorname{tr}\big[(\mathbf{W}' + \mathbf{W})\mathbf{U}\big] + \frac{2}{p^2n^2}\operatorname{tr}\mathbf{A}^2\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U} - \frac{2}{pn^2}\operatorname{tr}(\mathbf{W}\mathbf{A})\operatorname{tr}\mathbf{U}\\
&\qquad - \frac{2}{pn^2}\operatorname{tr}(\mathbf{U}\mathbf{A})\operatorname{tr}\mathbf{W} + \frac{\tau-3}{n^2}\operatorname{tr}(\mathbf{W}\circ\mathbf{U}) + \frac{\tau-3}{p^2n^2}\operatorname{tr}\mathbf{W}\operatorname{tr}\mathbf{U}\operatorname{tr}(\mathbf{A}\circ\mathbf{A})\\
&\qquad - \frac{\tau-3}{pn^2}\operatorname{tr}\mathbf{W}\operatorname{tr}(\mathbf{U}\circ\mathbf{A}) - \frac{\tau-3}{pn^2}\operatorname{tr}\mathbf{U}\operatorname{tr}(\mathbf{W}\circ\mathbf{A}) + o\big(p^{-1}\big). \qquad \text{(A.14)}
\end{aligned}
$$

The proof is then complete.

A.3. Proof of Lemma A.3

Using the identity

$$\frac{1}{s} = 2 - s + (1-s)^2 + s^{-1}(1-s)^3$$

we have

$$\mathbf{\Sigma} = \mathbb{E}\frac{1}{s}\mathbf{A}^{\frac{1}{2}}\mathbf{z}\mathbf{z}'\mathbf{A}^{\frac{1}{2}} = \mathbb{E}\mathbf{A}^{\frac{1}{2}}\mathbf{z}\mathbf{z}'\mathbf{A}^{\frac{1}{2}}\big(2 - s + (1-s)^2 + s^{-1}(1-s)^3\big), \qquad \text{(A.15)}$$

where $s = \mathbf{z}'\mathbf{A}\mathbf{z}/p$. First, we show that

$$\big\|\mathbb{E}\mathbf{A}^{\frac{1}{2}}\mathbf{z}\mathbf{z}'\mathbf{A}^{\frac{1}{2}}s^{-1}(1-s)^3\big\| = o\big(p^{-1}\big). \qquad \text{(A.16)}$$

Define an event $A = \{|s-1| > 1/2\}$ then, by Markov's inequality and (A.2), we have $\mathrm{P}(A) = o(n^{-s})$ for any $s > 0$. Therefore,

$$
\begin{aligned}
\big\|\mathbb{E}\mathbf{A}^{\frac{1}{2}}\mathbf{z}\mathbf{z}'\mathbf{A}^{\frac{1}{2}}s^{-1}(1-s)^3\big\| &\le K\big\|\mathbb{E}\mathbf{z}\mathbf{z}'s^{-1}(1-s)^3I(A)\big\| + K\big\|\mathbb{E}\mathbf{z}\mathbf{z}'s^{-1}(1-s)^3I\big(A^c\big)\big\|\\
&\le K\big\|\mathbb{E}\mathbf{z}\mathbf{z}'|1-s|^3\big\| + o\big(n^{-s}\big).
\end{aligned}
$$

Applying Hölder's inequality and (A.2), we have

$$
\begin{aligned}
\big\|\mathbb{E}\mathbf{z}\mathbf{z}'|1-s|^3\big\| &= \max_{\alpha\in\mathbb{R}^p,\|\alpha\|=1}\mathbb{E}\alpha'\mathbf{z}\mathbf{z}'\alpha|1-s|^3 \le \max_{\alpha\in\mathbb{R}^p,\|\alpha\|=1}\mathbb{E}\big|\mathbf{z}'\alpha\alpha'\mathbf{z} - 1\big||1-s|^3 + \mathbb{E}|1-s|^3\\
&\le \max_{\alpha\in\mathbb{R}^p,\|\alpha\|=1}\mathbb{E}^{\frac{1}{2}}\big|\mathbf{z}'\alpha\alpha'\mathbf{z} - 1\big|^2\mathbb{E}^{\frac{1}{2}}|1-s|^6 + o\big(p^{-1}\big),
\end{aligned}
$$

which is $o(p^{-1})$ from (A.2) and the fact $\mathbb{E}|\mathbf{z}'\alpha\alpha'\mathbf{z} - 1|^2 = O(1)$. Therefore, (A.16) is verified, which together with (A.15) give

$$\mathbf{\Sigma} = \mathbb{E}\mathbf{A}^{\frac{1}{2}}\mathbf{z}\mathbf{z}'\mathbf{A}^{\frac{1}{2}}\big(2 - s + (1-s)^2\big) + o\big(p^{-1}\big) = \mathbf{A}^{\frac{1}{2}}\big[\mathbb{E}\mathbf{z}\mathbf{z}'\big(3 - 3s + s^2\big)\big]\mathbf{A}^{\frac{1}{2}} + o\big(p^{-1}\big), \qquad \text{(A.17)}$$

where the "$o(p^{-1})$" is in terms of spectral norm.

Next, we deal with the terms $\mathbb{E}\mathbf{z}\mathbf{z}'s$ and $\mathbb{E}\mathbf{z}\mathbf{z}'s^2$. For $\mathbb{E}\mathbf{z}\mathbf{z}'s$, we have its (i, j)-th entry given by

$$[\mathbb{E}\mathbf{z}\mathbf{z}'s]_{(i,j)} = \frac{1}{p}\mathbb{E}z_i z_j \sum_{k,\ell} z_k z_\ell \mathbf{A}_{k\ell} = \begin{cases} \frac{1}{p}\mathbf{A}_{ij} + \frac{1}{p}\mathbf{A}_{ji} & i \neq j, \\ 1 + \frac{1}{p}\big(\tau - 1 + o(1)\big)\mathbf{A}_{ii} & i = j, \end{cases}$$

which gives

$$\mathbb{E}\mathbf{z}\mathbf{z}'s = \mathbf{I}_p + \frac{2}{p}\mathbf{A} + \frac{1}{p}\big(\tau - 3 + o(1)\big)\operatorname{diag}(\mathbf{A})$$

and

$$\mathbb{E}\mathbf{A}^{\frac{1}{2}}\mathbf{z}\mathbf{z}'\mathbf{A}^{\frac{1}{2}}s = \mathbf{A} + \frac{2}{p}\mathbf{A}^2 + \frac{1}{p}(\tau - 3)\mathbf{A}^{\frac{1}{2}}\operatorname{diag}(\mathbf{A})\mathbf{A}^{\frac{1}{2}} + o\big(p^{-1}\big). \tag{A.18}$$

For the term $\mathbb{E}\mathbf{z}\mathbf{z}'s^2$, similar to the derivation of (A.5), we have its (i, j)-th entry is given by

$$\begin{aligned}[\mathbb{E}\mathbf{z}\mathbf{z}'s^2]_{(i,j)} &= \frac{1}{p^2}\mathbb{E}z_i z_j \sum_{k,\ell,s,u} z_k z_\ell z_s z_u \mathbf{A}_{k\ell}\mathbf{A}_{su} \\ &= \begin{cases} \frac{4}{p}\mathbf{A}_{ij} - \frac{4}{p^2}\mathbf{A}_{ii}\mathbf{A}_{ij} - \frac{4}{p^2}\mathbf{A}_{ij}\mathbf{A}_{jj} + o\big(p^{-2}\big) & i \neq j, \\ \frac{1}{p^2}(\tau - 3)\sum_k \mathbf{A}_{kk}^2 + \frac{2}{p}(\tau - 1)\mathbf{A}_{ii} + \frac{2}{p^2}\operatorname{tr}\mathbf{A}^2 + 1 + o\big(p^{-1}\big) & i = j. \end{cases}\end{aligned}$$

Therefore, we get

$$\mathbb{E}\mathbf{z}\mathbf{z}'s^2 = \frac{4}{p}\mathbf{A} + \frac{1}{p^2}(\tau - 3)\operatorname{tr}(\mathbf{A}\circ\mathbf{A})\mathbf{I}_p + \mathbf{I}_p + \frac{2}{p}(\tau - 3)\operatorname{diag}(\mathbf{A}) + \frac{2}{p^2}\operatorname{tr}\mathbf{A}^2\cdot\mathbf{I}_p + o\big(p^{-1}\big),$$

which further gives that

$$\begin{aligned}\mathbb{E}\mathbf{A}^{\frac{1}{2}}\mathbf{z}\mathbf{z}'\mathbf{A}^{\frac{1}{2}}s^2 &= \frac{4}{p}\mathbf{A}^2 + \frac{1}{p^2}(\tau - 3)\operatorname{tr}(\mathbf{A}\circ\mathbf{A})\mathbf{A} + \mathbf{A} \\ &\quad + \frac{2}{p}(\tau - 3)\mathbf{A}^{\frac{1}{2}}\operatorname{diag}(\mathbf{A})\mathbf{A}^{\frac{1}{2}} + \frac{2}{p^2}\operatorname{tr}\mathbf{A}^2\cdot\mathbf{A} + o\big(p^{-1}\big).\end{aligned} \tag{A.19}$$

Collecting (A.17), (A.18) and (A.19), we obtain

$$\begin{aligned}\boldsymbol{\Sigma} &= \mathbf{A} - \frac{\tau - 3}{p}\mathbf{A}\operatorname{diag}(\mathbf{A})\mathbf{A}' - \frac{2}{p}\mathbf{A}^2 + \left(\frac{\tau - 3}{p^2}\operatorname{tr}(\mathbf{A}\circ\mathbf{A}) + \frac{2}{p^2}\operatorname{tr}\mathbf{A}^2\right)\mathbf{A} + o\big(p^{-1}\big) \\ &= \mathbf{T} + o\big(p^{-1}\big).\end{aligned}$$

The proof is thus complete.

A.4. Proof of Lemma A.4

This lemma can be obtained from similar arguments for the proof of Lemma 6 in [24]. We omit the details.

Acknowledgements

Weiming Li's research is partially supported by NSFC (No. 11971293) and Program of IRTSHUFE. Qinwen Wang acknowledges support from a NSFC Grant (No. 11801085) and the Shanghai Sailing Program (No. 18YF1401500). Jianfeng Yao's research is partly supported by a HKSAR RGC Grant (GRF 17306918). Wang Zhou's research is partially supported by a grant R-155-000-192-114 at the National University of Singapore.

References

[1] Anderson, G.W. and Zeitouni, O. (2006). A CLT for a band matrix model. *Probab. Theory Related Fields* **134** 283–338. MR2222385 https://doi.org/10.1007/s00440-004-0422-3

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

[3] Bai, Z. and Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

[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 https://doi.org/10.1214/aop/1022855421

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

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

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

[8] Billingsley, P. (1979). *Probability and Measure*. *Wiley Series in Probability and Mathematical Statistics*. New York–Chichester–Brisbane: Wiley. MR0534323

[9] Chakraborty, A. and Chaudhuri, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann. Statist.* **45** 771–799. MR3650400 https://doi.org/10.1214/16-AOS1467

[10] Dürre, A., Tyler, D.E. and Vogel, D. (2016). On the eigenvalues of the spatial sign covariance matrix in more than two dimensions. *Statist. Probab. Lett.* **111** 80–85. MR3474786 https://doi.org/10.1016/j.spl.2016.01.009

[11] Dürre, A., Vogel, D. and Fried, R. (2015). Spatial sign correlation. *J. Multivariate Anal.* **135** 89–105. MR3306428 https://doi.org/10.1016/j.jmva.2014.12.002

[12] Feng, L. and Sun, F. (2016). Spatial-sign based high-dimensional location test. *Electron. J. Stat.* **10** 2420–2434. MR3544292 https://doi.org/10.1214/16-EJS1176

[13] Feng, L., Zou, C. and Wang, Z. (2016). Multivariate-sign-based high-dimensional tests for the two-sample location problem. *J. Amer. Statist. Assoc.* **111** 721–735. MR3538700 https://doi.org/10.1080/01621459.2015.1035380

[14] Gao, J., Han, X., Pan, G. and Yang, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402 https://doi.org/10.1111/rssb.12189

[15] Hu, J., Li, W., Liu, Z. and Zhou, W. (2019). High-dimensional covariance matrices in elliptical distributions with application to spherical test. *Ann. Statist.* **47** 527–555. MR3909941 https://doi.org/10.1214/18-AOS1699

[16] Hu, J., Li, W. and Zhou, W. (2019). Central limit theorem for mutual information of large MIMO systems with elliptically correlated channels. *IEEE Trans. Inf. Theory* **65** 7168–7180. MR4030880 https://doi.org/10.1109/TIT.2019.2913760

[17] Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** 151–204. MR1487983 https://doi.org/10.1215/S0012-7094-98-09108-6

[18] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** 1–38. MR0650926 https://doi.org/10.1016/0047-259X(82)90080-X

[19] Li, Y., Wang, Z. and Zou, C. (2016). A simpler spatial-sign-based two-sample test for high-dimensional data. *J. Multivariate Anal.* **149** 192–198. MR3507323 https://doi.org/10.1016/j.jmva.2016.04.004

[20] Locantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T. and Cohen, K.L. (1999). Robust principal component analysis for functional data. *TEST* **8** 1–73. MR1707596 https://doi.org/10.1007/BF02595862

[21] Lytova, A. and Pastur, L. (2009). Central limit theorem for linear eigenvalue statistics of the Wigner and the sample covariance random matrices. *Ann. Probab.* **37** 1778-1840.

[22] Magyar, A.F. and Tyler, D.E. (2014). The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions. *Biometrika* **101** 673–688. MR3254908 https://doi.org/10.1093/biomet/asu020

[23] Pastur, L. and Marčenko, V. (1967). The distribution of eigenvalues in certain sets of random matrices. *Math. USSR, Sb.* **1** 457–483.

[24] Morales-Jimenez, D., Johnstone, I., McKay, M. and Yang, J. (2021). Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models. *Statist. Sinica* **31** 571–601.

[25] 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 https://doi.org/10.1214/07-AAP477

[26] Silverstein, J.W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408 https://doi.org/10.1006/jmva.1995.1083

[27] Sinai, Ya. and Soshnikov, A. (1998). Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Bras. Mat.* **29** 1–24. MR1620151 https://doi.org/10.1007/BF01245866

[28] Visuri, S., Koivunen, V. and Oja, H. (2000). Sign and rank covariance matrices. *J. Statist. Plann. Inference* **91** 557–575. MR1814801 https://doi.org/10.1016/S0378-3758(00)00199-3

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

[30] Yin, Y.Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** 50–68. MR0862241 https://doi.org/10.1016/0047-259X(86)90019-9

[31] Zheng, S., Bai, Z. and Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292

Received January 2021 and revised April 2021