J Theor Probab (2010) 23: 1–20
DOI 10.1007/s10959-009-0270-2

Almost Sure Limit of the Smallest Eigenvalue of Some Sample Correlation Matrices

Han Xiao · Wang Zhou

Received: 24 October 2006 / Revised: 18 July 2009 / Published online: 1 January 2010

Abstract Let $X^{(n)} = (X_{ij})$ be a $p \times n$ data matrix, where the n columns form a random sample of size n from a certain p-dimensional distribution. Let $R^{(n)} = (\rho_{ij})$ be the $p \times p$ sample correlation coefficient matrix of $X^{(n)}$, and $S^{(n)} = (1/n)X^{(n)}(X^{(n)})^* - \bar{X}\bar{X}^*$ be the sample covariance matrix of $X^{(n)}$, where $\bar{X}$ is the mean vector of the n observations. Assuming that X_{ij} are independent and identically distributed with finite fourth moment, we show that the smallest eigenvalue of $R^{(n)}$ converges almost surely to the limit $(1-\sqrt{c})^2$ as $n \to \infty$ and $p/n \to c \in (0, \infty)$. We accomplish this by showing that the smallest eigenvalue of $S^{(n)}$ converges almost surely to $(1-\sqrt{c})^2$.

Keywords Random matrix · Sample correlation coefficient matrix · Sample covariance matrix · Smallest eigenvalue

Mathematics Subject Classification (2000) Primary 60H15 · Secondary 62H99

1 Introduction

Suppose that $X^{(n)} = (X_{ij})$ is a $p \times n$ data matrix, where the n columns form a random sample of size n from a certain p-dimensional distribution. Let $R^{(n)} = (\rho_{ij})$ be the $p \times p$ sample correlation coefficient matrix of $X^{(n)}$, where ρ_{ij} is the usual Pearson

Wang Zhou was partially supported by an NUS grant R-155-000-083-112.

H. Xiao
Department of Statistics, The University of Chicago, 5734 S University Ave, Chicago, IL 60637, USA
e-mail: xiao@galton.uchicago.edu

W. Zhou (✉)
Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore
e-mail: stazw@nus.edu.sg

correlation coefficient between the ith and jth rows of $X^{(n)}$. We are interested in the strong limits of the extreme eigenvalues of this matrix as its dimensions tend to infinity.

There are two random matrices which are closely related with the sample correlation matrix. One is the sample covariance matrix $S^{(n)}$ defined by

$$S^{(n)} = \big(S_{ij}^{(n)}\big) = \frac{1}{n} X^{(n)}\big(X^{(n)}\big)^* - \bar{X}\bar{X}^*,$$

where $\bar{X}$ is the mean vector of the n observations. Let

$$D^{(n)} = \operatorname{diag}\Big\{\sqrt{S_{11}^{(n)}}, \sqrt{S_{22}^{(n)}}, \ldots, \sqrt{S_{pp}^{(n)}}\Big\}.$$

Then $R^{(n)}$ can be expressed as

$$R^{(n)} = \big(D^{(n)}\big)^{-1} S^{(n)} \big(D^{(n)}\big)^{-1}. \tag{1.1}$$

The other one is the simplified version of the sample covariance matrix given by

$$\mathbb{S}^{(n)} = (1/n) X^{(n)}\big(X^{(n)}\big)^*.$$

Remarks (1) For notational economy, we will omit the superindex (n) from now on when there is no confusion. (2) In the literature, $\mathbb{S}$ is often referred under the name "sample covariance matrix." However, in this paper, we rename it by *simplified sample covariance matrix* to avoid confusion.

Suppose that $\lambda_1(\mathbb{S}), \lambda_2(\mathbb{S}), \ldots, \lambda_p(\mathbb{S})$ are the p eigenvalues of $\mathbb{S}$ in increasing order. While the definition of the largest eigenvalue is clear, one needs to examine that of the smallest one.

Since $\operatorname{rank}(\mathbb{S}) \le n$ when $p \ge n$, the $(p-n)$ smallest eigenvalues are all zero. Hence we define the smallest eigenvalue of the matrix $\mathbb{S}$ as

$$\lambda_{\min}(\mathbb{S}) = \begin{cases} \lambda_1(\mathbb{S}) & \text{if } p < n, \\ \lambda_{p-n+1}(\mathbb{S}) & \text{if } p \ge n. \end{cases} \tag{1.2}$$

Since the *empirical spectral distribution* $F^{\mathbb{S}}$ of $\mathbb{S}$ almost surely converges to the *Marčenko–Pastur law* F_c with the density

$$F_c'(x) = \begin{cases} \frac{1}{2\pi c x}\sqrt{(b-x)(x-a)} & \text{if } a \le x \le b, \\ 0 & \text{otherwise}, \end{cases} \tag{1.3}$$

and the point mass $1-1/c$ at the origin if $c>1$, where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$ (see Chap. 3 of Bai and Silverstein [3]), we have

$$\liminf \lambda_{\max}(\mathbb{S}) \ge b = (1+\sqrt{c})^2 \quad \text{a.s.},$$

$$\limsup \lambda_{\min}(\mathbb{S}) \le a = (1-\sqrt{c})^2 \quad \text{a.s.}$$

However, the converse assertions

$$\limsup \lambda_{\max}(\mathbb{S}) \le b = (1+\sqrt{c})^2 \quad \text{a.s.,} \tag{1.4}$$

$$\liminf \lambda_{\min}(\mathbb{S}) \ge a = (1-\sqrt{c})^2 \quad \text{a.s.} \tag{1.5}$$

are not trivial.

Yin et al. [9] established (1.4). The following modified version is an immediate consequence of their original result.

Theorem 1.1 *Let X be the up-left $p \times n$ corner of a double array $\{X_{uv} : u, v = 1, 2, \ldots\}$ of independent and identically distributed (i.i.d.) complex random variables (r.v.s.) with zero mean and unit variance. If $E|X_{11}|^4 < \infty$, then, as $n \to \infty$ and $p/n \to c \in (0, \infty)$,*

$$\lim \lambda_{\max}(\mathbb{S}) = b = (1+\sqrt{c})^2 \quad \text{a.s.}$$

It is much more difficult to establish (1.5) than (1.4). Bai and Yin [4] devised a unified approach to prove (1.4) and (1.5) at the same time. As an immediate consequence of their result, we have the following theorem.

Theorem 1.2 *Under the conditions of Theorem 1.1,*

$$\lim \lambda_{\min}(\mathbb{S}) = a = (1-\sqrt{c})^2 \quad \text{a.s.}$$

More than ten years later, Jiang [6] proved that the largest eigenvalue of the sample correlation matrix R converges to the limit $(1+\sqrt{c})^2$ with probability one as $n \to \infty$ and $p/n \to c \in (0, \infty)$. Jiang [6] also conjectured that the smallest eigenvalue of R converges to $(1-\sqrt{c})^2$ a.s.

Since $\operatorname{rank}(R) \le n-1$, the smallest eigenvalue of the matrix R can be defined as

$$\lambda_{\min}(R) = \begin{cases} \lambda_1(R) & \text{if } p < n, \\ \lambda_{p-n+2}(R) & \text{if } p \ge n. \end{cases} \tag{1.6}$$

In this paper, we prove Jiang's conjecture.

Theorem 1.3 *Let X be the up-left $p \times n$ corner of a double array $\{X_{uv} : u, v = 1, 2, \ldots\}$ of i.i.d. complex r.v.s. with unit variance. If $E|X_{11}|^4 < \infty$, then, as $n \to \infty$ and $p/n \to c \in (0, \infty)$,*

$$\lim \lambda_{\min}(R) = a = (1-\sqrt{c})^2 \quad \text{a.s.}$$

We accomplish the proof of Theorem 1.3 by establishing the following result on the sample covariance matrix S. Note that the definition of the smallest eigenvalue of S is given by replacing R in (1.6) by S.

Theorem 1.4 *Under the conditions of Theorem 1.3,*

$$\lim \lambda_{\min}(S) = a = (1-\sqrt{c})^2 \quad \text{a.s.}$$

The paper is organized as follows. In Sect. 2, we show how Theorem 1.4 implies Theorem 1.3. The proof of Theorem 1.4 will be completed in Sect. 3. The auxiliary lemmas are collected in the last section.

2 From Sample Covariance Matrix to Sample Correlation Matrix

Our task in this section is to prove Theorem 1.3 by Theorem 1.4. Actually the argument here parallels that in [6]. We repeat it for the completeness of the whole proof.

Since we are interested in the sample covariance matrix and sample correlation matrix, we can assume that $EX_{11}=0$. According to Theorem 1.4, it suffices to show that

$$\left|\sqrt{\lambda_{\min}(R)}-\sqrt{\lambda_{\min}(S)}\,\right|\to 0\quad \text{a.s.} \tag{2.1}$$

Note that the sample covariance matrix S could be written as $S=n^{-1}XPX^*$, where P is the $n\times n$ projection matrix defined as $I-\frac{1}{n}\mathbf{1}\mathbf{1}^T$, and $\mathbf{1}$ is the $n\times 1$ vector whose entries are all 1's. Since $R=D^{-1}SD^{-1}$ (see (1.1)), by Lemma 4.1 we have

$$\begin{aligned}\left|\sqrt{\lambda_{\min}(R)}-\sqrt{\lambda_{\min}(S)}\,\right| &\le \left\|D^{-1}\frac{1}{\sqrt{n}}XP-\frac{1}{\sqrt{n}}XP\right\| \\ &\le \left\|D^{-1}-I\right\|\cdot\left\|\frac{1}{\sqrt{n}}X\right\|. \end{aligned}\tag{2.2}$$

Since $E|X_{11}|^4<\infty$, due to Lemma 4.4, we know that

$$\max_{1\le i\le p}\left|\frac{\sum_{j=1}^n X_{ij}^2}{n}-1\right|\to 0\quad \text{a.s.} \tag{2.3}$$

and

$$\max_{1\le i\le p}\bar{X}_i\to 0\quad \text{a.s.,} \tag{2.4}$$

where $\bar{X}_i$ is the ith entry of the mean vector $\bar{X}$. Combining (2.3) and (2.4), we have

$$\max_{1\le i\le p}\left|\frac{\sum_{j=1}^n (X_{ij}-\bar{X}_i)^2}{n}-1\right|\le \max_{1\le i\le p}\left|\frac{\sum_{j=1}^n X_{ij}^2}{n}-1\right|+\max_{1\le i\le p}\bar{X}_i^2\to 0\quad \text{a.s.,}$$

and this implies that

$$\left\|D^{-1}-I\right\|=\max_{1\le i\le p}\left|\frac{\sqrt{n}}{\sqrt{\sum_{j=1}^n (X_{ij}-\bar{X}_i)^2}}-1\right|\to 0\quad \text{a.s.} \tag{2.5}$$

By Theorem 1.1,

$$\left\|\frac{1}{\sqrt{n}}X\right\|=\sqrt{\lambda_{\max}(\mathbb{S})}\to 1+\sqrt{c}\quad \text{a.s.}$$

This, together with (2.2) and (2.5), proves (2.1).

3 Proof of Theorem 1.4

We first derive the limiting spectral distribution of S. Since $S = \mathbb{S} - \bar{X}\bar{X}^*$, by Lemma 4.2, we know that

$$\left\| F^S - F^{\mathbb{S}} \right\| \le \frac{1}{p} \operatorname{rank}\bigl(\bar{X}\bar{X}^*\bigr) = \frac{1}{p}.$$

Since the convergence of the distribution functions in the sup norm implies their weak convergence, we know that F^S also converges to the Marčenko–Pastur law, and hence

$$\limsup \lambda_{\min}(S) \le a = (1 - \sqrt{c}\,)^2 \quad \text{a.s.}$$

Therefore, in order to prove Theorem 1.4, it suffices to show that

$$\liminf \lambda_{\min}(S) \ge a = (1 - \sqrt{c})^2 \quad \text{a.s.} \tag{3.1}$$

Note that when $c = 1$, the situation is trivial. When $c > 1$, p is larger than n when n is very large. In this case we will consider $\lambda_{\min}(S) = \lambda_{p-n+2}(S)$, which is the $(p - n + 2)$th smallest eigenvalue of S. According to Theorem 4.3.4 of Horn and Johnson [5], we have

$$\lambda_{p-n+2}(S) \ge \lambda_{p-n+1}(\mathbb{S}).$$

As an immediate consequence of this fact and Theorem 1.2, we know that (3.1) holds when $c > 1$. Now we shall prove (3.1) when $0 < c < 1$, and the long proof will be divided into several steps.

3.1 Truncation

We use the truncation technique given in [2] to bound the underlying variables. For $C > 0$, let $Y_{ij} = X_{ij} I_{\{|X_{ij}| \le C\}} - E X_{ij} I_{\{|X_{ij}| \le C\}}$, $Y = (Y_{ij})$, and $\tilde{S} = (1/n) Y P Y^*$. Denote the eigenvalues of S and $\tilde{S}$ by λ_k and $\tilde{\lambda}_k$ (in increasing order). Since these are the squares of the kth smallest singular values of $(1\sqrt{n}) X P$ and $(1\sqrt{n}) Y P$ (respectively), it follows from Lemma 4.1 that

$$\max_{k \le n} \bigl| \lambda_k^{1/2} - \tilde{\lambda}_k^{1/2} \bigr| \le \frac{1}{\sqrt{n}} \| X - Y \|.$$

Since $X_{ij} - Y_{ij} = X_{ij} I_{\{|X_{ij}| > C\}} - E X_{ij} I_{\{|X_{ij}| > C\}}$, from Theorem 1.1 we have, with probability one,

$$\limsup_{n \to \infty} \max_{k \le n} \bigl| \lambda_k^{1/2} - \tilde{\lambda}_k^{1/2} \bigr| \le (1 + \sqrt{c}) E^{1/2} |X_{11}|^2 I_{\{|X_{ij}| > C\}}.$$

Since $E|X_{11}|^2 = 1$, we can make the above bound arbitrarily small by choosing C sufficiently large. Thus, in the following investigation, we can assume that the underlying variables are uniformly bounded. It is easy to verify that we can rescale the variables so that the assumption $E|X_{11}|^2 = 1$ still holds.

3.2 An Equivalent Problem

There exists a nonzero vector β such that

$$S\beta = \big(\mathbb{S} - \bar{X}\bar{X}^*\big)\beta = \lambda_{\min}(S)\beta,$$

which is equivalent to

$$\big(\mathbb{S} - \lambda_{\min}(S)I\big)\beta = \bar{X}\bar{X}^*\beta. \tag{3.2}$$

Suppose that $\liminf \lambda_{\min}(S)$ is smaller than $a = (1-\sqrt{c})^2$. Then we can imagine that $\lambda_{\min}(S)$ is not an eigenvalue of $\mathbb{S}$ as n gets large, since $\lambda_{\min}(\mathbb{S})$ converges to a (see (1.2)). In this case the matrix $(\mathbb{S} - \lambda_{\min}(S)I)$ is nonsingular, and hence (3.2) can be inverted to give

$$\beta = \big(\mathbb{S} - \lambda_{\min}(S)I\big)^{-1}\bar{X}\bar{X}^*\beta.$$

If we multiply both sides of the above equation by $\bar{X}^*$, we will get

$$\bar{X}^*\beta = \bar{X}^*\big(\mathbb{S} - \lambda_{\min}(S)I\big)^{-1}\bar{X}\bar{X}^*\beta.$$

Since $\bar{X}^*\beta \neq 0$, we can obtain that

$$\bar{X}^*\big(\mathbb{S} - \lambda_{\min}(S)I\big)^{-1}\bar{X} = 1. \tag{3.3}$$

The above arguments (especially (3.3)) provide the basic idea that we will make use of to state the current problem in an equivalent form given by the following lemma.

Lemma 3.1 *If $P(\liminf \lambda_{\min}(S) < a) > 0$, then for some $0 < \lambda < a$,*

$$P\big(\limsup \bar{X}^*(\mathbb{S} - \lambda I)^{-1}\bar{X} \geq 1\big) > 0.$$

In other words, if

$$\limsup \bar{X}^*(\mathbb{S} - \lambda I)^{-1}\bar{X} < 1 \quad a.s.\ \forall 0 < \lambda < a, \tag{3.4}$$

then we have the desirable property

$$\liminf \lambda_{\min}(S) \geq a \quad a.s.$$

Proof If $P(\liminf \lambda_{\min}(S) < a) > 0$, then there exists a small $\epsilon > 0$ such that $P(\liminf \lambda_{\min}(S) < a - 3\epsilon) > 0$. For simplicity, we denote the event $\{\liminf \lambda_{\min}(S) < a - 3\epsilon\}$ by E_0. Let B_n denote the event $\{\lambda_{\min}(\mathbb{S}) \leq a - \epsilon\}$. From Lemma 4.5 we know that $P(B_n) = o(n^{-l})$ for any $l > 0$. Hence it is easy to see that for some N large enough, $P(E_0 \setminus \bigcup_{n=N}^{\infty} B_n) > 0$. We use E to denote the event $E_0 \setminus \bigcup_{n=N}^{\infty} B_n$. For each $\omega \in E$, the following two properties hold:

$$\liminf \lambda_{\min}(S(\omega)) < a - 3\epsilon, \quad \lambda_{\min}(\mathbb{S}(\omega)) > a - \epsilon,\ \forall n \geq N.$$

Since $\liminf \lambda_{\min}(S(\omega)) < a - 3\epsilon$, we can find a subsequence n_k such that

$$\liminf \lambda_{\min}^{(n_k)}\big(S(\omega)\big) \to \lambda(\omega) < a - 3\epsilon.$$

When k is large enough, $\lambda_{\min}^{(n_k)}(S(\omega)) < a - 2\epsilon$, and hence from (3.3) and from the fact that $\bar{X}^*(\mathbb{S} - \lambda I)^{-1}\bar{X}$ is an increasing function of λ we have

$$\big[\bar{X}^*\big(\mathbb{S}^{(n_k)} - (a - 2\epsilon)I\big)^{-1}\bar{X}\big]\big|_\omega \geq 1,$$

and this means that

$$\limsup\big[\bar{X}^*\big(\mathbb{S}^{(n)} - (a - 2\epsilon)I\big)^{-1}\bar{X}\big]\big|_\omega \geq 1.$$

Therefore, we know that for $\lambda = a - 2\epsilon$,

$$P\big(\limsup \bar{X}^*(\mathbb{S} - \lambda I)^{-1}\bar{X} \geq 1\big) \geq P(E) > 0,$$

which completes the proof. □

Now our target is to prove (3.4) when $0 < c < 1$. Suppose $0 < \lambda < a$, and let $2\epsilon = a - \lambda$. We expand $\bar{X}^*(\mathbb{S} - \lambda I)^{-1}\bar{X}$ as

$$\begin{aligned}\bar{X}^*(\mathbb{S} - \lambda I)^{-1}\bar{X} &= \frac{1}{n}(X_1 + \cdots + X_n)^*(\mathbb{S} - \lambda I)^{-1}\frac{1}{n}(X_1 + \cdots + X_n)\\ &= \frac{1}{n^2}\sum_{i=1}^n X_i^*(\mathbb{S} - \lambda I)^{-1}X_i + \frac{1}{n^2}\sum_{i\neq j} X_i^*(\mathbb{S} - \lambda I)^{-1}X_j.\end{aligned}$$

Let

$$T_1 = \frac{1}{n^2}\sum_{i=1}^n X_i^*(\mathbb{S} - \lambda I)^{-1}X_i, \tag{3.5}$$

$$T_2 = \frac{1}{n^2}\sum_{i\neq j} X_i^*(\mathbb{S} - \lambda I)^{-1}X_j. \tag{3.6}$$

We will consider T_1 and T_2 respectively. For the time being, let us define $T_1 = T_2 = 0$ when the matrix $(\mathbb{S} - \lambda I)$ is singular. Actually Assumption (i) provides justification of this definition.

3.3 Nonnegative Terms

Let $S_i = \mathbb{S} - (1/n)X_iX_i^*$. Using Lemma 4.3, we may write T_1 as

$$\begin{aligned}T_1 &= \frac{1}{n^2}\sum_{i=1}^n \frac{X_i^*(S_i - \lambda I)^{-1}X_i}{1 + \frac{1}{n}X_i^*(S_i - \lambda I)^{-1}X_i}\\ &= \frac{1}{n}\sum_{i=1}^n \frac{\frac{1}{n}X_i^*(S_i - \lambda I)^{-1}X_i}{1 + \frac{1}{n}X_i^*(S_i - \lambda I)^{-1}X_i}.\end{aligned}$$

We use E_i to denote the event $\{\lambda_{\min}(S_i) > a - \epsilon\}$, and let $E = \bigcap_{i=1}^n E_i$. Again from Lemma 4.5 we know that $P(E_i^c) = o(n^{-l})$ for any $l > 0$, and hence $P(E^c) = o(n^{-l})$ for any $l > 0$. On the event E, $\|(S_i - \lambda I)^{-1}\| \le 1/\epsilon$, and the matrix $(S_i - \lambda I)^{-1}$ is nonnegative definite. Since all the entries of X are bounded by a constant C, we have

$$\left|\frac{1}{n} X_i^*(S_i - \lambda I)^{-1} X_i\right| \le \frac{1}{n}\|X_i^*\|\,\|(S_i - \lambda I)^{-1}\|\,\|X_i\| \le \frac{C^2}{\epsilon}.$$

Therefore we know that on the event E,

$$T_1 \le \frac{\frac{C^2}{\epsilon}}{1 + \frac{C^2}{\epsilon}} < 1.$$

Since $P(E^c) = o(n^{-l})$ for any $l > 0$, by the Borel–Cantelli lemma we know that

$$\limsup T_1 \le \frac{\frac{C^2}{\epsilon}}{1 + \frac{C^2}{\epsilon}} < 1 \quad \text{a.s.} \tag{3.7}$$

3.4 Cross Terms

Now we will focus on T_2, and it suffices to show that

$$\lim T_2 = 0 \quad \text{a.s.} \tag{3.8}$$

Let $S_{ij} = \mathbb{S} - (1/n)X_i X_i^* - (1/n)X_j X_j^*$. By Lemma 4.3, we can write T_2 as

$$T_2 = \frac{1}{n^2}\sum_{i \ne j} \frac{X_i^*(S_{ij} - \lambda I)^{-1} X_j}{[1 + \frac{1}{n} X_i^*(S_i - \lambda I)^{-1} X_i][1 + \frac{1}{n} X_j^*(S_{ij} - \lambda I)^{-1} X_j]}.$$

This expression plays the central role in our investigation.

Previously, we have defined the matrices S_i and S_{ij}. Similarly, we can define such matrices with additional subindices, such as $S_{i_1 i_2 j_1}$, $S_{i_1 i_2 j_1 j_2}$, etc. In general, let $\Lambda \subset [n]$ be a finite index set, then S_Λ is defined as

$$S_\Lambda = \mathbb{S} - \frac{1}{n}\sum_{i \in \Lambda} X_i X_i^*.$$

For simplicity, we use the notation A_Λ to denote the matrix

$$A_\Lambda = (S_\Lambda - \lambda I)^{-1}.$$

3.4.1 Change S_i to S_{ij} in the Denominator

Motivated by the symmetry, we first change S_i in the denominator of T_2 to S_{ij} and denote the new term by T_3:

$$T_3 = \frac{1}{n^2}\sum_{i\neq j}\frac{X_i^*(S_{ij}-\lambda I)^{-1}X_j}{[1+\frac{1}{n}X_i^*(S_{ij}-\lambda I)^{-1}X_i][1+\frac{1}{n}X_j^*(S_{ij}-\lambda I)^{-1}X_j]}$$
$$= \frac{1}{n^2}\sum_{i\neq j}\frac{X_i^*A_{ij}X_j}{(1+\frac{1}{n}X_i^*A_{ij}X_i)(1+\frac{1}{n}X_j^*A_{ij}X_j)}.$$

Our task in this step is to show that

$$D_{23} = T_2 - T_3 \to 0 \quad \text{a.s.} \tag{3.9}$$

According to Lemma 4.3, we can write D_{23} as

$$D_{23} = \frac{1}{n^2}\sum_{i\neq j}\frac{X_i^*A_{ij}X_j(\frac{1}{n}X_i^*A_{ij}X_i - \frac{1}{n}X_i^*A_iX_i)}{(1+\frac{1}{n}X_i^*A_iX_i)(1+\frac{1}{n}X_i^*A_{ij}X_i)(1+\frac{1}{n}X_j^*A_{ij}X_j)}$$
$$= \frac{1}{n^4}\sum_{i\neq j}\frac{X_i^*A_{ij}X_jX_i^*A_{ij}X_jX_j^*A_{ij}X_i}{(1+\frac{1}{n}X_i^*A_iX_i)(1+\frac{1}{n}X_i^*A_{ij}X_i)(1+\frac{1}{n}X_j^*A_{ij}X_j)^2}.$$

In order to control the norm of the matrix A_{ij}, we restrict it on the event $E_{ij} = \{\lambda_{\min}(S_{ij}) > a-\epsilon\}$ and consider

$$\bar{D}_{23} = \frac{1}{n^4}\sum_{i\neq j}\frac{X_i^*A_{ij}X_jX_i^*A_{ij}X_jX_j^*A_{ij}X_i}{(1+\frac{1}{n}X_i^*A_iX_i)(1+\frac{1}{n}X_i^*A_{ij}X_i)(1+\frac{1}{n}X_j^*A_{ij}X_j)^2}I_{ij},$$

where I_{ij} stands for the indicator function of the event E_{ij}. By the Borel–Cantelli lemma and Lemma 4.5, it is not difficult to see that, with probability one, $D_{23} = \bar{D}_{23}$ for all n large enough. Hence it suffices to consider $\bar{D}_{23}$ in the sequel. On the event E_{ij}, A_i and A_{ij} are positive definite, and hence

$$\left(1+\frac{1}{n}X_i^*A_iX_i\right)\left(1+\frac{1}{n}X_i^*A_{ij}X_i\right)\left(1+\frac{1}{n}X_j^*A_{ij}X_j\right)^2 \geq 1;$$

so in order to prove (3.9), it is enough to show that

$$D'_{23} = \frac{1}{n^4}\sum_{i\neq j}\left|X_i^*A_{ij}X_jX_i^*A_{ij}X_jX_j^*A_{ij}X_i\right|I_{ij} \to 0 \quad \text{a.s.} \tag{3.10}$$

Since on the event E_{ij}, the norm of A_{ij} is bounded by $1/\epsilon$, due to Lemma 4.6, we know that

$$
\begin{aligned}
E\left|X_i^* A_{ij} I_{ij} X_j\right|^r &= E\left[E\left(\left|X_i^* A_{ij} I_{ij} X_j\right|^r \mid A_{ij}\right)\right] \\
&\le K_r n^{r/2} \quad \text{for any } r \ge 2,
\end{aligned}
$$

and similarly

$$
E\left|X_j^* A_{ij} I_{ij} X_i\right|^r \le K_r n^{r/2} \quad \text{for any } r \ge 2,
$$

where K_r is a constant only depending on r. Making use of these orders, together with Hölder inequality, we can compute the third absolute moment of D_{23}', and the result is given by

$$
E\left|D_{23}'\right|^3 = O\left(n^{-3/2}\right).
$$

Therefore, (3.10) follows from the Borel–Cantelli lemma.

Remark When we move from D_{23} to $\bar{D}_{23}$, we restrict the matrix A_{ij} on the event $E_{ij} = \{\lambda_{\min}(S_{ij}) > a - \epsilon\}$ so that its norm could be controlled by $1/\epsilon$. Note that $A_{ij} I_{ij}$ is still independent of X_i and X_j. In general, for any index set $\Lambda \subset [n]$, we could restrict the matrix A_Λ on the event $E_\Lambda = \{\lambda_{\min}(S_\Lambda) > a - \epsilon\}$ to control its norm. In the subsequent investigation, we need to use this kind of restriction again and again. Fortunately, due to Lemma 4.5 and the Borel–Cantelli lemma, none of these restrictions will change the strong limit under consideration. Instead of writing down $A_\Lambda I_{E_\Lambda}$ every time, we will use A_Λ and make the following assumptions:

Assumption (i). A_Λ is nonnegative definite, and $0 < \|A_\Lambda\| \le (1/\epsilon)$ for any finite index set Λ;

Assumption (ii). A_Λ and $\{X_i, i \in \Lambda\}$ are independent.

3.4.2 *Remove X_i and X_j in the Denominator*

We first show that in the denominator, $X_i^* A_{ij} X_i$ can be replaced by $\operatorname{tr} A_{ij}$. Let

$$
T_4 = \frac{1}{n^2} \sum_{i \ne j} \frac{X_i^* A_{ij} X_j}{(1 + \frac{1}{n} \operatorname{tr} A_{ij})(1 + \frac{1}{n} X_j^* A_{ij} X_j)}.
$$

Our task is to show that

$$
D_{43} = T_4 - T_3 \to 0 \quad \text{a.s.}
$$

We write D_{43} as

$$
D_{43} = \frac{1}{n^2} \sum_{i \ne j} \frac{X_i^* A_{ij} X_j (\frac{1}{n} X_i^* A_{ij} X_i - \frac{1}{n} \operatorname{tr} A_{ij})}{(1 + \frac{1}{n} \operatorname{tr} A_{ij})(1 + \frac{1}{n} X_i^* A_{ij} X_i)(1 + \frac{1}{n} X_j^* A_{ij} X_j)}.
$$

It is convenient to consider

$$
\bar{D}_{43} = \frac{1}{n^2} \sum_{i \ne j} \frac{X_i^* A_{ij} X_j (\frac{1}{n} X_i^* A_{ij} X_i - \frac{1}{n} \operatorname{tr} A_{ij})}{(1 + \frac{1}{n} \operatorname{tr} A_{ij})(1 + \frac{1}{n} \operatorname{tr} A_{ij})(1 + \frac{1}{n} X_j^* A_{ij} X_j)}
$$

instead of D_{43}. The reason is that by computation (again due to Lemma 4.6 and the Hölder inequality) we can find that

$$E|D_{43} - \bar{D}_{43}|^3 = O\big(n^{-3/2}\big),$$

and hence $D_{43} - \bar{D}_{43} \to 0$ almost surely. For similar reasons, we can consider simply

$$\tilde{D}_{43} = \frac{1}{n^2} \sum_{i \neq j} \frac{X_i^* A_{ij} X_j (\frac{1}{n} X_i^* A_{ij} X_i - \frac{1}{n} \operatorname{tr} A_{ij})}{(1 + \frac{1}{n} \operatorname{tr} A_{ij})^3}.$$

For simplicity, we use $S(i, j)$ to denote

$$S(i, j) = \frac{X_i^* A_{ij} X_j (\frac{1}{n} X_i^* A_{ij} X_i - \frac{1}{n} \operatorname{tr} A_{ij})}{(1 + \frac{1}{n} \operatorname{tr} A_{ij})^3}.$$

Now we will compute the fourth absolute moment of $\tilde{D}_{43}$, and we first expand $E|\tilde{D}_{43}|^4$ as

$$E|\tilde{D}_{43}|^4 = \frac{1}{n^8} \sum_{\substack{i_1 \neq j_1 \\ i_2 \neq j_2 \\ i_3 \neq j_3 \\ i_4 \neq j_4}} E\big[S(i_1, j_1)\bar{S}(i_2, j_2)S(i_3, j_3)\bar{S}(i_4, j_4)\big],$$

where $\bar{S}(i, j)$ is the complex conjugate of $S(i, j)$. Totally, we need to use eight subindices here, although some of them may have the same value. According to Assumption (i), the Hölder inequality, and Lemma 4.6, we know that

$$\begin{aligned} E\big|S(i, j)\big|^r &\le \big(E\big(X_i^* A_{ij} X_j\big)^{2r}\big)^{1/2} \left(E\left(\frac{1}{n} X_i^* A_{ij} X_i - \frac{1}{n} \operatorname{tr} A_{ij}\right)^{2r}\right)^{1/2} \\ &\le K_r n^{r/2} n^{-r/2} = O(1) \quad \text{for any } r \ge 2. \end{aligned} \tag{3.11}$$

Now it is easy to verify that the contribution of those terms with less than or equal to six different subindices in $E|\tilde{D}_{43}|^4$ is of order $O(n^{-2})$, which is summable. Therefore, in order to show that

$$\tilde{D}_{43} \to 0 \quad \text{a.s.,} \tag{3.12}$$

we only need to consider the following two cases.

Case 1: Seven different indices

When there are seven different indices, the summand has finite different forms depending on which two indices are the same. We only deal with the following kind of summands here:

$$E\big[S(i_1, j_1)\bar{S}(i_2, j_2)S(i_3, j_3)\bar{S}(i_3, j_4)\big]. \tag{3.13}$$

The other forms of the summand can be treated similarly.

Now, for convenience, we define a useful operator Δ_i. Let $f(A_\Lambda)$ be a function which involves the matrix A_Λ, and assume that $i \notin \Lambda$. Δ_i is defined as

$$\Delta_i\big(f(A_\Lambda)\big) = f(A_\Lambda) - f(A_{\Lambda\cup\{i\}}).$$

For the term in (3.13), in the ideal situation, if X_{j_1} is independent of other parts, then the conditional expectation of X_{j_1}, given all the other observations, is zero, which means that the expectation in (3.13) is zero. Unfortunately, this is not the case, because X_{j_1} is involved in the matrices $A_{i_2 j_2}$, $A_{i_3 j_3}$, and $A_{i_3 j_4}$. However, motivated by this idea, we can consider the following term:

$$\begin{aligned} E\Bigg(&\frac{X_{i_1}^* A_{i_1 j_1} X_{j_1}(\frac{1}{n} X_{i_1}^* A_{i_1 j_1} X_{i_1} - \frac{1}{n}\operatorname{tr} A_{i_1 j_1})}{(1+\frac{1}{n}\operatorname{tr} A_{i_1 j_1})^3} \\ &\times \frac{X_{j_2}^* A_{i_2 j_1 j_2} X_{i_2}(\frac{1}{n} X_{i_2}^* A_{i_2 j_1 j_2} X_{i_2} - \frac{1}{n}\operatorname{tr} A_{i_2 j_1 j_2})}{(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_1 j_2})^3} \\ &\times \frac{X_{i_3}^* A_{i_3 j_1 j_3} X_{j_3}(\frac{1}{n} X_{i_3}^* A_{i_3 j_1 j_3} X_{i_3} - \frac{1}{n}\operatorname{tr} A_{i_3 j_1 j_3})}{(1+\frac{1}{n}\operatorname{tr} A_{i_3 j_1 j_3})^3} \\ &\times \frac{X_{j_4}^* A_{i_3 j_1 j_4} X_{i_3}(\frac{1}{n} X_{i_3}^* A_{i_3 j_1 j_4} X_{i_3} - \frac{1}{n}\operatorname{tr} A_{i_3 j_1 j_4})}{(1+\frac{1}{n}\operatorname{tr} A_{i_3 j_1 j_4})^3}\Bigg). \end{aligned} \tag{3.14}$$

For simplicity, we introduce the notation $S_k(i,j)$:

$$S_k(i,j) = \frac{X_i^* A_{ijk} X_j(\frac{1}{n} X_i^* A_{ijk} X_i - \frac{1}{n}\operatorname{tr} A_{ijk})}{(1+\frac{1}{n}\operatorname{tr} A_{ijk})^3},$$

and (3.14) can be written as

$$E\big[S(i_1,j_1)\bar{S}_{j_1}(i_2,j_2)S_{j_1}(i_3,j_3)\bar{S}_{j_1}(i_3,j_4)\big].$$

Note that now all the matrices involved in (3.14) are independent of X_{j_1}, so the expectation in (3.14) is zero, and hence subtracting (3.14) from (3.13) will not change the expectation in (3.13). This leads us to consider

$$\begin{aligned} &S(i_1,j_1)\bar{S}(i_2,j_2)S(i_3,j_3)\bar{S}(i_3,j_4) - S(i_1,j_1)\bar{S}_{j_1}(i_2,j_2)S_{j_1}(i_3,j_3)\bar{S}_{j_1}(i_3,j_4) \\ &\quad = S(i_1,j_1)\big[\Delta_{j_1}\bar{S}(i_2,j_2)\big]S(i_3,j_3)\bar{S}(i_3,j_4) \\ &\qquad + S(i_1,j_1)\bar{S}_{j_1}(i_2,j_2)\big[\Delta_{j_1}S(i_3,j_3)\big]\bar{S}(i_3,j_4) \\ &\qquad + S(i_1,j_1)\bar{S}_{j_1}(i_2,j_2)S_{j_1}(i_3,j_3)\big[\Delta_{j_1}\bar{S}(i_3,j_4)\big]. \end{aligned} \tag{3.15}$$

The explicit formula of $[\Delta_{j_1}S(i_2,j_2)]$ can be expressed as

$$
\begin{aligned}
\Delta_{j_1} S(i_2, j_2) &= S(i_2, j_2) - S_{j_1}(i_2, j_2)\\
&= \frac{[-\Delta_{j_1}(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2})^3] X_{i_2}^* A_{i_2 j_2} X_{j_2} (\frac{1}{n} X_{i_2}^* A_{i_2 j_2} X_{i_2} - \frac{1}{n}\operatorname{tr} A_{i_2 j_2})}{(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2})^3 (1+\frac{1}{n}\operatorname{tr} A_{i_2 j_1 j_2})^3}\\
&\quad + \frac{[\Delta_{j_1}(X_{i_2}^* A_{i_2 j_2} X_{j_2})](\frac{1}{n} X_{i_2}^* A_{i_2 j_2} X_{i_2} - \frac{1}{n}\operatorname{tr} A_{i_2 j_2})}{(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_1 j_2})^3}\\
&\quad + \frac{X_{i_2}^* A_{i_2 j_1 j_2} X_{j_2} [\Delta_{j_1}(\frac{1}{n} X_{i_2}^* A_{i_2 j_2} X_{i_2} - \frac{1}{n}\operatorname{tr} A_{i_2 j_2})]}{(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_1 j_2})^3},
\end{aligned}
\tag{3.16}
$$

where

$$
\begin{aligned}
\Delta_{j_1}\left(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2}\right)^3 &= \left[\Delta_{j_1}\left(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2}\right)\right]\left(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2}\right)^2\\
&\quad + \left(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_1 j_2}\right)\left[\Delta_{j_1}\left(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2}\right)\right]\\
&\quad \times \left(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2}\right)\\
&\quad + \left(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_1 j_2}\right)^2\left[\Delta_{j_1}(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2})\right],
\end{aligned}
\tag{3.17}
$$

$$
\Delta_{j_1}\left(1+\frac{1}{n}\operatorname{tr} A_{i_2 j_2}\right) = -\frac{1}{n}\frac{\frac{1}{n} X_{j_1}^* A_{i_2 j_1 j_2}^2 X_{j_1}}{1+\frac{1}{n} X_{j_1}^* A_{i_2 j_1 j_2} X_{j_1}},
\tag{3.18}
$$

$$
\Delta_{j_1}\left(X_{i_2}^* A_{i_2 j_2} X_{j_2}\right) = -\frac{1}{n}\frac{X_{i_2}^* A_{i_2 j_1 j_2} X_{j_1} X_{j_1}^* A_{i_2 j_1 j_2} X_{j_2}}{1+\frac{1}{n} X_{j_1}^* A_{i_2 j_1 j_2} X_{j_1}},
\tag{3.19}
$$

$$
\begin{aligned}
\Delta_{j_1}\left(\frac{1}{n} X_{i_2}^* A_{i_2 j_2} X_{i_2} - \frac{1}{n}\operatorname{tr} A_{i_2 j_2}\right) &= \frac{1}{n}\frac{\frac{1}{n} X_{j_1}^* A_{i_2 j_1 j_2}^2 X_{j_1}}{1+\frac{1}{n} X_{j_1}^* A_{i_2 j_1 j_2} X_{j_1}}\\
&\quad - \frac{1}{n^2}\frac{X_{i_2}^* A_{i_2 j_1 j_2} X_{j_1} X_{j_1}^* A_{i_2 j_1 j_2} X_{i_2}}{1+\frac{1}{n} X_{j_1}^* A_{i_2 j_1 j_2} X_{j_1}}.
\end{aligned}
\tag{3.20}
$$

Combining (3.16) to (3.20), again by Lemma 4.6 and the Hölder inequality, we find that

$$
E\left|\Delta_{j_1} S(i_2, j_2)\right|^r = O\left(n^{-r/2}\right) \quad \text{for any } r \geq 2.
\tag{3.21}
$$

We can verify that $E|\Delta_{j_1} S(i_3, j_3)|^r$ and $E|\Delta_{j_1} S(i_3, j_4)|^r$ also have the above order. Therefore, by (3.11) and (3.15), the order of (3.13) is $O(n^{-1/2})$. Furthermore, the same order can be verified for all the other terms with seven different subindices. Since the number of the terms with seven different sub-indices is at most $O(n^7)$, we know that the contribution of these terms in $E|\tilde{D}_{43}|^4$ is of order $O(n^{-3/2})$, which is summable.

Note that if we compute the order of (3.13) directly by Lemma 4.6 and the Hölder inequality, the result will be $O(1)$. By taking the difference between (3.13) and (3.14), the order is reduced by $n^{1/2}$. Since this order reduction method will be used frequently in the subsequent discussions, we present it in a slightly more general form than that in Lemma 3.2.

In order to simplify the long expressions, we introduce another operator Ψ_i and extend the definition of Δ_i. Let $f(A_{\Lambda_1}, A_{\Lambda_2}, \ldots, A_{\Lambda_m})$ be a function which involves the matrix $A_{\Lambda_1}, A_{\Lambda_2}, \ldots, A_{\Lambda_m}$; Ψ_i is defined as

$$\Psi_i\bigl(f(A_{\Lambda_1}, A_{\Lambda_2}, \ldots, A_{\Lambda_m})\bigr) = f(A_{\Lambda_1'}, A_{\Lambda_2'}, \ldots, A_{\Lambda_m'}),$$

where

$$\Lambda_k' = \Lambda_k \cup \{i\}, \quad k = 1, 2, \ldots, m.$$

As an example of this operator, note that $\Psi_k S(i, j) = S_k(i, j)$. Δ_i is defined as

$$\Delta_i f = f - \Psi_i f.$$

Consider the following typical expressions:

$$X_i^* A_\Lambda X_j \quad \text{with } i, j \in \Lambda, \tag{3.22}$$

$$\frac{1}{n}\bigl(X_i^* A_\Lambda X_i - \operatorname{tr} A_\Lambda\bigr) \quad \text{with } i \in \Lambda, \tag{3.23}$$

$$\frac{1}{1 + \frac{1}{n}\operatorname{tr} A_\Lambda}, \tag{3.24}$$

$$\frac{1}{1 + \frac{1}{n} X_i^* A_\Lambda X_i} \quad \text{with } i \in \Lambda, \tag{3.25}$$

$$X_i^* A_{\Lambda_1} A_{\Lambda_2} X_i \quad \text{with } i \in \Lambda_1 \cap \Lambda_2, \tag{3.26}$$

$$X_i^* A_{\Lambda_1} A_{\Lambda_2} X_j \quad \text{with } i \in \Lambda_1,\ j \in \Lambda_1 \cap \Lambda_2, \tag{3.27}$$

where $i \neq j$. The orders of (3.22)–(3.26) can be easily computed by Lemma 4.6. For (3.27), we use the following inequality about the operator norm:

$$\begin{aligned} E\bigl|X_i^* A_{\Lambda_1} A_{\Lambda_2} X_j\bigr|^r &\le E\bigl(\|X_i^*\| \|A_{\Lambda_1}\| \|A_{\Lambda_2}\| \|X_j\|\bigr)^r \\ &\le \left(C\sqrt{n} \cdot \frac{1}{\epsilon^2} \cdot C\sqrt{n}\right)^r = O\bigl(n^r\bigr) \end{aligned}$$

for $r \ge 1$, because the entries of X_i and X_j are uniformly bounded. Now we pick some index $k \neq i, j$ and apply the operator Δ_k to them. For (3.22)–(3.24), the results can be found from (3.18)–(3.20). The other results are the following:

$$\Delta_k\left(\frac{1}{1 + \frac{1}{n} X_i^* A_\Lambda X_i}\right) = \frac{1}{n^2} \frac{X_i^* A_{\Lambda\cup\{k\}} X_k X_k^* A_{\Lambda\cup\{k\}} X_i}{(1 + \frac{1}{n} X_i^* A_\Lambda X_i)(1 + \frac{1}{n} X_i^* A_{\Lambda\cup\{k\}} X_i)} I_{\{k\notin\Lambda\}}, \tag{3.28}$$

$$\begin{aligned}
&\Delta_k\big(X_i^* A_{\Lambda_1} A_{\Lambda_2} X_j\big)\\
&\quad = X_i^*(A_{\Lambda_1} - A_{\Lambda_1\cup\{k\}})A_{\Lambda_2}X_j + X_i^* A_{\Lambda_1\cup\{k\}}(A_{\Lambda_2} - A_{\Lambda_2\cup\{k\}})X_j\\
&\quad = -\frac{\frac{1}{n}X_i^* A_{\Lambda_1\cup\{k\}} X_k X_k^* A_{\Lambda_1\cup\{k\}} A_{\Lambda_2} X_j}{1+\frac{1}{n}X_k^* A_{\Lambda_1\cup\{k\}} X_k} I_{\{k\notin\Lambda_1\}}\\
&\qquad - \frac{\frac{1}{n}X_i^* A_{\Lambda_1\cup\{k\}} A_{\Lambda_2\cup\{k\}} X_k X_k^* A_{\Lambda_2\cup\{k\}} X_j}{1+\frac{1}{n}X_k^* A_{\Lambda_2\cup\{k\}} X_k} I_{\{k\notin\Lambda_2\}},
\end{aligned}\tag{3.29}$$

where $I_{k\notin\Lambda}$ is 1 when $k \notin \Lambda$ and 0 otherwise. The one for (3.26) is very similar with (3.29). We can also compute the orders of these differences and summarize the results in the following table. For $r \geq 2$,

C	$E\lvert C\rvert^r$	$E\lvert\Delta_k C\rvert^r$
(3.22)	$O(n^{r/2})$	$O(1)$
(3.23)	$O(n^{-r/2})$	$O(n^{-r})$
(3.24)	$O(1)$	$O(n^{-r})$
(3.25)	$O(1)$	$O(n^{-r})$
(3.26)	$O(n^r)$	$O(n^{r/2})$
(3.27)	$O(n^r)$	$O(n^{r/2})$

Now we are ready to present the lemma.

Lemma 3.2 *Let X be the up-left $p \times n$ corner of a double array $\{X_{uv} : u, v = 1, 2, \ldots\}$ of i.i.d. bounded complex r.v.s with zero mean and unit variance. Assume that Assumptions (i) and (ii) hold. We consider a typical product,*

$$\prod_{k=1}^m C_k, \tag{3.30}$$

which satisfies

(i) *each C_k is (or the conjugate of) one of (3.22)–(3.27), so there are at most two indices i_k, j_k and at most two index sets $\Lambda_{k1}, \Lambda_{k2}$ corresponding to i, j and Λ_1, Λ_2 in (3.22)–(3.27) for each C_k;*
(ii) *there is an index, say j_1, which satisfies*
   (a) *j_1 comes from C_1, and C_1 has the form (3.22): $X_{i_1}^* A_{\Lambda_1} X_{j_1}$;*
   (b) *j_1 appears only once among $\{i_k, j_k, k = 1, 2, \ldots, m\}$.*

Then the order of $\lvert E\prod_{k=1}^m C_k\rvert$ can be reduced by $n^{1/2}$ from the one computed by Lemma 4.6 directly.

Proof Because of (ii), we know that

$$E\left[\prod_{k=1}^m \Psi_{j_1} C_k\right] = 0.$$

Therefore, we can express the expectation as the following telescoping sum:

$$E\prod_{k=1}^{m} C_k = E\left(\prod_{k=1}^{m} C_k - \prod_{k=1}^{m} \Psi_{j_1} C_k\right)$$
$$= E\left[\sum_{k=1}^{m}\left(\prod_{l<k} \Psi_{j_1} C_l\right)\Delta_{j_1} C_k\left(\prod_{l>k} C_l\right)\right]. \tag{3.31}$$

Let us consider the kth summand in (3.31). For $l < k$, we know that $E|\Psi_{j_1} C_l|^r$ has the same order with $E|C_l|^r$. For the difference term $\Delta_{j_1} C_k$, if $j_1 \in \Lambda_{k1} \cap \Lambda_{k2}$, it is 0 according to the definition of the operator Δ_{j_1}; otherwise the orders are given in the preceding table, which completes the proof together with the Hölder inequality. □

Remark Let us make two observations here. By applying Δ_k to (3.30), we get a sum of several products.

Observation (i). Each product still satisfies condition (i) of Lemma 3.2.
Observation (ii). If there is another index, say j_2, which also satisfies condition (ii), then for each product, j_2 still satisfies condition (ii).

Therefore, it is possible to apply Lemma 3.2 several times when computing $|E\prod_{k=1}^{m} C_k|$.

Case 2: Eight different indices

Now we consider the terms with eight different indices which have the form

$$E\left[S(i_1, j_1)\bar{S}(i_2, j_2)S(i_3, j_3)\bar{S}(i_4, j_4)\right]. \tag{3.32}$$

Due to (3.11),

$$\left|E\left[S(i_1, j_1)\bar{S}(i_2, j_2)S(i_3, j_3)\bar{S}(i_4, j_4)\right]\right| = O(1).$$

However, $[S(i_1, j_1)\bar{S}(i_2, j_2)S(i_3, j_3)\bar{S}(i_4, j_4)]$ satisfies condition (i) of Lemma 3.2, and j_1, j_2, j_3, j_4 all satisfy condition (ii). Therefore, we can apply Lemma 3.2 four times to obtain

$$\left|E\left[S(i_1, j_1)\bar{S}(i_2, j_2)S(i_3, j_3)\bar{S}(i_4, j_4)\right]\right| = O\left(n^{-2}\right). \tag{3.33}$$

Now the contribution of all the terms with eight different indices in $E|\tilde{D}_{43}|^4$ is of order $O(n^{-2})$, which is summable.

With the results from the above two cases, we can complete the proof of (3.12), which leads us to consider

$$T_4 = \frac{1}{n^2}\sum_{i\neq j} \frac{X_i^* A_{ij} X_j}{(1+\frac{1}{n}\operatorname{tr} A_{ij})(1+\frac{1}{n}X_j^* A_{ij} X_j)}.$$

Similarly, in the denominator of T_4, $X_j^* A_{ij} X_j$ can also be replaced by $\operatorname{tr} A_{ij}$. In the following parts, we will focus on

$$T_5 = \frac{1}{n^2} \sum_{i \neq j} \frac{X_i^* A_{ij} X_j}{(1 + \frac{1}{n} \operatorname{tr} A_{ij})^2}, \tag{3.34}$$

and our task is to show that

$$T_5 \to 0 \quad \text{a.s.} \tag{3.35}$$

3.4.3 Proof of (3.35)

Let

$$T(i, j) = \frac{X_i^* A_{ij} X_j}{(1 + \frac{1}{n} \operatorname{tr} A_{ij})^2}.$$

Then we can simplify the expression of T_5:

$$T_5 = \frac{1}{n^2} \sum_{i \neq j} T(i, j).$$

Since T_5 is real, in order to prove (3.35), it suffices to show that $E(T_5^4)$ is summable. We expand T_5^4 as

$$T_5^4 = \frac{1}{n^8} \sum_{\substack{i_1 \neq j_1 \\ i_2 \neq j_2 \\ i_3 \neq j_3 \\ i_4 \neq j_4}} T(i_1, j_1) T(i_2, j_2) T(i_3, j_3) T(i_4, j_4).$$

By Lemma 4.6, we know that

$$E|T(i, j)|^4 = O(n^2);$$

then by the Hölder inequality, we have

$$\bigl|ET(i_1, j_1)T(i_2, j_2)T(i_3, j_3)T(i_4, j_4)\bigr| = O(n^2). \tag{3.36}$$

This means that the contribution by terms with less than or equal to 4 indices is of order $O(n^{-2})$, which is summable. Now we consider the following 4 cases.

- When there are 5 different indices, there are at least two of them which appear only once. So we can apply Lemma 3.2 two times to reduce the order of (3.4.3) to be $O(n)$. Then the contribution of terms with 5 different indices are of order $O(n^{-2})$, which is summable.
- When there are 6 different indices, there are at least four of them which appear only once, so the order of (3.4.3) can be reduced to $O(1)$, and the contribution of these terms is of order $O(n^{-2})$.

- When there are 7 different indices, there are six of them which appear only once, so the order of (3.4.3) can be reduced to $O(n^{-1})$, and the contribution of these terms is of order $O(n^{-2})$.
- When there are 8 different indices, we can apply Lemma 3.2 eight times to reduce the order of (3.4.3) to $O(n^{-2})$, and the contribution of these terms is also of order $O(n^{-2})$.

Therefore, we have proved (3.35)

Now we are in the position to conclude the proof of Theorem 1.4. By the discussion in Sects. 3.4.1 and 3.4.2, we know that (3.35) leads to (3.8), which is

$$\lim T_2 = 0 \quad \text{a.s.}$$

In Sect. 3.3, we show that (see (3.7))

$$\limsup T_1 < 1 \quad \text{a.s.}$$

Combining these two results, we have succeeded in proving (3.4), that is,

$$\limsup \bar{X}^*(\mathbb{S} - \lambda I)^{-1}\bar{X} < 1 \quad \text{a.s. } \forall 0 < \lambda < a$$

when $0 < c < 1$. As a result of Lemma 3.1, this means that we have established (3.1), that is,

$$\liminf \lambda_{\min}(S) \ge a = (1 - \sqrt{c})^2 \quad \text{a.s.}$$

when $0 < c < 1$. The proof of Theorem 1.4 is now completed.

4 Some Lemmas

We first introduce a classical result in linear algebra. In fact, it is Corollary 7.3.8 of Horn and Johnson [5].

Lemma 4.1 *Suppose that A and B are $m \times n$ complex matrices, and let $q = \min\{m, n\}$. If $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_q$ are the singular values of A and $\tau_1 \ge \tau_2 \ge \cdots \ge \tau_q$ are the singular values of B, then*

$$|\sigma_i - \tau_i| \le \|A - B\| \quad \text{for all } i = 1, 2, \ldots, q,$$

where $\|A\|$ denotes the spectrum norm of the complex matrix A, which is defined as the largest singular value of A.

The following *rank inequality*, which helps us to measure the difference between two empirical distributions, was proved in Silverstein and Bai [8].

Lemma 4.2 *For $n \times n$ Hermitian matrices A and B,*

$$\left\| F^A - F^B \right\| \le \frac{1}{n}\operatorname{rank}(A - B),$$

where $\|f\| = \sup_x |f(x)|$.

In the subsequent lemma, we list three equalities, which are used frequently in our proof. They can be proved by simple linear algebra.

Lemma 4.3 *Suppose that A is an $n \times n$ complex matrix and $\beta \in \mathbb{C}^n$, where $\mathbb{C}$ is the complex plane. If both A and $(A + \beta\beta^*)$ are nonsingular, then $1 + \beta^* A^{-1}\beta \neq 0$, and*

$$\left(A + \beta\beta^*\right)^{-1}\beta = \frac{A^{-1}\beta}{1 + \beta^* A^{-1}\beta}, \tag{4.1}$$

$$\beta^*\left(A + \beta\beta^*\right)^{-1} = \frac{\beta^* A^{-1}}{1 + \beta^* A^{-1}\beta}, \tag{4.2}$$

$$A^{-1} - \left(A + \beta\beta^*\right)^{-1} = \frac{A^{-1}\beta\beta^* A^{-1}}{1 + \beta^* A^{-1}\beta}. \tag{4.3}$$

The following lemma, which could be viewed as a generalization of the Marcinkiewicz strong law of large numbers (see [7], pp. 242–243), was proved in [4].

Lemma 4.4 *Let $\{X_{ij}, i, j = 1, 2, \ldots\}$ be a double array of i.i.d. complex r.v.s. Let $\alpha > 1/2$, $\beta \geq 0$, and $M > 0$ be constants. Then, as $n \to \infty$,*

$$\max_{j \leq Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} (X_{ij} - c) \right| \to 0 \quad a.s.$$

if and only if the following conditions are true:

(i) $E|X_{11}|^{(1+\beta)/\alpha} < \infty$;

(ii) $c = \begin{cases} EX_{11} & \text{if } \alpha \leq 1, \\ \text{any value in } \mathbb{C} & \text{if } \alpha > 1. \end{cases}$

The next result was proved in [1] (see (1.9b) and the theorem in the Appendix).

Lemma 4.5 *Under the conditions of Theorem 1.1, if the underlying variables are uniformly bounded, then we have, for $c \in (0, 1)$,*

$$P\left(\lambda_{\min}(\mathbb{S}) \leq \eta\right) = o\left(n^{-l}\right)$$

for any $0 < \eta < (1 - \sqrt{c})^2$ and any positive l.

The first two inequalities in the following lemma were originally proved in [2] (Lemma 2.7 and Lemma A.1) by martingale inequalities. We also state some simple consequences for our purpose.

Lemma 4.6 *Let $Y = (Y_1, Y_2, \ldots, Y_n)^T$ be a random vector containing i.i.d. standardized complex entries, B be an $n \times n$ nonnegative definite Hermitian matrix, and C be an $n \times n$ complex matrix. Then*

$$E\left|Y^* B Y\right|^p \leq K_p\left((\operatorname{tr} B)^p + E|Y_1|^{2p} \operatorname{tr} B^p\right) \quad \text{for any } p \geq 1, \tag{4.4}$$

$$E\left|Y^*CY - \operatorname{tr} C\right|^p \le K_p\big(\big(E|Y_1|^4 \operatorname{tr} CC^*\big)^{p/2} + E|Y_1|^{2p} \operatorname{tr}\big(CC^*\big)^{p/2}\big) \quad \textit{for any } p \ge 2. \tag{4.5}$$

In addition, if all the entries of Y are bounded by a constant M_1 and the norm of B and C are bounded by another constant M_2, then we have the following immediate consequences:

$$E\left|Y^*AY\right|^p \le K_p n^p \quad \textit{for any } p \ge 1, \tag{4.6}$$

$$E\left|Y^*AY - \operatorname{tr} A\right|^p \le K_p n^{p/2} \quad \textit{for any } p \ge 2, \tag{4.7}$$

$$E\left|Y^*CY\right|^p \le K_p n^p \quad \textit{for any } p \ge 1, \tag{4.8}$$

and if Z is i.i.d. with Y, then

$$E\left|Y^*CZ\right|^p \le K_p n^{p/2} \quad \textit{for any } p \ge 2. \tag{4.9}$$

These K_p are constants only depending on p, and they do not need to have the same value.

Acknowledgements The author would like to thank Zhidong Bai for his helpful suggestions. We would also like to thank two referees for their very helpful comments and criticisms, which have helped to improve the paper a great deal.

References

1. Bai, Z.D., Silverstein, J.W.: CLT for linear spectral statistics of large-dimensional sample covariance matrices. Ann. Probab. **32**, 553–605 (2004)
2. Bai, Z.D., Silverstein, J.W.: No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. Ann. Probab. **26**, 316–345 (1998)
3. Bai, Z.D., Silverstein, J.W.: Spectral Analysis of Large Dimensional Random Matrices. Mathematics Monograph Series, vol. 2. Science Press, Beijing (2006)
4. Bai, Z.D., Yin, Y.Q.: Limit of the smallest eigenvalue of large dimensional covariance matrix. Ann. Probab. **21**, 1275–1294 (1993)
5. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1985)
6. Jiang, T.: The limiting distributions of eigenvalues of sample correlation matrices. Sankhyā **66**, 35–48 (2004)
7. Loève, M.: Probability Theory, 3rd edn. Van Nostrand, Princeton (1963)
8. Silverstein, J.W., Bai, Z.D.: On the empirical distribution of eigenvalues of a class of large dimensional random matrices. J. Multivar. Anal. **54**, 175–192 (1995)
9. Yin, Y.Q., Bai, Z.D., Krishnaiah, P.R.: On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. Probab. Theory Relat. Fields **78**, 509–521 (1988)