○ Journal of Inequalities and Applications
a SpringerOpen Journal

A result on the limiting spectral distribution of random matrices with unequal variance entries

Shaojia Jin[1] and Junshan Xie[2*]

*Correspondence:
xjsnwpu@163.com
[2]School of Mathematics and
Statistics, Henan University, Kaifeng,
China
Full list of author information is
available at the end of the article

Abstract

A classical result in random matrix theory reveals that the limiting spectral distribution of a Wigner matrix whose entries have a common variance and satisfy other regular assumptions almost surely converges to the semicircular law. In the paper, we will relax the assumption of uniform covariance of each entry, when the average of the normalized sums of the variances in each row of the data matrix converges to a constant, we prove that the same limiting spectral distribution holds. A similar result on a sample covariance matrix is also established. The proofs mainly depend on the Stein equation and the generalized Stein equation of independent random variables.

MSC: Primary 60E10; secondary 60F15

Keywords: Limiting spectral distribution; Unequal variance; Stein equation

1 Introduction

Suppose A_n is an $n \times n$ Hermitian matrix and $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote the real eigenvalues of A_n. The empirical spectral distribution function (ESD) of A_n can be defined as

$$F^{A_n}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{\lambda_i \le x\}},$$

where I_A represents the indicator function on the set A. The limit distribution of $F^{A_n}(x)$ as $n \to \infty$, if it exists, will be called the limiting spectral distribution (LSD) of A_n. Since most of the global spectral limiting properties of A_n can be determined by its LSD, the LSD of large dimensional random matrices has attracted considerable interest among mathematicians, probabilists, and statisticians, one can refer to Wigner [15, 16], Grenander and Silverstein [7], Jonsson [8], Yin and Krishnaiah [18], Bai and Yin [4] and so on.

The Wigner matrix is one of the most basic and popular objects in the random matrix theory. A Wigner matrix is a symmetric (or Hermitian in the complex case) random matrix whose entries on or above the diagonal are independent random variables. When a Wigner matrix X_n whose entries are i.i.d. real (or complex) random variables with mean zero and variance 1, Wigner [16] proved that the expected ESD of $W_n = \frac{1}{\sqrt{n}} X_n$, tends to the limiting

Springer

distribution F_{sc}, whose density is given by

$$f_{sc}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}I_{\{|x|\leq 2\}} & \text{if } |x| \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

The LSD F_{sc} is usually called the semicircular law in the literature. Grenander [6] proved that $\|F^{W_n} - F_{sc}\| \to 0$ in probability. Arnold [1, 2] obtained the result that F^{W_n} converges to F_{sc} almost surely. Pastur [12] removed the identically distributed assumption, and considered that when the entries above or on the diagonal of X_n are independent real or complex random variables with mean zero and variance 1, may not necessarily be identically distributed, but satisfy the following Lindeberg type assumption, for any constant $\eta > 0$:

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j}^{n} E|X_{ij}|^2 I\big(|X_{ij}| \geq \eta\sqrt{n}\big) = 0. \tag{1.1}$$

Then the ESD of W_n converges almost surely to the semicircular law.

　　Among the results above, the assumption that the entries of the Wigner matrix have a common variance is necessary. However, in a practical application, the uniform variance assumption is a strong condition. In the paper, we will remove the uniform variance assumption and establish the same semicircular law result under a milder assumption on the variances of the entries, in particular, we assume that the covariances of the entries may not be equal to a constant, but only the average of the normalized sums of in each row of the data matrix converges to a positive constant. The result reads as follows.

Theorem 1.1 *Let $W_n = \frac{1}{\sqrt{n}}X_n$ be a Wigner matrix, and the entries above or on the diagonal of X_n be independent real or complex random variables, but they may not be necessarily identically distributed. Assume that all the entries of X_n are of mean zero, and the variance $E|X_{ij}|^2 = \sigma_{ij}^2$, where σ_{ij} satisfies $\frac{1}{n}\sum_{i=1}^{n} |\frac{1}{n}\sum_{j=1}^{n}\sigma_{ij}^2 - 1| \to 0$ as $n \to \infty$, and the assumption (1.1) holds. Then, almost surely, the ESD of W_n converges weakly to the semicircular law.*

Remark 1.1 The result of Theorem 1.1 can be extended to a general one: when the average of the normalized sums in each row converges weakly to a positive constant σ^2, then almost surely the LSD of W_n is the general semicircular law with density

$$f_{sc,\sigma}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4\sigma^2 - x^2}I_{\{|x|\leq 2\sigma\}} & \text{if } |x| \leq 2\sigma, \\ 0 & \text{otherwise.} \end{cases}$$

　　Now, we will consider the LSD of a sample covariance matrix, which is also an important object in random matrix theory and multivariate statistics. Suppose $Y_n = (Y_{ij})_{n\times N}$ is a real or complex random matrix, whose entries Y_{ij} $(i = 1, \ldots, n, j = 1, \ldots, N)$ are i.i.d. real or complex random variables with mean zero and variance 1. Write $Y_j = (Y_{1j}, \ldots, Y_{nj})'$ and $\mathbf{Y}_n = (Y_1, \ldots, Y_N)$. Define $\bar{Y} = \frac{1}{N}\sum_{k=1}^{N}Y_k$. Since $\tilde{S}_n = \frac{1}{N-1}\sum_{k=1}^{N}(Y_k - \bar{Y})(Y_k - \bar{Y})^*$ shares the same the LSD with $S_n = \frac{1}{N}\sum_{k=1}^{N}Y_kY_k^* = \frac{1}{N}\mathbf{Y}_n\mathbf{Y}_n^*$, where $*$ represents the conjugate transpose symbol, we usually consider the sample covariance matrix defined by $S_n = \frac{1}{N}\mathbf{Y}_n\mathbf{Y}_n^*$, The limiting spectral properties of large sample covariance matrices have generated a considerable amount of interest in statistics, signal processing and other disciplines. The

first result on the LSD of S_n is due to Marčenko and Pastur [10], who proved that when $\lim_{n\to\infty}\frac{n}{N} = y \in (0,\infty)$, the LSD of S_n is M-P law $F_y^{MP}(x)$ with density

$$f_y^{MP}(x) = \begin{cases} \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}, & \text{if } a \le x \le b, \\ 0, & \text{otherwise,} \end{cases}$$

and has a point mass $1 - 1/y$ at the origin if $y > 1$, where $a = (1 - \sqrt{y})^2$ and $b = (1 + \sqrt{y})^2$. There is also some work on the discussion of the M-P law of sample covariance matrices, such as Bai and Yin [4], Grenander and Silverstein [7], Jonsson [8], Yin [17], Silverstein [13] and Silverstein and Bai [14]. A typical result (see Theorem 3.9 of Bai and Silverstein [3]) states that when the entries of Y_n are independent random variables with mean zero and variance 1, $n/N \to y \in (0,\infty)$, and for any $\eta > 0$,

$$\frac{1}{\eta^2 nN}\sum_{i,j} E\big(|Y_{ij}|^2 I\big(|Y_{ij}| \ge \eta\sqrt{N}\big)\big) \to 0, \tag{1.2}$$

then the ESD of S_n tends to the M-P law F_y^{MP} almost surely. Note that the entries of Y_n having a uniform variance 1 is also a necessary condition in the proof. By the same motivation as Theorem 1.1, we will also consider removing the equal covariance condition. Similarly, we can get the following result.

Theorem 1.2 *Assume that the entries of the random matrix Y_n defined above are independent variables with mean zero and variance $E|Y_{ij}|^2 = \sigma_{ij}^2$, where σ_{ij} satisfies $\frac{1}{n}\sum_{i=1}^n |\frac{1}{N}\sum_{j=1}^N \sigma_{ij}^2 - 1| \to 0$. Assume that $n/N \to y \in (0,\infty)$ as $n \to \infty$ and the assumption (1.2) holds. Then, almost surely, the ESD of the sample covariance matrix $S_n = \frac{1}{N}Y_n Y_n^*$ converges weakly to the M-P law.*

Remark 1.2 Likewise, if there exists a positive constant $\sigma^2 > 0$ satisfying $\frac{1}{n}\sum_{i=1}^n \times |\frac{1}{N}\sum_{j=1}^N \sigma_{ij}^2 - \sigma^2| \to 0$, the other assumptions remain unchanged, we also get, almost surely, for the LSD of $S_n = \frac{1}{N}Y_n Y_n^*$ the general M-P law with density

$$f_{y,\sigma}^{MP}(x) = \begin{cases} \frac{1}{2\pi xy\sigma^2}\sqrt{(\tilde{b}-x)(x-\tilde{a})}, & \text{if } \tilde{a} \le x \le \tilde{b}, \\ 0 & \text{otherwise,} \end{cases}$$

and it has a point mass $1 - 1/y$ at the origin if $y > 1$, where $\tilde{a} = \sigma^2(1 - \sqrt{y})^2$ and $\tilde{b} = \sigma^2(1 + \sqrt{y})^2$.

The rest of the paper is organized as follows. The proofs of the main results are presented in Sect. 2. In the Appendix, some useful lemmas are listed. In the sequel, when there is no confusion, we may get rid of the subscript n in the notation of matrices for brevity. A^* denotes the conjugate transpose of matrix A, and tr(A) denotes the trace of A, and C denotes positive constant, which may be different in different cases.

2 Proofs

The Stieltjes transform method is mainly adopted to complete the proofs. For a distribution function $F(x)$, its Stieltjes transform can be defined as

$$s_F(z) = \int \frac{1}{x-z}\, dF(x), \quad z \in \mathbb{C}^+ = \{z = u + iv | u \in \mathbb{R}, v > 0\}.$$

Obviously, we can write the Stieltjes transform of ESD $F^{A_n}(x)$ as

$$s_{F^{A_n}}(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z} = \frac{1}{n}\operatorname{tr}(A_n - zI_n)^{-1},$$

where I_n is the identity matrix with order n. The continuity theorem of Stieltjes transform states that, for a sequence of functions of bounded variation $\{G_n\}$ with the Stieltjes transform $s_{G_n}(z)$, $G_n(-\infty) = 0$ for all n, and a function of bounded variation G, $G(-\infty) = 0$, with the Stieltjes transform $s_G(z)$, G_n converges vaguely to G if and only if $s_{G_n}(z)$ converges to $s_G(z)$ for $z \in \mathbb{C}^+$. In view of the fact that the sequence of the ESD of Wigner matrix is tight (see Lytova and Pastur [9]), the weak convergence of the ESD can be obtained by the convergence of their corresponding Stieltjes transform. Furthermore, if the LSD is a deterministic probability density function, then the almost surely convergence of the ESD can be achieved by the almost surely convergence of the Stieltjes transform, which is a basic idea in the following proofs.

2.1 Proof of Theorem 1.1

Define F^{W_n} to be the ESD of W_n and $s_n(z)$ the Stieltjes transforms of F^{W_n}. Then by the continuity theorem of the Stieltjes transform, we complete the proof of Theorem 1.1 by showing

$$s_n(z) \to s(z), \quad z \in \mathbb{C}^+ \text{ a.s.,} \tag{2.1}$$

where $s(z)$ is the Stieltjes transform of the semicircular law F_{sc}.

The proofs of the real-valued Wigner matrix are almost the same as those of the complex-valued Wigner matrix, that is, all the results as well as the main ingredients of the proofs in the real symmetric matrices case remain valid in the Hermitian case with natural modifications. For the sake of simplicity, we will confine ourselves to a real symmetric Wigner matrix. To this end, we will write $\hat{W}_n = \frac{1}{\sqrt{n}}\hat{X}_n$ to be a Wigner matrix independent of W_n, and the entries of $\hat{X}_n = (\hat{X}_{ij})_{n \times n}$ are independent $N(0,1)$ random variables. Define $F^{\hat{W}_n}$ to be the ESD of $\hat{W}_n$, and $\hat{s}_n(z)$ the Stieltjes transforms of $F^{\hat{W}_n}$. By Theorem 2.9 of Bai and Silverstein [3], we know that, almost surely, the LSD of $\hat{W}_n$ is a semicircular law $F_{sc}(x)$, which means

$$\hat{s}_n(z) \to s(z), \quad z \in \mathbb{C}^+, \text{ a.s.}$$

Thus, (2.1) can be achieved by

$$\hat{s}_n(z) - s_n(z) \to 0, \quad z \in \mathbb{C}^+, \text{ a.s.} \tag{2.2}$$

In the sequel, we will complete the proof of (2.2) by the following two steps.

(i) For any fixed $z \in \mathbb{C}^+$, $\hat{s}_n(z) - E\hat{s}_n(z) \to 0, a.s.$ and $s_n(z) - Es_n(z) \to 0, a.s.$

(ii) For any fixed $z \in \mathbb{C}^+$, $Es_n(z) - E\hat{s}_n(z) \to 0$.

We begin with step (i). Define W_k to be a major submatrix of order $(n-1)$, obtained from W_n with the kth row and column removed, and α_k to be the vector from the kth column of W_n by deleting the kth entry. Denote by $E_k(\cdot)$ conditional expectation with respect to the σ-field generated by the random variables $\{X_{i,j}, i, j > k\}$, with the convention that $E_n s_n(z) = E s_n(z)$ and $E_0 s_n(z) = s_n(z)$. Then

$$s_n(z) - Es_n(z) = \sum_{k=1}^{n} \left[E_{k-1}(s_n(z)) - E_k(s_n(z)) \right] := \sum_{k=1}^{n} \gamma_k.$$

By Theorem A.5 of Bai and Silverstein [3], we know

$$\gamma_k = \frac{1}{n} \left(E_{k-1} \operatorname{tr}(W_n - zI)^{-1} - E_k \operatorname{tr}(W_n - zI)^{-1} \right)$$

$$= \frac{1}{n} \left(E_{k-1} \left(\operatorname{tr}(W_n - zI)^{-1} - \operatorname{tr}(W_k - zI_{n-1})^{-1} \right) \right.$$

$$\left. - E_k \left(\operatorname{tr}(W_n - zI)^{-1} - \operatorname{tr}(W_k - zI_{n-1})^{-1} \right) \right)$$

$$= \frac{1}{n} \left(E_{k-1} \frac{1 + \alpha_k^*(W_k - zI_{n-1})^{-2}\alpha_k}{-z - \alpha_k^*(W_k - zI_{n-1})^{-1}\alpha_k} - E_k \frac{1 + \alpha_k^*(W_k - zI_{n-1})^{-2}\alpha_k}{-z - \alpha_k^*(W_k - zI_{n-1})^{-1}\alpha_k} \right).$$

Note that

$$\left| 1 + \alpha_k^*(W_k - zI_{n-1})^{-2}\alpha_k \right|$$

$$\leq 1 + \alpha_k^*(W_k - zI_{n-1})^{-1}(W_k - \bar{z}I_{n-1})^{-1}\alpha_k$$

$$= v^{-1} \operatorname{Im}\left(z + \alpha_k^*(W_k - zI_{n-1})^{-1}\alpha_k \right),$$

which implies $|\gamma_k| \leq 2/nv$. Since $\{\gamma_k, k \geq 1\}$ forms a martingale difference sequence, then it follows by Lemma A.1 with $p = 4$ that

$$E\left| s_n(z) - Es_n(z) \right|^4 \leq K_4 E\left(\sum_{k=1}^{n} |\gamma_k|^2 \right)^2 \leq K_4 \left(\sum_{k=1}^{n} \frac{4}{n^2 v^2} \right)^2 \leq \frac{16K_4}{n^2 v^4} = O(n^{-2}),$$

which, together with the Borel–Cantelli lemma, yields

$$s_n(z) - Es_n(z) \to 0, \quad \text{a.s.}$$

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

Similarly, we also get

$$\hat{s}_n(z) - E\hat{s}_n(z) \to 0, \quad \text{a.s.}$$

for every fixed $z \in \mathbb{C}^+$. Therefore, step (i) is completed.

Now we come to step (ii). We firstly introduce some notation:

$$X(s) = s^{1/2}X + (1-s)^{1/2}\hat{X}, \quad 0 \leq s \leq 1, \qquad G(z) = \left(\frac{1}{\sqrt{n}}X - zI \right)^{-1},$$

$$\hat{G}(z) = \left(\frac{1}{\sqrt{n}}\hat{X} - zI\right)^{-1}, \qquad G(s, z) = \left(\frac{1}{\sqrt{n}}X(s) - zI\right)^{-1}.$$

By the facts that $X(1) = X$, $X(0) = \hat{X}$, we can write

$$
\begin{aligned}
&Es_n(z) - E\hat{s}_n(z) \\
&\quad = \frac{1}{n}E\big(trG(z) - \operatorname{tr}\hat{G}(z)\big) \\
&\quad = \int_0^1 \frac{\partial}{\partial s}E\left(\frac{1}{n}trG(s, z)\right)ds \\
&\quad = \frac{1}{2n^{3/2}}\int_0^1 Etr\frac{\partial}{\partial z}G(s, z)\big((1 - s)^{-1/2}\hat{X} - s^{-1/2}X\big)ds.
\end{aligned}
\tag{2.3}
$$

Denote $G' = \frac{\partial}{\partial z}G(s, z)$. Write the (i, j)-entry of G' by G'_{ij} and the (i, j)-entry of $X(s)$ by $X_{ij}(s)$. Since the random variables $\hat{X}_{ij}$ are independent $N(0, 1)$ random variables, applying the Stein equation in Lemma A.2 with $\Phi = G'_{ij}$, we have

$$
\begin{aligned}
&Etr\frac{\partial}{\partial z}G(s, z)(1 - s)^{-1/2}\hat{X} \\
&\quad = (1 - s)^{-1/2}\sum_{i,j}E\big(\hat{X}_{ij}G'_{ij}\big) \\
&\quad = (1 - s)^{-1/2}\sum_{i,j}\big(E|\hat{X}_{ij}|^2 E\big(D_{ij}(s)G'_{ij}\big)(1 - s)^{1/2}n^{-1/2}\big) \\
&\quad = n^{-1/2}\sum_{i,j}E\big(D_{ij}(s)G'_{ij}\big),
\end{aligned}
$$

where $D_{ij}(s) = \partial/\partial X_{ij}(s)$.

On the other hand, as the random variables X_{ij} are independent, we will adopt the generalized Stein equation in Lemma A.3 to rewrite the second term in the parentheses of the r.h.s. of (2.3). To this end, we will take $p = 1$ and $\Phi = G'_{ij}$ in Lemma A.3. Note that $\kappa_1 = EX_{ij} = 0$ and $\kappa_2 = E|X_{ij}|^2$. Then we have

$$
\begin{aligned}
&Etr\frac{\partial}{\partial z}G(s, z)s^{-1/2}X \\
&\quad = s^{-1/2}\sum_{i,j}E\big(X_{ij}G'_{ji}\big) \\
&\quad = s^{-1/2}\sum_{i,j}\big(E|X_{ij}|^2 E\big(D_{ij}(s)G'_{ji}\big)s^{1/2}n^{-1/2} + \varepsilon_{ij}\big) \\
&\quad = n^{-1/2}\sum_{i,j}E|X_{ij}|^2 E\big(D_{ij}(s)G'_{ji}\big) + s^{-1/2}\varepsilon
\end{aligned}
$$

and

$$|\varepsilon| := \left|\sum_{i,j}\varepsilon_{ij}\right| \le C\frac{s}{n}\sum_{i,j}E|X_{ij}|^3 \sup_{X(s)\in\wp_n}\big|D_{ij}^2(s)G'_{ji}\big|,$$

where $\wp_n$ is the set of $n \times n$ real symmetric matrices.

Thus

$$
\begin{aligned}
&\left| E s_n(z) - E \hat{s}_n(z) \right| \\
&\quad \le \frac{1}{2n^{3/2}} \int_0^1 \left(\frac{1}{n^{1/2}} \left| \sum_{i,j} \left\{ \left(1 - E|X_{ij}|^2\right) E\left(D_{ij}(s) G'_{ji}\right) \right\} \right| + \frac{1}{s^{1/2}} |\varepsilon| \right) ds \\
&\quad = \frac{1}{2n^2} \int_0^1 \left| \sum_{i,j} \left\{ \left(1 - E|X_{ij}|^2\right) E\left(D_{ij}(s) G'_{ji}\right) \right\} \right| ds + \frac{1}{2n^{3/2}} \int_0^1 \frac{1}{s^{1/2}} |\varepsilon| \, ds \\
&\quad := I + II.
\end{aligned}
\tag{2.4}
$$

By (3.25) in Lytova and Pastur [9],

$$
\left| D_{ij}^{(l)} G'_{ij} \right| \le c_l / v^{(l+2)},
\tag{2.5}
$$

where c_l is an absolute constant for every l. Let $l = 1$, then $|D_{ij} G'_{ij}| \le c_1 / v^3$, and

$$
E\left(D_{ij}(s) G'_{ji}\right) \le \sup_s \left| D_{ij}(s) G'_{ij} \right| \le c_1 / v^3.
$$

Since $E|X_{ij}|^2 = \sigma_{ij}^2$, and σ_{ij} satisfies $\frac{1}{n} \sum_{i=1}^n \left| \frac{1}{n} \sum_{j=1}^n \sigma_{ij}^2 - 1 \right| \to 0$ based on the condition in Theorem 1.1, we easily get

$$
\frac{1}{n^2} \left| \sum_{i,j} \left(1 - E|X_{ij}|^2\right) \right| = \frac{1}{n^2} \left| n^2 - \sum_{i,j} \sigma_{ij}^2 \right| \le \frac{1}{n} \sum_i^n \left| 1 - \frac{1}{n} \sum_j^n \sigma_{ij}^2 \right| \to 0,
$$

and then

$$
I = o(1).
\tag{2.6}
$$

We have

$$
II \le \frac{C}{2n^{3/2}} \int_0^1 \frac{s^{1/2}}{n} \sum_{i,j} E|X_{ij}|^3 \sup_{X(s) \in \wp_n} \left| D_{ij}^2(s) G'_{ij} \right| ds.
$$

By the assumption (1.1), we select a sequence $\eta_n \downarrow 0$ as $n \to \infty$, such that

$$
\lim_{n \to \infty} \frac{1}{n^2 \eta_n^2} \sum_{i,j} E|X_{ij}|^2 I\left(|X_{ij}| \ge \eta_n \sqrt{n}\right) = 0.
$$

And the convergence rate of η_n can be as slow as desired. For definiteness, we may assume that $\eta_n > 1/\log n$ and $\eta_n \to 0$. Then we have

$$
II \le \frac{C \eta_n}{2n^2} \int_0^1 s^{1/2} \sum_{i,j} E|X_{ij}|^2 \sup_{X(s) \in \wp_n} \left| D_{ij}^2(s) G'_{ij} \right| ds.
$$

Since

$$
\frac{1}{n^2} \sum_{i,j} E|X_{ij}|^2 - 1 \le \frac{1}{n^2} \left| \sum_{i,j} \sigma_{ij}^2 - n^2 \right|
$$

$$= \frac{1}{n}\left|n - \frac{1}{n}\sum_{i,j}\sigma_{ij}^2\right| \leq \frac{1}{n}\sum_i^n\left|1 - \frac{1}{n}\sum_j^n\sigma_{ij}^2\right| \to 0,$$

we obtain $\frac{1}{n^2}\sum_{i,j}E|X_{ij}|^2 = 1 + o(1)$. And by (2.5), let $l = 2$, then

$$\sup_{X(s)\in\wp_n}\left|D_{ij}^2(s)G_{ij}'\right| \leq c_2/v^4.$$

So we have

$$II \leq \frac{1}{2}\int_0^1 s^{1/2}\eta_n\left(\frac{1}{n^2}\sum_{i,j}E|X_{ij}|^2\right)\sup_{X(s)\in\wp_n}\left|D_{ij}^2(s)G_{ji}'\right|ds \leq C\eta_n + o(1).$$

By $\eta_n \to 0$ as $n \to \infty$, we have $II = o(1)$. This, together with (2.4) and (2.6), means that

$$Es_n(z) - E\hat{s}_n(z) \to 0$$

for any fixed $z \in \mathbb{C}^+$. Step (ii) is completed.

Combining steps (i) and (ii), we see that (2.2) is proved. Therefore, we have

$$F^{W_n} \xrightarrow{w} F_{sc}, \quad \text{a.s.},$$

By (2.1)–(2.2), we complete the proof of Theorem 1.1.

2.2 Proof of Theorem 1.2

We will also consider the real-valued sample covariance matrix case and take a similar procedure to Theorem 1.1 to complete the proof. To this end, we also firstly define $\hat{Y}_n = (\hat{Y}_{ij})_{n\times N}$ to be a $n \times N$ random matrix independent of Y_n, and the entries $\hat{Y}_{ij}s$ to be i.i.d. $N(0,1)$ random variables. Write $\hat{S}_n = \frac{1}{N}\hat{Y}_n\hat{Y}_n^*$. We will use F^{S_n} and $F^{\hat{S}_n}$ to denote the ESD of S_n and $\hat{S}_n$, respectively. Let $m_n(z)$ and $\hat{m}_n(z)$ be the Stieltjes transforms of F^{S_n} and $F^{\hat{S}_n}$, respectively.

By Theorem 3.10 of Bai and Silverstein [3], we have obtained

$$\hat{m}_n(z) \to m(z), \quad \text{a.s.}$$

where $m(z)$ is the Stieltjes transform of standard M-P law F_y^{MP}. Thus, by the continuity theorem of the Stieltjes transform again, we complete the proof by showing that, for any fixed $z \in \mathbb{C}^+$,

(i) $m_n(z) - Em_n(z) \to 0$, a.s. and $\hat{m}_n(z) - E\hat{m}_n(z) \to 0$, a.s.

(ii) $Em_n(z) - E\hat{m}_n(z) \to 0$.

For (i), we prove it by a similar argument to Bai and Silverstein [3]. For the sake of completeness, we will also give the proof. Let $\widetilde{E}_k(\cdot)$ denote the conditional expectation given by $\{Y_{k+1}, \ldots, Y_N\}$, with the convention that $\widetilde{E}_N m_n(z) = Em_n(z)$ and $\widetilde{E}_0 m_n(z) = m_n(z)$. Then

$$m_n(z) - Em_n(z) = \sum_{k=1}^N\left(\widetilde{E}_k m_n(z) - \widetilde{E}_{k-1}m_n(z)\right) := \sum_{k=1}^N\widetilde{\gamma}_k,$$

where

$$\widetilde{\gamma}_k = \frac{1}{n}\left(\widetilde{E}_k \operatorname{tr}(S_n - zI_n)^{-1} - \widetilde{E}_{k-1}\operatorname{tr}(S_n - zI_n)^{-1}\right)$$

$$= \frac{1}{n}\left(\left(\widetilde{E}_k - \widetilde{E}_{k-1}\right)\left(\operatorname{tr}(S_n - zI_n)^{-1} - \operatorname{tr}(S_{nk} - zI_n)^{-1}\right)\right)$$

$$= -\frac{1}{n}\left(\left(\widetilde{E}_k - \widetilde{E}_{k-1}\right)\frac{Y_k^*(S_{nk} - zI_n)^{-2}Y_k}{1 + Y_k^*(S_{nk} - zI_n)^{-1}Y_k}\right).$$

Here $S_{nk} = S_n - Y_k Y_k^*$ and Y_k is the kth column of Y_n with the kth element removed, and

$$\left|\frac{Y_k^*(S_{nk} - zI_n)^{-2}Y_k}{1 + Y_k^*(S_{nk} - zI_n)^{-1}Y_k}\right| \le \frac{Y_k^*((S_{nk} - uI_n)^2 + v^2 I_n)^{-1}Y_k}{\operatorname{Im}(1 + Y_k^*(S_{nk} - zI_n)^{-1}Y_k)} = \frac{1}{v}.$$

Note that $\{\widetilde{\gamma}_k, k \ge 1\}$ forms a sequence of bounded martingale differences.

By Lemma A.1 with $p = 4$, we have

$$E\left|m_n(z) - Em_n(z)\right|^4 \le K_4 E\left(\sum_{k=1}^{N}|\gamma_k|^2\right)^2 \le \frac{K_4 N^2}{v^4 n^4} = O(N^{-2}).$$

By the Borel–Cantelli lemma again, we see that almost surely $m_n(z) - Em_n(z) \to 0$. By the same argument, we get $\hat{m}_n(z) - E\hat{m}_n(z) \to 0$, a.s., which means (i) is completed.

Then we come to the proof of (ii). We firstly introduce some notation. For $0 \le s \le 1$,

$$H(s) = s^{1/2}Y + (1-s)^{1/2}\hat{Y}, \qquad V(s) = \frac{1}{\sqrt{n}}H(s),$$

$$J(s) = V(s)V^*(s), \qquad U(z,s) = \left(J(s) - zI\right)^{-1},$$

$$M_0(z,s) = \frac{1}{n}\operatorname{tr}U(z,s), \qquad U' = \frac{\partial}{\partial z}U(z,s).$$

By the same procedure in (2.3), we have

$$Em_n(z) - E\hat{m}_n(z)$$

$$= \int_0^1 \frac{\partial}{\partial s}E\left(M_0(z,s)\right) ds$$

$$= \frac{1}{2nN^{1/2}}\int_0^1 Etr\left(\left((1-s)^{-1/2}\hat{Y} - s^{-1/2}Y\right)V^*(s)U'\right) ds. \tag{2.7}$$

It follows by Lemma A.2 with $\Phi = (V^*(s)U')_{ij}$ that

$$Etr\left((1-s)^{-1/2}\hat{Y}V^*(s)U'\right)$$

$$= (1-s)^{-1/2}\sum_{i,j}E\left(\hat{Y}_{ij}\left(V^*(s)U'\right)_{ij}\right)$$

$$= \frac{1}{N^{1/2}}\sum_{i,j}E|\hat{Y}_{ij}|^2 E\left(D_{ij}(s)\left(V^*(s)U'\right)_{ij}\right),$$

where $D_{ij}(s) = \partial/\partial V_{ij}(s)$.

By Lemma A.3 with $p = 1$ and $\Phi = (V^*(s)U')_{ij}$ again, we can see by $\kappa_1 = EY_{ij} = 0$ and $\kappa_2 = E|Y_{ij}|^2$ that

$$
\begin{aligned}
Etr&\left(s^{-1/2}Y\left(V^*(s)U'\right)\right) \\
&= s^{-1/2}\sum_{i,j}E\left(Y_{ij}\left(V^*(s)U'\right)_{ij}\right) \\
&= s^{-1/2}\sum_{i,j}\left(\frac{1}{N^{1/2}}s^{1/2}E|Y_{ij}|^2 E\left(D_{ij}\left(V^*(s)U'\right)_{ij}\right) + \varepsilon_{ij}\right) \\
&= \frac{1}{N^{1/2}}\sum_{i,j}E|Y_{ij}|^2 E\left(D_{ij}\left(V^*(s)U'\right)_{ij}\right) + s^{-1/2}\varepsilon_0
\end{aligned}
$$

and

$$
|\varepsilon_0| := \left|\sum_{i,j}\varepsilon_{ij}\right| \le C\frac{s}{N}\sum_{i,j}E|Y_{ij}|^3 \sup_{V\in\mathcal{M}_{n,N}}\left|D_{ij}^2\left(V^*U'\right)_{ij}\right|,
$$

where $\mathcal{M}_{n,N}$ is the set of $n \times N$ real matrices.

By (2.7), we have

$$
\begin{aligned}
\left|Em_n(z) - E\hat{m}_n(z)\right| \\
&\le \frac{1}{2nN^{1/2}}\int_0^1\left(\frac{1}{N^{1/2}}\left|\sum_{i,j}\left[(1 - E|Y_{ij}|^2)E\left(D_{ij}\left(V^*U'\right)_{ij}\right)\right]\right| + \frac{1}{s^{1/2}}|\varepsilon_0|\right)ds \\
&= \frac{1}{2nN}\int_0^1\left|\sum_{i,j}\left\{(1 - E|Y_{ij}|^2)E\left(D_{ij}\left(V^*U'\right)_{ij}\right)\right\}\right|ds + \frac{1}{2nN^{1/2}}\int_0^1\frac{1}{s^{1/2}}|\varepsilon_0|\,ds \\
&:= \hat{I} + \hat{II}.
\end{aligned}
\tag{2.8}
$$

The bound of $|D_{ij}^r(V^*U')_{ji}|, r = 1,2$, is critical for the proof. Since $(V^*U')_{ij}$ is analytic in $z \in \mathbb{C}^+$, by the Cauchy inequality for the bound of derivatives of analytic functions in Lemma A.4, to get the bound of $D_{ij}^r(V^*U')_{ij}, r = 1,2$, on any compact set of $\mathbb{C}^+$, it suffices to find the bound of $D_{ij}^r(V^*U)_{ij}$ on the compact set. By elementary calculations, we can get the derivatives of V^*U with respect to the entries $V_{ij}, i = 1,2,\ldots,n, j = 1,2,\ldots N$,

$$
D_{ij}\left(V^*U\right)_{ij} = D_{ij}(UV)_{ij} = U_{ii} - \left(V^*UV\right)_{jj}U_{ii} - (UV)_{ij}^2,
$$
$$
D_{ij}^2\left(V^*U\right)_{ij} = -6U_{ii}(UV)_{ij} + 6U_{ii}(UV)_{ii}\left(V^*UV\right)_{jj} + 2(UV)_{ij}^3.
$$

As $U = (VV^* - zI_{n\times n})^{-1}$, this induces $\|U\| \le \frac{1}{v}, |U_{ii}| \le \frac{1}{v}$. Define $\widetilde{U} = (V^*V - zI_{N\times N})^{-1}$. We also have $\|\widetilde{U}\| \le \frac{1}{v}$. By the facts that $V\widetilde{U} = UV$ and $V^*V\widetilde{U} = V^*UV$, we get

$$
V^*V\widetilde{U} = I_{N\times N} + z\widetilde{U}
$$

and

$$
\left|\left(V^*UV\right)_{jj}\right| = \left|\left(V^*V\widetilde{U}\right)_{jj}\right| \le 1 + \frac{|z|}{v}.
$$

By the Schwartz inequality, we have

$$\left|(UV)_{ij}\right| \le \left(U^*VV^*U^*\right)_{ii}^{1/2} \le \left(\left(1 + \frac{|z|}{v}\right)\frac{1}{v}\right)^{1/2}.$$

It follows by the Cauchy inequality in Lemma A.4 that

$$\sup_{V \in \mathcal{M}_{n,N}} \left|D_{ij}\left(V^*U'\right)_{ij}\right| = O(1)$$

and

$$\sup_{V \in \mathcal{M}_{n,N}} \left|D_{ij}^2\left(V^*U'\right)_{ij}\right| = O(1) \tag{2.9}$$

hold uniformly on any compact set of $\mathbb{C}^+$.

Since $\frac{1}{n}\sum_{i=1}^n |\frac{1}{N}\sum_{j=1}^N \sigma_{ij}^2 - 1| \to 0$ by the assumption in Theorem 1.2,

$$\frac{1}{nN}\left|\sum_{i,j}\left(1 - E|Y_{ij}|^2\right)\right| = \frac{1}{nN}\left|nN - \sum_{i,j}\sigma_{ij}^2\right| \le \frac{1}{n}\sum_i^n\left|1 - \frac{1}{N}\sum_j^N \sigma_{ij}^2\right| \to 0,$$

which, together with (2.8) and (2.9), yields $\widehat{I} = o(1)$.

By the assumption (1.2), without loss of generality, we select a sequence $\eta_n \downarrow 0$ and $\eta_n > 1/\log N$ as $n \to \infty$. Then

$$\widehat{II} \le \frac{C\eta_n}{2nN}\int_0^1 s^{1/2}\sum_{i,j}E|Y_{ij}|^2 \sup_{V \in \mathcal{M}_{n,N}}\left|D_{ij}^2\left(V^*U'\right)_{ij}\right|ds.$$

We also easily get $\frac{1}{nN}\sum_{i,j}E|Y_{ij}|^2 = 1 + o(1)$. Using (2.9) again, we can see

$$\left|Em_n(z) - E\hat{m}_n(z)\right| \le C\eta_n + o(1).$$

As $\eta_n \to 0$, we have

$$Em_n(z) - E\hat{m}_n(z) \to 0$$

for any $z \in \mathbb{C}^+$, which completes the proof of (ii).

Based on steps (i) and (ii), we conclude that

$$F^{S_n} \xrightarrow{w} F_y^{MP}, \quad \text{a.s.}$$

The proof of Theorem 1.2 is complete.

Appendix: Some lemmas

We will list several important lemmas in our proofs. The first one is the Burkholder inequality for a complex martingale difference sequence, which can be found in Burkholder [5].

Lemma A.1 *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $p > 1$,*

$$E\left|\sum X_k\right|^p \le K_p E\left(\sum |X_k|^2\right)^{\frac{p}{2}},$$

where K_p is a constant related to p.

The second one is the well-known Stein equation for independent Gaussian random variables. It should be noted that a similar result holds for independent complex-valued Gaussian random variables.

Lemma A.2 *Let ξ be a Gaussian random variable of zero mean, and $\Phi : \mathbb{R} \to \mathbb{C}$ be a differentiable function with bounded derivative Φ'. Then we have*

$$E\big(\xi \Phi(\xi)\big) = E\big(\xi^2\big) E\big(\Phi'(\xi)\big).$$

The third one is the generalized Stein equation for independent random variables, which can be found in Lytova and Pastur [9, Proposition 3.1].

Lemma A.3 *Let ξ be a random variable such that $E(|\xi|^{p+2}) < \infty$ for a certain nonnegative integer p. Then, for any function $\Phi : \mathbb{R} \to \mathbb{C}$ of the class C^{p+1} with bounded derivatives $\Phi^{(l)}$, $l = 1, 2, \dots, p + 1$, we have*

$$E\big(\xi \Phi(\xi)\big) = \sum_{l=0}^{p} \frac{\kappa_{l+1}}{l!} E\big(\Phi^{(l)}(\xi)\big) + \varepsilon_p,$$

where κ_l is the lth cumulant of ξ, and the remainder term ε_p admits the bound

$$|\varepsilon_p| \le c_p E\{|\xi|^{p+2} \sup\big|\Phi^{(p+1)}(t)\big|, \quad c_p \le \frac{1 + (3 + 2p)^{p+2}}{(p+1)!}.$$

At last, a useful result on the bound of analytic function will be introduced, which is quoted from Markushevich and Silverstein [11, Theorem 14.7].

Lemma A.4 (Cauchy inequality) *Let $f(z)$ be analytic in the simply connected domain D that contains the circle $C_R(z_0) = \{z : |z - z_0| = R\}$. If $|f(z)| \le M$ holds for all points $z \in C_R(z_0)$, then*

$$\big|f^{(n)}(z_0)\big| \le \frac{n!M}{R^n} \quad (n = 1, 2, \dots).$$

Author details
[1]School of Mathematics and Statistics, Wuhan University, Wuhan, China. [2]School of Mathematics and Statistics, Henan University, Kaifeng, China.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Arnold, L.: On the asymptotic distribution of the eigenvalues of random matrices. J. Math. Anal. Appl. **20**, 262–268 (1967)
2. Arnold, L.: On Wigner's semicircle law for the eigenvalues of random matrices. Probab. Theory Relat. Fields **19**, 191–198 (1971)
3. Bai, Z., Silverstein, J.: Spectral Analysis of Large Dimensional Random Matrices. Science Press, Beijing (2010)
4. Bai, Z., Yin, Y.: Convergence to the semicircle law. Ann. Probab. **16**, 863–875 (1988)
5. Burkholder, D.: Distribution function inequalities for martingales. Ann. Probab. **1**, 19–42 (1973)
6. Grenander, U.: Probabilities on Algebraic Structures. Wiley, New York (1963)
7. Grenander, U., Silverstein, J.: Spectral analysis of networks with random topologies. SIAM J. Appl. Math. **32**, 499–519 (1977)
8. Jonsson, D.: Some limit theorems for the eigenvalues of a sample covariance matrix. J. Multivar. Anal. **12**, 1–38 (1982)
9. Lytova, A., Pastur, L.: Central limit theorem for linear eigenvalue statistics of random matrices with independent entrie. Ann. Probab. **37**, 1778–1840 (2009)
10. Marčenko, V., Pastur, L.: Distribution for some sets of random matrices. Math. USSR Sb. **1**, 457–483 (1967)
11. Markushevich, A., Silverstein, R.: Theorem of Functions of a Complex Variable. AMS Chelsea Series. AMS, Providence (2005)
12. Pastur, L.: On the spectrum of random matrices. Theor. Math. Phys. **10**, 67–74 (1972)
13. Silverstein, J.: Strong convergence of the limiting distribution of the eigrnvalues of large dimensional random matrices. J. Multivar. Anal. **55**, 331–339 (1995)
14. Silverstein, J., Bai, Z.: On the empirical distribution of eigenvalues of a class of large dimensional random matrices. J. Multivar. Anal. **54**, 175–192 (1995)
15. Wigner, E.: Characteristic vectors bordered matrices with infinite dimensions. Ann. Math. **62**, 548–564 (1955)
16. Wigner, E.: On the distributions of the roots of certain symmetric matrices. Ann. Math. **67**, 325–327 (1958)
17. Yin, Y.: Limiting spectral distribution for a class of random matrices. J. Multivar. Anal. **20**, 50–68 (1986)
18. Yin, Y., Krishnaiahm, P.: A limit theorem for the eigenvalues of product of two random matrices. J. Multivar. Anal. **13**, 489–507 (1983)