Submitted to Bernoulli

Central limit theorem for linear spectral statistics of large dimensional separable sample covariance matrices

ZHIDONG BAI[1] HUIQIN LI[2] and GUANGMING PAN[3]
*

[1] *KLASMOE and School of Mathematics & Statistics, Northeast Normal University, Changchun, P.R.C., 130024. E-mail:* baizd@nenu.edu.cn

[2] *School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, P.R.C., 221116. E-mail:* huiqinli@jsnu.edu.cn

[3] *Division of Mathematical Sciences, School of Physical and Mathmatical Sciences, Nanyang Technological University, Singapore 637371. E-mail:* GMPAN@ntu.edu.sg

Suppose that $\mathbf{X}_n = (x_{jk})$ is $N \times n$ whose elements are independent complex variables with mean zero, variance 1. The separable sample covariance matrix is defined as $\mathbf{B}_n = \frac{1}{N}\mathbf{T}_{2n}^{1/2}\mathbf{X}_n\mathbf{T}_{1n}\mathbf{X}_n^*\mathbf{T}_{2n}^{1/2}$ where $\mathbf{T}_{1n}$ is a Hermitian matrix and $\mathbf{T}_{2n}^{1/2}$ is a Hermitian square root of the nonnegative definite Hermitian matrix $\mathbf{T}_{2n}$. Its linear spectral statistics (LSS) are shown to have Gaussian limits when n/N approaches a positive constant under some conditions.

Keywords: Central limit theorem, Separable sample covariance matrix, Linear spectral statistics, Random matrix theory. .

1. Introduction

Random matrix theory has found many applications in physics, statistics and engineering since its inception. Although early developments were motivated by practical experimental problems, random matrices are now used in fields as diverse as stochastic differential equations, condensed matter physics, chaotic systems, numerical linear algebra, neural networks, multivariate statistics, information theory, signal processing and so on.

In the last few years, a considerable body of work has emerged in the communications and information theory literature on the fundamental limits of communication channels that makes substantial use of results in random matrix theory (see Li et al. (2006); Nadakuditi and Edelman (2008); Tulino and Verdu (2004)). Most of the information theoretic literature that studies the

*Z. D. Bai was partially supported by NSFC 11571067 and 11471140; H. Q. Li was partially supported by NSFC 11701234, China Scholarship Council and the Priority Academic Program Development of Jiangsu Higher Education Institutions; Guangming Pan was supported in part by by a MOE Tier 2 grant 2014-T2-2-060 and by a MOE Tier 1 Grant RG25/14 at the Nanyang Technological University, Singapore.

effect of those features on channel capacity deals with linear vector memoryless channels of the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1.1}$$

where $\mathbf{x}$ is the K-dimensional input vector, $\mathbf{y}$ is the N-dimensional output vector, and the N-dimensional vector $\mathbf{n}$ models the additive circularly symmetric Gaussian noise. All these quantities are, in general, complex-valued. In addition to input constraints, and the degree of knowledge of the channel at receiver and transmitter, (1.1) is characterized by the distribution of the $N \times K$ random channel matrix $\mathbf{H}$ whose entries are also complex-valued. Especially, define $\mathbf{H} = \mathbf{CSA}$ where $\mathbf{S}$ is an $N \times K$ matrix whose entries are independent complex random variables satisfying the Lindeberg condition, i.e., for any $\delta > 0$,

$$\frac{1}{K} \sum_{j,k} \mathrm{E}\left(s_{jk}^2 I(|h_{jk}| \geq \delta) \right) \to 0$$

with identical means and variance $\frac{1}{N}$. Let $\mathbf{C}$ and $\mathbf{A}$ be, respectively, $N \times N$ and $K \times K$ random matrices such that the asymptotic spectra of $\mathbf{D} = \mathbf{CC}^*$ and $\mathbf{T} = \mathbf{AA}^*$ converge almost surely to compactly supported measures. If $\mathbf{C}, \mathbf{A}$ and $\mathbf{S}$ are independent, as $K, N \to \infty$ with $K/N \to \beta$. Then, the most concerned problem is the asymptotic properties of $\mathbf{HH}^*$ in communication channels. For details, please refer to Tulino and Verdu (2004). $\mathbf{HH}^*$ can be treated as the extension of the common sample covariance matrix.

The sample covariance matrix is one of the most commonly studied random matrices in Random Matrix Theory, which can be traced back to Wishart (1928) (see Wishart (1928)). It plays an important role in multivariate analysis because many statistics in traditional multivariate statistical analysis (e.g., principle component analysis, factor analysis and multivariate regression analysis) can be written as functionals of the eigenvalues of sample covariance matrices.

Large dimensional data now appear in various fields such as finance and genetic experiments due to different reasons. To deal with such large-dimensional data, a new area in asymptotic statistics has been developed where the data dimension p is no more fixed but tends to infinity together with the sample size n. The random matrices proves to be a powerful tool for such large dimensional statistical problems. One may refer to the latest book in this area by Yao, Zheng and Bai (2015), the recent work by Ledoit and Wolf (2002); Jiang and Yang (2013).

So far, most work focus on the sample covariance matrices of the form

$$\mathbf{S}_n = \frac{1}{N} \mathbf{T}_n^{1/2} \mathbf{X}_n \mathbf{X}_n^* \mathbf{T}_n^{1/2}$$

where $\mathbf{X}_n$ is a $N \times n$ matrix with independent entries and $\mathbf{T}_n$ is a nonnegative definite Hermitian matrix. As we know $\mathbf{S}_n$ can be viewed as a sample covariance matrix formed from n samples of the random vector $\mathbf{T}_n^{1/2} \mathbf{x}_1$(where $\mathbf{x}_1$ denotes the first column of $\mathbf{X}_n$, which has population covariance matrix $\mathbf{T}_n$. Much work has been done on the central limit theorem (CLT) for linear eigenvalues statistics of $\mathbf{S}_n$ under different assumptions. Among others we mention Bai and Silverstein (2004); Jonsson (1982); Najim and Yao (2016); Pan and Zhou (2008); Shcherbina (2011). One of the key features of the above sample covariance matrices $\mathbf{S}_n$ is that the sample are independent. As far as we know there is no CLT available for the sample covariance matrices generated from the dependent sample.

In view of the above we consider a kind of general sample covariance matrices

$$\mathbf{B}_n = \frac{1}{N}\mathbf{T}_{2n}^{1/2}\mathbf{X}_n\mathbf{T}_{1n}\mathbf{X}_n^*\mathbf{T}_{2n}^{1/2}, \tag{1.2}$$

where $\mathbf{T}_{2n}$ is $N \times N$ nonnegative definite Hermitian matrix and $\mathbf{T}_{1n}$ is $n \times n$ Hermitian. This model finds applications in the diverse fields including spatio-temporal statistics, wireless communications and econometrics. For example, the data matrix can be represented as

$$\mathbf{Y}_n = \mathbf{T}_{2n}^{1/2}\mathbf{X}_n\mathbf{T}_{1n}^{1/2} \tag{1.3}$$

if $\mathbf{T}_{1n}$ is nonnegative definite Hermitian. Denote by $\mathrm{vec}(\mathbf{Y}_n)$ the vector operator that stacks the columns of $\mathbf{Y}_n$ into a column vector. This model is referred to as a separable covariance model because the covariance of $\mathrm{vec}(\mathbf{Y}_n)$ is the Kronecker product of $\mathbf{T}_{1n}$ and $\mathbf{T}_{2n}$. The rows of the data matrix $\mathbf{Y}_n$ correspond to indices of spatial locations and the column indices correspond to points in time in the field of spatio-temporal statistics. This covariance structure implies that the entries of $\mathbf{Y}_n$ are correlated in time (column), but the pattern of temporal correlation does not change with location (row). One may see Paul and Silverstein (2009) and the references therein.

In econometrics, when determining the number of factors in the approximate factor models Onatski (2010) assumes that the idiosyncratic components of the data is of the form $\mathbf{Y}_n$. This allows the idiosyncratic terms to be non-trivially correlated both cross-sectionally and over time. The cross-sectional correlation is caused by matrix $\mathbf{T}_{2n}^{1/2}$ linearly combining different rows of $\mathbf{X}_n$, whereas the correlation over time is caused by matrix $\mathbf{T}_{1n}^{1/2}$ linearly combining different columns of $\mathbf{X}_n$.

For any Hermitian matrix $\mathbf{A}$ of size $n \times n$ its empirical spectral distribution (ESD) is defined by

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

where $\{\lambda_j\}$ are eigenvalues of $\mathbf{A}$. For $\mathbf{B}_n$ defined in (1.2), a number of papers (Boutet de Mondvel et al. (1996) and Zhang (2006)) investigated its empirical spectral distribution $F_{\mathbf{B}_n}$ and the weakest assumption is given in Zhang (2006), which is specified below. To characterize its limit define the Stieltjes transform of any distribution function $F^{\mathbf{A}}(x)$ to be

$$m_{F^{\mathbf{A}}}(z) = \int \frac{1}{x-z}dF^{\mathbf{A}}(x) = \frac{1}{n}\mathrm{tr}(\mathbf{A} - z\mathbf{I})^{-1}, \quad z \in \mathbb{C}^+.$$

Throughout the paper we make the following assumption.

***Condition* 1.1.**

(i) $\mathbf{X}_n = (x_{jl})$ is $N \times n$ consisting of independent complex random variables with $\mathrm{E}x_{jl} = 0, \mathrm{E}|x_{jl}|^2 = 1$, satisfying for each $\delta > 0$, as $n \to \infty$

$$\frac{1}{\delta^2 nN}\sum_{j,l}\mathrm{E}\left(|x_{jl}|^2 I\left(|x_{jl}| > \delta\sqrt{n}\right)\right) \to 0.$$

(ii) $\mathbf{T}_{1n}$ is $n \times n$ Hermitian matrix (without loss of generality, we assume that $\mathbf{T}_{1n}$ is not semi-negative definite) and $\mathbf{T}_{2n}$ is $N \times N$ nonnegative definite Hermitian matrix.
(iii) With probability 1, as $n \to \infty$, the empirical spectral distributions of $\mathbf{T}_{1n}$ and $\mathbf{T}_{2n}$, denoted by H_{1n} and H_{2n} respectively, converge weakly to two probability functions H_1 and H_2, respectively.
(iv) $N = N(n)$ with $n/N \to c > 0$.
(v) $\mathbf{X}_n, \mathbf{T}_{1n}, \mathbf{T}_{2n}$ are independent.

Zhang (2006) establishes the following conclusion under Condition 1.1. For $\mathbf{B}_n$ defined in (1.2), with probability 1, as $n \to \infty$, the ESD of $\mathbf{B}_n$ converges weakly to a non-random probability distribution function F for which if $H_1 = 1_{[0,\infty)}$ or $H_2 = 1_{[0,\infty)}$, then $F = 1_{[0,\infty)}$; otherwise the Stieltjes transform $m(z)$ of F is determined by the following system of equations (1.4), where for each $z \in \mathbb{C}^+$,

$$\begin{cases} s(z) = -z^{-1}(1-c) - z^{-1}c \int \frac{1}{1+q(z)x} dH_1(x) \\ s(z) = -z^{-1} \int \frac{1}{1+p(z)y} dH_2(y) \\ s(z) = -z^{-1} - p(z)q(z). \end{cases} \tag{1.4}$$

Then, the Stieltjes transfrom $m(z)$ of F, together with the two other functions, denoted by $g_1(z)$ and $g_2(z)$, $(m(z), g_1(z), g_2(z))$ is the unique solution to (1.4) in the set

$$U = \left\{ (s(z), p(z), q(z)) : \Im s(z) > 0, \Im(zp(z)) > 0, \Im q(z) > 0 \right\}$$

where $\Im h(z)$ stands for the imaginary part of $h(z)$. Denote $\underline{\mathbf{B}}_n = \frac{1}{N}\mathbf{T}_{1n}\mathbf{X}_n^*\mathbf{T}_{2n}\mathbf{X}_n$. Then we have the following relationship between the empirical distributions of $\mathbf{B}_n$ and $\underline{\mathbf{B}}_n$

$$F^{\mathbf{B}_n}(x) = c_n F^{\underline{\mathbf{B}}_n}(x) + (1-c_n)I_{[0,\infty)}(x),$$

and hence

$$m_n(z) = c_n \underline{m}_n(z) + z^{-1}(c_n - 1). \tag{1.5}$$

where $c_n = n/N$, $m_n(z) = m_{F^{\mathbf{B}_n}}(z)$ and $\underline{m}_n(z) = m_{F^{\underline{\mathbf{B}}_n}}(z)$. Denote by $\underline{F}$ the limiting distribution of $F^{\underline{\mathbf{B}}_n}$. Then F and $\underline{F}$ must satisfy

$$F(x) = c\underline{F}(x) + (1-c)I_{[0,\infty)}(x),$$

and

$$m(z) = c\underline{m}(z) - z^{-1}(1-c) \tag{1.6}$$

where $\underline{m}(z) = m_{\underline{F}}(z)$. If we let F^{c,H_1,H_2} denote F, then $F^{c_n,H_{1n},H_{2n}}$ is obtained from F^{c,H_1,H_2} with c, H_1, H_2 replaced by c_n, H_{1n}, H_{2n} respectively. Let $m_n^0(z) = m_{F^{c_n,H_{1n},H_{2n}}}(z)$ for simplicity. Moreover $g_{1n}^0(z)$ and $g_{2n}^0(z)$ are similarly obtained from $g_1(z)$ and $g_2(z)$ respectively. Then $\left(m_n^0(z), g_{1n}^0(z), g_{2n}^0(z)\right)$ satisfies the equations (1.4). In other words

$$\underline{m}_n^0(z) = -z^{-1} \int \frac{1}{1 + g_{2n}^0(z)x} dH_{1n}(x) \tag{1.7}$$

$$m_n^0(z) = -z^{-1}\int \frac{1}{1+g_{1n}^0(z)y}dH_{2n}(y) \tag{1.8}$$

$$m_n^0(z) = -z^{-1} - g_{1n}^0(z)g_{2n}^0(z). \tag{1.9}$$

Furthermore,

$$zg_{1n}^0(z) = -c_n\int \frac{x}{1+g_{2n}^0(z)x}dH_{1n}(x) \tag{1.10}$$

$$zg_{2n}^0(z) = -\int \frac{y}{1+g_{1n}^0(z)y}dH_{2n}(y). \tag{1.11}$$

Couillet and Hachem (2014) further investigated the limiting spectral measure of $\mathbf{B}_n$ and Paul and Silverstein (2009) proved that no eigenvalues exist outside the support of limiting empirical spectral distribution of $\mathbf{B}_n$. But Paul and Silverstein (2009) required $\mathbf{T}_{2n}$ in $\mathbf{B}_n$ to be diagonal (with positive diagonal entries). It is well known that many important statistics in multivariate analysis can be written as functionals of the ESD of some random matrices. In view of this the aim of this paper is to establish the central limit theorem for linear spectral statistics (LSS) of $\mathbf{B}_n$. LSS of general sample covariance matrices are quantities of the form

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

where f is some continuous and bounded real function on $(-\infty,\infty)$.

This paper is organized as follows. Section 2 establishes the main result about the CLT for LSS of $\mathbf{B}_n$. By the Stieltjes transform method, we complete the proof of theorem when the entries of matrix are Gaussian variables in Section 3. Section 4 extends the result from the Gaussian case to the general case through comparing their characteristic functions.

The crucial step in proving the main result in this paper is Lemma 2.5, whose proof is divided into the random part and the nonrandom part. Since the latter is lengthy, we postpone it to Supplement A. Supplement B shows the analysis of the remainder term for general case in Section 4. Some useful lemmas are listed in Supplement C and Supplement D.

2. Main result

Define

$$G_n(x) = N\Big(F^{\mathbf{B}_n}(x) - F^{c_n,H_{1n},H_{2n}}(x)\Big).$$

The main result is stated in the following theorem.

Theorem 2.1. *Denote by $s_1 \geq \cdots \geq s_n$ ($s_1 > 0$) the eigenvalues of $\mathbf{T}_{1n}$. Let $f_1,\cdots,f_\kappa$ be functions on $\mathbb{R}$ analytic on an open interval containing*

$$\Bigg[\liminf_n s_n\left(\lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c)\left(1-\sqrt{c}\right)^2 I(s_n\geq 0) + \lambda_{\max}^{\mathbf{T}_{2n}}\left(1+\sqrt{c}\right)^2 I(s_n<0)\right),$$

$$\limsup_n s_1\left(\lambda_{\max}^{\mathbf{T}_{2n}}\left(1+\sqrt{c}\right)^2\right)\Big]. \tag{2.1}$$

In addition to Condition 1.1, we further suppose that for each $\delta > 0$, as $n \to \infty$

$$\frac{1}{\delta^4 n^2}\sum_{j,l} \mathrm{E}\left(\left|x_{jl}\right|^4 I\left(|x_{jl}| > \delta\sqrt{n}\right)\right) \to 0.$$

Also suppose that $\mathbf{T}_{1n}$ and $\mathbf{T}_{2n}$ are nonrandom matrices, and their spectral norms are both bounded in n. Then

(i) *If $\mathbf{X}_n = (x_{jk})$, $\mathbf{T}_{1n}$, $\mathbf{T}_{2n}$ are real and $\mathrm{E}|x_{jk}|^4 = 3$, $j = 1, \cdots, N, k = 1, \cdots, n$, then*

$$\left(\int f_1(x)dG_n(x), \cdots, \int f_\kappa(x)dG_n(x)\right) \tag{2.2}$$

converges weakly to a Gaussian vector $\left(X_{f_1}, \cdots, X_{f_\kappa}\right)$ with mean

$$\begin{aligned}\mathrm{E}X_f = &-\frac{1}{2\pi i}\oint_C f(z)\Bigg\{(d_1(z)-d_2(z))\Bigg\{1-z^{-1}\left[\int\frac{x}{(1+xg_2(z))^2}dH_1(x)\right]^{-1}\\ &\times\int\frac{x^2}{(1+xg_2(z))^2}dH_1(x)\int\frac{t}{(1+g_1(z)t)^2}dH_2(t)\Bigg\}^{-1}\Bigg\}dz\end{aligned} \tag{2.3}$$

and covariance function

$$\mathrm{Cov}\left(X_f, X_g\right) = -\frac{1}{2\pi^2}\oint_{C_1}\oint_{C_2} f(z_1)g(z_2)\frac{\partial^2}{\partial z_2\partial z_1}\int_0^{d(z_1,z_2)}\frac{1}{1-z}dz dz_1 dz_2 \tag{2.4}$$

where $f, g \in \{f_1, \cdots, f_\kappa\}$. Here

$$\begin{aligned}d_1(z) = &-cz^{-3}\int\frac{x^2}{(1+xg_2(z))^2}dH_1(x)\int\frac{t^2}{(g_1(z)t+1)^3}dH_2(t)\\ &\times\left[1-cz^{-2}\int\frac{x^2}{(1+xg_2(z))^2}dH_1(x)\int\frac{t^2}{(g_1(z)t+1)^2}dH_2(t)\right]^{-1}\\ &-cz^{-4}\int\frac{x^3}{(1+xg_2(z))^3}dH_1(x)\int\frac{t}{(g_1(z)t+1)^2}dH_2(t)\int\frac{t^2}{(g_1(z)t+1)^2}dH_2(t)\\ &\times\left[1-cz^{-2}\int\frac{x^2}{(1+xg_2(z))^2}dH_1(x)\int\frac{t^2}{(g_1(z)t+1)^2}dH_2(t)\right]^{-1},\end{aligned}$$

$$\begin{aligned}d_2(z) = &-cz^{-4}\int\frac{x^2}{(1+xg_2(z))^3}dH_1(x)\int\frac{t^2}{(g_1(z)t+1)^2}dH_2(t)\\ &\times\left[\int\frac{x}{(1+xg_2(z))^2}dH_1(x)\right]^{-1}\int\frac{x^2}{(1+xg_2(z))^2}dH_1(x)\int\frac{t}{(1+g_1(z)t)^2}dH_2(t)\end{aligned}$$

$$\times\left[1-cz^{-2}\int\frac{x^2}{(1+xg_2(z))^2}dH_1(x)\int\frac{t^2}{(g_1(z)t+1)^2}dH_2(t)\right]^{-1},$$

and

$$d(z_1,z_2)=\frac{1}{z_1z_2}\frac{z_1g_1(z_1)-z_2g_1(z_2)}{g_2(z_1)-g_2(z_2)}\frac{z_1g_2(z_1)-z_2g_2(z_2)}{g_1(z_1)-g_1(z_2)}.$$

The contours in (2.3) and (2.4) (two contours in (2.4), which we may assume to be nonoverlapping) are closed and are taken in the positive direction in the complex plane, each enclosing the support of F^{c,H_1,H_2}.

(ii) If $\mathbf{X}_n=(x_{jk})=(u_{jk}+iv_{jk}),u_{jk},v_{jk}\in\mathbb{R}$, is a complex matrix with $\mathrm{E}(u_{jk})=\mathrm{E}(v_{jk})=0$, $\mathrm{E}(u_{jk}^2)=\mathrm{E}(v_{jk}^2)=\frac{1}{2}$, $\mathrm{E}(u_{jk}^4)=\mathrm{E}(v_{jk}^4)=\frac{3}{4}$, and u_{jk} and v_{jk} are independent, then (2.2) also holds, except the means are zero and the covariance function is $1/2$ the function given in (2.4).

***Remark* 2.2.** Notice that the sample covariance matrix can be written as $\widetilde{\mathbf{B}}_n=\frac{1}{N}\sum_{k=1}^n s_k\mathbf{y}_k\mathbf{y}_k^*$ (see (3.1)) in the Gaussian case which is analogous to the common matrix $\frac{1}{N}\sum_{k=1}^n\mathbf{y}_k\mathbf{y}_k^*$ (see Bai and Silverstein (2004)). Hence the overall strategy of handling $\widetilde{\mathbf{B}}_n$ in the Gaussian case is similar to that used in Bai and Silverstein (2004). Since the limit of $m_{F^{\widetilde{\mathbf{B}}_n}}(z)$ satisfies three equations (see (1.4)) other than only one this brings us additional difficulty.

***Remark* 2.3.** We only assume that $x_{jk},j=1,\cdots,N,k=1,\cdots,n$ are independent instead of independent and identically distributed as in Bai and Silverstein (2004) and the identically distributed assumption on x_{jk} can replaced by the moment assumptions that $Ex_{jl}=0,Ex_{jl}^2=1$, $Ex_{jl}^4=3$, and the Lindeberg assumption for the real case.

***Remark* 2.4.** It is worth mentioning that our result is consistent with that in Bai and Silverstein (2004). We distinguish two cases to show the consistency according to whether $\mathbf{T}_{2n}$ or $\mathbf{T}_{1n}$ reduces to the identity matrix.

When $\mathbf{T}_{2n}=\mathbf{I}$ and $\mathbf{T}_{1n}$ is a nonnegative definite Hermitian matrix, $\mathbf{B}_n=\frac{1}{N}\mathbf{X}_n\mathbf{T}_{1n}\mathbf{X}_n^*$. Then (1.4) is transformed into

$$\begin{cases}\underline{m}(z)=-z^{-1}\int\frac{1}{1+m(z)x}dH_1(x)\\ g_1(z)=-\frac{1}{zm(z)}-1\\ g_2(z)=m(z).\end{cases}$$

It follows that

$$\mathrm{E}X_f=-\frac{1}{2\pi i}\oint_{\mathcal{C}}f(z)\int\frac{cm(z)^3x^2}{(1+xm(z))^3}dH_1(x)\left\{1-\int\frac{cm(z)^2x^2}{(1+xm(z))^2}dH_1(x)\right\}^{-2}dz$$

and

$$d(z_1,z_2)=1+\frac{m(z_1)m(z_2)(z_1-z_2)}{m(z_2)-m(z_1)}.$$

These are the same as those in Bai and Silverstein (2004).

If $\mathbf{T}_{1n} = \mathbf{I}$ then $\mathbf{B}_n = \frac{1}{N}\mathbf{T}_{2n}^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}_{2n}^{1/2}$. Let $\widetilde{\mathbf{B}}_n = \frac{1}{n}\mathbf{T}_{2n}^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}_{2n}^{1/2}$ and $\widetilde{\underline{\mathbf{B}}_n} = \frac{1}{n}\mathbf{X}_n^*\mathbf{T}_{2n}\mathbf{X}_n$. We use $\widetilde{m}_n(z)$ and $\widetilde{\underline{m}}_n(z)$ to denote the Stieltjes transforms of $F^{\widetilde{\mathbf{B}}_n}$ and $F^{\widetilde{\underline{\mathbf{B}}_n}}$ respectively. Denote by $\widetilde{F}^{c^{-1},H_2}$ the limiting distribution of $\widetilde{F}^{\widetilde{\mathbf{B}}_n}$. Moreover $\widetilde{F}^{c_n^{-1},H_{2n}}$ is obtained from $\widetilde{F}^{c^{-1},H_2}$ with c, H_2 replaced by c_n, H_{2n} respectively. Let $\widetilde{m}(z) = \lim_{n\to\infty}\widetilde{m}_n(z), \widetilde{\underline{m}}(z) = \lim_{n\to\infty}\widetilde{\underline{m}}_n(z)$ and $\widetilde{m}_n^0(z) = m_{\widetilde{F}^{c_n^{-1},H_{2n}}}(z)$. Due to (2.8) below we only need to consider the limiting distribution of $\widetilde{M}_n(z) = N[\widetilde{m}_n(z) - \widetilde{m}_n^0(z)]$. Firstly, (1.4) becomes

$$\begin{cases} m(z) = -z^{-1}\int \frac{1}{1+c\underline{m}(z)x}dH_2(x) \\ g_1(z) = c\underline{m}(z) \\ g_2(z) = -\frac{1}{z\underline{m}(z)} - 1 \end{cases} .$$

By Lemma 2.5 below and the above equations, we have

$$\mathrm{E}M(z) = \int \frac{c\underline{m}(z)^3x^2}{\left(1+cx\underline{m}(z)\right)^3}dH_2(x)\left\{1 - \int \frac{c\underline{m}(z)^2x^2}{\left(1+cx\underline{m}(z)\right)^2}dH_2(x)\right\}^{-2} \tag{2.5}$$

and

$$d(z_1, z_2) = 1 + \frac{\underline{m}(z_1)\underline{m}(z_2)\,(z_1 - z_2)}{\underline{m}(z_2) - \underline{m}(z_1)}.$$

Note that $\mathbf{B}_n = c_n\widetilde{\mathbf{B}}_n$. It can be verified that

$$\widetilde{\underline{m}}_n(z/c_n) = c_n\underline{m}_n(z)$$

and

$$M_n(z) = c_n^{-1}\widetilde{M}_n(z/c_n).$$

These imply that

$$\widetilde{\underline{m}}(z/c) = c\underline{m}(z) \tag{2.6}$$

and

$$M(z) = c^{-1}\widetilde{M}(z/c) \tag{2.7}$$

where $\widetilde{M}(z)$ is a two-dimensional Gaussian process, the limit of weak convergence of $\widetilde{M}_n(z)$. Plugging (2.6) and (2.7) into (2.5), one has

$$\mathrm{E}\widetilde{M}(z/c) = c^{-1}\int \frac{\widetilde{\underline{m}}(z/c)^3x^2}{\left(1+x\widetilde{\underline{m}}(z/c)\right)^3}dH_2(x)\left\{1 - c^{-1}\int \frac{\widetilde{\underline{m}}(z/c)^2x^2}{\left(1+x\widetilde{\underline{m}}(z/c)\right)^2}dH_2(x)\right\}^{-2}$$

and

$$d(z_1, z_2) = 1 + \frac{\widetilde{\underline{m}}(z_1/c)\widetilde{\underline{m}}(z_2/c)\,(z_1/c - z_2/c)}{\widetilde{\underline{m}}(z_2/c) - \widetilde{\underline{m}}(z_1/c)}.$$

Hence the expectation and covariance are the same as those in Bai and Siverstein (2004).

By Cauchy's formula

$$\int f(x)dG(x) = -\frac{1}{2\pi i}\oint f(z)m_G(z)dz \tag{2.8}$$

where G is a cumulative distribution function (c.d.f.) and f is analytic on an open set containing the support of G. The complex integral on the right-hand side is over any positively oriented contour enclosing the support of G and on which f is analytic. Hence, the proof of Theorem 2.1 relies on establishing limiting results on

$$M_n(z) = N\left[m_n(z) - m_n^0(z)\right].$$

The contour $\mathcal{C}$ is defined as follows.

By the assumption of Theorem 2.1, we may suppose $\max\{\|\mathbf{T}_{1n}\|, \|\mathbf{T}_{2n}\|\} \leq \tau$. Let v_0 be any positive number. Let x_r be any positive number if the right end point of interval (2.1) is zero. Otherwise choose

$$x_r \in (\limsup_n s_1\lambda_{\max}^{\mathbf{T}_{2n}}\left(1+\sqrt{c}\right)^2, \infty).$$

Let x_l be any negative number if the left end point of interval (2.1) is zero. Otherwise choose

$$x_l \in \begin{cases} (0, \liminf_n s_n\lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c)\left(1-\sqrt{c}\right)^2), & \text{if } \liminf_n s_n\lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c) > 0, \\ (-\infty, \liminf_n s_n\lambda_{\max}^{\mathbf{T}_{2n}}\left(1+\sqrt{c}\right)^2), & \text{if } \liminf_n s_n\lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c) \leq 0. \end{cases}$$

Let

$$\mathcal{C}_u = \{x + iv_0 : x \in [x_l, x_r]\}.$$

Define the contour $\mathcal{C}$

$$\mathcal{C} = \{x_l + iv : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + iv : v \in [0, v_0]\}.$$

To avoid dealing with the small $\Im z$, we truncate $M_n(z)$ on a contour $\mathcal{C}$ of the complex plane. We define now the subsets $\mathcal{C}_n$ of $\mathcal{C}$ on which $M_n(\cdot)$ agrees with $\widehat{M}_n(\cdot)$. Choose sequence $\{\varepsilon_n\}$ decreasing to zero satisfying for some $\alpha \in (0,1)$

$$\varepsilon_n \geq n^{-\alpha}.$$

Let

$$\mathcal{C}_l = \left\{x_l + iv : v \in [n^{-1}\varepsilon_n, v_0]\right\} \quad \text{and} \quad \mathcal{C}_r = \left\{x_r + iv : v \in [n^{-1}\varepsilon_n, v_0]\right\}.$$

Then $\mathcal{C}_n = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_r$. For $z = x + iv$, the process $\widehat{M}_n(\cdot)$ can now be defined as

$$\widehat{M}_n(\cdot) = \begin{cases} M_n(z), & \text{for } z \in \mathcal{C}_n, \\ M_n(x_l + in^{-1}\varepsilon_n), & \text{for } x = x_l, v \in [0, n^{-1}\varepsilon_n], \\ M_n(x_r + in^{-1}\varepsilon_n), & \text{for } x = x_r, v \in [0, n^{-1}\varepsilon_n]. \end{cases} \tag{2.9}$$

The central limit theorem of $\widehat{M}_n(z)$ is specified below.

***Lemma* 2.5.** Under the conditions of Theorem 2.1, $\widehat{M}_n(z)$ converges weakly to a two-dimensional Gaussian process $M(\cdot)$ satisfying for $z \in C$ under the assumptions in (i)

$$\mathrm{E}M(z) = (d_1(z) - d_2(z))\left\{1 - z^{-1}\left[\int \frac{x}{(1 + xg_2(z))^2} dH_1(x)\right]^{-1}\right. \tag{2.10}$$

$$\left.\times \int \frac{x^2}{(1 + xg_2(z))^2} dH_1(x) \int \frac{t}{(1 + g_1(z)t)^2} dH_2(t)\right\}^{-1} \tag{2.11}$$

and for $z_1, z_2 \in C \cup \overline{C}$ with $\overline{C} = \{\bar{z} : z \in C\}$,

$$\mathrm{Cov}\,(M(z_1), M(z_2)) = 2\frac{\partial^2}{\partial z_2 \partial z_1} \int_0^{d(z_1,z_2)} \frac{1}{1-z} dz,$$

while under the assumptions in (ii) $\mathrm{E}M(z) = 0$ and the covariance function analogous to (2.10) is $1/2$ the right-hand side of (2.10).

Proof of Theorem 2.1. From Yin, Bai and Krishnaiah (1988) and Bai and Yin (1993), we conclude that

$$\lambda_{\max}\left(\frac{1}{N}\mathbf{X}_n^*\mathbf{X}_n\right) \to \left(1 + \sqrt{c}\right)^2 \quad \text{a.s.} \tag{2.12}$$

and

$$\lambda_{\min}\left(\frac{1}{N}\mathbf{X}_n^*\mathbf{X}_n\right) \to \left(1 - \sqrt{c}\right)^2 \quad \text{a.s.}$$

The upper and lower bounds of the extreme eigenvalues of $\mathbf{B}_n$ depends largely on the signs of s_1 and s_n. Since $s_1 > 0$, we have

$$\lambda_{\max}(\mathbf{B}_n) \le s_1 \lambda_{\max}^{\mathbf{T}_{2n}} \lambda_{\max}\left(\frac{1}{N}\mathbf{X}_n^*\mathbf{X}_n\right) \le s_1 \lambda_{\max}^{\mathbf{T}_{2n}}\left(1 + \sqrt{c}\right)^2 \quad \text{a.s.}$$

If $s_n > 0$, then we have

$$\lambda_{\min}(\mathbf{B}_n) \ge s_n \lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c) \lambda_{\min}\left(\frac{1}{N}\mathbf{X}_n^*\mathbf{X}_n\right) \ge s_n \lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c)\left(1 - \sqrt{c}\right)^2 \quad \text{a.s.}$$

Otherwise, we get

$$\lambda_{\min}(\mathbf{B}_n) \ge s_n \lambda_{\max}^{\mathbf{T}_{2n}} \lambda_{\max}\left(\frac{1}{N}\mathbf{X}_n^*\mathbf{X}_n\right) \ge s_n \lambda_{\max}^{\mathbf{T}_{2n}}\left(1 + \sqrt{c}\right)^2 \quad \text{a.s.}$$

Combining the definitions of x_l, x_r, we find with probability 1

$$\liminf_{n\to\infty} \min\,(x_r - \lambda_{\max}(\mathbf{B}_n), \lambda_{\min}(\mathbf{B}_n) - x_l) > 0.$$

Since $F^{\mathbf{B}_n} \to F^{c,H_1,H_2}$ with probability 1 the support of $F^{c_n,H_{1n},H_{2n}}$ is contained in interval (2.1) with probability 1. Thus, by (2.8), for $f \in \{f_1, \cdots, f_\kappa\}$ and large n, with probability 1,

$$\int f(x)dG_n(x) = -\frac{1}{2\pi i}\oint f(z)M_n(z)dz$$

where the complex integral is over $\mathcal{C} \cup \overline{\mathcal{C}}$. For $v \in [0, n^{-1}\varepsilon_n]$, note that

$$\left|M_n(x_r + iv) - M_n(x_r + in^{-1}\varepsilon_n)\right| \le 4n\left|\max\left(\lambda_{\max}(\mathbf{B}_n), e_r\right) - x_r\right|^{-1}$$

and

$$\left|M_n(x_l + iv) - M_n(x_l + in^{-1}\varepsilon_n)\right| \le 4n\left|\min\left(\lambda_{\min}(\mathbf{B}_n), e_l\right) - x_l\right|^{-1}.$$

It follows that for large n, with probability 1,

$$\begin{aligned}&\left|\oint f(z)\left(M_n(z) - \widehat{M}_n(z)\right)dz\right|\\ \le& 8K\varepsilon_n\left[\left|\max\left(\lambda_{\max}(\mathbf{B}_n), e_r\right) - x_r\right|^{-1} + \left|\min\left(\lambda_{\min}(\mathbf{B}_n), e_l\right) - x_l\right|^{-1}\right] \to 0\end{aligned}$$

where e_l (e_r) is the left endpoint (right endpoint) of interval (2.1) and K is the bound on f over $\mathcal{C}$.

Note that the mapping

$$\widehat{M}_n(\cdot) \to \left(-\frac{1}{2\pi i}\oint f_1(z)\widehat{M}_n(z)dz, \cdots, -\frac{1}{2\pi i}\oint f_\kappa(z)\widehat{M}_n(z)dz\right)$$

is continuous. Using Lemma 2.5, we complete the proof of Theorem 2.1. □

3. The Gaussian case

This section is to prove Lemma 2.5 under the Gaussian case, i.e., $\{x_{jk}\}$, $j = 1, \cdots, N, k = 1, \cdots, n$ are standard complex normal random variables. Since $\mathbf{T}_{1n}$ is Hermitian there exists an unitary matrix $\mathbf{U}$ such that

$$\mathbf{T}_{1n} = \mathbf{U}\mathrm{diag}\left(s_1, \cdots, s_n\right)\mathbf{U}^*.$$

Note that $\mathbf{X}_n$ has the same distribution as $\mathbf{X}_n\mathbf{U}$. It then suffices to consider

$$\widetilde{\mathbf{B}}_n = \frac{1}{N}\sum_{k=1}^{n} s_k\mathbf{T}_{2n}^{1/2}\mathbf{x}_k\mathbf{x}_k^*\mathbf{T}_{2n}^{1/2} \triangleq \frac{1}{N}\sum_{k=1}^{n} s_k\mathbf{y}_k\mathbf{y}_k^* \tag{3.1}$$

where $\mathbf{x}_k$ is the k-th column of $\mathbf{X}_n$. In what follows, we omit the symbol $\widetilde{\cdot}$ from the notation of $\widetilde{\mathbf{B}}_n$ in order to simplify notation. Rewrite for $z \in \mathcal{C}_n$

$$M_n(z) = N[m_n(z) - \mathrm{E}m_n(z)] + N[\mathrm{E}m_n(z) - m_n^0(z)] \triangleq M_{n1}(z) + M_{n2}(z).$$

We below consider the random part $M_{n1}(z)$. The nonrandom part $M_{n2}(z)$ are postponed to Supplement A.

In the sequel we assume $x_{jk}, j = 1, \cdots, N, k = 1, \cdots, n$ are truncated at $\delta_n \sqrt{n}$, centralized and re-normalized. The details are omitted which is similar to Bai and Silverstein (2004). Furthermore, $\mathrm{E}x_{jk}^2 = o(N^{-1})$, $\mathrm{E}|x_{jk}|^4 = 2 + o(1)$ under the complex case and $\mathrm{E}x_{jk}^4 = 3 + o(1)$ under the real case.

We start with two probability inequalities for extreme eigenvalues of $\mathbf{B}_n$. It is well known (see Bai and Silverstein (2004); Yin, Bai and Krishnaiah (1988)) that for any l, $\eta_1 > (1+\sqrt{c})^2$ and $\eta_2 < (1-\sqrt{c})^2$

$$\mathrm{P}\left(\lambda_{\max}\left(\frac{1}{N}\mathbf{X}_n^*\mathbf{X}_n\right) \geq \eta_1\right) = o(n^{-l})$$

and

$$\mathrm{P}\left(\lambda_{\min}\left(\frac{1}{N}\mathbf{X}_n^*\mathbf{X}_n\right) \leq \eta_2\right) = o(n^{-l}).$$

Thus, letting

$$\eta_r \in \begin{cases} (0, x_r), & c \geq 1, \\ (\limsup_n s_1 \lambda_{\max}^{\mathbf{T}_{2n}} \left(1+\sqrt{c}\right)^2, x_r), & \text{otherwise,} \end{cases}$$

we have for any $l > 0$

$$\mathrm{P}\left(\lambda_{\max}\left(\mathbf{B}_n\right) \geq \eta_r\right) = o(n^{-l}). \tag{3.2}$$

Likewise, we have

$$\mathrm{P}\left(\lambda_{\min}\left(\mathbf{B}_n\right) \leq \eta_l\right) = o(n^{-l}). \tag{3.3}$$

where

$$\eta_l \in \begin{cases} (x_l, 0), & c \geq 1, \\ (x_l, \liminf_n s_n \lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c) \left(1-\sqrt{c}\right)^2), & \text{if } \liminf_n s_n \lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c) > 0, \\ (x_l, \liminf_n s_n \lambda_{\max}^{\mathbf{T}_{2n}} \left(1+\sqrt{c}\right)^2), & \text{if } \liminf_n s_n \lambda_{\min}^{\mathbf{T}_{2n}} I_{(0,1)}(c) \leq 0. \end{cases}$$

Here η_l, η_r, x_l, x_r can be chosen such that

$$x_r - \eta_r > 2\tau^2 \quad \text{and} \quad \eta_l - x_l > 2\tau^2, \tag{3.4}$$

where τ are the upper bound of the spectral norms of $\mathbf{T}_{1n}$ and $\mathbf{T}_{2n}$ defined before.

3.1. The limiting distribution of $M_{n1}(z)$

The aim of this part is to find the limiting distribution of $M_{n1}(z)$. That is to say, we show for any positive integer r, the sum

$$\sum_{j=1}^{r} \alpha_j M_{n1}(z_j) \qquad \Im z_j \neq 0$$

converges in distribution to a Gaussian random variable. Since

$$\lim_{v_0 \downarrow 0} \limsup_{n\to\infty} \mathrm{E} \left| \int_{C_l \cup C_r} f(z) M_{n1}(z) dz \right|^2 \to 0,$$

it suffices to consider $z = u + iv_0 \in C_u$. This result can proceed in two steps.

The first step is to use the central limit theorem for martingale difference sequences, so we can accomplish the goal by finding the in probability limit of (3.14). Introduce

$$\begin{aligned}
&\mathbf{D}(z) = \mathbf{B}_n - z\mathbf{I}_N,\ \mathbf{D}_k(z) = \mathbf{D}(z) - \frac{1}{N} s_k \mathbf{y}_k \mathbf{y}_k^*,\\
&\mathbf{B}_{nk} = \mathbf{B}_n - \frac{1}{N} s_k \mathbf{y}_k \mathbf{y}_k^*,\ g_{2n}(z) = \frac{1}{N} \mathrm{tr}(\mathbf{D}^{-1}(z)\mathbf{T}_{2n}),
\end{aligned} \tag{3.5}$$

and

$$\varepsilon_k(z) = \mathbf{y}_k^* \mathbf{D}_k^{-1}(z)\mathbf{y}_k - \mathrm{tr}(\mathbf{D}_k^{-1}(z)\mathbf{T}_{2n}),\ \gamma_k(z) = \mathbf{y}_k^* \mathbf{D}_k^{-2}(z)\mathbf{y}_k - \mathrm{tr}(\mathbf{D}_k^{-2}(z)\mathbf{T}_{2n})$$

$$\beta_k(z) = \frac{1}{1 + N^{-1} s_k \mathbf{y}_k^* \mathbf{D}_k^{-1}(z)\mathbf{y}_k},\ \widetilde{\beta}_k(z) = \frac{1}{1 + N^{-1} s_k \mathrm{tr}(\mathbf{D}_k^{-1}(z)\mathbf{T}_{2n})}, \tag{3.6}$$

$$b_k(z) = \frac{1}{1 + N^{-1} s_k \mathrm{E}\mathrm{tr}(\mathbf{D}_k^{-1}(z)\mathbf{T}_{2n})},\ \psi_k(z) = \frac{1}{1 + s_k \mathrm{E} g_{2n}(z)}. \tag{3.7}$$

Note that

$$m_n(z) = \frac{1}{N} \mathrm{tr}\,(\mathbf{B}_n - z\mathbf{I}_N)^{-1} \triangleq \frac{1}{N} \mathrm{tr} \mathbf{D}^{-1}(z).$$

Let $\mathrm{E}_0(\cdot)$ denote mathematical expectation and $\mathrm{E}_k(\cdot)$ denote conditional expectation with respect to the σ-field given by $\mathbf{x}_1, \cdots, \mathbf{x}_k$. By the formula

$$(\boldsymbol{\Sigma} + q\boldsymbol{\alpha}\boldsymbol{\beta}^*)^{-1} = \boldsymbol{\Sigma}^{-1} - \frac{q\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}\boldsymbol{\beta}^*\boldsymbol{\Sigma}^{-1}}{1 + q\boldsymbol{\beta}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}}, \tag{3.8}$$

we have

$$\begin{aligned}
M_{n1}(z) &= \sum_{k=1}^{n} \mathrm{tr}\left\{\mathrm{E}_k \mathbf{D}^{-1}(z) - \mathrm{E}_{k-1}\mathbf{D}^{-1}(z)\right\} = \sum_{k=1}^{n} (\mathrm{E}_k - \mathrm{E}_{k-1})\, \mathrm{tr}\left[\mathbf{D}(z)^{-1} - \mathbf{D}_k^{-1}(z)\right]\\
&= -\frac{1}{N} \sum_{k=1}^{n} (\mathrm{E}_k - \mathrm{E}_{k-1})\, s_k \beta_k(z) \mathbf{y}_k^* \mathbf{D}_k^{-2}(z)\mathbf{y}_k
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{N}\sum_{k=1}^{n}(\mathrm{E}_k - \mathrm{E}_{k-1})\, s_k\beta_k(z)\gamma_k(z) - \frac{1}{N}\sum_{k=1}^{n}(\mathrm{E}_k - \mathrm{E}_{k-1})\, s_k\beta_k(z)\mathrm{tr}(\mathbf{D}_k^{-2}(z)\mathbf{T}_{2n}) \\
&\triangleq \mathcal{I}_1 + \mathcal{I}_2.
\end{aligned} \tag{3.9}
$$

From the identity

$$
\beta_k(z) - \widetilde{\beta}_k(z) = -\frac{1}{N}s_k\widetilde{\beta}_k(z)\beta_k(z)\varepsilon_k(z), \tag{3.10}
$$

we have

$$
\mathcal{I}_1 = -\frac{1}{N}\sum_{k=1}^{n}\mathrm{E}_k s_k\widetilde{\beta}_k(z)\gamma_k(z) + \frac{1}{N^2}\sum_{k=1}^{n}(\mathrm{E}_k - \mathrm{E}_{k-1})\, s_k^2\widetilde{\beta}_k(z)\beta_k(z)\varepsilon_k(z)\gamma_k(z).
$$

By Lemma 4.1 and Lemma 4.9

$$
\begin{aligned}
&\frac{1}{N^4}\sum_{k=1}^{n}\mathrm{E}|\,(\mathrm{E}_k - \mathrm{E}_{k-1})\, s_k^2\widetilde{\beta}_k(z)\beta_k(z)\varepsilon_k(z)\gamma_k(z)|^2 \\
\le&\frac{C}{N^4}\sum_{k=1}^{n}\mathrm{E}^{1/2}|\widetilde{\beta}_k(z)\beta_k(z)|^2\mathrm{E}^{1/2}|\varepsilon_k(z)\gamma_k(z)|^4 \\
\le&\frac{C}{N^4}\sum_{k=1}^{n}\mathrm{E}^{1/4}|\varepsilon_k(z)|^8\mathrm{E}^{1/4}|\gamma_k(z)|^8 \le \frac{C}{N} \to 0.
\end{aligned}
$$

This implies

$$
\mathcal{I}_1 = -\frac{1}{N}\sum_{k=1}^{n}\mathrm{E}_k s_k\widetilde{\beta}_k(z)\gamma_k(z) + o_p(1). \tag{3.11}
$$

Using the same argument and

$$
\beta_k(z) - \widetilde{\beta}_k(z) = -\frac{1}{N}s_k\widetilde{\beta}_k^2(z)\varepsilon_k(z) + \frac{1}{N^2}s_k^2\beta_k(z)\widetilde{\beta}_k^2(z)\varepsilon_k^2(z), \tag{3.12}
$$

one gets

$$
\mathcal{I}_2 = \frac{1}{N^2}\sum_{k=1}^{n}\mathrm{E}_k s_k^2\widetilde{\beta}_k^2(z)\varepsilon_k(z)\mathrm{tr}(\mathbf{D}_k^{-2}(z)\mathbf{T}_{2n}) + o_p(1). \tag{3.13}
$$

From (3.9), (3.11), and (3.13), we conclude that

$$
M_{n1}(z) = -\frac{1}{N}\sum_{k=1}^{n}\mathrm{E}_k s_k\widetilde{\beta}_k(z)\gamma_k(z) + \frac{1}{N^2}\sum_{k=1}^{n}\mathrm{E}_k s_k^2\widetilde{\beta}_k^2(z)\varepsilon_k(z)\mathrm{tr}(\mathbf{D}_k^{-2}(z)\mathbf{T}_{2n}) + o_p(1).
$$

Define

$$
h_k(z) = -\frac{1}{N}\mathrm{E}_k s_k\widetilde{\beta}_k(z)\gamma_k(z) + \frac{1}{N^2}\mathrm{E}_k s_k^2\widetilde{\beta}_k^2(z)\varepsilon_k(z)\mathrm{tr}(\mathbf{D}_k^{-2}(z)\mathbf{T}_{2n})
$$

$$= - N^{-1}\frac{d}{dz}\mathrm{E}_k s_k\widetilde{\beta}_k(z)\varepsilon_k(z).$$

Thus we only need to prove that $\sum_{j=1}^{r}\alpha_j\sum_{k=1}^{n}h_k(z_j) = \sum_{k=1}^{n}\sum_{j=1}^{r}\alpha_j h_k(z_j)$ converges in distribution to a Gaussian random variable. By Lemma 4.10, it suffices to verify condition (i) and (ii). It follows from Lemma 4.1 and Lemma 4.9 that

$$\sum_{k=1}^{n}\mathrm{E}\left|\sum_{j=1}^{r}\alpha_j h_k(z_j)\right|^4 \leq \frac{C}{N^4}\sum_{k=1}^{n}\sum_{j=1}^{r}\alpha_j^4\Big[\mathrm{E}^{1/2}|\widetilde{\beta}_k|^8\mathrm{E}^{1/2}|\gamma_k(z_j)|^8 + \mathrm{E}^{1/2}|\widetilde{\beta}_k|^{16}\mathrm{E}^{1/2}|\varepsilon_k(z_j)|^8\Big] \leq \frac{C}{N}\to 0$$

which implies that conditions (ii) of Lemma 4.10 is satisfied. The goal turns into finding a limit in probability of

$$\Phi(z_1,z_2) \triangleq \sum_{k=1}^{n}\mathrm{E}_{k-1}\left[h_k(z_1)h_k(z_2)\right] \tag{3.14}$$

for z_1, z_2 with nonzero fixed imaginary parts.

The second step is to achieve the above goal through (1.7)-(1.11). It is obvious that

$$\Phi(z_1,z_2) = N^{-2}\frac{\partial^2}{\partial z_2\partial z_1}\sum_{k=1}^{n}\mathrm{E}_{k-1}\left[\mathrm{E}_k\left(s_k\widetilde{\beta}_k(z_1)\varepsilon_k(z_1)\right)\mathrm{E}_k\left(s_k\widetilde{\beta}_k(z_2)\varepsilon_k(z_2)\right)\right].$$

Due to the analysis on page 571 in Bai and Silverstein (2004), it is enough to prove that

$$N^{-2}\sum_{k=1}^{n}s_k^2\mathrm{E}_{k-1}\left[\mathrm{E}_k\left(\widetilde{\beta}_k(z_1)\varepsilon_k(z_1)\right)\mathrm{E}_k\left(\widetilde{\beta}_k(z_2)\varepsilon_k(z_2)\right)\right]$$

converges in probability to a constant. Similar to (4.32) in the supplement, it can be verified that $|\widetilde{\beta}_k(z)|$ and $|b_k(z)|$ has the same bound as $\beta_k(z)$. From Lemma 4.2, we get

$$\begin{aligned}
&\mathrm{E}\Bigg|N^{-2}\sum_{k=1}^{n}s_k^2\mathrm{E}_{k-1}\Big[\mathrm{E}_k\left(\widetilde{\beta}_k(z_1)\varepsilon_k(z_1)\right)\mathrm{E}_k\left(\widetilde{\beta}_k(z_2)\varepsilon_k(z_2)\right)\\
&- \mathrm{E}_k\left(b_k(z_1)\varepsilon_k(z_1)\right)\mathrm{E}_k\left(b_k(z_2)\varepsilon_k(z_2)\right)\Big]\Bigg|\\
\leq& CN^{-2}\sum_{k=1}^{n}\Big[\mathrm{E}|\mathrm{E}_k\left((\widetilde{\beta}_k(z_1)-b_k(z_1))\varepsilon_k(z_1)\right)\mathrm{E}_k\left(\widetilde{\beta}_k(z_2)\varepsilon_k(z_2)\right)|\\
&+ \mathrm{E}|\mathrm{E}_k\left(b_k(z_1)\varepsilon_k(z_1)\right)\mathrm{E}_k\left((\widetilde{\beta}_k(z_2)-b_k(z_2))\varepsilon_k(z_2)\right)|\Big]\\
\leq& CN^{-2}\sum_{k=1}^{n}\Big[\mathrm{E}^{1/2}|(\widetilde{\beta}_k(z_1)-b_k(z_1))\varepsilon_k(z_1)|^2\mathrm{E}^{1/2}|\widetilde{\beta}_k(z_2)\varepsilon_k(z_2)|^2
\end{aligned}$$

$$+\mathrm{E}^{1/2}|b_k(z_1)\varepsilon_k(z_1)|^2\mathrm{E}^{1/2}|\left((\widetilde{\beta}_k(z_2)-b_k(z_2))\varepsilon_k(z_2)\right)|^2\Big]$$

$$\le CN^{-1}\sum_{k=1}^{n}\left[\mathrm{E}^{1/2}|\widetilde{\beta}_k(z_1)-b_k(z_1)|^2+\mathrm{E}^{1/2}|(\widetilde{\beta}_k(z_2)-b_k(z_2)|^2\right]\to 0,$$

which yields

$$N^{-2}\sum_{k=1}^{n}s_k^2\mathrm{E}_{k-1}\Big[\mathrm{E}_k\left(\widetilde{\beta}_k(z_1)\varepsilon_k(z_1)\right)\mathrm{E}_k\left(\widetilde{\beta}_k(z_2)\varepsilon_k(z_2)\right)$$
$$-\mathrm{E}_k\left(b_k(z_1)\varepsilon_k(z_1)\right)\mathrm{E}_k\left(b_k(z_2)\varepsilon_k(z_2)\right)\Big]\xrightarrow{i.p.}0.$$

Therefore, our goal is to find the limit in probability of

$$N^{-2}\sum_{k=1}^{n}s_k^2b_k(z_1)b_k(z_2)\mathrm{E}_{k-1}\left[\mathrm{E}_k\left(\varepsilon_k(z_1)\right)\mathrm{E}_k\left(\varepsilon_k(z_2)\right)\right].$$

Using the moments of the complex random variables, we have

$$\mathrm{E}_{k-1}\left[\mathrm{E}_k\left(\varepsilon_k(z_1)\right)\mathrm{E}_k\left(\varepsilon_k(z_2)\right)\right]=\mathrm{tr}(\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(z_1)\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(z_2))+o(1)A_n$$

where

$$|A_n|\le\left[\mathrm{tr}\left(\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(z_1)\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(\bar{z}_1)\right)\mathrm{tr}\left(\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(\bar{z}_2)\right)\right]^{1/2}=O(N).$$

For the real case, it follows that

$$\mathrm{E}_{k-1}\left[\mathrm{E}_k\left(\varepsilon_k(z_1)\right)\mathrm{E}_k\left(\varepsilon_k(z_2)\right)\right]=2\mathrm{tr}(\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(z_1)\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(z_2))+o(N).$$

Consequently, it suffices to study

$$N^{-2}\sum_{k=1}^{n}s_k^2b_k(z_1)b_k(z_2)\mathrm{tr}(\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(z_1)\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_k^{-1}(z_2)).\tag{3.15}$$

Let $\mathbf{R}_k(z)=z\mathbf{I}-\frac{1}{N}\sum_{j\neq k}s_j\psi_j(z)\mathbf{T}_{2n}$,

$$\beta_{jk}(z)=\frac{1}{1+N^{-1}s_j\mathbf{y}_j^*\mathbf{D}_{jk}^{-1}(z)\mathbf{y}_j}\quad\text{and}\quad b_{jk}(z)=\frac{1}{1+N^{-1}s_j\mathrm{E}\mathrm{tr}(\mathbf{D}_{jk}^{-1}(z)\mathbf{T}_{2n})}.$$

Write

$$\mathbf{D}_k(z_1)+\mathbf{R}_k(z_1)=\frac{1}{N}\sum_{j\neq k}s_j\mathbf{y}_j\mathbf{y}_j^*-\frac{1}{N}\sum_{j\neq k}s_j\psi_j(z_1)\mathbf{T}_{2n}$$

which implies that

$$\mathbf{R}_k^{-1}(z_1)+\mathbf{D}_k^{-1}(z_1)=\frac{1}{N}\sum_{j\neq k}s_j\mathbf{R}_k^{-1}(z_1)\mathbf{y}_j\mathbf{y}_j^*\mathbf{D}_k^{-1}(z_1)-\frac{1}{N}\sum_{j\neq k}s_j\psi_j(z_1)\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_1).$$

Using the formula

$$\left(\boldsymbol{\Sigma}+q\boldsymbol{\alpha}\boldsymbol{\beta}^*\right)^{-1}\boldsymbol{\alpha}=\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}}{1+q\boldsymbol{\beta}^*\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}},\tag{3.16}$$

we have

$$\begin{aligned}\mathbf{R}_k^{-1}(z_1)+\mathbf{D}_k^{-1}(z_1)=&\frac{1}{N}\sum_{j\neq k}s_j\psi_j(z_1)\mathbf{R}_k^{-1}(z_1)\left(\mathbf{y}_j\mathbf{y}_j^*-\mathbf{T}_{2n}\right)\mathbf{D}_{jk}^{-1}(z_1)\\&+\frac{1}{N}\sum_{j\neq k}s_j\left(\beta_{jk}(z_1)-\psi_j(z_1)\right)\mathbf{R}_k^{-1}(z_1)\mathbf{y}_j\mathbf{y}_j^*\mathbf{D}_{jk}^{-1}(z_1)\\&+\frac{1}{N}\sum_{j\neq k}s_j\psi_j(z_1)\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\left(\mathbf{D}_{jk}^{-1}(z_1)-\mathbf{D}_k^{-1}(z_1)\right)\\\triangleq&\mathbf{A_1}(z_1)+\mathbf{A_2}(z_1)+\mathbf{A_3}(z_1).\end{aligned}\tag{3.17}$$

By a direct calculation, we have for any positive number $t\geq 0$

$$\begin{aligned}\Im\left(z-\frac{1}{N}\sum_{j\neq k}s_j\psi_j(z)t\right)=&v_0-\frac{1}{N}\sum_{j\neq k}\frac{s_j^2t}{|1+s_j\mathrm{E}g_{2n}(z)|^2}\Im\mathrm{E}g_{2n}(\bar{z})\\=&v_0\left(1+\frac{1}{N^2}\sum_{j\neq k}\frac{s_j^2t}{|1+s_j\mathrm{E}g_{2n}(z)|^2}\mathrm{Etr}\left(\mathbf{D}^{-1}(z)\mathbf{D}^{-1}(\bar{z})\mathbf{T}_{2n}\right)\right)\geq v_0\end{aligned}$$

which yields

$$\left\|\mathbf{R}_k^{-1}(z)\right\|\leq\frac{1}{v_0}.$$

Let $\mathbf{M}$ be a $N\times N$ matrix with a nonrandom bound on the spectral norm of $\mathbf{M}$ for all parameters governing $\mathbf{M}$ and under all realizations of $\mathbf{M}$. By the Cauchy-Schwarz inequality, one gets

$$\begin{aligned}\mathrm{E}\left|\mathrm{tr}(\mathbf{A_1}(z_1)\mathbf{M})\right|\leq C\mathrm{E}^{1/2}&\left|\mathbf{y}_j^*\mathbf{D}_{jk}^{-1}(z_1)\mathbf{R}_k^{-1}(z_1)\mathbf{y}_j\right.\\&\left.-\mathrm{tr}\left(\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_{jk}^{-1}(z_1)\right)\right|^2=O(N^{1/2}).\end{aligned}\tag{3.18}$$

Let $\widetilde{\beta}_{jk}(z)=\frac{1}{1+N^{-1}s_j\mathrm{tr}(\mathbf{D}_{jk}^{-1}(z)\mathbf{T}_{2n})}$. From Lemma 4.2,

$$\mathrm{E}|\widetilde{\beta}_{jk}(z)-b_{jk}(z)|=O(N^{-1}).$$

Applying the above inequality, Lemma 4.9 and Lemma 4.11, we obtain

$$\begin{aligned}\mathrm{E}|\beta_{jk}(z)-\psi_j(z)|^2 \le & C\left[\mathrm{E}|\beta_{jk}(z)-\widetilde{\beta}_{jk}(z)|^2+|b_{jk}(z)-\psi_j(z)|^2\right]+O(N^{-1}) \\ \le & \frac{C}{N^2}\mathrm{E}^{1/2}|\mathbf{y}_j^*\mathbf{D}_{jk}^{-1}(z)\mathbf{y}_j-\mathrm{tr}(\mathbf{D}_{jk}^{-1}(z)\mathbf{T}_{2n})|^4 \\ & +\frac{C}{N^2}\mathrm{E}|\mathrm{tr}(\mathbf{D}_{jk}^{-1}(z)\mathbf{T}_{2n})-\mathrm{tr}(\mathbf{D}^{-1}(z)\mathbf{T}_{2n})|^2+O(N^{-1}) \\ \le & \frac{C}{N}+\frac{C}{N^2}+O(N^{-1})=O(N^{-1})\end{aligned} \tag{3.19}$$

which implies that

$$\begin{aligned}\mathrm{E}\left|\mathrm{tr}(\mathbf{A_2}(z_1)\mathbf{M})\right| \le & \frac{C}{N}\sum_{j\neq k}\mathrm{E}^{1/2}\left|\beta_{jk}(z_1)-\psi_j(z_1)\right|^2 \\ & \times\mathrm{E}^{1/2}\left|\mathbf{y}_j^*\mathbf{D}_{jk}^{-1}(z_1)\mathbf{M}\mathbf{R}_k^{-1}(z_1)\mathbf{y}_j\right|^2=O(N^{1/2}).\end{aligned} \tag{3.20}$$

Lemma 4.11 implies that

$$\mathrm{E}\left|\mathrm{tr}\left(\mathbf{A_3}(z_1)\mathbf{M}\right)\right| \le \frac{C}{N}\sum_{j\neq k}\mathrm{tr}\left[\left(\mathbf{D}_{jk}^{-1}(z_1)-\mathbf{D}_k^{-1}(z_1)\right)\times\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\right]\le C. \tag{3.21}$$

Using (3.17), (3.20), and (3.21), one gets

$$\begin{aligned}\mathrm{tr}(\mathrm{E}_k\left(\mathbf{D}_k(z_1)\right)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}) =& -\mathrm{tr}(\mathrm{E}_k\left(\mathbf{R}_k^{-1}(z_1)\right)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}) \\ & +\mathrm{tr}(\mathrm{E}_k\left(\mathbf{A}_1(z_1)\right)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n})+a(z_1,z_2)\end{aligned} \tag{3.22}$$

where $\mathrm{E}|a(z_1,z_2)|\le O(N^{1/2})$. Furthermore, write

$$\begin{aligned}&\mathrm{tr}(\mathrm{E}_k\left(\mathbf{A}_1(z_1)\right)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}) \\ =&-\frac{1}{N^2}\sum_{j<k}s_j^2\psi_j(z_1)\beta_{jk}(z_2)\left[\mathbf{y}_j^*\mathrm{E}_k\mathbf{D}_{jk}^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_{jk}^{-1}(z_2)\mathbf{y}_j-\mathrm{tr}\left(\mathrm{E}_k\mathbf{D}_{jk}^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\right)\right] \\ &\times\left[\mathbf{y}_j^*\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_1)\mathbf{y}_j-\mathrm{tr}\left(\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\right)\right] \\ &-\frac{1}{N^2}\sum_{j<k}s_j^2\psi_j(z_1)\beta_{jk}(z_2)\left(\mathbf{y}_j^*\mathrm{E}_k\mathbf{D}_{jk}^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_{jk}^{-1}(z_2)\mathbf{y}_j-\mathrm{tr}\left(\mathrm{E}_k\mathbf{D}_{jk}^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\right)\right) \\ &\times\mathrm{tr}\left(\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\right) \\ &-\frac{1}{N^2}\sum_{j<k}s_j^2\psi_j(z_1)\beta_{jk}(z_2)\mathrm{tr}\left(\mathrm{E}_k\mathbf{D}_{jk}^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\right) \\ &\times\left[\mathbf{y}_j^*\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_1)\mathbf{y}_j-\mathrm{tr}\left(\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\right)\right] \\ &-\frac{1}{N^2}\sum_{j<k}s_j^2\psi_j(z_1)\beta_{jk}(z_2)\mathrm{tr}\left(\mathrm{E}_k\mathbf{D}_{jk}^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\right)\mathrm{tr}\left(\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\right)\end{aligned}$$

$$+\frac{1}{N}\sum_{j<k} s_j\psi_j(z_1)\mathrm{tr}\left[\mathbf{R}_k^{-1}(z_1)\left(\mathbf{y}_j\mathbf{y}_j^*-\mathbf{T}_{2n}\right)\mathrm{E}_k\mathbf{D}_{jk}^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_{jk}^{-1}(z_2)\mathbf{T}_{2n}\right]$$
$$-\frac{1}{N}\sum_{j<k} s_j\psi_j(z_1)\mathrm{tr}\left[\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathrm{E}_k\mathbf{D}_{jk}^{-1}(z_1)\mathbf{T}_{2n}\left(\mathbf{D}_k^{-1}(z_2)-\mathbf{D}_{jk}^{-1}(z_2)\right)\mathbf{T}_{2n}\right]$$
$$\triangleq a_1(z_1,z_2)+a_2(z_1,z_2)+a_3(z_1,z_2)+a_4(z_1,z_2)+a_5(z_1,z_2)+a_6(z_1,z_2).$$

It follows from Lemma 4.1 and Lemma 4.9 that

$$\mathrm{E}|a_1(z_1,z_2)+a_2(z_1,z_2)+a_3(z_1,z_2)+a_5(z_1,z_2)|\le CN^{1/2}.$$

In addition, Lemma 4.11 yields that

$$\mathrm{E}|a_6(z_1,z_2)|\le C$$

and that

$$\mathrm{E}|a_4(z_1,z_2)+\frac{1}{N^2}\sum_{j<k} s_j^2\psi_j(z_1)\psi_j(z_2)$$
$$\mathrm{tr}\left(\mathrm{E}_k\mathbf{D}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}\right)\mathrm{tr}\left(\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\right)|\le CN^{-1}$$

where the last inequality uses (3.19). (3.22), together with the above three inequalities, ensures that

$$\mathrm{tr}(\mathrm{E}_k\mathbf{D}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n})\left[1+\frac{1}{N^2}\sum_{j<k} s_j^2\psi_j(z_1)\psi_j(z_2)\mathrm{tr}\left(\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\right)\right]$$
$$=-\mathrm{tr}\left(\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}\right)+a_7(z_1,z_2)$$

where $\mathrm{E}|a_7(z_1,z_2)|\le CN^{1/2}$. Combining (3.18), (3.20) with (3.21), one has

$$\begin{aligned}&\mathrm{tr}(\mathrm{E}_k\mathbf{D}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n})\\&\quad\times\left[1-\frac{1}{N^2}\sum_{j<k} s_j^2\psi_j(z_1)\psi_j(z_2)\mathrm{tr}\left(\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_2)\mathbf{T}_{2n}\right)\right]\\&=\mathrm{tr}\left(\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_2)\mathbf{T}_{2n}\right)+a_8(z_1,z_2)\end{aligned}\tag{3.23}$$

where $\mathrm{E}|a_8(z_1,z_2)|\le CN^{1/2}$. From Zhang (2006)

$$g_{2n}(z)\to g_2(z)\quad\text{a.s.}\quad\text{as } n\to\infty.$$

It follows that

$$|\psi_j(z)-\frac{1}{1+s_jg_{2n}^0(z)}|\le C\left(|\mathrm{E}g_{2n}(z)-g_2(z)|+|g_{2n}^0(z)-g_2(z)|\right)=o(1),\tag{3.24}$$

where $g_{2n}(z)$ is defined at (3.5). Note that by (1.10)

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{n} s_j^2 \frac{1}{\left(1+s_jg_{2n}^0(z_1)\right)\left(1+s_jg_{2n}^0(z_2)\right)} \\
=&\frac{1}{g_{2n}^0(z_1)-g_{2n}^0(z_2)}\left[\frac{1}{N}\sum_{j=1}^{n}\frac{s_j}{1+s_jg_{2n}^0(z_2)}-\frac{1}{N}\sum_{j=1}^{n}\frac{s_j}{1+s_jg_{2n}^0(z_1)}\right] \\
=&\frac{z_1g_{1n}^0(z_1)-z_2g_{1n}^0(z_2)}{g_{2n}^0(z_1)-g_{2n}^0(z_2)} \qquad (3.25)
\end{aligned}$$

and by (1.11)

$$\int \frac{t^2}{\left(1+g_{1n}^0(z_1)t\right)\left(1+g_{1n}^0(z_2)t\right)}dH_{2n}(t) = \frac{z_1g_{2n}^0(z_1)-z_2g_{2n}^0(z_2)}{g_{1n}^0(z_1)-g_{1n}^0(z_2)}. \qquad (3.26)$$

Using the fact from (1.10) and (3.24) that

$$\frac{1}{N}\sum_{j\neq k} s_j\psi_j(z) + zg_{1n}^0(z) = o(1),$$

we deduce that

$$\begin{aligned}
\operatorname{tr}\left(\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_2)\mathbf{T}_{2n}\right) =&\frac{N}{z_1z_2}\int\frac{t^2}{\left(1+g_{1n}^0(z_1)t\right)\left(1+g_{1n}^0(z_2)t\right)}dH_{2n}(t) \\
=&\frac{N}{z_1z_2}\frac{z_1g_{2n}^0(z_1)-z_2g_{2n}^0(z_2)}{g_{1n}^0(z_1)-g_{1n}^0(z_2)}.
\end{aligned}$$

We now deal with $\frac{1}{N^2}\sum_{j<k} s_j^2\psi_j(z_1)\psi_j(z_2)$ in (3.23). For any $\varepsilon \in (0, 1/100)$, we now distinguish the following two cases.

Case 1 : When $k \leq n^{1-\varepsilon}$, one gets

$$\begin{aligned}
&\frac{1}{N^2}\sum_{j<k}\Bigg|\left[\frac{s_j^2}{(1+s_jg_{2n}^0(z_1))(1+s_jg_{2n}^0(z_2))} - c_n^{-1}\frac{z_1g_{1n}^0(z_1)-z_2g_{1n}^0(z_2)}{g_{2n}^0(z_1)-g_{2n}^0(z_2)}\right] \\
&\times\operatorname{tr}\left(\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_2)\mathbf{T}_{2n}\right)\Bigg| \leq CN^{-\varepsilon} = o(1),
\end{aligned}$$

Case 2 : When $k > n^{1-\varepsilon}$, one gets by (3.25)

$$\begin{aligned}
&\frac{1}{N^2}\Bigg|\sum_{j<k}\left[\frac{s_j^2}{(1+s_jg_{2n}^0(z_1))(1+s_jg_{2n}^0(z_2))} - c_n^{-1}\frac{z_1g_{1n}^0(z_1)-z_2g_{1n}^0(z_2)}{g_{2n}^0(z_1)-g_{2n}^0(z_2)}\right] \\
&\times\operatorname{tr}\left(\mathbf{R}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{R}_k^{-1}(z_2)\mathbf{T}_{2n}\right)\Bigg|
\end{aligned}$$

$$\leq \frac{1}{N}\Big|\sum_{j<k}\Big[\frac{s_j^2}{(1+s_j g_{2n}^0(z_1))(1+s_j g_{2n}^0(z_2))} - c_n^{-1}\frac{z_1 g_{1n}^0(z_1) - z_2 g_{1n}^0(z_2)}{g_{2n}^0(z_1) - g_{2n}^0(z_2)}\Big]\Big|$$
$$= o(1).$$

It follows that

$$\begin{aligned}
&\operatorname{tr}(\mathrm{E}_k \mathbf{D}_k^{-1}(z_1)\mathbf{T}_{2n}\mathbf{D}_k^{-1}(z_2)\mathbf{T}_{2n}) \\
&\times\Big[1 - \frac{k-1}{nz_1z_2}\frac{z_1 g_{1n}^0(z_1) - z_2 g_{1n}^0(z_2)}{g_{2n}^0(z_1) - g_{2n}^0(z_2)}\frac{z_1 g_{2n}^0(z_1) - z_2 g_{2n}^0(z_2)}{g_{1n}^0(z_1) - g_{1n}^0(z_2)}\Big] \\
=&\frac{N}{z_1z_2}\frac{z_1 g_{2n}^0(z_1) - z_2 g_{2n}^0(z_2)}{g_{1n}^0(z_1) - g_{1n}^0(z_2)} + a_9(z_1,z_2)
\end{aligned}$$

where $\mathrm{E}|a_9(z_1,z_2)| = o(N)$. Applying Lemma 4.11 and (3.24), one gets

$$|b_k(z) - \frac{1}{1+s_k g_{2n}^0(z)}| \leq \frac{C}{N}\mathrm{E}\Big|\operatorname{tr}\Big(\mathbf{D}_k^{-1}(z) - \mathbf{D}^{-1}(z)\Big)\mathbf{T}_{2n}\Big| + o(1) = o(1).$$

Set

$$d_n(z_1,z_2) = \frac{1}{z_1z_2}\frac{z_1 g_{1n}^0(z_1) - z_2 g_{1n}^0(z_2)}{g_{2n}^0(z_1) - g_{2n}^0(z_2)}\frac{z_1 g_{2n}^0(z_1) - z_2 g_{2n}^0(z_2)}{g_{1n}^0(z_1) - g_{1n}^0(z_2)}$$

and

$$r_{nk}(z_1,z_2) = \frac{s_k^2}{\Big(1+s_k g_{2n}^0(z_1)\Big)\Big(1+s_k g_{2n}^0(z_2)\Big)}.$$

By (3.25) and (3.26), we obtain

$$\begin{aligned}
|d_n(z_1,z_2)| \leq &\left[c_n\int\frac{x^2}{|1+g_{2n}^0(z_1)x|^2}dH_{1n}(x)\int\frac{t^2}{|z_1\left(1+g_{1n}^0(z_1)t\right)|^2}dH_{2n}(t)\right]^{1/2} \\
&\times\left[c_n\int\frac{x^2}{|1+g_{2n}^0(z_2)x|^2}dH_{1n}(x)\int\frac{t^2}{|z_2\left(1+g_{1n}^0(z_2)t\right)|^2}dH_{2n}(t)\right]^{1/2} \\
=&\left[\frac{\Im\left(z_1 g_{1n}^0(z_1)\right)}{\Im g_{2n}^0(\bar z_1)}\frac{\Im g_{2n}^0(z_1) - v\int\frac{t}{|z_1(1+g_{1n}^0(z_1)t))|^2}dH_{2n}(t)}{\Im\left(\bar z_1 g_{1n}^0(\bar z_1)\right)}\right]^{1/2} \\
&\times\left[\frac{\Im\left(z_2 g_{1n}^0(z_2)\right)}{\Im g_{2n}^0(\bar z_2)}\frac{\Im g_{2n}^0(z_2) - v\int\frac{t}{|z_2(1+g_{1n}^0(z_2)t))|^2}dH_{2n}(t)}{\Im\left(\bar z_2 g_{1n}^0(\bar z_2)\right)}\right]^{1/2} < 1.
\end{aligned} \tag{3.27}$$

Using (3.27), (3.15) can be rewritten as for large n

$$\frac{1}{Nz_1z_2}\frac{z_1 g_{2n}^0(z_1) - z_2 g_{2n}^0(z_2)}{g_{1n}^0(z_1) - g_{1n}^0(z_2)}\sum_{k=1}^n r_{nk}(z_1,z_2)\Big(1 - \frac{k-1}{n}d_n(z_1,z_2)\Big)^{-1} + o_p(1).$$

Applying Lemma 4.12 and (3.25), we have

$$\begin{aligned}
&\frac{1}{N}\sum_{k=1}^{n} r_{nk}(z_1,z_2)\left(1-\frac{k-1}{n}d_n(z_1,z_2)\right)^{-1}\\
&=(1-d_n(z_1,z_2))^{-1}\,\frac{1}{N}\sum_{k=1}^{n} r_{nk}(z_1,z_2)-\frac{1}{N}\sum_{k=1}^{n}\sum_{j=1}^{k} r_{nj}(z_1,z_2)\\
&\times\left[\frac{1}{1-n^{-1}kd_n(z_1,z_2)}-\frac{1}{1-n^{-1}(k-1)d_n(z_1,z_2)}\right]\\
&=(1-d_n(z_1,z_2))^{-1}\,\frac{z_1g_{1n}^0(z_1)-z_2g_{1n}^0(z_2)}{g_{2n}^0(z_1)-g_{2n}^0(z_2)}-d_n(z_1,z_2)\frac{1}{N}\sum_{k=1}^{n}\frac{\sum_{j=1}^{k} r_{nj}(z_1,z_2)}{k}\\
&\times\frac{n^{-1}k}{(1-n^{-1}kd_n(z_1,z_2))\,(1-n^{-1}(k-1)d_n(z_1,z_2))}.
\end{aligned}$$

We next develop the above limit by Abel's lemma. To this end, consider the following two cases, for any $\varepsilon\in(0,1/100)$ and large n.

Case 1 : When $k\le n^{1-\varepsilon}$, one gets

$$\begin{aligned}
&\frac{1}{N}\sum_{k\le n^{1-\varepsilon}}\left|\left[\frac{\sum_{j=1}^{k} r_{nj}(z_1,z_2)}{k}-c_n^{-1}\frac{z_1g_{1n}^0(z_1)-z_2g_{1n}^0(z_2)}{g_{2n}^0(z_1)-g_{2n}^0(z_2)}\right]\right.\\
&\left.\times\frac{n^{-1}kf_n(z_1,z_2)}{(1-n^{-1}kd_n(z_1,z_2))\,(1-n^{-1}(k-1)d_n(z_1,z_2))}\right|\le CN^{-\varepsilon}=o(1).
\end{aligned}$$

Case 2 : When $k>n^{1-\varepsilon}$, one gets by (3.25)

$$\begin{aligned}
&\frac{1}{N}\sum_{k\ge n^{1-\varepsilon}}\left|\left[\frac{\sum_{j=1}^{k} r_{nj}(z_1,z_2)}{k}-c_n^{-1}\frac{z_1g_{1n}^0(z_1)-z_2g_{1n}^0(z_2)}{g_{2n}^0(z_1)-g_{2n}^0(z_2)}\right]\right.\\
&\left.\times\frac{n^{-1}kd_n(z_1,z_2)}{(1-n^{-1}kd_n(z_1,z_2))\,(1-n^{-1}(k-1)d_n(z_1,z_2))}\right|\\
&\le\frac{C}{N}\sum_{k\ge n^{1-\varepsilon}}\left|\frac{\sum_{j=1}^{k} r_{nj}(z_1,z_2)}{k}-c_n^{-1}\frac{z_1g_{1n}^0(z_1)-z_2g_{1n}^0(z_2)}{g_{2n}^0(z_1)-g_{2n}^0(z_2)}\right|=o(1).
\end{aligned}$$

Hence, (3.15) can be transformed into

$$\begin{aligned}
&d_n(z_1,z_2)\,(1-d_n(z_1,z_2))^{-1}-d_n^2(z_1,z_2)\\
&\times\frac{1}{n}\sum_{k=1}^{n}\frac{n^{-1}k}{(1-n^{-1}kd_n(z_1,z_2))\,(1-n^{-1}(k-1)d_n(z_1,z_2))}+o_p(1).
\end{aligned}$$

Thus,

$$(3.15)\xrightarrow{i.p.}\frac{d(z_1,z_2)}{1-d(z_1,z_2)}-d^2(z_1,z_2)\int_0^1\frac{t}{(1-td(z_1,z_2))^2}dt=\int_0^{d(z_1,z_2)}\frac{1}{1-z}dz.$$

We conclude that

$$\Phi(z_1, z_2) \xrightarrow{i.p.} \frac{\partial^2}{\partial z_2 \partial z_1} \int_0^{d(z_1,z_2)} \frac{1}{1-z} dz.$$

3.2. Tightness of $M_{n1}(z)$

This section is to prove tightness of the sequence of random functions $\widehat{M}_{n1}(z)$ for $z \in \mathcal{C}$ defined in (2.9). Similar to Section 3 of Bai and Silverstein (2004) (see Bai and Silverstein (2004)), it suffices to show that

$$\sup_{n; z_1, z_2 \in \mathcal{C}_n} \frac{\mathrm{E}\,|M_{n1}(z_1) - M_{n1}(z_2)|^2}{|z_1 - z_2|^2}$$

is finite.

We claim that the moments of $\|\mathbf{D}^{-1}(z)\|$, $\|\mathbf{D}_j^{-1}(z)\|$, and $\|\mathbf{D}_{jk}^{-1}(z)\|$ are bounded in n and $z \in \mathcal{C}_n$. Without loss of generality, we only give the proof for $\mathrm{E}\|\mathbf{D}_1^{-1}(z)\|^p$ and the others are similar. In fact, it is obvious for $z = u + iv \in \mathcal{C}_u$. For $z \in \mathcal{C}_l$ or $z \in \mathcal{C}_r$, using (3.2) and (3.3), we have for any positive p and suitably large l

$$\begin{aligned}
\mathrm{E}\|\mathbf{D}_1^{-1}(z)\|^p =& \mathrm{E}\|\mathbf{D}_1^{-1}(z)\|^p I(\eta_l \le \lambda^{\mathbf{B}_1(z)} \le \eta_r) \\
&+ \mathrm{E}\|\mathbf{D}_1^{-1}(z)\|^p I(\lambda_{\min}^{\mathbf{B}_1(z)} < \eta_l \text{ or } \lambda_{\max}^{\mathbf{B}_1(z)} > \eta_r) \\
\le& \max\{\frac{1}{|x_r - \eta_r|^p}, \frac{1}{|\eta_l - x_l|^p}\} + v^{-p}\mathrm{P}(\lambda_{\min}^{\mathbf{B}_1(z)} < \eta_l \text{ or } \lambda_{\max}^{\mathbf{B}_1(z)} > \eta_r) \\
\le& C_1 + C_2 n^p \varepsilon_n^{-p} n^{-l} \le C_p.
\end{aligned}$$

Write

$$m_n(z_1) - m_n(z_2) = \frac{1}{N}\mathrm{tr}\left(\mathbf{D}^{-1}(z_1) - \mathbf{D}^{-1}(z_2)\right) = \frac{1}{N}(z_1 - z_2)\,\mathrm{tr}\mathbf{D}^{-1}(z_1)\mathbf{D}^{-1}(z_2).$$

We then have

$$\begin{aligned}
\frac{M_n(z_1) - M_n(z_2)}{z_1 - z_2} &= \sum_{j=1}^N \left(\mathrm{E}_j - \mathrm{E}_{j-1}\right) \mathrm{tr}\mathbf{D}^{-1}(z_1)\mathbf{D}^{-1}(z_2) \\
&= \sum_{j=1}^N \left(\mathrm{E}_j - \mathrm{E}_{j-1}\right) \mathrm{tr}\left(\mathbf{D}^{-1}(z_1)\mathbf{D}^{-1}(z_2) - \mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\right) \\
&= \sum_{j=1}^N \left(\mathrm{E}_j - \mathrm{E}_{j-1}\right) \mathrm{tr}\left(\mathbf{D}^{-1}(z_1) - \mathbf{D}_j^{-1}(z_1)\right)\left(\mathbf{D}^{-1}(z_2) - \mathbf{D}_j^{-1}(z_2)\right) \\
&\quad + \sum_{j=1}^N \left(\mathrm{E}_j - \mathrm{E}_{j-1}\right) \mathrm{tr}\left(\mathbf{D}^{-1}(z_1) - \mathbf{D}_j^{-1}(z_1)\right)\mathbf{D}_j^{-1}(z_2)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=1}^{N}\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\mathrm{tr}\mathbf{D}_j^{-1}(z_1)\left(\mathbf{D}^{-1}(z_2)-\mathbf{D}_j^{-1}(z_2)\right)\\
=&\frac{1}{N^2}\sum_{j=1}^{N}s_j^2\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\beta_j(z_1)\beta_j(z_2)\left(\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right)^2\\
&-\frac{1}{N}\sum_{j=1}^{N}s_j\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\beta_j(z_1)\mathbf{y}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\\
&-\frac{1}{N}\sum_{j=1}^{N}s_j\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\beta_j(z_2)\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-2}(z_2)\mathbf{y}_j\\
\triangleq&\mathcal{P}_1+\mathcal{P}_2+\mathcal{P}_3.
\end{aligned}
$$

Thus, it suffices to show that $\mathrm{E}\,|\mathcal{P}_1+\mathcal{P}_2+\mathcal{P}_3|^2$ is bounded. Denote $\rho_j(z)=\mathbf{y}_j^*\mathbf{D}_j^{-1}(z)\mathbf{y}_j-\mathrm{E}\mathrm{tr}\left(\mathbf{T}_{2n}\mathbf{D}_j^{-1}(z)\right)$. Note that

$$
\begin{aligned}
\beta_j(z)=&b_j(z)-\frac{1}{N}s_j\beta_j(z)b_j(z)\rho_j(z) &(3.28)\\
=&b_j(z)-\frac{1}{N}s_jb_j^2(z)\rho_j(z)+\frac{1}{N^2}s_j^2\beta_j(z)b_j^2(z)\rho_j^2(z). &(3.29)
\end{aligned}
$$

Applying (3.28), Lemma 4.9, and Lemma 4.4, we deduce for all large n

$$
\begin{aligned}
|b_j(z)|\le&|\mathrm{E}\beta_j(z)|+\frac{1}{N}|s_jb_j(z)\mathrm{E}(\beta_j(z)\rho_j(z))|\\
\le&C_1+C_2|b_j(z)|N^{-1/2}\le\frac{C_1}{1-C_2N^{-1/2}}.
\end{aligned}
$$

Hence $|b_j(z)|$ is bounded for all n. Using (3.28), write

$$
\begin{aligned}
\mathcal{P}_1=&\frac{1}{N^2}\sum_{j=1}^{N}s_j^2b_j(z_1)b_j(z_2)\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\left(\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right)^2\\
&-\frac{1}{N^3}\sum_{j=1}^{N}s_j^3b_j(z_1)b_j(z_2)\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\beta_j(z_2)\rho_j(z_2)\left(\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right)^2\\
&-\frac{1}{N^3}\sum_{j=1}^{N}s_j^3b_j(z_1)\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\beta_j(z_1)\beta_j(z_2)\rho_j(z_1)\left(\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right)^2\\
\triangleq&\mathcal{P}_{11}+\mathcal{P}_{12}+\mathcal{P}_{13}.
\end{aligned}
$$

By Lemma 4.9, we deduce that

$$
\mathrm{E}|\mathcal{P}_{11}|^2=\frac{1}{N^4}\mathrm{E}\left|\sum_{j=1}^{N}s_j^2b_j(z_1)b_j(z_2)\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\left[\left(\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right)^2\right.\right.
$$

$$
\begin{aligned}
&- \left(\mathrm{tr}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}_{2n}\right)^2 \Big] \Bigg|^2 \\
\leq & \frac{C}{N^4}\sum_{j=1}^{N} \mathrm{E}\left|\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j - \mathrm{tr}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}_{2n}\right|^4 \\
&+ \frac{C}{N^2}\sum_{j=1}^{N} \mathrm{E}\left|\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j - \mathrm{tr}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}_{2n}\right|^2 \\
\leq & \frac{C}{N} + C \leq C.
\end{aligned}
$$

Using Lemma 4.4 and Lemma 4.9, one finds

$$
\begin{aligned}
\mathrm{E}|\mathcal{P}_{12}|^2 =& \frac{1}{N^6}\mathrm{E}\left|\sum_{j=1}^{N} s_j^3 b_j(z_1)b_j(z_2)\left(\mathrm{E}_j - \mathrm{E}_{j-1}\right)\beta_j(z_2)\rho_j(z_2)\left(\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right)^2\right|^2 \\
\leq & \frac{C}{N^6}\sum_{j=1}^{N} \mathrm{E}\left|\beta_j(z_2)\rho_j(z_2)\left(\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j - \mathrm{tr}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}_{2n}\right)^2\right|^2 \\
&+ \frac{C}{N^2}\sum_{j=1}^{N} \mathrm{E}\left|\beta_j(z_2)\rho_j(z_2)\right|^2 \\
\leq & \frac{C}{N^6}\sum_{j=1}^{N} \mathrm{E}^{1/2}\left|\rho_j(z_2)\left(\mathbf{y}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j - \mathrm{tr}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}_{2n}\right)^2\right|^4 \\
&\times \mathrm{E}^{1/2}\left|\beta_j(z_2)\right|^4 + \frac{C}{N^2}\sum_{j=1}^{N} \mathrm{E}^{1/2}\left|\beta_j(z_2)\right|^4 \mathrm{E}^{1/2}\left|\rho_j(z_2)\right|^4 \\
\leq & \frac{C}{N^2} + C \leq C.
\end{aligned}
$$

By the same argument, we get $\mathrm{E}|\mathcal{P}_{13}|^2 \leq C$. Hence, we obtain

$$
\mathrm{E}|\mathcal{P}_1|^2 \leq C.
$$

For $\mathcal{P}_2$ and $\mathcal{P}_3$, we only need to analyze one of them due to their similarity. From (3.28), it is obvious that

$$
\begin{aligned}
\mathcal{P}_2 =& -\frac{1}{N}\sum_{j=1}^{N} s_j b_j(z_1)\left(\mathrm{E}_j - \mathrm{E}_{j-1}\right)\mathbf{y}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j \\
&+ \frac{1}{N^2}\sum_{j=1}^{N} s_j^2 b_j(z_1)\left(\mathrm{E}_j - \mathrm{E}_{j-1}\right)\beta_j(z_1)\rho_j(z_1)\mathbf{y}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j.
\end{aligned}
$$

This yields that

$$
\begin{aligned}
\mathrm{E}|\mathcal{P}_2|^2 =&\frac{1}{N^2}\mathrm{E}\left|\sum_{j=1}^{N} s_j b_j(z_1)\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\mathbf{y}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right|^2 \\
&+\frac{1}{N^4}\mathrm{E}\left|\sum_{j=1}^{N} s_j^2 b_j(z_1)\left(\mathrm{E}_j-\mathrm{E}_{j-1}\right)\beta_j(z_1)\rho_j(z_1)\mathbf{y}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right|^2 \\
\le&\frac{C}{N^2}\sum_{j=1}^{N}\mathrm{E}\left|\mathbf{y}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j-\mathrm{tr}\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}_{2n}\right|^2 \\
&+\frac{C}{N^4}\sum_{j=1}^{N}\mathrm{E}\left|\beta_j(z_1)\rho_j(z_1)\mathbf{y}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j\right|^2 \\
\le& C+\frac{C}{N^4}\sum_{j=1}^{N}\mathrm{E}\left|\beta_j(z_1)\rho_j(z_1)\left(\mathbf{y}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{y}_j-\mathrm{tr}\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}_{2n}\right)\right|^2 \\
&+\frac{C}{N^2}\sum_{j=1}^{N}\mathrm{E}\left|\beta_j(z_1)\rho_j(z_1)\right|^2 \le C
\end{aligned}
$$

where the first inequality is from Lemma 4.9 and the last inequality is from Lemma 4.4. Therefore, we conclude that

$$
\sup_{n;z_1,z_2\in\mathcal{C}_n}\frac{\mathrm{E}\,|M_{n1}(z_1)-M_{n1}(z_2)|^2}{|z_1-z_2|^2}\le\sup_{n;z_1,z_2\in\mathcal{C}_n}\mathrm{E}\,|\mathcal{P}_1+\mathcal{P}_2+\mathcal{P}_3|^2\le C.
$$

This implies that $\widehat{M}_{n1}(z)$ is tight.

4. Non-Gaussian case

It has been verified that Lemma 2.5 is true when the entries of the matrix are independent Gaussian variables. This section is to show this conclusion still holds in the general case. The strategy is to compare the characteristic functions of the linear spectral statistics under the normal case and the general case.

It is worth mentioning that for the complex case, u_{jk}, the real part of x_{jk} and v_{jk}, the imaginary part of x_{jk} are independent. Thus it is enough to consider the real case only.

We below assume that $x_{jk}, j=1,\cdots,N, k=1,\cdots,n$ are truncated at $\delta_n\sqrt{n}$, centralized and renormalized as in the last section. That is to say,

$$
|x_{jl}|\le\delta_n\sqrt{n},\ \mathrm{E}x_{jl}=0,\ \mathrm{E}x_{jl}^2=1,\ \mathrm{E}x_{jl}^4=3+o(1).
$$

Denote $\mathbf{A}_n=\frac{1}{N}\mathbf{T}_{2n}^{1/2}\mathbf{Y}_n\mathbf{T}_{1n}\mathbf{Y}_n'\mathbf{T}_{2n}^{1/2}$ where the entries of $\mathbf{Y}_n=(y_{jk})$ are independent real Gaussian random variables such that

$$
\mathrm{E}y_{jk}=0,\quad \mathrm{E}y_{jk}^2=1,\quad \text{for } j=1\cdots N, k=1,\cdots,n.
$$

Moreover, suppose that $\mathbf{X}_n$ and $\mathbf{Y}_n$ be independent random matrices. As in Götze et al. (2015) for any $\theta \in [0, \pi/2]$, we introduce the following matrices

$$\mathbf{W}_n(\theta) = \mathbf{X}_n \sin\theta + \mathbf{Y}_n \cos\theta \quad \text{and} \quad \mathbf{G}_n(\theta) = \frac{1}{N}\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\mathbf{W}_n'\mathbf{T}_{2n}^{1/2} \tag{4.1}$$

where

$$(\mathbf{W}_n(\theta))_{jk} = w_{jk} = x_{jk}\sin\theta + y_{jk}\cos\theta.$$

Furthermore, let

$$\begin{aligned} &\mathbf{H}_n(t,\theta) = e^{it\mathbf{G}_n(\theta)},\ S(\theta) = \operatorname{tr} f(\mathbf{G}_n(\theta)), \\ &S^0(\theta) = S(\theta) - N\int f(x)dF^{c_n,H_{1n},H_{2n}}(x),\ Z_n(x,\theta) = \mathrm{E}e^{ixS^0(\theta)}. \end{aligned} \tag{4.2}$$

For simplicity, we omit the argument θ from the notations of $\mathbf{W}_n(\theta), \mathbf{G}_n(\theta), \mathbf{H}_n(t,\theta)$ and denote them by $\mathbf{W}_n, \mathbf{G}_n, \mathbf{H}_n(t)$ respectively.

Note that

$$Z_n(x,\pi/2) - Z_n(x,0) = \int_0^{\pi/2} \frac{\partial Z_n(x,\theta)}{\partial\theta} d\theta. \tag{4.3}$$

The aim is to prove that $\frac{\partial Z_n(x,\theta)}{\partial\theta}$ converges to zero uniformly in θ over the interval $[0,\pi/2]$, which ensures Lemma 2.5.

To this end, let $f(\lambda)$ be a smooth function with the Fourier transform given by

$$\widehat{f}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\lambda)e^{-it\lambda}d\lambda.$$

From Lemma 4.7, we have

$$\frac{\partial Z_n(x,\theta)}{\partial\theta} = \frac{2xi}{N}\sum_{j=1}^{N}\sum_{k=1}^{n}\mathrm{E}w_{jk}'\left[\mathbf{T}_{2n}^{1/2}\widetilde{f}(\mathbf{G}_n)\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\right]_{jk} e^{ixS^0(\theta)}$$

where

$$w_{jk}' = \frac{dw_{jk}}{d\theta} = x_{jk}\cos\theta - y_{jk}\sin\theta$$

and

$$\widetilde{f}(\mathbf{G}_n) = i\int_{-\infty}^{\infty} u\widehat{f}(u)\mathbf{H}_n(u)du. \tag{4.4}$$

Let $\mathbf{W}_{njk}(x,\theta)$ denote the corresponding matrix $\mathbf{W}_n$ with w_{jk} replaced by w. And let

$$\begin{aligned} &\mathbf{G}_{njk}(w,\theta) = \frac{1}{N}\mathbf{T}_{2n}^{1/2}\mathbf{W}_{njk}(w,\theta)\mathbf{T}_{1n}\mathbf{W}_{njk}'(w,\theta)\mathbf{T}_{2n}^{1/2}, \quad \mathbf{H}_{njk}(w,t,\theta) = e^{it\mathbf{G}_{njk}(w,\theta)} \\ &S(w,\theta) = \operatorname{tr} f(\mathbf{G}_{njk}(w,\theta)) \quad S^0(w,\theta) = S(w,\theta) - N\int f(x)dF^{c_n,H_{1n},H_{2n}}(x) \end{aligned}$$

and

$$\widetilde{f}(\mathbf{G}_{njk}(w,\theta)) = i\int_{-\infty}^{\infty} u\widehat{f}(u)\mathbf{H}_{njk}(w,u,\theta)du$$
$$\varphi_{jk}(w) = \left[\mathbf{T}_{2n}^{1/2}\widetilde{f}(\mathbf{G}_{njk}(w,\theta))\mathbf{T}_{2n}^{1/2}\mathbf{W}_{njk}(w,\theta)\mathbf{T}_{1n}\right]_{jk} e^{ixS^0(w,\theta)}.$$

By Taylor's formula, one finds

$$\varphi_{jk}(w_{jk}) = \sum_{l=0}^{3}\frac{1}{l!}w_{jk}^{l}\varphi_{jk}^{(l)}(0) + \frac{1}{4!}w_{jk}^{4}\varphi_{jk}^{(4)}(\varrho w_{jk}) \quad \varrho\in(0,1)$$

which implies that

$$\frac{\partial Z_n(x,\theta)}{\partial\theta} = \frac{2xi}{N}\sum_{l=0}^{3}\frac{1}{l!}\sum_{j=1}^{N}\sum_{k=1}^{n}\mathrm{E}w'_{jk}w_{jk}^{l}\mathrm{E}\varphi_{jk}^{(l)}(0) + \frac{2xi}{4!N}\sum_{j=1}^{N}\sum_{k=1}^{n}\mathrm{E}w'_{jk}w_{jk}^{4}\varphi_{jk}^{(4)}(\varrho w_{jk}).$$

It is easy to obtain

$$\begin{aligned}\mathrm{E}w'_{jk}w_{jk}^{0} &= 0, & \mathrm{E}w'_{jk}w_{jk}^{1} &= 0,\\ \mathrm{E}w'_{jk}w_{jk}^{2} &= \mathrm{E}w_{jk}^{3}\sin^2\theta\cos\theta, & \mathrm{E}w'_{jk}w_{jk}^{3} &= o(1)\sin^3\theta\cos\theta.\end{aligned}$$

It follows that

$$\begin{aligned}\frac{\partial Z_n(x,\theta)}{\partial\theta} =& \frac{xi}{N}\sum_{j=1}^{N}\sum_{k=1}^{n}\mathrm{E}w_{jk}^{3}\sin^2\theta\cos\theta\mathrm{E}\varphi_{jk}^{(2)}(0) + \frac{xi}{12N}\sum_{j=1}^{N}\sum_{k=1}^{n}\mathrm{E}w'_{jk}w_{jk}^{4}\varphi_{jk}^{(4)}(\varrho w_{jk})\\ \triangleq& \mathcal{I}_1 + \mathcal{I}_2.\end{aligned}$$

We here only consider $\mathcal{I}_1$. The analysis of $\mathcal{I}_2$ is in Supplement B. A direct calculation yields that

$$\begin{aligned}\varphi_{jk}^{(2)}(w_{jk}) =& \left[\mathbf{T}_{2n}^{1/2}\frac{\partial^2\widetilde{f}(\mathbf{G}_n)}{\partial w_{jk}^2}\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\right]_{jk} e^{ixS^0(\theta)} + 2\left[\mathbf{T}_{2n}^{1/2}\frac{\partial\widetilde{f}(\mathbf{G}_n)}{\partial w_{jk}}\mathbf{T}_{2n}^{1/2}\right]_{jj}[\mathbf{T}_{1n}]_{kk}\, e^{ixS^0(\theta)}\\ &+ \frac{6xi}{N}\left[\mathbf{T}_{2n}^{1/2}\frac{\partial\widetilde{f}(\mathbf{G}_n)}{\partial w_{jk}}\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\right]_{jk}\left[\mathbf{T}_{2n}^{1/2}\widetilde{f}(\mathbf{G}_n)\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\right]_{jk} e^{ixS^0(\theta)}\\ &+ \frac{6xi}{N}\left[\mathbf{T}_{2n}^{1/2}\widetilde{f}(\mathbf{G}_n)\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\right]_{jk}\left[\mathbf{T}_{2n}^{1/2}\widetilde{f}(\mathbf{G}_n)\mathbf{T}_{2n}^{1/2}\right]_{jj}[\mathbf{T}_{1n}]_{kk}\, e^{ixS^0(\theta)}\\ &- \frac{4x^2}{N^2}\left[\mathbf{T}_{2n}^{1/2}\widetilde{f}(\mathbf{G}_n)\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\right]_{jk}^{3} e^{ixS^0(\theta)}\\ \triangleq& \mathcal{J}_{jk}^{1} + \mathcal{J}_{jk}^{2} + \mathcal{J}_{jk}^{3} + \mathcal{J}_{jk}^{4} + \mathcal{J}_{jk}^{5}.\end{aligned}$$

Using Lemma 4.7, one finds

$$\mathcal{J}_{jk}^{1} = -\frac{2}{N}\int_{-\infty}^{\infty} u\widehat{f}(u)[\mathbf{T}_{1n}]_{kk}[\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}]_{jj} * [\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}]_{jk}(u)e^{ixS^0(\theta)}du$$

$$
\begin{aligned}
&-\frac{6i}{N^2}\int_{-\infty}^{\infty} u\widehat{f}(u)[\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}]_{jj} * [\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}]_{jk} \\
&* [\mathbf{T}_{1n}\mathbf{W}_n'\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}]_{kk}(u)e^{ixS^0(\theta)}du \\
&-\frac{2i}{N^2}\int_{-\infty}^{\infty} u\widehat{f}(u)[\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}]_{jk} * [\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}]_{jk} \\
&* [\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}]_{jk}(u)e^{ixS^0(\theta)}du.
\end{aligned}
$$

It is straightforward to check that the moments of $\left\|\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\right\|$, $\frac{1}{\sqrt{N}}\left\|\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\right\|$ and

$$
\frac{1}{N}\left\|\mathbf{T}_{1n}\mathbf{W}_n'\mathbf{T}_{2n}^{1/2}\mathbf{H}_n\mathbf{T}_{2n}^{1/2}\mathbf{W}_n\mathbf{T}_{1n}\right\| \tag{4.5}
$$

are bounded. Applying Lemma 4.8, we obtain

$$
\left|\frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{n}\mathrm{E}\mathcal{J}_{jk}^{1}\right| \le \frac{C}{N^{1/4}}\int_{-\infty}^{\infty}\left(|u|^2+|u|^3\right)|\widehat{f}(u)|du \le CN^{-1/4}.
$$

By the same argument, we get

$$
\left|\frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{n}\mathrm{E}\left(\mathcal{J}_{jk}^{2}+\mathcal{J}_{jk}^{3}+\mathcal{J}_{jk}^{4}+\mathcal{J}_{jk}^{5}\right)\right| \le CN^{-1/4}.
$$

Hence,

$$
|\mathcal{I}_1| \to 0 \quad \text{as } n\to\infty.
$$

References

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

Bai, Z. D. and Silverstein, J. W. (2010) *Spectral analysis of large dimensional random matrices*. Springer.

Bai, Z. D. and Yin, Y. Q. (1993) Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, **21**(3), 1275–1294.

Billingsley, P. (1999) *Convergence of Probability Measures*. Wiley, New York.

Boutet de Mondvel, A., Khorunzhy, A., and Vasilchuk, V. (1996) Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Processes and Related Fields*, **2**, 607–636.

Couillet, R. and Hachem, W. (2014) Analysis of the limiting spectral measure of large random matrices of the separble covariance type. *Random Matrices: Theory and Applications*, **3**(4), 1450016.

Götze, F., Naumov, A. and Tikhomirov, A. (2015) Distribution of linear statistics of singular values of the product of random matrices. *arXiv preprint arXiv:1412.3314v2*.

Hachem, W., Loubaton, P. and Najim, J. (2006) The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile. *Annales De L'institut Henri Poincare-probabilities et Statistiques*, **42** (6), 649–670.

Jiang, T. and Yang, F. (2013) Central Limit Theorems for Classical Likelihood Ratio Tests for High-Dimensional Normal Distributions. *The Annals of Statistics*, **41** (4), 2029–2074.

Jonsson, D. (1982) Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, **12**, 1–38.

Ledoit, O. and Wolf, M. (2002) Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, **30** (4), 1081–1102.

Leng, C. and Tang, C. Y. (2012) Sparse matrix graphical models. *Journal of the American Statistical Association,* **107**, 1287–1300.

Li, P., Paul, D., Narasimhan, R. and Cioffi, J. (2006) On the distribution of SINR for the MMSE MIMO receiver and performance analysis. *IEEE Transactions on Information Theory*, **52** (1), 271–286.

Lytova, A. and Pastur, L. (2009) Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *The Annals of Probability*, **37** (5), 1778–1840.

Nadakuditi, R. R. and Edelman, A. (2008) Sample Eigenvalue Based Detection of High-Dimensional Signals in White Noise Using Relatively Few Samples. *IEEE Transactions on Signal Processing*, **56** (7), 2625–2638.

Najim, J. and Yao, J. F. (2016) Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *The Annals of Applied Probability*, **26** (3), 1837–1887.

Onatski, A. (2010) Determining the Number of Factors From Empirical Distribution of Eigenvalues. *Review of Economics and Statistics*, **92** (4), 1004–1016.

Pan, G. M. and Zhou, W. (2008) Central limit theorem for signal-to-inferference ratio of reduced rank linear receiver. *The Annals of Applied Probability*, **18**, 1232–1270.

Paul, D. and Silverstein, J. (2009) No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix. *Journal of Multivariate Analysis*, **100** (1), 37–57.

Shcherbina, M. (2011) Central limit theorem for linear eigenvalues statistics of Wigner and sample covariance random matrices. *Mathematical Physics, Analysis and Geometry*, **72**, 176–192.

Silverstein, J. W. and Bai, Z. D. (1995) On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate analysis*, **54** (2), 175–192.

Tulino, A. M. and Verdu, S. (2004) Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, **1** (1), 1–182.

Wishart, J. (1928) The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, **20** (1/2), 32–52.

Yao, J. F., Zheng, S. R. and Bai, Z. D. (2015) *Sample covariance matrices and high-dimensional data analysis*. Cambridge University Press.

Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. (1988) On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, **78** (4), 509–521.

Zhang, L. X. (2006) *Spectral analysis of large dimensional random matrices*. The phd thesis, National University of Singapore.