Ann Inst Stat Math (2016) 68:765–785
DOI 10.1007/s10463-015-0514-0

Convergence of empirical spectral distributions of large dimensional quaternion sample covariance matrices

Huiqin Li · Zhi Dong Bai · Jiang Hu

Received: 4 June 2014 / Revised: 27 December 2014 / Published online: 22 March 2015

Abstract In this paper, we establish the limit of empirical spectral distributions of quaternion sample covariance matrices. Motivated by Bai and Silverstein (Spectral analysis of large dimensional random matrices, Springer, New York, 2010) and Marčenko and Pastur (Matematicheskii Sb, 114:507–536, 1967), we can extend the results of the real or complex sample covariance matrix to the quaternion case. Suppose $\mathbf{X}_n = (x_{jk}^{(n)})_{p\times n}$ is a quaternion random matrix. For each n, the entries $\{x_{ij}^{(n)}\}$ are independent random quaternion variables with a common mean μ and variance $\sigma^2 > 0$. It is shown that the empirical spectral distribution of the quaternion sample covariance matrix $\mathbf{S}_n = n^{-1}\mathbf{X}_n\mathbf{X}_n^*$ converges to the Marčenko–Pastur law as $p \to \infty$, $n \to \infty$ and $p/n \to y \in (0, +\infty)$.

Keywords Empirical spectral distribution · LSD · Quaternion matrices · Sample covariance matrix

---

Z. D. Bai was partially supported by CNSF 11171057, the Fundamental Research Funds for the Central Universities, and PCSIRT; J. Hu was partially supported by a grant CNSF 11301063.

H. Li · Z. D. Bai (✉) · J. Hu
KLASMOE and School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, People's Republic of China
e-mail: baizd@nenu.edu.cn

H. Li
e-mail: lihq118@nenu.edu.cn

J. Hu
e-mail: huj156@nenu.edu.cn

1 Introduction

In 1843, Hamilton described the hyper-complex number of rank 4, to which he gave the name of quaternion (see Kuipers 1999). Research on the quaternion matrices can be traced back to Wolf (1936). After a long blank period, people gradually discovered that quaternions and quaternion matrices play important roles in quantum physics, robot technology and artificial satellite attitude control, among other applications, see Adler (1995) and Finkelstein et al. (1962). Consequently, studies on quaternions have attracted considerable attention in recent years, see So et al. (1994), Zhang (1995), Kanzieper (2002), Akemann (2005), and Akemann and Phillips (2013), among others. In the following, we introduce the quaternion notation. A quaternion can be represented as a 2×2 complex matrix

$$x = a \cdot \mathbf{e} + b \cdot \mathbf{i} + c \cdot \mathbf{j} + d \cdot \mathbf{k} = \begin{pmatrix} a + bi & c + di \\ -c + di & a - bi \end{pmatrix}, \quad a, b, c, d \in \mathbb{R} \tag{1}$$

where i denotes the imaginary unit and the quaternion units can be represented as

$$\mathbf{e} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

The conjugate of x is defined as

$$\bar{x} = a \cdot \mathbf{e} - b \cdot \mathbf{i} - c \cdot \mathbf{j} - d \cdot \mathbf{k} = \begin{pmatrix} a - bi & -c - di \\ c - di & a + bi \end{pmatrix}$$

and its norm as

$$\|x\| = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

More details can be found in Kuipers (1999), Zhang (1997), and Mehta (2004). Using the matrix representation (1) of quaternions, an $n \times n$ quaternion matrix $\mathbf{X}$ can be rewritten as a $2n \times 2n$ complex matrix $\psi(\mathbf{X})$, and so we can deal with quaternion matrices as complex matrices for convenience. Denote $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^*$ and $\psi(\mathbf{S}) = \frac{1}{n}\psi(\mathbf{X})\psi(\mathbf{X})^*$. It is known (see Zhang 1997) that the multiplicities of all the eigenvalues (obviously they are all real) of $\psi(\mathbf{S})$ are even. Taking one from each of the n pairs of eigenvalues of $\psi(\mathbf{S})$, the n values are defined to be the eigenvalues of $\mathbf{S}$.

In addition, wide application of computer science has increased a thousand fold in terms of computing speed and storage capability in the recent decades. Due to the failure of the applications of many classical conclusions, we need a new theory to analyze very large data sets with high dimensions. Luckily, the theory of random matrices (RMT) might be a possible route for dealing with these problems. The sample covariance matrix is one of the most important random matrices in RMT, which can be traced back to Wishart (1928). Marčenko and Pastur (1967) proved that the empirical spectral distribution (ESD) of a large dimensional complex sample covariance matrix tends to the Marčenko–Pastur (M–P) law. Since then, several successive studies on large dimensional complex or real sample covariance matrices have been completed. Here, the readers are referred to three books Anderson et al. (2010), Bai and Silverstein (2010), and Mehta (2004) for more details.

Under the normality assumption, there are three classic random matrix models: Gaussian orthogonal ensemble (GOE), for which all entries of the matrix are real normal random variables, Gaussian unitary ensemble (GUE), for which all entries of the matrix are complex normal random variables, and Gaussian symplectic ensemble (GSE), for which all entries of the matrix are normal quaternion random variables. Benefiting from the density function of the ensemble and the joint density of the eigenvalues, the results have gotten their own style. If we remove the normality assumption, the corresponding first two models have already had satisfactory results. For quaternion matrices, there are only a few references (see Yin and Bai 2014; Yin et al. 2013, 2014).

In this paper, we prove that the ESD of the quaternion sample covariance matrix also converges to the M–P law. However, due to the multiplication of quaternions is not commutative, when the entries of $\mathbf{X}_n$ are quaternion random variables, few works on the spectral properties are found in the literature unless the random variables are normality distributed, because in this case the joint density of the eigenvalues is available. Thanks to the tool provided by Yin et al. (2013), it makes the quaternion case possible. For the proof of this result, we first introduce two definitions about ESD and Stieltjes transform. Let $\mathbf{A}$ be a $p \times p$ Hermitian matrix and denote its eigenvalues by $s_j, j = 1, 2, \ldots, p$. The ESD of $\mathbf{A}$ is defined by

$$F^{\mathbf{A}}(x) = \frac{1}{p}\sum_{j=1}^{p} I(s_j \leq x),$$

where $I(D)$ is the indicator function of an event D and the Stieltjes transform of $F^{\mathbf{A}}(x)$ is given by

$$m(z) = \int_{-\infty}^{+\infty} \frac{1}{x-z}\,\mathrm{d}F^{\mathbf{A}}(x),$$

where $z = u + \upsilon i \in \mathbb{C}^+$. Let $g(x)$ and $\mathbf{m}_g(x)$ denote the density function and the Stieltjes transform of the M–P law, which are

$$g(x) = \begin{cases} \frac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)}, & a \leq x \leq b; \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

and

$$m_g(z) = \frac{\sigma^2(1-y) - z + \sqrt{(z - \sigma^2 - y\sigma^2)^2 - 4y\sigma^2}}{2yz\sigma^2}, \tag{3}$$

respectively, where $a = \sigma^2(1-\sqrt{y})^2$, $b = \sigma^2(1+\sqrt{y})^2$. Here, the constant y is the limit of dimension p to sample size n ratio and σ^2 is the scale parameter. If $y > 1$, $G(x)$, the distribution function of $g(x)$, has a point mass $1 - 1/y$ at the origin.

Now, our main theorem can be described as follows.

Theorem 1 *Let* $\mathbf{X}_n = \left(x_{jk}^{(n)}\right)$, $j = 1, \ldots, p, k = 1, \ldots, n$. *Suppose for each* n, $\left\{x_{jk}^{(n)}\right\}$ *are independent quaternion random variables with a common mean* μ *and variance* σ^2. *Assume that* $y_n = p/n \to y \in (0, \infty)$ *and for any constant* $\eta > 0$,

$$\frac{1}{np}\sum_{jk} E\|x_{jk}^{(n)}\|^2 I\left(\|x_{jk}^{(n)}\| > \eta\sqrt{n}\right) \to 0. \tag{4}$$

Then, with probability one, the ESD of the sample covariance matrix $\mathbf{S}_n = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*$ *converges to the M–P law in distribution which has density function* (2) *and a point mass* $1 - 1/y$ *at the origin when* $y > 1$. *Here, superscript* * *stands for the complex conjugate transpose.*

Remark 2 Without loss of generality, in the proof of Theorem 1, we assume that $\sigma^2 = 1$. Furthermore, one can see that removing the common mean of the entries of $\mathbf{X}_n$ does not alter the LSD of sample covariance matrices. In fact, let

$$\mathbf{T}_n = \frac{1}{n}(\mathbf{X}_n - E\mathbf{X}_n)(\mathbf{X}_n - \mathrm{E}\mathbf{X}_n)^*.$$

By Lemma 17, we have, for all large p,

$$\|F^{\mathbf{S}_n} - F^{\mathbf{T}_n}\|_{KS} \le \frac{1}{2p}\mathrm{rank}(E\mathbf{X}_n) \le \frac{1}{p} \to 0$$

where $\|f\|_{KS} = \sup_x |f(x)|$. Consequently, we assume that $\mu = 0$.

The paper is organized as follows. In Sect. 2, the structure of the inverse of some matrices about quaternions is established which is the key tool of proving Theorem 1. Section 3 demonstrates the proof of the main theorem by two steps and in Sect. 4, we outline some auxiliary lemmas that can be used in last section.

2 Preliminaries

We shall use Lemma 2.5 of Yin et al. (2013) to prove our main result in next section. To keep this work self-contained, the lemma is now stated as follows.

Definition 3 A matrix is of Type-I, if it has the following structure:

$$\begin{pmatrix} t_1 & 0 & a_{12} & b_{12} & \cdots & a_{1n} & b_{1n} \\ 0 & t_1 & c_{12} & d_{12} & \cdots & c_{1n} & d_{1n} \\ d_{12} & -b_{12} & t_2 & 0 & \cdots & a_{2n} & b_{2n} \\ -c_{12} & a_{12} & 0 & t_2 & \cdots & c_{2n} & d_{2n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ d_{1n} & -b_{1n} & d_{2n} & -b_{2n} & \cdots & t_n & 0 \\ -c_{1n} & a_{1n} & -c_{2n} & a_{2n} & \ldots & 0 & t_n \end{pmatrix}.$$

Here, all the entries are complex.

Definition 4 A matrix is of Type-II, if it has the following structure:

$$\begin{pmatrix} t_1 & 0 & a_{12}+c_{12}i & b_{12}+d_{12}i & \cdots & a_{1n}+c_{1n}i & b_{1n}+d_{1n}i \\ 0 & t_1 & -\bar{b}_{12}-\bar{d}_{12}i & \bar{a}_{12}+\bar{c}_{12}i & \cdots & -\bar{b}_{1n}-\bar{d}_{1n}i & \bar{a}_{1n}+\bar{c}_{1n}i \\ \bar{a}_{12}+\bar{c}_{12}i & -b_{12}-d_{12}i & t_2 & 0 & \cdots & a_{2n}+c_{2n}i & b_{2n}+d_{2n}i \\ \bar{b}_{12}+\bar{d}_{12}i & a_{12}+c_{12}i & 0 & t_2 & \cdots & -\bar{b}_{2n}-\bar{d}_{2n}i & \bar{a}_{2n}+\bar{c}_{2n}i \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{a}_{1n}+\bar{c}_{1n}i & -b_{1n}-d_{1n}i & \bar{a}_{2n}+\bar{c}_{2n}i & -b_{2n}-d_{2n}i & \cdots & t_n & 0 \\ \bar{b}_{1n}+\bar{d}_{1n}i & a_{1n}+c_{1n}i & \bar{b}_{2n}+\bar{d}_{2n}i & a_{2n}+c_{2n}i & \ldots & 0 & t_n \end{pmatrix}.$$

Here, $i=\sqrt{-1}$ denotes the usual imaginary unit and all the other entries are complex numbers.

Definition 5 A matrix is of Type-III, if it has the following structure:

$$\begin{pmatrix} t_1 & 0 & a_{12} & b_{12} & \cdots & a_{1n} & b_{1n} \\ 0 & t_1 & -\bar{b}_{12} & \bar{a}_{12} & \cdots & -\bar{b}_{1n} & \bar{a}_{1n} \\ \bar{a}_{12} & -b_{12} & t_2 & 0 & \cdots & a_{2n} & b_{2n} \\ \bar{b}_{12} & a_{12} & 0 & t_2 & \cdots & -\bar{b}_{2n} & \bar{a}_{2n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{a}_{1n} & -b_{1n} & \bar{a}_{2n} & -b_{2n} & \cdots & t_n & 0 \\ \bar{b}_{1n} & a_{1n} & \bar{b}_{2n} & a_{2n} & \ldots & 0 & t_n \end{pmatrix}.$$

Here, all the entries are complex.

Lemma 6 *For all $n \geq 1$, if a complex matrix Ω_n is invertible and of Type-II, then Ω_n^{-1} is a Type-I matrix.*

The following corollary is immediate.

Corollary 7 *For all $n \geq 1$, if a complex matrix Ω_n is invertible and of Type-III, then Ω_n^{-1} is a Type-I matrix.*

3 Proof of Theorem 1

In this section, we present the proof in two steps. The first one is to truncate, centralize and rescale the random variables $\{x_{ij}^{(n)}\}$, then we may assume the additional conditions which are given in Remark 12. The other is the proof of the Theorem 1 under the additional conditions. Throughout the remainder of this paper, a local constant C may take different value at different place.

3.1 Truncation, centralization and rescaling

3.1.1 Truncation

Note that, condition (4) is equivalent to: for any $\eta > 0$,

$$\lim_{n\to\infty} \frac{1}{\eta^2 np} \sum_{jk} E\|x_{jk}^{(n)}\|^2 I\left(\|x_{jk}^{(n)}\| > \eta\sqrt{n}\right) = 0. \tag{5}$$

Applying Lemma 15, one can select a sequence $\eta_n \downarrow 0$ such that (5) remains true when η is replaced by η_n.

Lemma 8 *Suppose that the assumptions of Theorem* refth:1 *hold. Truncate the variables* $x_{jk}^{(n)}$ *at* $\eta_n\sqrt{n}$*, and denote the resulting variables by* $\widehat{x}_{jk}^{(n)}$*, i.e.,* $\widehat{x}_{jk}^{(n)} = x_{jk}^{(n)} I(\|x_{jk}^{(n)}\| \le \eta_n\sqrt{n})$*. Also denote*

$$\widehat{\mathbf{X}}_n = \left(\widehat{x}_{jk}^{(n)}\right) \quad \textit{and} \quad \widehat{\mathbf{S}}_n = \frac{1}{n}\widehat{\mathbf{X}}_n\widehat{\mathbf{X}}_n^*.$$

Then, with probability 1,

$$\|F^{\mathbf{S}_n} - F^{\widehat{\mathbf{S}}_n}\|_{KS} \to 0.$$

Proof Using Lemma 17, one has

$$\begin{aligned}\|F^{\mathbf{S}_n} - F^{\widehat{\mathbf{S}}_n}\|_{KS} &\le \frac{1}{2p}\text{rank}\left(\frac{1}{\sqrt{n}}\mathbf{X}_n - \frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_n\right) \\ &\le \frac{1}{2p}\sum_{jk} I\left(\|x_{jk}^{(n)}\| > \eta_n\sqrt{n}\right).\end{aligned} \tag{6}$$

Taking condition (5) into consideration, we get

$$E\left(\frac{1}{2p}\sum_{jk} I\left(\|x_{jk}^{(n)}\| > \eta_n\sqrt{n}\right)\right) \le \frac{1}{2\eta_n^2 np}\sum_{jk} E\|x_{jk}^{(n)}\|^2 I\left(\|x_{jk}^{(n)}\| > \eta_n\sqrt{n}\right) = o(1)$$

and

$$\operatorname{Var}\left(\frac{1}{2p}\sum_{jk} I\left(\|x_{jk}^{(n)}\| > \eta_n\sqrt{n}\right)\right)$$
$$\leq \frac{1}{4\eta_n^2 p^2 n}\sum_{jk} E\|x_{jk}^{(n)}\|^2 I\left(\|x_{jk}^{(n)}\| > \eta_n\sqrt{n}\right) = o\left(\frac{1}{p}\right).$$

Then, by Bernstein's inequality (see Lemma 18), for all small $\varepsilon > 0$ and large n, we obtain

$$\mathrm{P}\left(\frac{1}{2p}\sum_{jk} I\left(\|x_{jk}^{(n)}\| > \eta_n\sqrt{n}\right) \geq \varepsilon\right) \leq 2\mathrm{e}^{-\varepsilon p/2}$$

which is summable. Combining (6), the above inequality with the Borel–Cantelli lemma, it follows that

$$\|F^{\mathbf{S}_n} - F^{\widehat{\mathbf{S}}_n}\|_{KS} \xrightarrow{\text{a.s.}} 0.$$

This completes the proof of the lemma. □

3.1.2 Centralization

Lemma 9 *Suppose that the assumptions of Lemma 8 hold. Denote*

$$\widetilde{x}_{jk}^{(n)} = \widehat{x}_{jk}^{(n)} - E\widehat{x}_{jk}^{(n)},\ \widetilde{\mathbf{X}}_n = (\widetilde{x}_{jk}^{(n)}) \quad \textit{and} \quad \widetilde{\mathbf{S}}_n = \frac{1}{n}\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^*.$$

Then, we obtain

$$L(F^{\widehat{\mathbf{S}}_n}, F^{\widetilde{\mathbf{S}}_n}) = o(1),$$

where $L(\cdot,\cdot)$ denotes the Lévy distance.

Proof Using Lemma 16 and condition (5), we have

$$\begin{aligned}
L^4(F^{\widehat{\mathbf{S}}_n}, F^{\widetilde{\mathbf{S}}_n}) &\leq \frac{1}{2p^2}\left(\operatorname{tr}(\widehat{\mathbf{S}}_n + \widetilde{\mathbf{S}}_n)\right)\left(\operatorname{tr}\left(\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_n - \frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n\right)\left(\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_n - \frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n\right)^*\right)\\
&= \frac{1}{2n^2p^2}\left(\sum_{jk}\left(\|\widehat{x}_{jk}^{(n)}\|^2 + \|\widehat{x}_{jk}^{(n)} - E\widehat{x}_{jk}^{(n)}\|^2\right)\right)\left(\sum_{jk}\|E\widehat{x}_{jk}^{(n)}\|^2\right)\\
&= \left(\frac{1}{np}\sum_{jk}\left(\|\widehat{x}_{jk}^{(n)}\|^2 + \|\widehat{x}_{jk}^{(n)} - E\widehat{x}_{jk}^{(n)}\|^2\right)\right)\left(\frac{1}{2np}\sum_{jk}\|E\widehat{x}_{jk}^{(n)}\|^2\right).
\end{aligned} \tag{7}$$

To complete the proof of this lemma, we need to show that the first parentheses of the right-hand side of (7) is almost surely bounded. Applying Lemma 22, one has

$$E\left|\frac{1}{np}\sum_{jk}\left(\|\widehat{x}_{jk}^{(n)}\|^2 - E\|\widehat{x}_{jk}^{(n)}\|^2\right)\right|^4 \le \frac{C}{n^4p^4}\left[\sum_{jk}E\|\widehat{x}_{jk}^{(n)}\|^8 + \left(\sum_{j,k}E\|\widehat{x}_{jk}^{(n)}\|^4\right)^2\right]$$
$$\le Cn^{-2}\left(\eta_n^6 n^{-1}y_n^{-3} + \eta_n^4 y_n^{-2}\right).$$

This indicates by the Borel–Cantelli lemma

$$\frac{1}{np}\sum_{jk}\left(\|\widehat{x}_{jk}^{(n)}\|^2 - E\|\widehat{x}_{jk}^{(n)}\|^2\right) \xrightarrow{\text{a.s.}} 0.$$

Moreover, we can similarly obtain

$$\frac{1}{np}\sum_{jk}\left(\|\widehat{x}_{jk}^{(n)} - E\widehat{x}_{jk}^{(n)}\|^2 - E\|\widehat{x}_{jk}^{(n)} - E\widehat{x}_{jk}^{(n)}\|^2\right) \xrightarrow{\text{a.s.}} 0. \tag{8}$$

Now, turning to (7), for all large n,

$$\begin{aligned}
&L^4(F^{\widehat{\mathbf{S}}_n}, F^{\widetilde{\mathbf{S}}_n})\\
&\le \left(\frac{1}{np}\sum_{jk}\left(E\|\widehat{x}_{jk}^{(n)}\|^2 + E\|\widehat{x}_{jk}^{(n)} - E\widehat{x}_{jk}^{(n)}\|^2\right) + o_{\text{a.s.}}(1)\right)\left(\frac{1}{2np}\sum_{jk}\|E\widehat{x}_{jk}^{(n)}\|^2\right)\\
&\le \frac{C}{np}\sum_{jk}\|E\widehat{x}_{jk}^{(n)}\|^2\\
&\le \frac{C}{np}\sum_{jk}E\|x_{jk}^{(n)}\|^2 I\left(\|x_{jk}^{(n)}\| > \eta_n\sqrt{n}\right) \to 0.
\end{aligned}$$

The proof of the lemma is complete. □

3.1.3 Rescaling

Define

$$\widetilde{\sigma}_{jk}^2 = E\|\widetilde{x}_{jk}^{(n)}\|^2,\ \ \xi_{jk} = \begin{cases}\zeta_{jk},\ \widetilde{\sigma}_{jk}^2 < 1/2\\ \widetilde{x}_{jk}^{(n)},\ \widetilde{\sigma}_{jk}^2 \ge 1/2\end{cases},\ \ \mathbf{\Lambda} = \frac{1}{\sqrt{n}}(\xi_{jk}),\ \ \sigma_{jk}^2 = E\|\xi_{jk}\|^2,$$

where ζ_{jk} is a bounded quaternion random variable with $E\zeta_{jk} = 0$, $\text{Var}\zeta_{jk} = 1$ and independent with all other variables.

Lemma 10 *Write*

$$\breve{x}_{jk}^{(n)} = \sigma_{jk}^{-1}\xi_{jk}, \quad \breve{\mathbf{X}}_n = \left(\breve{x}_{jk}^{(n)}\right), \quad \text{and} \quad \breve{\mathbf{S}}_n = \frac{1}{n}\breve{\mathbf{X}}_n\breve{\mathbf{X}}_n^*.$$

Under the conditions assumed in Lemma 9, we have

$$L\left(F^{\breve{\mathbf{S}}_n}, F^{\widetilde{\mathbf{S}}_n}\right) = o(1).$$

Proof (a): Our first goal is to show that

$$L\left(F^{\widetilde{\mathbf{S}}_n}, F^{\mathbf{\Lambda\Lambda}^*}\right) \xrightarrow{\text{a.s.}} 0.$$

Let $\mathcal{E}_n$ be the set of pairs $(j,k) : \widetilde{\sigma}_{jk}^2 < \frac{1}{2}$ and $N_n = \sum_{(j,k)\in\mathcal{E}_n} I\left(\widetilde{\sigma}_{jk}^2 < 1/2\right)$ be the number of such pairs. Due to $\frac{1}{np}\sum_{jk}\widetilde{\sigma}_{jk}^2 \to 1$, we conclude that $N_n = o(np)$. Owing to Lemma 16 and (8), we get

$$\begin{aligned}
L^4(F^{\widetilde{\mathbf{S}}_n}, F^{\mathbf{\Lambda\Lambda}^*}) &\leq \frac{1}{2p^2}(\mathrm{tr}(\widetilde{\mathbf{S}}_n + \mathbf{\Lambda\Lambda}))\left(\mathrm{tr}\left(\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n - \mathbf{\Lambda}\right)\left(\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n - \mathbf{\Lambda}\right)^*\right) \\
&= \frac{1}{2n^2p^2}\left(\sum_{jk}\left(\|\widetilde{x}_{jk}^{(n)}\|^2 + \|\xi_{jk}\|^2\right)\right)\left(\sum_{jk}\|\xi_{jk} - \widetilde{x}_{jk}^{(n)}\|^2\right) \\
&= \frac{1}{2n^2p^2}\left(\sum_{jk} E\left(\|\widetilde{x}_{jk}^{(n)}\|^2 + \|\xi_{jk}\|^2\right) + o_{\text{a.s.}}(1)\right)\left(\sum_{jk}\|\xi_{jk} - \widetilde{x}_{jk}^{(n)}\|^2\right) \\
&\leq \frac{C}{np}\sum_{jk}\|\xi_{jk} - \widetilde{x}_{jk}^{(n)}\|^2 := \frac{C}{np}\sum_{h=1}^{K} u_h \qquad (9)
\end{aligned}$$

where $K = N_n$ and $u_h = \|\xi_{jk} - \widetilde{x}_{jk}^{(n)}\|^2$. Using the fact that for all $l \geq 1$, $l! \geq (l/3)^l$, we have

$$\begin{aligned}
E\left(\frac{1}{np}\sum_{h=1}^{K} u_h\right)^m &= \frac{1}{n^m p^m}\sum_{m_1+\cdots+m_K=m}\frac{m!}{m_1!\ldots m_K!}Eu_1^{m_1}\ldots Eu_K^{m_K} \\
&\leq \frac{1}{n^m p^m}\sum_{l=1}^{m}\sum_{\substack{m_1+\cdots+m_l=m \\ m_t\geq 1}}\frac{m!}{l!m_1!\ldots m_l!}\prod_{t=1}^{l}\left(\sum_{h=1}^{K} Eu_h^{m_t}\right) \\
&\leq C\sum_{l=1}^{m} n^{-m}p^{-m}l^m(l!)^{-1}(2\eta_n^2 n)^{m-l}2^l K^l \\
&\leq C\sum_{l=1}^{m}\left(\frac{6K}{np}\right)^l\left(\frac{2\eta_n^2 l}{p}\right)^{m-l} \leq C\left(\frac{6K}{np} + \frac{2\eta_n^2 m}{p}\right)^m.
\end{aligned}$$

By selecting $m = [\log p]$ that implies $\frac{2\eta_n^2 m}{p} \to 0$, and noticing $\frac{6K}{np} \to 0$, one obtains for any fixed $t, \varepsilon > 0$,

$$E\left(\frac{1}{\varepsilon np}\sum_{h=1}^{K} u_h\right)^m \leq o\left(p^{-t}\right).$$

From the inequality above with $t = 2$ and (9), it follows that

$$L\left(F^{\widetilde{\mathbf{S}}_n}, F^{\mathbf{\Lambda\Lambda}^*}\right) \xrightarrow{\text{a.s.}} 0.$$

(b): Our next goal is to show that

$$L\left(F^{\breve{\mathbf{S}}_n}, F^{\mathbf{\Lambda\Lambda}^*}\right) \xrightarrow{\text{a.s.}} 0.$$

Applying Lemma 16, we have

$$\begin{aligned}
L^4(F^{\breve{\mathbf{S}}_n}, F^{\mathbf{\Lambda\Lambda}^*}) &\leq \frac{1}{2p^2}\left(\operatorname{tr}\left(\breve{\mathbf{S}}_n + \mathbf{\Lambda\Lambda}\right)\right)\left(\operatorname{tr}\left(\frac{1}{\sqrt{n}}\breve{\mathbf{X}}_n - \mathbf{\Lambda}\right)\left(\frac{1}{\sqrt{n}}\breve{\mathbf{X}}_n - \mathbf{\Lambda}\right)^*\right) \\
&= \frac{1}{2n^2p^2}\left(\sum_{jk}\left(\|\breve{x}_{jk}^{(n)}\|^2 + \|\xi_{jk}\|^2\right)\right)\left(\sum_{jk}\|\xi_{jk} - \breve{x}_{jk}^{(n)}\|^2\right) \\
&= \frac{1}{2n^2p^2}\left(\sum_{jk}\left(1 + \sigma_{jk}^{-2}\right)E\|\xi_{jk}\|^2 + o_{\text{a.s.}}(1)\right) \\
&\quad \times \left(\sum_{jk}\left(1 - \sigma_{jk}^{-1}\right)^2\|\xi_{jk}\|^2\right) \\
&\leq \frac{C}{np}\sum_{jk}\left(1 - \sigma_{jk}^{-1}\right)^2\|\xi_{jk}\|^2.
\end{aligned}$$

Using the fact

$$\begin{aligned}
&E\left(\frac{C}{np}\sum_{jk}\left(1 - \sigma_{jk}^{-1}\right)^2\|\xi_{jk}\|^2\right) = \frac{C}{np}\sum_{jk}\left(1 - \sigma_{jk}\right)^2 \leq \frac{C}{np}\sum_{jk}\left(1 - \sigma_{jk}^2\right) \\
&\leq \frac{C\eta_n^2}{\eta_n^2 np}\sum_{(j,k)\notin\mathcal{E}_n}\left[E\|x_{jk}^{(n)}\|^2 I\left(\|x_{jk}^{(n)}\| \geq \eta_n\sqrt{n}\right) + \left(E\|x_{jk}^{(n)}\| I\left(\|x_{jk}^{(n)}\| \geq \eta_n\sqrt{n}\right)\right)^2\right] \\
&\to 0
\end{aligned}$$

and Lemma 22, one gets

$$E\left|\frac{C}{np}\sum_{jk}\left(1-\sigma_{jk}^{-1}\right)^2\left(\|\xi_{jk}\|^2-E\|\xi_{jk}\|^2\right)\right|^4$$
$$\leq\frac{C}{n^4p^4}\left[\sum_{j,k}E\|x_{jk}^{(n)}\|^8I(\|x_{jk}^{(n)}\|\leq\eta_n\sqrt{n})+\left(\sum_{j,k}E\|x_{jk}^{(n)}\|^4I\left(\|x_{jk}^{(n)}\|\leq\eta_n\sqrt{n}\right)\right)^2\right]$$
$$\leq Cn^{-2}\left[n^{-1}\eta_n^6y_n^{-3}+\eta_n^4y_n^{-2}\right]$$

which is summable. Together with the Borel–Cantelli lemma, it follows that

$$L\left(F^{\check{\mathbf{S}}_n},F^{\mathbf{\Lambda\Lambda}^*}\right)\xrightarrow{\text{a.s.}}0.$$

(c): Finally, from (a) and (b), we can easily get the lemma. □

Combining the results of Lemmas 8, 9, and 10, we have the following remarks.

Remark 11 For brevity, we shall drop the superscript (n) from the variables. Also the truncated and renormalized variables are still denoted by x_{jk}.

Remark 12 Under the conditions assumed in Theorem 1, we can further assume that
(1) $\|x_{jk}\|\leq\eta_n\sqrt{n}$,
(2) $E(x_{jk})=0$ and $\mathrm{Var}(x_{jk})=1$.

3.2 Completion of the proof

Denote

$$m_n(z)=\frac{1}{2p}\mathrm{tr}\left(\mathbf{S}_n-z\mathbf{I}_{2p}\right)^{-1},\tag{10}$$

where $z=u+\upsilon i\in\mathbb{C}^+$.

3.2.1 Random part

First, we should show that

$$m_n(z)-Em_n(z)\xrightarrow{\text{a.s.}}0.\tag{11}$$

Let π_j denote the jth column of $\mathbf{X}_n$, $\mathbf{S}_n^k=\mathbf{S}_n-\frac{1}{n}\boldsymbol{\pi}_k\boldsymbol{\pi}_k^*$ and $E_k(\cdot)$ denote the conditional expectation given $\{\boldsymbol{\pi}_{k+1},\boldsymbol{\pi}_{k+2},\ldots,\boldsymbol{\pi}_{2n}\}$. Then,

$$m_n(z)-Em_n(z)=\frac{1}{2p}\sum_{k=1}^{2n}\gamma_k,$$

where

$$\gamma_k = E_{k-1}\mathrm{tr}\big(\mathbf{S}_n - z\mathbf{I}_{2p}\big)^{-1} - E_k\mathrm{tr}\big(\mathbf{S}_n - z\mathbf{I}_{2p}\big)^{-1}$$
$$= (E_{k-1} - E_k)\left[\mathrm{tr}\big(\mathbf{S}_n - z\mathbf{I}_{2p}\big)^{-1} - \mathrm{tr}\Big(\mathbf{S}_n^k - z\mathbf{I}_{2p}\Big)^{-1}\right].$$

1. When $k = 2t - 1(t = 1, 2, \ldots, n)$, due to $(k + 1)$th column is a function of the kth column, we obtain

$$\gamma_k = \mathrm{E}_{k-1}\mathrm{tr}\big(\mathbf{S}_n - z\mathbf{I}_{2p}\big)^{-1} - E_k\mathrm{tr}\big(\mathbf{S}_n - z\mathbf{I}_{2p}\big)^{-1} = 0.$$

2. When $k = 2t\,(t = 0, 1, \ldots, n)$, together with the formula

$$(\mathbf{A} + \boldsymbol{\alpha}\boldsymbol{\beta}^*)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\boldsymbol{\alpha}\boldsymbol{\beta}^*\mathbf{A}^{-1}}{1 + \boldsymbol{\beta}^*\mathbf{A}^{-1}\boldsymbol{\alpha}},$$

one finds

$$\gamma_k = (E_{k-1} - E_k)\left[\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_{2p})^{-1} - \mathrm{tr}\Big(\mathbf{S}_n^k - z\mathbf{I}_{2p}\Big)^{-1}\right]$$
$$= (E_{k-1} - E_k)\frac{\frac{1}{n}\boldsymbol{\pi}_k^*\big(\mathbf{S}_n^k - z\mathbf{I}_{2p}\big)^{-2}\boldsymbol{\pi}_k}{1 + \frac{1}{n}\boldsymbol{\pi}_k^*\big(\mathbf{S}_n^k - z\mathbf{I}_{2p}\big)^{-1}\boldsymbol{\pi}_k}.$$

Since

$$\left|\frac{\frac{1}{n}\boldsymbol{\pi}_k^*\big(\mathbf{S}_n^k - z\mathbf{I}_{2p}\big)^{-2}\boldsymbol{\pi}_k}{1 + \frac{1}{n}\boldsymbol{\pi}_k^*\big(\mathbf{S}_n^k - z\mathbf{I}_{2p}\big)^{-1}\boldsymbol{\pi}_k}\right|$$
$$\le \frac{\frac{1}{n}\boldsymbol{\pi}_k^*\Big(\big(\mathbf{S}_n^k - u\mathbf{I}_{2p}\big)^2 + \upsilon^2\mathbf{I}_{2p}\Big)^{-1}\boldsymbol{\pi}_k}{\Im\Big(1 + \frac{1}{n}\boldsymbol{\pi}_k^*(\mathbf{S}_n^k - z\mathbf{I}_{2p})^{-1}\boldsymbol{\pi}_k\Big)}$$
$$= \frac{1}{\upsilon},$$

we can easily get

$$|\gamma_k| \le \frac{2}{\upsilon}.$$

Using Lemma 21, it follows that

$$E|m_n(z) - Em_n(z)|^4 \le \frac{K_4}{(2p)^4}E\left(\sum_{k=1}^{2n}|\gamma_k|^2\right)^2 \le \frac{4K_4n^2}{p^4\upsilon^4} = O\left(n^{-2}\right).$$

Combining the Borel–Cantelli lemma with the Chebyshev inequality, we conclude

$$m_n(z) - Em_n(z) \xrightarrow{\text{a.s.}} 0.$$

3.2.2 Mean convergence

When $\sigma^2 = 1$, (3) turns into

$$m(z) = \frac{1 - y - z + \sqrt{(1 - z - y)^2 - 4yz}}{2yz}. \tag{12}$$

Next, we show that

$$Em_n(z) \to m(z).$$

By Lemma 20 and (10), one has

$$m_n(z) = \frac{1}{2p} \sum_{k=1}^{p} \operatorname{tr} \left(\frac{1}{n} \boldsymbol{\phi}_k' \bar{\boldsymbol{\phi}}_k - z\mathbf{I}_2 - \frac{1}{n^2} \boldsymbol{\phi}_k' \mathbf{X}_{nk}^* \left(\frac{1}{n} \mathbf{X}_{nk} \mathbf{X}_{nk}^* - z\mathbf{I}_{2p-2} \right)^{-1} \mathbf{X}_{nk} \bar{\boldsymbol{\phi}}_k \right)^{-1} \tag{13}$$

where $\mathbf{X}_{nk}$ is the matrix resulting from deleting the kth quaternion row of $\mathbf{X}_n$, and $\boldsymbol{\phi}_k'$ is the vector obtained from the kth quaternion row of $\mathbf{X}_n$. Here, superscript $'$ only stands for the transpose and $\boldsymbol{\phi}_k'$ is a $1 \times n$ quaternion matrix. Write

$$\begin{aligned} \boldsymbol{\varepsilon}_k = & \frac{1}{n} \boldsymbol{\phi}_k' \bar{\boldsymbol{\phi}}_k - z\mathbf{I}_2 - \frac{1}{n^2} \boldsymbol{\phi}_k' \mathbf{X}_{nk}^* \left(\frac{1}{n} \mathbf{X}_{nk} \mathbf{X}_{nk}^* - z\mathbf{I}_{2p-2} \right)^{-1} \mathbf{X}_{nk} \bar{\boldsymbol{\phi}}_k \\ & - (1 - z - y_n - y_n z Em_n(z))\mathbf{I}_2 \end{aligned} \tag{14}$$

and

$$\begin{aligned} \delta_n = & -\frac{1}{2p\,(1 - z - y_n - y_n z Em_n(z))} \\ & \times \sum_{k=1}^{p} E\operatorname{tr} \left\{ \boldsymbol{\varepsilon}_k ((1 - z - y_n - y_n z Em_n(z))\, \mathbf{I}_2 + \boldsymbol{\varepsilon}_k)^{-1} \right\} \end{aligned} \tag{15}$$

where $y_n = p/n$. This implies that

$$Em_n(z) = \frac{1}{1 - z - y_n - y_n z Em_n(z)} + \delta_n.$$

Solving $Em_n(z)$ from the equation above, we get

$$Em_n(z) = \frac{1}{2y_n z} \left(1 - z - y_n + y_n z \delta_n \pm \sqrt{(1 - z - y_n - y_n z \delta_n)^2 - 4y_n z} \right).$$

As proved in the Eq. (3.17) of Bai (1993), we can assert that

$$Em_n(z) = \frac{1 - z - y_n + y_n z\delta_n + \sqrt{(1 - z - y_n - y_n z\delta_n)^2 - 4y_n z}}{2y_n z}. \tag{16}$$

Comparing (12) with (16), it suffices to show that

$$\delta_n \to 0.$$

For this purpose, we need the following two lemmas.

Lemma 13 *Under the conditions of Remark 12, for any $z = u + vi$ with $v > 0$ and for any $k = 1, \ldots, p$, we have*

$$|E\mathrm{tr}\boldsymbol{\varepsilon}_k| \to 0. \tag{17}$$

Proof By calculation, we have

$$\begin{aligned}
|E\mathrm{tr}\boldsymbol{\varepsilon}_k| &= \left| -\frac{1}{n^2} E\mathrm{tr}\mathbf{X}_{nk}^* \left(\frac{1}{n}\mathbf{X}_{nk}\mathbf{X}_{nk}^* - z\mathbf{I}_{2p-2}\right)^{-1} \mathbf{X}_{nk} + 2y_n + 2y_n z Em_n(z) \right| \\
&= \left| -\frac{1}{n} E\mathrm{tr}\left(\frac{1}{n}\mathbf{X}_{nk}\mathbf{X}_{nk}^* - z\mathbf{I}_{2p-2}\right)^{-1} \frac{1}{n}\mathbf{X}_{nk}\mathbf{X}_{nk}^* + 2y_n + 2y_n z Em_n(z) \right| \\
&\le \frac{2}{n} + \frac{|z|}{n} \left| E\left[\mathrm{tr}\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^* - z\mathbf{I}_{2p}\right)^{-1} - \mathrm{tr}\left(\frac{1}{n}\mathbf{X}_{nk}\mathbf{X}_{nk}^* - z\mathbf{I}_{2p-2}\right)^{-1}\right]\right| \\
&\le \frac{2}{n} + \frac{2|z|}{nv} \to 0
\end{aligned}$$

where the last inequality has used Lemma 19 twice. Then, the proof is complete. □

Lemma 14 *Under the conditions of Remark 12, for any $z = u + vi$ with $v > 0$ and any $k = 1, \ldots, p$, we have*

$$E|\mathrm{tr}\boldsymbol{\varepsilon}_k^2| \to 0.$$

Proof Write the form of $(\mathbf{S}_n - z\mathbf{I}_{2p})$

$$\begin{pmatrix}
t_1 & 0 & a_{12} & b_{12} & \cdots \\
0 & t_1 & -\bar{b}_{12} & \bar{a}_{12} & \cdots \\
\bar{a}_{12} & -b_{12} & t_2 & 0 & \cdots \\
\bar{b}_{12} & a_{12} & 0 & t_2 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}.$$

By Corollary 7 and (13), we have $\frac{1}{n}\boldsymbol{\phi}_k'\bar{\boldsymbol{\phi}}_k - z\mathbf{I}_2 - \frac{1}{n^2}\boldsymbol{\phi}_k'\mathbf{X}_{nk}^*\mathbf{R}_k\mathbf{X}_{nk}\bar{\boldsymbol{\phi}}_k$ is a scalar matrix. Let $\mathbf{R}_k = \left(\frac{1}{n}\mathbf{X}_{nk}\mathbf{X}_{nk}^* - z\mathbf{I}_{2p-2}\right)^{-1}$. Denote by $\boldsymbol{\alpha}_k$ the first column of $\boldsymbol{\phi}_k$ and by $\boldsymbol{\beta}_k$ the second column of $\boldsymbol{\phi}_k$, then combining (14) we have

$$\boldsymbol{\varepsilon}_k = \theta_k \mathbf{I}_2 \tag{18}$$

where

$$\begin{aligned}\theta_k &= \frac{1}{n}\boldsymbol{\alpha}_k'\bar{\boldsymbol{\alpha}}_k - z - \frac{1}{n^2}\boldsymbol{\alpha}_k'\mathbf{X}_{nk}^*\mathbf{R}_k\mathbf{X}_{nk}\bar{\boldsymbol{\alpha}}_k - (1 - z - y_n - y_n z E m_n(z))\\ &= \frac{1}{n}\boldsymbol{\beta}_k'\bar{\boldsymbol{\beta}}_k - z - \frac{1}{n^2}\boldsymbol{\beta}_k'\mathbf{X}_{nk}^*\mathbf{R}_k\mathbf{X}_{nk}\bar{\boldsymbol{\beta}}_k - (1 - z - y_n - y_n z E m_n(z)).\end{aligned}$$

Let $\widetilde{E}(\cdot)$ denote the conditional expectation given $\{x_{jl}, j = 1, \ldots, p, l = 1, \ldots, n; l \neq k\}$, then we get

$$E|\mathrm{tr}\boldsymbol{\varepsilon}_k^2| = \frac{1}{2}E|\mathrm{tr}\boldsymbol{\varepsilon}_k|^2 \leq \quad 2\left[E|\mathrm{tr}\boldsymbol{\varepsilon}_k - \widetilde{E}\mathrm{tr}\boldsymbol{\varepsilon}_k|^2 + E|\widetilde{E}\mathrm{tr}\boldsymbol{\varepsilon}_k - E\mathrm{tr}\boldsymbol{\varepsilon}_k|^2 + |E\mathrm{tr}\boldsymbol{\varepsilon}_k|^2\right]. \tag{19}$$

According to the inequality above, we proceed to complete the estimation of $E|\mathrm{tr}\boldsymbol{\varepsilon}_k^2|$ by the following three steps.

(a) For the first term of the right-hand side of (19), denote $\mathbf{T} = (t_{jl}) = \mathbf{I}_{2n} - \frac{1}{n}\mathbf{X}_{nk}^*\mathbf{R}_k\mathbf{X}_{nk}$ where $t_{jl} = \begin{pmatrix} e_{jl} & f_{jl} \\ h_{jl} & g_{jl}\end{pmatrix}$. Then, rewrite

$$\begin{aligned}\mathrm{tr}\boldsymbol{\varepsilon}_k - \widetilde{E}\mathrm{tr}\boldsymbol{\varepsilon}_k &= \mathrm{tr}\left(\frac{1}{n}\boldsymbol{\phi}_k'\bar{\boldsymbol{\phi}}_k - \frac{1}{n^2}\boldsymbol{\phi}_k'\mathbf{X}_{nk}^*\mathbf{R}_k\mathbf{X}_{nk}\bar{\boldsymbol{\phi}}_k\right) - \mathrm{tr}\left(\mathbf{I}_2 - \frac{1}{n^2}\mathbf{X}_{nk}^*\mathbf{R}_k\mathbf{X}_{nk}\right)\\ &= \frac{1}{n}\mathrm{tr}\left(\boldsymbol{\phi}_k'\mathbf{T}\bar{\boldsymbol{\phi}}_k - \mathbf{T}\right)\\ &= \frac{1}{n}\left(\sum_{j=1}^n \mathrm{tr}(\|x_{kj}\|^2 - 1)t_{jj} + \sum_{j\neq l}\mathrm{tr}(x_{kl}^* x_{kj} t_{jl})\right).\end{aligned}$$

By elementary calculation, we obtain

$$\begin{aligned}&\widetilde{E}|\mathrm{tr}\boldsymbol{\varepsilon}_k - \widetilde{E}\mathrm{tr}\boldsymbol{\varepsilon}_k|^2\\ &\quad= \frac{1}{n^2}\Bigg(\sum_{j=1}^n \widetilde{E}|\mathrm{tr}(\|x_{kj}\|^2 - 1)t_{jj}|^2 + \sum_{j\neq l}\widetilde{E}\Big[\mathrm{tr}\left(x_{kl}^* x_{kj} t_{jl}\right)\mathrm{tr}(x_{kj}^* x_{kl} t_{jl}^*)\\ &\qquad + \mathrm{tr}\left(x_{kl}^* x_{kj} t_{jl}\right)\mathrm{tr}\left(x_{kl}^* x_{kj} t_{lj}^*\right)\Big]\Bigg)\\ &\quad\leq \frac{1}{n^2}\left(\sum_{j=1}^n \widetilde{E}(\|x_{kj}\|^2 - 1)^2|e_{jj} + g_{jj}|^2 + 2\sum_{j\neq l}\widetilde{E}|\mathrm{tr}(x_{kl}^* x_{kj} t_{jl})|^2\right)\\ &\quad\leq \frac{C}{n^2}\left(\eta_n^2 n\sum_{j=1}^n(|e_{jj}|^2 + |g_{jj}|^2) + \sum_{j\neq l}(|e_{jl}|^2 + |f_{jl}|^2 + |g_{jl}|^2 + |h_{jl}|^2)\right)\end{aligned}$$

$$\leq \frac{C\eta_n^2}{n}\sum_{j=1}^{n}(|e_{jj}|^2+|g_{jj}|^2)+\frac{C}{n^2}\sum_{j,l}(|e_{jl}|^2+|f_{jl}|^2+|g_{jl}|^2+|h_{jl}|^2)$$
$$\leq \frac{C\eta_n^2}{n}\mathrm{tr}\mathbf{T}\mathbf{T}^*+\frac{C}{n^2}\mathrm{tr}\mathbf{T}\mathbf{T}^*. \tag{20}$$

For $\frac{1}{\sqrt{n}}\mathbf{X}_{nk}$, there exists a $(2p-2)\times q$ orthonormal matrix $\mathbf{U}$ and a $2n\times q$ orthonormal matrix $\mathbf{V}$ such that

$$\frac{1}{\sqrt{n}}\mathbf{X}_{nk}=\mathbf{U}\mathrm{diag}(s_1,\ldots,s_q)\mathbf{V}^*$$

where $s_1,\ldots,s_q$ are the singular values of $\frac{1}{\sqrt{n}}\mathbf{X}_{nk}$ and $q=\min\{(2p-2),2n\}$. Then, we get

$$\mathbf{I}_{2n}-\mathbf{T}=\left(\frac{1}{\sqrt{n}}\mathbf{X}_{nk}^*\right)\mathbf{R}_k\left(\frac{1}{\sqrt{n}}\mathbf{X}_{nk}\right)$$
$$=\mathbf{V}\mathrm{diag}\left(\frac{s_1^2}{s_1^2-z},\cdots,\frac{s_q^2}{s_q^2-z}\right)\mathbf{V}^*$$

which implies that

$$\mathbf{T}=\mathbf{V}\mathrm{diag}\left(\frac{-z}{s_1^2-z},\cdots,\frac{-z}{s_q^2-z}\right)\mathbf{V}^*.$$

Consequently, it follows that

$$\mathrm{tr}\mathbf{T}\mathbf{T}^*=\sum_{j=1}^{q}\frac{|z|^2}{|s_j^2-z|^2}\leq\frac{2n|z|^2}{\upsilon^2}. \tag{21}$$

By (20) and (21), we obtain

$$E|\mathrm{tr}\boldsymbol{\varepsilon}_k-\widetilde{E}\mathrm{tr}\boldsymbol{\varepsilon}_k|^2\to 0. \tag{22}$$

(b) Next, the second term of right-hand side of (19) is estimated. Note that

$$\widetilde{E}\mathrm{tr}\boldsymbol{\varepsilon}_k-E\mathrm{tr}\boldsymbol{\varepsilon}_k=\frac{z}{n}\left(E\mathrm{tr}\mathbf{R}_k-\mathrm{tr}\mathbf{R}_k\right).$$

Using the martingale decomposition method, we have

$$E|\widetilde{E}\mathrm{tr}\boldsymbol{\varepsilon}_k-E\mathrm{tr}\boldsymbol{\varepsilon}_k|^2=\frac{|z|^2}{n^2}E|E\mathrm{tr}\mathbf{R}_k-\mathrm{tr}\mathbf{R}_k|^2\leq\frac{4|z|^2}{n\upsilon^2}\to 0. \tag{23}$$

(c) Finally, combining (17), (19), (22), and (23), we conclude that

$$E|\mathrm{tr}\boldsymbol{\varepsilon}_k^2| \to 0.$$

These indicate that we complete the proof of the lemma. □

Now, we are in a position to show that

$$\delta_n \to 0.$$

By (15) and (18), we can write

$$\delta_n = -\frac{1}{2p(1-z-y_n-y_nzEm_n(z))^2}\sum_{k=1}^{p} E\mathrm{tr}\boldsymbol{\varepsilon}_k \\ + \frac{1}{2p(1-z-y_n-y_nzEm_n(z))^2}\sum_{k=1}^{p} E\frac{\mathrm{tr}\boldsymbol{\varepsilon}_k^2}{(1-z-y_n-y_nzEm_n(z))+\theta_k}.$$

Note that

$$\Im(1-z-y_n-y_nzEm_n(z)) < -\upsilon,$$

which implies that

$$|1-z-y_n-y_nzEm_n(z)| > \upsilon. \tag{24}$$

By Lemma 13 and (24), we have

$$\left|\frac{1}{2p(1-z-y_n-y_nzEm_n(z))^2}\sum_{k=1}^{p} E\mathrm{tr}\boldsymbol{\varepsilon}_k\right| \le \frac{1}{2p\upsilon^2}\sum_{k=1}^{p}|E\mathrm{tr}\boldsymbol{\varepsilon}_k| \to 0. \tag{25}$$

Together with Lemma 14, (24), and

$$\Im(\theta_k + (1-z-y_n-y_nzEm_n(z))) \\ = \Im\left(\frac{1}{n}\boldsymbol{\alpha}_k'\bar{\boldsymbol{\alpha}}_k - z - \frac{1}{n^2}\boldsymbol{\alpha}_k'\mathbf{X}_{nk}^*\left(\frac{1}{n}\mathbf{X}_{nk}\mathbf{X}_{nk}^* - z\mathbf{I}_{2p-2}\right)^{-1}\mathbf{X}_{nk}\bar{\boldsymbol{\alpha}}_k\right) \\ = -\upsilon\left(1 + \boldsymbol{\alpha}_k'\mathbf{X}_{nk}^*\left[\left(\frac{1}{n}\mathbf{X}_{nk}\mathbf{X}_{nk}^* - u\mathbf{I}_{2p-2}\right)^2 + \upsilon^2\mathbf{I}_{2p-2}\right]^{-1}\mathbf{X}_{nk}\bar{\boldsymbol{\alpha}}_k\right) < -\upsilon$$

one finds that

$$\left|\frac{1}{2p(1-z-y_n-y_nzEm_n(z))^2}\sum_{k=1}^{p}E\frac{\mathrm{tr}\boldsymbol{\varepsilon}_k^2}{(1-z-y_n-y_nzEm_n(z))+\theta_k}\right|$$
$$\leq \frac{1}{2p\upsilon^3}\sum_{k=1}^{p}E|\mathrm{tr}\boldsymbol{\varepsilon}_k^2| \to 0. \tag{26}$$

Combining (25) with (26), we get

$$|\delta_n| \leq \left|\frac{1}{2p\left(1-z-y_n-y_nzEm_n(z)\right)^2}\sum_{k=1}^{p}E\mathrm{tr}\boldsymbol{\varepsilon}_k\right|$$
$$+\left|\frac{1}{2p\left(1-z-y_n-y_nzEm_n(z)\right)^2}\sum_{k=1}^{p}E\frac{\mathrm{tr}\boldsymbol{\varepsilon}_k^2}{\left(1-z-y_n-y_nzEm_n(z)\right)+\theta_k}\right|$$
$$\to 0.$$

So far, we have completed the proof of the mean convergence

$$Em_n(z) \to m(z).$$

3.2.3 *Completion of the proof of Theorem 1*

By Sects. 3.2.1 and 3.2.2, for any fixed $z \in \mathbb{C}^+$, we have

$$m_n(z) \xrightarrow{\text{a.s.}} m(z).$$

To complete the proof Theorem 1, we need the last part of Chapter 2 of Bai and Silverstein (2010). For the readers convenience, we repeat here. That is, for each $z \in \mathbb{C}^+$, there exists a null set N_z (i.e., $\mathrm{P}(N_z) = 0$) such that

$$m_n(z, w) \to m(z), \quad \text{for all } w \in N_z^c.$$

Now, let $\mathbb{C}_0^+$ be a dense subset of $\mathbb{C}^+$ (e.g., all z of rational real and imaginary parts) and let $N = \bigcup_{z\in\mathbb{C}_0^+} N_z$. Then,

$$m_n(z, w) \to m(z), \quad \text{for all } w \in N^c \text{ and } z \in \mathbb{C}_0^+.$$

Let $\mathbb{C}_m^+ = \{z \in \mathbb{C}^+ : \Im z > 1/m,\ |z| \leq m\}$. When $z \in \mathbb{C}_m^+$, we have $|m_n(z)| \leq m$. Applying Lemma 23, we have

$$m_n(z, w) \to m(z), \quad \text{for all } w \in N^c \text{ and } z \in \mathbb{C}_m^+.$$

Since the convergence above holds for every m, we conclude that

$$m_n(z, w) \to m(z), \quad \text{for all } \; w \in N^c \text{ and } z \in \mathbb{C}^+.$$

Applying Lemma 24, we conclude that

$$F^{\mathbf{S}_n} \xrightarrow{w} F, \text{ a.s.}$$

4 Appendix

In this section, some results are listed which are used in the proof of the main theorem.

Lemma 15 *Suppose for any $\eta > 0$ $\sum_{n=1}^{\infty} f(\eta, n) < \infty$, then we can select a slowly decreasing sequence of constants $\eta_n \to 0$ such that*

$$\sum_{n=1}^{\infty} f(\eta_n, n) < \infty$$

where f is a nonnegative function.

Similarly, if $f(\eta, n) \to 0$ for any fixed $\eta > 0$, then there exists a decreasing sequence $\eta_n \to 0$ such that $f(\eta_n, n) \to 0$.

Proof Letting $\eta = \frac{1}{m}$, one has $\sum_{n=1}^{\infty} f\left(\frac{1}{m}, n\right) < \infty$. Moreover, there exists a increasing sequence N_m such that $\sum_{n=N_m}^{\infty} f\left(\frac{1}{m}, n\right) \le \frac{1}{2^m}$. Define a sequence $\eta_n = \frac{1}{m}$ when $N_m \le n < N_{m+1}$. We get

$$\sum_{n=1}^{\infty} f(\eta_n, n) = \sum_{m=1}^{\infty} \sum_{n=N_m}^{N_{m+1}-1} f\left(\frac{1}{m}, n\right) \le \sum_{m=1}^{\infty} \sum_{n=N_m}^{\infty} f\left(\frac{1}{m}, n\right)$$
$$\le \sum_{n=1}^{N_1-1} f(1, n) + \sum_{m=1}^{\infty} \frac{1}{2^m} < \infty.$$

This completes the proof of this lemma. □

Lemma 16 (Corollary A.42 of Bai and Silverstein 2010) *Let $\mathbf{A}$ and $\mathbf{B}$ be two $p \times n$ matrices and denote the ESD of $\mathbf{S} = \mathbf{AA}^*$ and $\widetilde{\mathbf{S}} = \mathbf{BB}^*$ by $F^{\mathbf{S}}$ and $F^{\widetilde{\mathbf{S}}}$, respectively. Then,*

$$L^4(F^{\mathbf{S}}, F^{\widetilde{\mathbf{S}}}) \le \frac{2}{p^2}(\mathrm{tr}(\mathbf{AA}^* + \mathbf{BB}^*))(\mathrm{tr}[(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^*]),$$

where $L(\cdot, \cdot)$ denotes the Lévy distance, that is,

$$L(F^{\mathbf{S}}, F^{\widetilde{\mathbf{S}}}) = \inf\{\varepsilon : F^{\mathbf{S}}(x - \varepsilon, y - \varepsilon) - \varepsilon \le F^{\widetilde{\mathbf{S}}} \le F^{\mathbf{S}}(x + \varepsilon, y + \varepsilon) + \varepsilon\}.$$

Lemma 17 (Theorem A.44 of Bai and Silverstein 2010) *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $p \times n$ *complex matrices. Then,*

$$\|F^{\mathbf{AA}^*} - F^{\mathbf{BB}^*}\|_{KS} \leq \frac{1}{p}\text{rank}\,(\mathbf{A} - \mathbf{B}).$$

Lemma 18 (Bernstein's inequality) *If* $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ *are independent random variables with means zero and uniformly bounded by b, then, for any* $\varepsilon > 0$*,*

$$\mathrm{P}\left(\left|\sum_{j=1}^{n} \boldsymbol{\tau}_j\right| \geq \varepsilon\right) \leq 2\exp\left(-\varepsilon^2/\left[2(B_n^2 + b\varepsilon)\right]\right)$$

where $B_n^2 = E(\boldsymbol{\tau}_1 + \cdots + \boldsymbol{\tau}_n)^2$.

Lemma 19 (see $(A.1.12)$ of Bai and Silverstein 2010) *Let* $z = u + iv$, $v > 0$, *and* $\mathbf{A}$ *be an* $n \times n$ *Hermitian matrix. Denote by* $\mathbf{A}_k$ *the kth major sub-matrix of* $\mathbf{A}$ *of order* $(n-1)$, *to be the matrix resulting from deleting the kth row and column from* $\mathbf{A}$. *Then,*

$$|\text{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \text{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1}| \leq \frac{1}{v}.$$

Lemma 20 *Suppose that the matrix* $\boldsymbol{\Sigma}$ *has the partition as given by* $\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$. *If* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Sigma}_{11}$ *are nonsingular, then the inverse of* $\boldsymbol{\Sigma}$ *has the form*

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1} \\ -\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{\Sigma}_{22.1}^{-1} \end{pmatrix}$$

where $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Lemma 21 (Burkholder's inequality) *Let* $\{\mathbf{X}_k\}$ *be a complex martingale difference sequence with respect to the increasing* σ*-field. Then, for* $p > 1$,

$$E\left|\sum_k \mathbf{X}_k\right|^p \leq K_p E\left(\sum_k |\mathbf{X}_k|^2\right)^{p/2}.$$

Lemma 22 (Rosenthal's inequality) *Let* $\mathbf{X}_i$ *be independent with zero means, then we have, for some constant* C_k,

$$E\left|\sum_i \mathbf{X}_i\right|^{2k} \leq C_k\left(\sum_i E\,|\mathbf{X}_i|^{2k} + \left(\sum_i E\,|\mathbf{X}_i|^2\right)^k\right).$$

Lemma 23 (Lemma 2.14 of Bai and Silverstein 2010) *Let* $f_1, f_2, \ldots$ *be analytic in D, a connected open set of* $\mathbb{C}$, *satisfying* $|f_n(z)| \leq M$ *for every n and z in D, and*

$f_n(z)$ converges as $n \to \infty$ for each z in a subset of D having a limit point in D. Then, there exists a function f analytic in D for which $f_n(z) \to f(z)$ and $f_n' \to f'(z)$ for all $z \in D$. Moreover, on any set bounded by a contour interior to D, the convergence is uniform and $\{f_n'(z)\}$ is uniformly bounded.

Lemma 24 (Theorem B.9 of Bai and Silverstein 2010) *Assume that $\{G_n\}$ is a sequence of functions of bounded variation and $G_n(-\infty) = 0$ for all n. Then,*

$$\lim_{n\to\infty} \mathbf{m}_{G_n}(z) = \mathbf{m}(z) \quad \forall z \in D$$

where $D \equiv \{z \in \mathbb{C} : \Im z > 0\}$ if and only if there is a function of bounded variation G with $G(-\infty) = 0$ and Stieltjes transform $\mathbf{m}(z)$ and such that $G_n \to G$ vaguely.

References

Adler, S. L. (1995). *Quaternionic quantum mechanics and quantum fields* (Vol. 1). Oxford: Oxford University Press.

Akemann, G. (2005). The complex Laguerre symplectic ensemble of non-Hermitian matrices. *Nuclear Physics B*, *730*(3), 253–299.

Akemann, G., Phillips, M. (2013). The interpolating airy kernels for the beta = 1 and beta = 4 elliptic Ginibre ensembles. arXiv:1308.3418.

Anderson, G. W., Guionnet, A., Zeitouni, O. (2010). *An introduction to random matrices* (Vol. 118). Cambridge: Cambridge University Press.

Bai, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices. *The Annals of Probability*, *21*(2), 649–672.

Bai, Z. D., Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices* (2nd ed.). New York: Springer.

Finkelstein, D., Jauch, J. M., Schiminovich, S., Speiser, D. (1962). Foundations of quaternion quantum mechanics. *Journal of Mathematical Physics*, *3*(2), 207.

Kanzieper, E. (2002). Eigenvalue correlations in non-Hermitean symplectic random matrices. *Journal of Physics A: Mathematical and General*, *35*(31), 6631.

Kuipers, J. B. (1999). *Quaternions and rotation sequences*. Princeton: Princeton University Press.

Marčenko, V. A., Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, *114*(4), 507–536.

Mehta, M. L. (2004). *Random matrices* (Vol. 142). Access online via Elsevier.

So, W., Thompson, R. C., Zhang, F. (1994). The numerical range of normal matrices with quaternion entries. *Linear and Multilinear Algebra*, *37*(1–3), 175–195.

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

Wolf, L. A. (1936). Similarity of matrices in which the elements are real quaternions. *Bulletin of the American Mathematical Society*, *42*(10), 737–743.

Yin, Y., Bai, Z. D. (2014). Convergence rates of the spectral distributions of large random quaternion self-dual Hermitian matrices. *Journal of Statistical Physics*, *157*(6), 1207–1224.

Yin, Y., Bai, Z. D., Hu, J. (2013). On the semicircular law of large dimensional random quaternion matrices. *Journal of Theoretical Probability* (to appear). arXiv:1309.6937.

Yin, Y., Bai, Z. D., Hu, J. (2014). On the limit of extreme eigenvalues of large dimensional random quaternion matrices. *Physics Letters A*, *378*(1617), 1049–1058.

Zhang, F. (1995). On numerical range of normal matrices of quaternions. *Journal of Mathmatical and Physical Science*, *29*(6), 235–251.

Zhang, F. (1997). Quaternions and matrices of quaternions. *Linear Algebra and Its Applications*, *251*, 21–57.