The Annals of Probability
1993, Vol. 21, No. 3, 1275–1294

LIMIT OF THE SMALLEST EIGENVALUE OF A LARGE DIMENSIONAL SAMPLE COVARIANCE MATRIX

BY Z. D. BAI AND Y. Q. YIN

Temple University and University of Massachusetts, Lowell

In this paper, the authors show that the smallest (if $p \le n$) or the $(p-n+1)$-th smallest (if $p > n$) eigenvalue of a sample covariance matrix of the form $(1/n)XX'$ tends almost surely to the limit $(1-\sqrt{y})^2$ as $n \to \infty$ and $p/n \to y \in (0, \infty)$, where X is a $p \times n$ matrix with iid entries with mean zero, variance 1 and fourth moment finite. Also, as a by-product, it is shown that the almost sure limit of the largest eigenvalue is $(1+\sqrt{y})^2$, a known result obtained by Yin, Bai and Krishnaiah. The present approach gives a unified treatment for both the extreme eigenvalues of large sample covariance matrices.

1. Introduction. Suppose A is a $p \times p$ matrix with real eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$. Then the spectral distribution of the matrix A is defined by

$$F^A(x) = \frac{1}{p}\#\{i \le p: \lambda_i \le x\}.$$

We are especially interested in the matrix of the form $S = S_n = (1/n)XX'$, where $X = X_n = (X_{ij})$, and where X_{ij}, $i = 1, \dots, p$; $j = 1, \dots, n$, are iid random variables with zero mean and variance σ^2. We will call it a sample covariance matrix.

There are many studies on the limiting behavior of the spectral distributions of sample covariance matrices. For example, under various conditions, Grenander and Silverstein (1977), Jonsson (1982) and Wachter (1978) prove that the spectral distribution $F^S(x)$ converges to

$$F_y(x) = (1 - y^{-1} \wedge 1)\delta(x) + \int_{-\infty}^{x} f_y(u)\,du,$$

where $\delta(x)$ is the distribution function with mass 1 at 0, and

$$f_y(x) = \begin{cases} \dfrac{1}{2\pi y x}\sqrt{(x-a)(b-x)}, & \text{if } a < x < b, \\ 0, & \text{otherwise,} \end{cases}$$

as $p = p(n) \to \infty$, $n \to \infty$ and $p/n \to y \in (0, \infty)$. Here

$$a = a(y) = \left(1 - \sqrt{y}\right)^2\sigma^2, \qquad b = b(y) = \left(1 + \sqrt{y}\right)^2\sigma^2.$$

As a consequence of Yin (1986), if the second moment of X_{11} is finite, the above convergence holds with probability 1. Note that σ^2 appears in the

---

Received April 1990; revised June 1992.
AMS 1991 *subject classifications*. Primary 60F15; secondary 62H99.
Key words and phrases. Random matrix, sample covariance matrix, smallest eigenvalue of a random matrix, spectral radius.

definition of $F_y(x)$. Thus, the condition on the existence of the second moment of X_{11} is also necessary.

It is not hard to see that if $F^S(x)$ converges to $F_y(x)$ a.s., then

$$\liminf \max_{1 \le i \le p} \lambda_i \ge b \quad \text{a.s.}$$

However, the converse assertion

$$\limsup \max_{1 \le i \le p} \lambda_i \le b \quad \text{a.s.}$$

is not trivial.

The first success in establishing the last relation ($\le$) was made by Geman (1980), who did it under the condition that

$$E|X_{11}|^k \le Mk^{\alpha k}, \quad \text{for some } M > 0,\ \alpha > 0, \text{ and all } k \ge 3.$$

Yin, Bai and Krishnaiah (1988) established the same conclusion under the condition that

$$E|X_{11}|^4 < \infty,$$

which is further proved to be necessary in Bai, Silverstein and Yin (1988) by showing that

$$E|X_{11}|^4 = \infty \quad \Rightarrow \quad \limsup \max_{1 \le i \le n} \lambda_i = \infty \quad \text{a.s.}$$

It is much harder to study the convergence of the smallest eigenvalue of a sample covariance matrix. The first breakthrough was given by Silverstein (1985), who proved that if $X_{11} \sim N(0,1)$, then

$$\min_{1 \le i \le p} \lambda_i \to a(y) \quad \text{a.s.},$$

as $p \to \infty$, $p/n \to y < 1$. However, it is hard to use his method to get the general result, since his proof depends heavily on the normality hypothesis.

In this paper, we shall prove the following theorems.

THEOREM 1. *Let* $[X_{uv};\ u, v = 1, 2, \dots]$ *be a double array of independent and identically distributed (iid) random variables with zero mean and unit variance, and let*

$$X = [X_{uv}\colon u = 1, \dots, p;\ v = 1, \dots, n], \qquad S = (1/n)XX'.$$

Then, if $E|X_{11}|^4 < \infty$, *as* $n \to \infty$, $p \to \infty$, $p/n \to y \in (0,1)$,

$$-2\sqrt{y} \le \liminf \lambda_{\min}(S - (1+y)I) \le \limsup \lambda_{\max}(S - (1+y)I) \le 2\sqrt{y} \quad a.s.$$

As an easy corollary of Theorem 1, we have the following.

THEOREM 2. *Under the conditions of Theorem* 1, *as* $n \to \infty$, $p \to \infty$, $p/n \to y \in (0,1)$,

$$\lim \lambda_{\min} = \left(1 - \sqrt{y}\right)^2 \quad a.s. \tag{1.1}$$

$$\lim \lambda_{\max} = \left(1 + \sqrt{y}\right)^2 \quad a.s. \tag{1.2}$$

REMARK 1. The assertion (1.1) is trivially true for $y = 1$. If $y > 1$, then $p > n$ for all large p, and the $p - n$ smallest eigenvalues of S must be 0. In this case, (1.1) is no longer true as it stands. However, if we redefine $\lambda_{\min}$ to be the $(p - n + 1)$-th smallest eigenvalue of S, then (1.1) is still true. In fact, for the case of $y > 1$, define $S^* = (1/p)X'X$ and $y^* = 1/y \in (0,1)$. By Theorem 2, we have

$$\lambda_{\min}(S^*) \to \left(1 - \sqrt{y^*}\right)^2 = \frac{1}{y}\left(1 - \sqrt{y}\right)^2 \quad \text{a.s.},$$

Therefore,

$$\lambda_{\min}(S) = \frac{p}{n}\lambda_{\min}(S^*) \to \left(1 - \sqrt{y}\right)^2 \quad \text{a.s.}$$

By a similar argument, one may easily show that the conclusion (1.2) is also true for $y \geq 1$.

REMARK 2. The conclusion (1.2) has already been proved in Yin, Bai and Krishnaiah (1988). Here, we prove it by an alternative approach as a by-product of our Theorem 1, which is the key step for the proof of the limit of the smallest eigenvalue.

REMARK 3. From the proof given later, one can see that if the condition $EX_{11}^4 < \infty$ is weakened to

$$n^2 P\left(|X_{11}| > \sqrt{n}\right) \to 0, \tag{1.3}$$

then the two limit relations (1.1) and (1.2) hold in probability.

In fact, if (1.3) is true, then for each $\varepsilon > 0$,

$$E|X_{11}|^{4-\varepsilon} < \infty$$

and there exists a sequence of positive constants $\delta = \delta_n \to 0$ such that

$$n^2 P\left(|X_{11}| > \delta\sqrt{n}\right) \to 0.$$

Here, we may assume that the rate of $\delta \to 0$ is sufficiently slow.

As done in Silverstein (1989) for the largest eigenvalue, one may prove that the probability of the event that the smallest eigenvalue of the sample covariance matrix constructed by the truncated variables at $\delta\sqrt{n}$ differs from the original by a quantity controlled by

$$n^2 P\left(|X_{11}| > \delta\sqrt{n}\right).$$

Also, employing von Neumann's inequality, one may conclude that the difference between the square root of the smallest eigenvalue of the truncated sample covariance matrix and that of the truncated and then centralized sample covariance matrix is controlled by

$$\sqrt{p}\,\left|EX_{11}I_{[|X_{11}| > \delta\sqrt{n}]}\right| \to 0$$

(For details of the application of von Neumann's inequality, see the beginning of Section 2.) Then the truncated and then centralized variables satisfy the conditions given in (2.1), and the desired result can be proved by the same lines of the proof of the main result.

REMARK 4. In Girko (1989), an attempt is made to prove the weak convergence of the smallest eigenvalue under a stronger condition. However, this proof contains some serious errors. Regardless, the result we get here is strong convergence under much weaker conditions.

2. Some lemmas. In this section we prove several lemmas. By the truncation lemma proved in Yin, Bai and Krishnaiah (1988), one may assume that the entries of X_n have already been truncated at $\delta\sqrt{n}$ for some slowly varying $\delta = \delta_n \to 0$. Let

$$V_{uv} = X_{uv} I(|X_{uv}| \le \delta\sqrt{n}) - EX_{uv} I(|X_{uv}| \le \delta\sqrt{n}).$$

In 1937, von Neumann proved that

$$\sum_{i=1}^{p} \lambda_i \tau_i \ge \operatorname{tr}(A'B),$$

if A and B are $p \times n$ matrices with singular values $\lambda_1 \ge \cdots \ge \lambda_p$ and $\tau_1 \ge \cdots \ge \tau_p$, respectively. Then, using von Neumann's inequality, we have

$$\begin{aligned}
&\left|\lambda_{\min}^{1/2}(n^{-1}\hat{X}_n\hat{X}_n') - \lambda_{\min}^{1/2}(n^{-1}V_nV_n')\right|^2 \\
&\quad \le \sum_{k=1}^{p} \left(\lambda_k^{1/2}(n^{-1}\hat{X}_n\hat{X}_n') - \lambda_k^{1/2}(n^{-1}V_nV_n')\right)^2 \\
&\quad \le \frac{1}{n}\operatorname{tr}(\hat{X}_n - V_n)(\hat{X}_n - V_n)' \\
&\quad \le pE^2|X_{11}|I_{[|X_{11}|>\delta\sqrt{n}]} \to 0,
\end{aligned}$$

where $\hat{X}_n$ and V are $n \times p$ matrices with (u,v)-th entries $X_{u,v}I_{[|X_{uv}|\le\delta\sqrt{n}]}$ and V_{uv}, respectively. In fact, the above convergence is true provided $n\delta^3 \to 0$. Therefore, we can assume that for each n the entries $X_{uv} = X_{uv}(n)$ of the matrix X_n are iid and satisfy

$$\begin{aligned}
&EX_{uv} = 0, \qquad EX_{uv}^2 \le 1 \quad \text{and} \quad EX_{uv}^2 \to 1 \quad \text{as } n \to \infty, \\
&E|X_{uv}|^l \le (\delta\sqrt{n})^{l-3}, \quad \text{for } l \ge 3,
\end{aligned} \tag{2.1}$$

where $\delta = \delta_n \to 0$ is nonrandom and sufficiently slow.

Replacing all diagonal elements of S by 0, we get a matrix T. Define $T(0) = I_p$ and $T(1) = T$, and, for each $l \ge 2$, let $T(l) = (T_{ab}(l))$ be a $p \times p$ matrix with

$$T_{ab}(l) = n^{-l} {\sum}' X_{av_1}X_{u_1v_1}X_{u_1v_2}X_{u_2v_2} \cdots X_{u_{l-1}v_l}X_{bv_l}, \tag{2.2}$$

where the summation Σ' runs over all integers $u_1, \dots, u_{l-1}$ from the set

$\{1, \ldots, p\}$ and $v_1, \ldots, v_l$ from $\{1, \ldots, n\}$, subject to the conditions that

$$\begin{aligned} &a \neq u_1, \qquad u_1 \neq u_2, \ldots, u_{l-1} \neq b, \\ &v_1 \neq v_2, \qquad v_2 \neq v_3, \ldots, v_{l-1} \neq v_l. \end{aligned} \tag{2.3}$$

LEMMA 1.

$$\limsup_{n \to \infty} \|T(l)\| \leq (2l + 1)(l + 1) y^{(l-1)/2} \quad a.s.,$$

where the matrix norm used here and throughout the paper is the operator norm, that is, the largest singular value of the matrix.

PROOF. For integers $u_0, \ldots, u_r$ from $\{1, \ldots, p\}$ and integers $v_1, \ldots, v_r$ from $\{1, \ldots, n\}$, define a graph

$$G\{u_0, v_1, u_1, \ldots, v_r, u_r\}$$

as follows. Let $u_0, \ldots, u_r$ be plotted on a straight line, and let $v_1, \ldots, v_r$ be plotted on another straight line. The two lines are supposed to be parallel. $u_0, \ldots, u_r, v_1, \ldots, v_r$ are vertices. The graph has $2r$ edges: $e_1, \ldots, e_{2r}$. The two ends of e_{2i-1} are u_{i-1}, v_i and those of e_{2i} are v_i, u_i. Two edges are said to be coincident if they have the same set of end vertices.

An edge e_i is said to be single up to e_j, $j \geq i$, if it does not coincide with any $e_1, \ldots, e_j$ other than itself.

If $e_{2i-1} = u_{i-1}v_i$ $(e_{2i} = v_i u_i)$ is such that

$$v_i \notin \{v_1, \ldots, v_{i-1}\} \ (u_i \notin \{u_1, \ldots, u_{i-1}\}),$$

then e_{2i-1} (e_{2i}) is called a column (row) innovation. T_1 denotes the set of all innovations. If e_j is such that there is an innovation e_i, $i < j$, and e_j is the first one to coincide with e_i, then we say $e_j \in T_3$. Other edges constitute the set T_4. Thus, edges are partitioned into three disjoint classes: T_1, T_3, T_4. Edges which are not innovations and single up to themselves form a set T_2. It is obvious that $T_2 \subset T_4$.

If e_j is a T_3 edge and there is more than one single (up to e_{j-1}) innovation among $e_1, \ldots, e_{j-1}$ which is adjacent to e_j, we say that e_j is a regular T_3 edge. We can prove that for a regular T_3 edge, the number of such innovations is bounded by $t + 1$, where t is the maximum number of noncoincident T_4 edges [see Yin, Bai and Krishnaiah (1988)], and the number of regular T_3 edges is bounded by twice the number of T_4 edges [see Yin, Bai and Krishnaiah (1988)].

In order to establish Lemma 1, we estimate $E \operatorname{tr} T^{2m}(l)$. By (2.2),

$$\begin{aligned} \operatorname{tr} T^{2m}(l) &= \sum T_{b_1 b_2}(l) T_{b_2 b_3}(l) \cdots T_{b_{2m} b_1}(l) \\ &= n^{-2ml} \sum_{(b_i)} \sum_1{}' \sum_2{}' \cdots \sum_{2m}{}' X_{b_1 v'_1} X_{u'_1 v'_1} X_{u'_1 v'_2} X_{u'_2 v'_2} \\ &\quad \cdots X_{u'_{l-1} v'_l} X_{b_2 v'_l} X_{b_2 v''_1} \cdots X_{b_{2m} v_l^{(2m-1)}} X_{b_{2m} v_1^{(2m)}} X_{u_1^{(2m)} v_2^{(2m)}} \\ &\quad \cdots X_{u_{l-1}^{(2m)} v_l^{(2m)}} X_{b_1 v_l^{(2m)}}. \end{aligned} \tag{2.4}$$

Here the summation Σ_i' is taken with respect to $u_1^{(i)},\dots,u_{l-1}^{(i)}$ running over $\{1,\dots,p\}$ and $v_1^{(i)},\dots,v_l^{(i)}$ running over $\{1,\dots,n\}$ subject to the condition that

$$b_i \neq u_1^{(i)}, \qquad u_1^{(i)} \neq u_2^{(i)},\dots,u_{l-1}^{(i)} \neq b_{i+1};$$
$$v_1^{(i)} \neq v_2^{(i)},\dots,v_{l-1}^{(i)} \neq v_l^{(i)},$$

for each $i = 1,2,\dots,2m$; and $\Sigma_{(b_i)}$ is the summation for b_i, $i = 1,\dots,2m$, running over $\{1,\dots,p\}$.

Now we can consider the sum (2.4) as a sum over all graphs G of the form

$$G = G\big[b_1, v_1', u_1', v_2',\dots,u_{l-1}', v_l', b_2, v_1'', u_1'',\dots,v_l'', b_3,\dots,b_{2m}, v_1^{(2m)}, u_1^{(2m)},\dots,u_{l-1}^{(2m)}, v_{l-1}^{(2m)}, b_1\big]. \tag{2.5}$$

At first we partition all these graphs into isomorphism classes. We take the sum within each isomorphism class, and then take the sum of all such sums over all isomorphism classes. (Here we say that two graphs are isomorphic, if equality of two vertex indices in one graph implies the equality of the corresponding vertex indices in the other graph.)

Within each isomorphism class, the ways of arranging the three different types of edges are all the same. In other words, if two graphs of the form (2.5) are isomorphic, the corresponding edges must have the same type. However, two graphs with the same arrangements of types are not necessarily isomorphic.

We claim that

$$\begin{aligned} &E\big[\operatorname{tr} T^{2m}(l)\big] \\ &\quad \le n^{-2ml}\sum{}^{*}\left(\prod_{i=1}^{2m}\binom{2l+1}{2a_i+1}\right)\binom{4ml-k}{k}\binom{4ml}{t} \\ &\qquad \times t^{4ml-2k}(t+1)^{8lm-4k}n^{k+1}\left(\frac{p}{n}\right)^{(k-t-2m)/2}\big(\sqrt{n}\,\delta\big)^{4ml-2k-t}, \end{aligned} \tag{2.6}$$

where the summation Σ^* is taken with respect to k, t and a_i, $i = 1,\dots,2m$, under some restrictions to be specified. Here:

(i) k $(= 1,\dots,2ml)$ is the total number of innovations in G.

(ii) t $(= 0,\dots,4ml-2k)$ is the number of noncoincident T_4 edges in G.

(iii) a_i $(= 0,\dots,l)$ is the number of pairs of consecutive edges (e,e') in the graph

$$G_i = G\big[b_i, v_1^{(i)}, u_1^{(i)},\dots,u_{l-1}^{(i)}, v_l^{(i)}, b_{i+1}\big] \tag{2.7}$$

in which e is an innovation but e' is not.

Now we explain the reasons why (2.6) is true:

(i) The factor n^{-2ml} is obvious.

(ii) If there is an overall single edge in a graph G, then the mean of the product of X_{ij} corresponding to this graph [denoted by $EX(G)$] is zero. Thus, in any graph corresponding to a nonzero term, we have $k \le 2ml$.

(iii) The number of T_3 edges is also k. Hence the number of T_4 edges is $4ml - 2k$, and $t \le 4ml - 2k$.

(iv) The graph G is split into $2m$ subgraphs $G_1, \dots, G_{2m}$ defined in (2.6). Obviously, $0 \le a_i \le l$.

(v) The number of sequences of consecutive innovations in G_i is either a_i or $a_i + 1$ (the latter happens when the last edge in G_i is an innovation). Hence the number of ways of arranging these consecutive sequences in G_i is at most

$$\binom{2l}{2a_i} + \binom{2l}{2a_i + 1} = \binom{2l+1}{2a_i+1}.$$

(vi) Given the position of innovations, there are at most $\binom{4ml-k}{k}$ ways to arrange T_3 edges.

(vii) Given the positions of innovations and T_3 edges, there are at most $\binom{4ml}{t}$ ways to choose t distinguishable positions for the t noncoincident T_4 edges. When these positions have been chosen, there are at most t^{4ml-2k} ways to distribute the $4ml - k$ T_4 edges into these t positions.

(viii) Yin, Bai and Krishnaiah (1988) proved that a regular T_3 edge e has at most $t + 1$ choices and that the number of regular T_3 edges is dominated by $2(4ml - 2k)$. Therefore, there are at most $(t+1)^{8ml-4k}$ different ways to arrange the T_3 edges.

(ix) Let r and c be the number of row and column innovations. Then $r + c = k$, and the number of graphs G within the isomorphism class is bounded by $n^c p^{r+1} = n^{k+1}(p/n)^{r+1}$.

Suppose that in the pair (e, e'), e is an innovation in G_i and e' is not an innovation in G_i. Then it is easy to see that e' is of type T_4 and is single up to itself. Therefore,

$$t \ge \sum_{i=1}^{2m} a_i. \tag{2.8}$$

In each G_i, there are at most $a_i + 1$ sequences of consecutive innovations. Therefore,

$$|r - c| \le \sum_{i=1}^{2m} a_i + 2m. \tag{2.9}$$

Since $r + c = k$, by (2.8) and (2.9) we obtain

$$r \ge \tfrac{1}{2}(k - t) - m,$$

by which we conclude that (by noticing that we can assume $p/n < 1$)

$$n^{k+1}\left(\frac{p}{n}\right)^{r+1} \le n^{k+1}\left(\frac{p}{n}\right)^{(k-t-2m)/2} \tag{2.10}$$

(x) By the same argument as in Yin, Bai and Krishnaiah (1988), we have

$$|EX(G)| \le \left(\sqrt{n}\,\delta\right)^{4ml-2k-t} \tag{2.11}$$

The above 10 points are discussed for $t > 0$, but they are still valid when $t = 0$, if we take the convention that $0^0 = 1$ for the term t^{4ml-2k}. Thus we have established (2.6). Now we begin to simplify the estimate (2.6). Note that

$$\binom{2l+1}{2a_i+1} \le (2l+1)^{2a_i+1}.$$

By (2.8), we have

$$\prod_{i=1}^{2m} \binom{2l+1}{2a_i+1} \le (2l+1)^{2\Sigma a_i + 2m} \le (2l+1)^{2t+2m} \tag{2.12}$$

The number of choices of the a_i's does not exceed $(l+1)^{2m}$. Therefore, by the elementary inequality $a^{-t}t^b \le (b/\log a)^b$, for all $a > 1$, $b > 0$, $t \ge 1$, and letting m be such that $m/\log n \to \infty$, $m\delta^{1/3}/\log n \to 0$ and $m/(\delta\sqrt{n}) \to 0$, we obtain from (2.6), for sufficiently large n,

$$\begin{aligned} E[\operatorname{tr} T^{2m}(l)] &\le \sum_{k=1}^{2ml} \sum_{t=0}^{4ml-2k} n(2l+1)^{2m}(l+1)^{2m} \\ &\quad \times \binom{4ml-k}{k} \left(\frac{4ml(2l+1)^2}{\sqrt{p}\,\delta} \right)^t \\ &\quad \times (t+1)^{12ml-6k} \left(\frac{p}{n} \right)^{(k-2m)/2} \delta^{4ml-2k} \\ &\le n^2(2l+1)^{2m}(l+1)^{2m} \left(\frac{p}{n} \right)^{-m} \\ &\quad \times \sum_{k=1}^{2ml} \binom{4ml-k}{k} \left(\frac{12ml-6k}{\left|\log[36ml^3/(\sqrt{p}\,\delta)]\right|} \right)^{12ml-6k} \\ &\quad \times \left(\frac{p}{n} \right)^{k/2} \delta^{4ml-2k} \\ &\le n^2((2l+1)(l+1))^{2m} \left(\frac{p}{n} \right)^{-m} \\ &\quad \times \sum_{k=1}^{2ml} \binom{4ml}{2k} \left(\frac{24ml\delta^{1/3}}{\frac{1}{2}\log n} \right)^{12ml-6k} \left(\frac{p}{n} \right)^{k/2} \\ &\le n^2((2l+1)(l+1))^{2m} \left(\frac{p}{n} \right)^{-m} \\ &\quad \times \left[\left(\frac{p}{n} \right)^{1/4} + \left(\frac{24ml\delta^{1/3}}{\frac{1}{2}\log n} \right)^3 \right]^{4ml}. \end{aligned} \tag{2.13}$$

Here, in the proof of the second inequality, we have used the fact that

$$\left(\frac{4ml(2l+1)}{\sqrt{p}\,\delta}\right)^{t} \le n\left(\frac{4ml(2l+1)}{\sqrt{p}\,\delta}\right)^{t+1}.$$

If we choose z to be

$$z = (2l+1)(l+1)y^{(l-1)/2}(1+\varepsilon),$$

where $\varepsilon > 0$, then

$$\sum z^{-2m} E \operatorname{tr} T^{2m}(l) < \infty.$$

Thus the lemma is proved. □

LEMMA 2. *Let $\{X_{ij},\ i, j, = 1, 2, \ldots,\}$ be a double array of iid random variables and let $\alpha > \frac{1}{2}$, $\beta \ge 0$ and $M > 0$ be constants. Then as $n \to \infty$,*

$$\max_{j \le Mn^{\beta}} \left| n^{-\alpha} \sum_{i=1}^{n} (X_{ij} - c) \right| \to 0 \quad a.s., \tag{2.14}$$

if and only if the following hold:

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

(ii) $$c = \begin{cases} EX_{11}, & \text{if } \alpha \le 1, \\ \text{any number}, & \text{if } \alpha > 1. \end{cases}$$

The proof of Lemma 2 is given in the Appendix.

LEMMA 3. *If $f > 0$ is an integer and $X^{(f)}$ is the $p \times n$ matrix $[X_{uv}^{f}]$, then*

$$\limsup \lambda_{\max}\{n^{-f} X^{(f)} X^{(f)\prime}\} \le 7 \quad a.s.$$

PROOF. When $f = 1$, we have

$$\begin{aligned} \left\| \frac{1}{n} X^{(1)} X^{(1)\prime} \right\| &\le \|T(1)\| + \left\| \operatorname{diag}\left[\frac{1}{n} \sum_{j=1}^{n} X_{ij}^2,\ i = 1, \ldots, p \right] \right\| \\ &\le \|T(1)\| + \frac{1}{n} \max_{i \le p} \sum_{j=1}^{n} X_{ij}^2. \end{aligned}$$

So, by Lemmas 1 and 2, we get

$$\limsup_n \|n^{-1} X^{(1)} X^{(1)\prime}\| \le 7 \quad \text{a.s.}$$

For $f = 2$, by the Geršgorin theorem and Lemma 2, we have

$$\begin{aligned}\lambda_{\max}\{n^{-2}X^{(2)}X^{(2)\prime}\} &\le \max_i n^{-2}\sum_{j=1}^{n} X_{ij}^4 + \max_i n^{-2}\sum_{k\ne i}\sum_{j=1}^{n} X_{ij}^2X_{kj}^2\\ &\le \max_i n^{-2}\sum_{j=1}^{n} X_{ij}^4 + \left(\max_i n^{-1}\sum_{j=1}^{n} X_{ij}^2\right)\left(\max_j n^{-1}\sum_{k=1}^{p} X_{kj}^2\right)\\ &\to y \quad \text{a.s.}\end{aligned}$$

For $f > 2$, the conclusion of Lemma 4 is stronger than that of this lemma. □

REMARK 5. For the case of $f = 1$, the result can easily follow from a result in Yin, Bai and Krishnaiah (1988) with a bound of 4. Here, we prove this lemma by using our Lemma 1 to avoid the use of the result of Yin, Bai and Krishnaiah (1988), since we want to get a unified treatment for limits of both the largest and the smallest eigenvalues of large covariance matrices, as we remarked after the statement of Theorem 2.

In the following, we say that a matrix is $o(1)$ if its operator norm tends to 0.

LEMMA 4. *Let $f > 2$ be an integer, and let $X^{(f)}$ be as defined in Lemma 3. Then*

$$\|n^{-f/2}X^{(f)}\| = o(1) \quad a.s.$$

PROOF. Note that, by Lemma 2, we have

$$\|n^{-f/2}X^{(f)}\|^2 \le n^{-f}\sum_{u,v} X_{uv}^{2f} \to 0 \quad \text{a.s.,}$$

since $E|X_{11}^{2f}|^{2/f} = EX_{11}^4 < \infty$. The proof is complete. □

LEMMA 5. *Let H be a $p \times n$ matrix. If $\|H\|$ is bounded a.s. and $f > 2$, or $H = o(1)$ a.s. and $f \ge 1$, then the following matrices are $o(1)$ a.s.:*

$$A(k,f) = \left(n^{-k+1-f/2}\sum_{\substack{a\ne u_1\ne\cdots\ne u_{k-1}\ne b\\ v_1\ne\cdots\ne v_k}} H_{av_1}X_{u_1v_1}^f X_{u_1v_2}\cdots X_{u_{k-1},v_k}X_{bv_k}\right);$$

$$B(k,f) = \left(n^{-k+1-f/2}\sum_{\substack{u_1\ne\cdots\ne u_{k-1}\ne b\\ v_1\ne\cdots\ne v_k}} H_{av_1}X_{u_1v_1}^f X_{u_1v_2}X_{u_2v_2}\cdots X_{bv_k}\right).$$

PROOF. For the case of $k = 1$, by Lemma 3, we have

$$B(1,f) = \left(n^{-f/2}\sum_{v_1} H_{av_1}X_{bv_1}^f\right) = n^{-f/2}HX^{(f)\prime} = Hn^{-f/2}X^{(f)\prime} = o(1)$$

and

$$A(1,f) = B(1,f) - n^{-f/2}\operatorname{diag}\left[\sum_{v_1} H_{av_1}X^f_{av_1}\right]$$
$$= B(1,f) - \operatorname{diag}(B(1,f)) = o(1),$$

where diag(B) denotes the diagonal matrix whose diagonal elements are the same as the matrix B. Here, in the proof of diag(B(1, f)) = $o(1)$, we have used the fact that $\|\operatorname{diag}(B)\| \le \|B\|$.

For the general case $k > 1$, by Lemma 1 and the assumptions, we have

$$(B_{ab}) = \left(n^{-k+1-f/2} \sum_{\substack{u_1 \neq \cdots \neq u_{k-1} \neq b \\ v_1 \neq \cdots \neq v_k}} H_{av_1} X^f_{u_1v_1} X_{u_1v_2} X_{u_2v_2} \cdots X_{bv_k}\right)$$
$$= \left(n^{-k+1-f/2} \sum_{\substack{u_1 \neq \cdots \neq u_{k-1} \neq b \\ v_2 \neq \cdots \neq v_k}} \left(\sum_{v_1} H_{av_1} X^f_{u_1v_1}\right) X_{u_1v_2} X_{u_2v_2} \cdots X_{bv_k}\right)$$
$$- \left(n^{-k+1-f/2} \sum_{\substack{u_1 \neq \cdots \neq u_{k-1} \neq b \\ v_2 \neq \cdots \neq v_k}} H_{av_2} X^{f+1}_{u_1v_2} X_{u_2v_2} \cdots X_{bv_k}\right)$$
$$= n^{-f/2} H X^{(f)\prime} T(k-1) - C = o(1) - C.$$

However, the entries of the matrix C satisfy

$$C_{ab} = n^{-k+1-f/2} \sum_{\substack{u_2 \neq \cdots \neq u_{k-1} \neq b \\ v_2 \neq \cdots \neq v_k}} H_{av_2} \sum_{u_1} X^{f+1}_{u_1v_2} X_{u_2v_2} \cdots X_{bv_k}$$
$$- n^{-k+1-f/2} \sum_{\substack{u_2 \neq \cdots \neq u_{k-1} \neq b \\ v_2 \neq \cdots \neq v_k}} H_{av_2} X^{f+2}_{u_2v_2} X_{u_2v_3} X_{u_3v_3} \cdots X_{bv_k}$$
$$= D_{ab} - E_{ab}.$$

Note that the matrix E is of the form of B with a smaller k-index. Thus, by the induction hypothesis, we have $E = o(1)$ a.s. The matrix D also has the same form as B with

$$1,\ k-1,\ H^* = \left(n^{-(f+1)/2} H_{av} \sum_u X^{f+1}_{uv}\right)$$

in place of f, k, H_{av}. Evidently, by Lemma 2, we have

$$H^* = H \operatorname{diag}\left\{n^{-(f+1)/2} \sum_u X^{f+1}_{uv};\ v = 1, \dots, n\right\} = o(1).$$

Thus, $D = o(1)$ and hence $B(k, f) = o(1)$.

For matrices A, it is seen that

$$A_{ab} = B_{ab} - n^{-k+1-f/2} \sum_{\substack{a \neq u_2 \neq \cdots \neq u_{k-1} \neq b \\ v_2 \neq \cdots \neq v_k}} \left(\sum_{v_1} H_{av_1} X_{av_1}^f \right) X_{av_2} X_{u_2 v_2} \cdots X_{bv_k}$$

$$+ n^{-k+1-f/2} \sum_{\substack{a \neq u_2 \neq \cdots \neq u_{k-1} \neq b \\ v_2 \neq \cdots \neq v_k}} H_{av_2} X_{av_2}^{f+1} X_{u_2 v_2} \cdots X_{bv_k}$$

$$= B_{ab} - F_{ab} + K_{ab}.$$

Note that

$$\|F\| = \left\| \left[\operatorname{diag}(H n^{-f/2} X^{(f)'}) \right] T(k-1) \right\| \leq \left\| H(n^{-f/2} X^{(f)'}) \right\| \|T(k-1)\| = o(1).$$

It is easy to see that the matrix K is of the form of A with

$$1, k-1, \hat{H} = \left(n^{-(f+1)/2} H_{uv} X_{uv}^{f+1} \right)$$

in places of f, k and H. Note that $\hat{H} = H \circ (n^{(f+1)/2} X^{(f+1)'})$, where $A \circ B = (A_{uv} B_{uv})$ denotes the Hadamard product of the matrices A and B. By the fact that $\|A \circ B\| \leq \|A\| \|B\|$ [when A and B are Hermitian and positive definite, this inequality can be found in Marcus (1964); a simple proof for the general case is given in the Appendix], we have $\hat{H} = o(1)$. Hence, by the induction hypothesis, $K = o(1)$. Thus, we conclude that $A = o(1)$ and the proof of this lemma is complete. □

LEMMA 6. *The following matrices are* $o(1)$ *a.s.*:

$$A_1 = \left(n^{-k-1} \sum_{\substack{u_1 \neq \cdots \neq u_{k-1} \neq b \\ v_1 \neq \cdots \neq v_k}} X_{av_1}^3 X_{u_1 v_1} X_{u_1 v_2} \cdots X_{u_{k-1} v_k} X_{bv_k} \right),$$

$$A_2 = \left(n^{-k-1} \sum_{\substack{a \neq u_2 \neq \cdots \neq u_{k-1} \neq b \\ v_1 \neq \cdots \neq v_k}} X_{av_1}^4 X_{av_2} X_{u_2 v_2} \cdots X_{bv_k} \right),$$

$$A_3 = \left(n^{-k-1} \sum_{\substack{u_1 \neq \cdots \neq b \\ v_1 \neq \cdots \neq v_k}} X_{av_1} X_{u_1 v_1}^3 X_{u_1 v_2} X_{u_2 v_2} \cdots X_{bv_k} \right),$$

$$A_4 = \left(n^{-k} \sum_{\substack{u_1 \neq \cdots \neq b \\ v_1 \neq \cdots \neq v_k}} X_{av_1} Z_{v_1} X_{u_1 v_1} X_{u_1 v_2} X_{u_2 v_2} \cdots X_{bv_k} \right),$$

$$A_5 = \left(n^{-k} \sum_{\substack{a \neq u_2 \neq \cdots \neq b \\ v_1 \neq \cdots \neq v_k}} W_a X_{av_1} X_{u_1 v_1} X_{u_2 v_2} \cdots X_{bv_k} \right),$$

where $Z = \operatorname{diag}[Z_1, \ldots, Z_n] = o(1)$ *and* $W = \operatorname{diag}[W_1, \ldots, W_p] = o(1)$.

PROOF. All are simple consequences of Lemmas 2–5. For instance, A_1 is a matrix of type B as in Lemma 5, with $f = 1$ and $H = n^{-3/2}X^{(3)} = o(1)$ a.s.

LEMMA 7. *For all* $k \geq 1$,

$$TT(k) = T(k+1) + yT(k) + yT(k-1) + o(1) \quad a.s.,$$

where $T = T(1)$ *and* $T(k)$ *are defined in* (2.2).

PROOF. Recall that $T(0) = I$ and $T(1) = T$. We have

$$\begin{aligned}
TT(k) &= \left[n^{-k-1} \sum_{\substack{a \neq u_0 \neq \cdots \neq u_{k-1} \neq b \\ v_0, v_1 \neq \cdots \neq v_k}} X_{av_0} X_{u_0 v_0} X_{u_0 v_1} X_{u_1 v_1} \cdots X_{u_{k-1} u_k} X_{b v_k} \right] \\
&= T(k+1) \\
&\quad + \left[n^{-k-1} \sum_{\substack{a \neq u_0 \neq \cdots \neq u_{k-1} \neq b \\ v_1 \neq \cdots \neq v_k}} X_{av_1} X_{u_0 v_1}^2 X_{u_1 v_1} \cdots X_{u_{k-1} v_k} X_{b v_k} \right] \\
&= T(k+1) + n^{-k-1} \\
&\quad \times \left[\sum_{\substack{u_1 \neq \cdots \neq u_{k-1} \neq b \\ v_1 \neq \cdots \neq v_k}} X_{av_1} \sum_{u_0} X_{u_0 v_1}^2 (1 - \delta_{u_0 a})(1 - \delta_{u_0 u_1}) X_{u_1 v_1} \cdots X_{b v_k} \right] \\
&= T(k+1) + n^{-k-1} \left[\sum\nolimits^* X_{av_1} \sum_u X_{uv_1}^2 X_{u_1 v_1} \cdots X_{b v_k} \right] \\
&\quad - n^{-k-1} \left[\sum\nolimits^* X_{av_1}^3 X_{u_1 v_1} \cdots X_{b v_k} \right] \\
&\quad - n^{-k-1} \left[\sum\nolimits^* X_{av_1} X_{u_1 v_1}^3 X_{u_1 v_2} X_{u_2 v_2} \cdots X_{b v_k} \right] \\
&\quad + n^{-k-1} \left[\sum_{\substack{a \neq u_2 \neq \cdots \neq u_{k-1} \neq b \\ v_1 \neq \cdots \neq v_k}} X_{av_1}^4 X_{av_2} X_{u_2 v_2} \cdots X_{b v_k} \right] \\
&= T(k+1) + R_1 - R_2 - R_3 + R_4,
\end{aligned}$$

where δ_{uv} is the Kronecker delta and

$$\sum\nolimits^* \quad \text{stands for} \quad \sum_{\substack{u_1 \neq \cdots \neq u_{k-1} \neq b \\ v_1 \neq \cdots \neq v_k}}.$$

By Lemmas 1 and 6, and the fact that $EX_{uv}^2 \to 1$, we obtain

$$R_1 = \Big[n^{-k-1} p \sum{}^* X_{av_1} \frac{1}{p} \sum_u \big(X_{uv_1}^2 - 1 \big) X_{u_1v_1} X_{u_1v_2} X_{u_2v_2} \cdots X_{bv_k}$$

$$+ n^{-k-1} p \sum{}^* X_{av_1} X_{u_1v_1} \cdots X_{bv_k} \Big]$$

$$= o(1) + yT(k) + n^{-k-1} p \Bigg[\sum_{\substack{a \neq u_2 \neq \cdots \neq u_{k-1} \neq b \\ v_1 \neq \cdots \neq v_k}} X_{av_1}^2 X_{av_2} X_{u_2v_2} \cdots X_{bv_k} \Bigg]$$

$$= o(1) + yT(k) + n^{-k} p$$

$$\times \Bigg[\sum_{\substack{a \neq u_2 \neq \cdots \neq u_{k-1} \neq b \\ v_2 \neq \cdots \neq v_k}} n^{-1} \sum_{v_1} \big(X_{av_1}^2 - 1 \big) X_{av_2} X_{u_2v_2} \cdots X_{bv_k} \Bigg]$$

$$- n^{-k-1} p \Bigg[\sum_{\substack{a \neq u_2 \neq \cdots \neq b \\ v_2 \neq \cdots \neq v_k}} X_{av_2}^3 X_{u_2v_2} \cdots X_{bv_k} \Bigg] + yT(k-1) + o(1)$$

$$= yT(k) + yT(k-1) + o(1)$$

and

$$R_2 = o(1), \qquad R_3 = o(1), \qquad R_4 = o(1).$$

LEMMA 8. $(T - yI)^k = \sum_{r=0}^k (-1)^{r+1} T(r) \sum_{i=0}^{[(k-r)/2]} C_i(k, r) y^{k-r-i} + o(1)$ *a.s., where*

$$|C_k(k, r)| \leq 2^k. \tag{2.15}$$

PROOF. We proceed with our proof by induction on k. When $k = 1$, it is trivial. Now suppose the lemma is true for k. By Lemma 7, we have

$$(T - yI)^{k+1} = \sum_{r=1}^k (-1)^{r+1} \{ T(r+1) + yT(r) + yT(r-1) \}$$

$$\times \sum_{i=0}^{[(k-r)/2]} C_i(k, r) y^{k-r-i} - T \sum_{i=0}^{[k/2]} C_i(k, 0) y^{k-i}$$

$$- \sum_{r=0}^k (-1)^{r+1} T(r) \sum_{i=0}^{[(k-r)/2]} C_i(k, r) y^{k-r-i+1} + o(1)$$

$$
\begin{aligned}
&= \sum_{r=1}^{k+1} (-1)^{r+1} T(r) \sum_{i=0}^{[(k+1-r)/2]} \{-C_i(k, r-1)\} y^{k+1-r-i} \\
&\quad + \sum_{r=0}^{k-1} (-1)^{r+1} T(r) \sum_{i=1}^{[(k+1-r)/2]} \{-C_{i-1}(k, r+1)\} y^{k+1-r-i} \\
&\quad + I \sum_{i=1}^{[k/2]} C_i(k, 0) y^{k-i} + o(1) \\
&= \sum_{r=0}^{k+1} (-1)^{r+1} T(r) \\
&\quad \times \sum_{i=0}^{[(k+1-r)/2]} C_i(k+1, r) y^{k+1-r-i} + o(1) \quad \text{a.s.}
\end{aligned}
$$

Here $C_i(k+1, r)$ is a sum of one or two terms of the form $-C_i(k, r+1)$ and $-C_i(k, r-1)$, which are also quantities satisfying (2.15). By induction, we conclude that (2.15) is true for all fixed k. Thus, the proof of this lemma is complete. □

3. Proof of Theorem 1. By Lemma 2, with $\alpha = \beta = 1$, we have

$$
\|S - I - T\| = \max_{i \le p} \left| n^{-1} \sum_{i=1}^{n} \left(X_{ij}^2 - 1 \right) \right| \to 0 \quad \text{a.s.} \tag{3.1}
$$

Therefore, to prove Theorem 1, we need only to show that

$$
\limsup \|T - yI\| \le 2\sqrt{y} \quad \text{a.s.} \tag{3.2}
$$

By Lemmas 1 and 8, we have, for any fixed even integer k,

$$
\begin{aligned}
\limsup_{n \to \infty} \|T - yI\|^k &\le \sum_{r=0}^{k} C k^2 y^{r/2} [(k-r)/2] 2^k y^{(k-r)/2} \\
&\le C k^4 2^k y^{k/2} \quad \text{a.s.}
\end{aligned}
$$

Taking the k-th root on both sides of this inequality and then letting $k \to \infty$, we obtain (3.2). The proof of Theorem 1 is complete. □

APPENDIX

Proof of Lemma 2 (*Sufficiency*). Without loss of generality, assume that $c = 0$. Since, for $\varepsilon > 0$ and $N \ge 1$,

$$
\begin{aligned}
P\left\{ \max_{j \le Mn^\beta} \left| \frac{1}{n^\alpha} \sum_{i=1}^{n} X_{ij} \right| \ge \varepsilon, \text{i.o.} \right\} &\le \sum_{k \ge N} P\left\{ \max_{2^{k-1} < n \le 2^k} \max_{j \le M 2^{k\beta}} \left| \sum_{i=1}^{n} X_{ij} \right| \ge \varepsilon' 2^{k\alpha} \right\} \\
&\le \sum_{k \ge N} M 2^{k\beta} P\left\{ \max_{2^{k-1} < n \le 2^k} \left| \sum_{i=1}^{n} X_{i1} \right| \ge 2^{k\alpha} \varepsilon' \right\},
\end{aligned}
$$

where $\varepsilon' = 2^{-\alpha}\varepsilon$, to conclude that the probability on the left-hand side of this inequality is equal to zero, it is sufficient to show that

$$\sum_{k=1}^{\infty} 2^{k\beta} P\left\{ \max_{n \le 2^k} \left| \sum_{i=1}^{n} X_{i1} \right| \ge 2^{k\alpha}\varepsilon \right\} < \infty. \tag{A.1}$$

Let $Y_{ik} = X_{i1} I(|X_{i1}| < 2^{k\alpha})$ and $Z_{ik} = Y_{ik} - EY_{ik}$. Then $|Z_{ik}| \le 2^{k\alpha+1}$ and $EZ_{ik} = 0$.

Let g be an even integer such that $g(\alpha - \frac{1}{2}) > \beta + 2\alpha$. Then, by the submartingale inequality, we have

$$\begin{aligned} P\left\{ \max_{n \le 2^k} \left| \sum_{i=1}^{n} Z_{ik} \right| \ge \varepsilon 2^{k\alpha} \right\} &\le C 2^{-kg\alpha} E \left| \sum_{i=1}^{2^k} Z_{ik} \right|^g \\ &\le C 2^{-kg\alpha} \left\{ 2^k E|Z_{1k}^g| + 2^{kg/2} \left(EZ_{1k}^2 \right)^{g/2} \right\}, \end{aligned} \tag{A.2}$$

where the last inequality follows from Lemma A.1, which will be given later. Hence

$$\sum_{k=N}^{\infty} 2^{k\beta} P\left\{ \max_{n \le 2^k} \left| \sum_{i=1}^{n} Z_{ik} \right| \ge \varepsilon 2^{k\alpha} \right\} < \infty, \tag{A.3}$$

which follows from the following bounds:

$$\begin{aligned} \sum_{k=1}^{\infty} 2^{k\beta - kg\alpha + k} E|Z_{1k}^g| &\le C \sum_{k=1}^{\infty} 2^{k(\beta - g\alpha + 1)} E|X_{11}^g| I\left[|X_{11}| < 2^{k\alpha} \right] \\ &\le C \sum_{k=1}^{\infty} 2^{k(\beta - g\alpha + 1)} \\ &\quad \times \left[\sum_{l=1}^{k} E|X_{11}|^g I\left[2^{\alpha(l-1)} \le |X_{11}| < 2^{\alpha l} \right] + 1 \right] \\ &\le C \sum_{l=1}^{\infty} E|X_{11}|^{(\beta+1)/\alpha} I\left[2^{\alpha(-1)} \le |X_{11}| < 2^{\alpha l} \right] + C_1 < \infty \end{aligned}$$

[note that $g\alpha - \beta - 1 > g(\alpha - \frac{1}{2}) - (\beta + 2\alpha) > 0$] and, when $(1 + \beta)/\alpha \ge 2$ (hence $EZ_{1k}^2 \le EX_{11}^2 < \infty$),

$$\sum_{k=1}^{\infty} 2^{k\beta - kg\alpha + kg/2} \left(EX_{1k}^2 \right)^{g/2} \le C \sum_{k=1}^{\infty} 2^{k(\beta + 2\alpha - g(\alpha - 1/2))} < \infty.$$

If $(1+\beta)/\alpha < 2$, we have

$$
\begin{aligned}
&\sum_{k=1}^{\infty} 2^{k\beta - kg\alpha + gk/2}\left(EZ_{1k}^2\right)^{g/2} \\
&\le \sum_{k=1}^{\infty} 2^{k\beta - kg\alpha + kg/2}\left(EZ_{1k}^{(1+\beta)/\alpha + 2 - (1+\beta)/\alpha}\right)^{g/2-1}\left(EZ_{1k}^2\right) \\
&\le C\sum_{k=1}^{\infty} 2^{k\beta - kg\alpha + kg/2} 2^{k\alpha(2-(1+\beta)/\alpha)(g/2-1)} E\left[X_{11}^2 I\left(|X_{11}| < 2^{k\alpha}\right)\right] \\
&\le C\sum_{k=1}^{\infty} 2^{k(\beta - g\alpha + g/2 + g\alpha - 2\alpha - (1+\beta)g/2 + 1 + \beta)} \\
&\quad \times\left[\sum_{l=1}^{k} E\left(X_{11}^2 I\left(2^{(l-1)\alpha} \le |X_{11}| < 2^{l\alpha}\right)\right) + 1\right] \\
&\le C\sum_{k=1}^{\infty} 2^{k(\beta - 2\alpha - \beta g/2 + 1 + \beta)} \sum_{l=1}^{k} E\left(|X_{11}|^2 I\left(2^{(l-1)\alpha} \le |X_{11}| < 2^{l\alpha}\right)\right) + C_1 \\
&= C\sum_{l=1}^{\infty} 2^{l(\beta - 2\alpha - \beta g/2 + 1 + \beta)} E\left(|X_{11}|^2 I\left(2^{(l-1)\alpha} \le |X_{11}| < 2^{l\alpha}\right)\right) + C_1 \\
&\le C\sum_{l=1}^{\infty} 2^{l(\beta - 2\alpha - \beta g/2 + 1 + \beta)} E|X_{11}|^{(1+\beta)/\alpha + 2 - (1+\beta)/\alpha} \\
&\quad \times I\left(2^{(l-1)\alpha} \le |X_{11}| < 2^{l\alpha}\right) + C_1 \\
&\le C\sum_{l=1}^{\infty} 2^{l(\beta - 2\alpha - \beta g/2 + 1 + \beta)} 2^{l\alpha(2-(1+\beta)/\alpha)} E|X_{11}|^{(1+\beta)/\alpha} \\
&\quad \times I\left(2^{(l-1)\alpha} \le |X_{11}| < 2^{l\alpha}\right) + C_1 \\
&< \infty.
\end{aligned}
$$

Now we estimate EY_{ik} for large k. We have

$$
\begin{aligned}
\max_{n \le 2^k}\left|\sum_{i=1}^{n} EY_{ik}\right| &\le 2^k |EY_{1k}| \\
&\le \begin{cases} 2^k E|X_{11}| I\left[|X_{11}| \ge 2^{k\alpha}\right] & \\ \quad \le 2^{k(\alpha-\beta)} E|X_{11}|^{(\beta+1)/\alpha} I\left[|X_{11}| \ge 2^{k\alpha}\right], & \text{if } \alpha \le 1, \\ 2^k \log k + 2^{k(\alpha-\beta)} E|X_{11}|^{(\beta+1)/\alpha} I\left[|X_{11}| > \log k\right], & \text{if } \alpha > 1, \end{cases} \\
&\le 2^{-1}\varepsilon 2^{k\alpha}.
\end{aligned} \tag{A.4}
$$

Because (A.3) and (A.4) are true for all $\varepsilon > 0$, the inequality (A.3) is still true if the Z_{ik}'s are replaced by Y_{ik}'s.

Finally, since $E|X_{11}|^{(\beta+1)/\alpha} < \infty$, we have

$$\text{(A.5)} \qquad \sum_{k=1}^{\infty} 2^{k\beta} P\left[\bigcup_{i=1}^{2^k} \{|X_{i1}| \geq 2^{k\alpha}\}\right] \leq \sum_{k=1}^{\infty} 2^{k(\beta+1)} P\left[|X_{11}| \geq 2^{k\alpha}\right] < \infty.$$

Then, (A.1) follows from (A.3)–(A.5).

(*Necessity.*) If $\beta = 0$, then it reduces to the Marcinkiewicz's strong law of large numbers, which is well known. We only need to prove the case of $\beta > 0$. By (2.14) we have

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

and

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

By changing to a smaller M, we may change $(n-1)^\beta$ to n^β for simplicity. Thus, we obtain

$$\max_{j \leq Mn^\beta} n^{-\alpha} |X_{nj}| \to 0 \quad \text{a.s.},$$

which, together with the Borel–Cantelli lemma, implies

$$\sum_n P\left[\max_{j \leq Mn^\beta} |X_{nj}| \geq n^\alpha\right] < \infty.$$

By the convergence theorem for an infinite product, this above inequality is equivalent to

$$\begin{aligned} \prod_{n=1}^{\infty} P\left[\max_{j \leq Mn^\beta} |X_{nj}| < n^\alpha\right] &= \prod_{n=1}^{\infty} \left(P(|X_{11}| < n^\alpha)\right)^{[Mn^\beta]} \\ &= \prod_{n=1}^{\infty} \left(1 - P(|X_{11}| \geq n^\alpha)\right)^{[Mn^\beta]} > 0, \end{aligned}$$

which, by using the same theorem again, implies that

$$\sum_{n=1}^{\infty} Mn^\beta P(|X_{11}| \geq n^\alpha) < \infty.$$

This routinely implies $E|X_{11}|^{(\beta+1)/\alpha} < \infty$. Then, applying the sufficiency part, condition (ii) in Lemma 2 follows. □

LEMMA A.1. *Suppose $X_1, \dots, X_n$ are iid random variables with mean 0 and finite g-th moment, where $g \geq 2$ is an even integer. Then, for some constant $C = C(g)$,*

$$\text{(A.6)} \qquad E\left|\sum_{i=1}^{n} X_1\right|^g \leq C\left[nEX_1^g + n^{g/2}\left(EX_1^2\right)^{g/2}\right].$$

PROOF. We need only to show (A.6) for $g > 2$. We have

$$E\left|\sum_{i=1}^{n} X_i\right|^g \le \sum_{l=1}^{g/2} \sum_{\substack{i_1+\cdots+i_l=g\\ i_1\ge 2,\ldots,i_l\ge 2}} \frac{n!g!E|X_1|^{i_1}\cdots E|X_1|^{i_l}}{l!(n-l)!i_1!\cdots i_l!}. \tag{A.7}$$

By Hölder's inequality, we have

$$E|X_1^{i_t}| \le \left(EX_1^g\right)^{(i_t-1)/(g-2)}\left(EX_1^2\right)^{(g-i_t)/(g-2)},$$

which, together with (A.7), implies that

$$\begin{aligned} E\left|\sum_{i=1}^{n} X_i\right|^g &\le C\sum_{l=1}^{g/2}\left(nEX_1^g\right)^{(g-2l)/(g-2)}\left(nEX_1^2\right)^{g(l-1)/(g-2)}\\ &\le \begin{cases} C\left(nEX_1^2\right)^{g/2}, & \text{if } \left(nEX_1^2\right)^{g/(g-2)} \ge \left(nEX_1^g\right)^{2/(g-2)},\\ CnEX_1^g, & \text{otherwise}.\end{cases}\end{aligned}$$

This implies (A.6), and the proof is finished. □

LEMMA A.2. *Let A and B be two $n \times p$ matrices with entries A_{uv} and B_{uv}, respectively. Denote by $A \circ B$ the Hadamard product of the matrices A and B. Then*

$$\|A \circ B\| \le \|A\|\,\|B\|.$$

PROOF. Let $\mathbf{x} = (x_1, \ldots, x_p)'$ be a unit p-vector. Then the lemma follows from

$$\begin{aligned} \|A \circ B\mathbf{x}\|^2 &= \sum_{u=1}^{n}\left(\sum_{v=1}^{p} A_{uv}B_{uv}x_v\right)^2\\ &\le \sum_{k=1}^{n}\sum_{u=1}^{n}\left(\sum_{v=1}^{p} A_{kv}B_{uv}x_v\right)^2\\ &= \operatorname{tr}(BXA'AXB')\\ &= \operatorname{tr}(XA'AXB'B) \le \|A\|^2\|B\|^2\operatorname{tr}(X^2) = \|A\|^2\|B\|^2,\end{aligned}$$

where $X = \operatorname{diag}(\mathbf{x})$. □

Recently, it was found that this result was proved in Horn and Johnson [(1991), page 332]. Because the proof is very simple, we still keep it here.

REFERENCES

BAI, Z. D., SILVERSTEIN, J. W., and YIN, Y. Q. (1988). A note on the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivariate Anal.* **26** 166–168.

GEMAN, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261.

GIRKO, V. L. (1989). Asymptotics of the distribution of the spectrum of random matrices. *Russian Math. Surveys* **44** 3–36.

GRENANDER, U. and SILVERSTEIN, J. (1977). Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.* **32** 499–519.
HORN, R. A. and JOHNSON, C. R. (1991). *Topics in Matrix Analysis*. Cambridge Univ. Press.
JONSSON, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** 1–38.
MARCUS, M. (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Boston.
SILVERSTEIN, J. W. (1985). The smallest eigenvalue of a large dimensional Wishart matrix. *Ann. Probab.* **13** 1364–1368.
SILVERSTEIN, J. W. (1989). On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivariate Anal.* **30** 307–311.
VON NEUMANN, J. (1937). Some matrix inequalities and metrization of matric space. *Tomsk Univ. Rev.* **1** 286–300.
WACHTER, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** 1–18.
YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** 50–68.
YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521.

DEPARTMENT OF STATISTICS
TEMPLE UNIVERSITY
PHILADELPHIA, PENNSYLVANIA 19122

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MASSACHUSETTS, LOWELL
LOWELL, MASSACHUSETTS 01854