Bernoulli **18**(4), 2012, 1405–1420
DOI: 10.3150/11-BEJ381

Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero

B.B. CHEN[*] and G.M. PAN[**]

Division of Mathematical Sciences, School of Physical and Mathematical Science, Nanyang Technological University, Singapore. E-mail: [*]*chen0635@e.ntu.edu.sg;* [**]*gmpan@ntu.edu.sg*

Let $\mathbf{X}_p = (\mathbf{s}_1, \ldots, \mathbf{s}_n) = (X_{ij})_{p\times n}$ where X_{ij}'s are independent and identically distributed (i.i.d.) random variables with $EX_{11} = 0$, $EX_{11}^2 = 1$ and $EX_{11}^4 < \infty$. It is showed that the largest eigenvalue of the random matrix $\mathbf{A}_p = \frac{1}{2\sqrt{np}}(\mathbf{X}_p\mathbf{X}_p' - n\mathbf{I}_p)$ tends to 1 almost surely as $p \to \infty, n \to \infty$ with $p/n \to 0$.

Keywords: empirical distribution; maximum eigenvalue; random matrices

1. Introduction

Consider the sample covariance type matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}_p\mathbf{X}_p'$, where $\mathbf{X}_p = (\mathbf{s}_1, \ldots, \mathbf{s}_n) = (X_{ij})_{p\times n}$ and $X_{ij}, i = 1, \ldots, p, j = 1, \ldots, n$, are i.i.d. random variables with mean zero and variance 1. For such a matrix, much attention has been paid to asymptotic properties of its eigenvalues in the setting of $p/n \to c > 0$ as $p \to \infty$ and $n \to \infty$. For example, its empirical spectral distribution (ESD) function $F^{\mathbf{S}}(x)$ converges with probability one to the famous Marčenko and Pastur law (see [9] and [8]). Here, the ESD for any matrix $\mathbf{A}$ with real eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_p$ is defined by

$$F^{\mathbf{A}}(x) = \frac{1}{p}\#\{i\colon \lambda_i \le x\},$$

where $\#\{\cdots\}$ denotes the number of elements of the set. Also, with probability one its maximum eigenvalue and minimum eigenvalue converge, respectively, to the left end point and right end point of the support of Marčenko and Pastur's law (see [7] and [3]).

In contrast with asymptotic behaviors of $\mathbf{S}$ in the case of $p/n \to c$, the asymptotic properties of $\mathbf{S}$ have not been well understood when $p/n \to 0$. The first breakthrough was made in Bai and Yin [2]. They considered the normalized matrix

$$\mathbf{A}_p = \frac{1}{2\sqrt{np}}(\mathbf{X}_p\mathbf{X}_p' - nI_p)$$

1350-7265 © 2012 ISI/BS

and proved with probability one

$$F^{\mathbf{A}_p} \to F(x),$$

which is the so-called semicircle law with a density

$$F'(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & \text{if } |x| \le 1, \\ 0, & \text{if } |x| > 1. \end{cases}$$

One should note that the semicircle law is also the limit of the empirical spectral distribution of a symmetric random matrix whose diagonal are i.i.d. random variables and above diagonal elements are also i.i.d. (see [10]). Second, when $X_{11} \sim N(0,1)$, El Karoui [5] proved that the largest eigenvalue of $\mathbf{X}_p\mathbf{X}_p'$ after properly centering and scaling converges to the Tracy–Widom law.

In this paper, for general X_{11}, we investigate the maximum eigenvalue of $\mathbf{A}_p$ under the setting of $p/n \to 0$ as $p \to \infty$ and $n \to \infty$. The main results are presented in the following theorems.

Theorem 1. *Let $\mathbf{X}_p = (X_{ij})_{p\times n}$ where $\{X_{ij}: i = 1,2,\ldots,p; j = 1,2,\ldots,n\}$ are i.i.d. real random variables with $EX_{11} = 0$, $EX_{11}^2 = 1$ and $EX_{11}^4 < \infty$. Suppose that $n = n(p) \to \infty$ and $p/n \to 0$ as $p \to \infty$. Define*

$$\mathbf{A}_p = (A_{ij})_{p\times p} = \frac{1}{2\sqrt{np}}(\mathbf{X}_p\mathbf{X}_p' - nI_p).$$

Then as $p \to \infty$

$$\lambda_{\max}(\mathbf{A}_p) \to 1 \qquad a.s.,$$

where $\lambda_{\max}(\mathbf{A}_p)$ represents the largest eigenvalue of $\mathbf{A}_p$.

Indeed, after truncation and normalization of the entries of the matrix $\mathbf{A}_p$, we may obtain a better result.

Theorem 2. *Let $n = n(p) \to \infty$ and $p/n \to 0$ as $p \to \infty$. Define a $p \times p$ random matrix $\mathbf{A}_p$:*

$$\mathbf{A}_p = (A_{ij})_{p\times p} = \frac{1}{2\sqrt{np}}(\mathbf{X}_p\mathbf{X}_p' - nI_p),$$

where $\mathbf{X}_p = (X_{ij})_{p\times n}$. Suppose that X_{ij}'s are i.i.d. real random variables and satisfy the following conditions

(1) *$EX_{11} = 0$, $EX_{11}^2 = 1$, $EX_{11}^4 < \infty$ and*
(2) *$|X_{ij}| \le \delta_p \sqrt[4]{np}$, where $\delta_p \downarrow 0$, but $\delta_p \sqrt[4]{np} \uparrow +\infty$, as $p \to \infty$.*

Then, for any $\epsilon > 0$, $\ell > 0$

$$p\bigl(\lambda_{\max}(\mathbf{A}_p) \ge 1 + \epsilon\bigr) = \mathrm{o}(p^{-\ell}).$$

So far we have considered the sample covariance type matrix $\mathbf{S}$. However, a common used sample covariance matrix in statistics is

$$\mathbf{S}_1 = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{s}_j - \bar{\mathbf{s}})(\mathbf{s}_j - \bar{\mathbf{s}})',$$

where

$$\bar{\mathbf{s}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{s}_j.$$

Similarly we renormalize it as

$$\mathbf{A}_{p1} = \frac{1}{2}\sqrt{\frac{n}{p}}(\mathbf{S}_1 - I_p).$$

Theorem 3. *Under assumptions of Theorem 1, as $p \to \infty$*

$$\lambda_{\max}(\mathbf{A}_{p1}) \to 1 \qquad a.s.,$$

where $\lambda_{\max}(\mathbf{A}_{p1})$ stands for the largest eigenvalues of $\mathbf{A}_{p1}$.

Estimating a population covariance matrix for high dimension data is a challenging task. Usually, one can not expect the sample covariance matrix to be a consistent estimate of a population covariance matrix when both p and n go to infinity, especially when the orders of p and n are very close to each other. In such circumstance, as argued in [4], operator norm consistent estimation of large population covariance matrix still has nice properties.

Suppose that Σ is a population covariance matrix, nonnegative definite symmetric matrix. Then $\Sigma^{1/2}\mathbf{s}_j$, $j = 1, \dots, n$, may be viewed as i.i.d. sample drawn from the population with covariance matrix Σ, where $(\Sigma^{1/2})^2 = \Sigma$. The corresponding sample covariance matrix is

$$\mathbf{S}_2 = \frac{1}{n}\sum_{j=1}^{n}(\Sigma^{1/2}\mathbf{s}_j - \Sigma^{1/2}\bar{\mathbf{s}})(\Sigma^{1/2}\mathbf{s}_j - \Sigma^{1/2}\bar{\mathbf{s}})'.$$

Theorem 3 indicates that the matrix $\mathbf{S}_2$ is an operator consistent estimation of Σ as long as $p/n \to 0$ when $p \to \infty$. Specifically, we have the following theorem.

Theorem 4. *In addition to the assumptions of Theorem 1, assume that $\|\Sigma\|$ is bounded. Then, as $p \to \infty$*

$$\|\mathbf{S}_2 - \Sigma\| = \mathrm{O}\left(\sqrt{\frac{p}{n}}\right) \qquad a.s.,$$

where $\|\cdot\|$ stands for the spectral norm of a matrix.

Remark 1. Related work is [1], where the authors investigated quantitative estimates of the convergence of the empirical covariance matrix in the Log-concave ensemble. Here we obtain a convergence rate of the empirical covariance matrix when the sample vectors are in the form of $\Sigma^{1/2}\mathbf{s}_j$.

Remark 2. Theorems 1–4 are stated for the real random matrix $\mathbf{X}_p$, but they also hold for the complex case under moment conditions $EX_{11} = 0$, $E|X_{11}|^2 = 1$ and $E|X_{11}|^4 < \infty$. The proofs are similar to those for the real case except some notation changes.

2. Proof of Theorem 1

Throughout the paper, C denotes a constant whose value may vary from line to line. Also, all limits in the paper are taken as $p \to \infty$.

It follows from Theorem in [2] that

$$\liminf_{p\to\infty} \lambda_{\max}(\mathbf{A}_p) \geq 1 \qquad \text{a.s.} \tag{1}$$

Thus, it suffices to show that

$$\limsup_{p\to\infty} \lambda_{\max}(\mathbf{A}_p) \leq 1 \qquad \text{a.s.} \tag{2}$$

Let $\hat{\mathbf{A}}_p = \frac{1}{2\sqrt{np}}(\hat{\mathbf{X}}_p\hat{\mathbf{X}}_p' - nI_p)$, where $\hat{\mathbf{X}}_p = (\hat{X}_{ij})_{p\times n}$ and $\hat{X}_{ij} = X_{ij}I(|X_{ij}| \leq \delta_p \sqrt[4]{np})$ where δ_p is chosen as the larger of δ_p constructed as in (3) and δ_p as in (5). On the one hand, since $EX_{11}^4 < \infty$ for any $\delta > 0$ we have

$$\lim_{p\to\infty} \delta^{-4}E|X_{11}|^4 I\big(|X_{11}| > \delta\sqrt[4]{np}\big) = 0.$$

Since the above is true for arbitrary positive δ there exists a sequence of positive δ_p such that

$$\lim_{p\to\infty} \delta_p = 0, \qquad \lim_{p\to\infty} \delta_p^{-4}E|X_{11}|^4 I\big(|X_{11}| > \delta_p\sqrt[4]{np}\big) = 0, \qquad \delta_p\sqrt[4]{np} \uparrow +\infty. \tag{3}$$

On the other hand, since $EX_{11}^4 < \infty$ for any $\nu > 0$

$$\sum_{k=1}^{\infty} 2^k P(|X_{11}| > \nu 2^{k/4}) < \infty.$$

In view of the arbitrariness of ν, there is a sequence of positive number ν_k such that

$$\nu_k \to 0, \text{ as } k \to \infty, \qquad \sum_{k=1}^{\infty} 2^k P(|X_{11}| > \nu_k 2^{k/4}) < \infty. \tag{4}$$

For each k, let p_k be the maximum p such that $n(p) \cdot p \le 2^k$. For $p_{k-1} < p \le p_k$, set

$$\delta_p = 2\nu_k. \tag{5}$$

Let $Z_t = X_{ij}, t = (i-1)n + j$ and obviously $\{Z_t\}$ are i.i.d. We then conclude from (4) and (5) that

$$\begin{aligned}
P(\mathbf{A}_p \ne \hat{\mathbf{A}}_p, i.o.) &\le \lim_{K\to\infty} P\left(\bigcup_{k=K}^{\infty} \bigcup_{p_{k-1}<p\le p_k} \bigcup_{i\le p, j\le n} \{|X_{ij}| > \delta_p \sqrt[4]{np}\}\right) \\
&\le \lim_{K\to\infty} \sum_{k=K}^{\infty} P\left(\bigcup_{p_{k-1}<p\le p_k} \bigcup_{t=1}^{2^k} \{|Z_t| > \nu_k 2^k\}\right) \\
&= \lim_{K\to\infty} \sum_{k=K}^{\infty} P\left(\bigcup_{t=1}^{2^k} \{|Z_t| > \nu_k 2^k\}\right) \\
&\le \lim_{K\to\infty} \sum_{k=K}^{\infty} 2^k P(|Z_1| > \nu_k 2^{k/4}) \\
&= 0 \qquad \text{a.s.}
\end{aligned}$$

It follows that $\lambda_{\max}(\mathbf{A}_p) - \lambda_{\max}(\hat{\mathbf{A}}_p) \to 0$ a.s. as $p \to \infty$.

From now on, we write δ for δ_p to simplify notation. Moreover, set $\tilde{\mathbf{A}}_p = \frac{1}{2\sqrt{np}}(\tilde{\mathbf{X}}_p \tilde{\mathbf{X}}'_p - nI_p)$, where $\tilde{\mathbf{X}}_p = (\tilde{X}_{ij})_{p\times n}$ and $\tilde{X}_{ij} = \frac{\hat{X}_{ij} - E\hat{X}_{11}}{\sigma}$. Here, $\sigma^2 = E(\hat{X}_{11} - E\hat{X}_{11})^2$ and $\sigma^2 \to 1$ as $p \to \infty$.

We obtain via (3)

$$|E\hat{X}_{11}| \le \frac{E|X_{11}|^4 I(|X_{11}| > \delta_p \sqrt[4]{np})}{\delta^3 (np)^{3/4}} \le \frac{C}{(np)^{3/4}} \tag{6}$$

and

$$|\sigma^2 - 1| \le CE|X_{11}|^2 I\big(|X_{11}| > \delta \sqrt[4]{np}\big) \le \frac{E|X_{11}|^4 I(|X_{11}| > \delta \sqrt[4]{np})}{\delta^2 \sqrt{np}} = \mathrm{o}\left(\frac{1}{\sqrt{np}}\right). \tag{7}$$

We conclude from the Rayleigh–Ritz theorem that

$$\begin{aligned}
&|\lambda_{\max}(\tilde{\mathbf{A}}_p) - \lambda_{\max}(\hat{\mathbf{A}}_p)| \\
&\quad \le \frac{1}{2\sqrt{np}} \left| \sup_{\|\mathbf{z}\|=1} \left(\sum_{i\ne j} z_i z_j \sum_{k=1}^{n} \hat{X}_{ik} \hat{X}_{jk} + \sum_{i=1}^{p} z_i^2 \sum_{k=1}^{n} (\hat{X}_{ik}^2 - 1) \right) \right. \\
&\qquad \left. - \sup_{\|\mathbf{z}\|=1} \left(\sum_{i\ne j} z_i z_j \sum_{k=1}^{n} \tilde{X}_{ik} \tilde{X}_{jk} + \sum_{i=1}^{p} z_i^2 \sum_{k=1}^{n} (\tilde{X}_{ik}^2 - 1) \right) \right|
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{2\sqrt{np}}\left|1-\frac{1}{\sigma^2}\right| \sup_{\|\mathbf{z}\|=1}\left|\sum_{i\neq j} z_i z_j \frac{1}{\sqrt{np}}\sum_{k=1}^{n}\hat{X}_{ik}\hat{X}_{jk} + \sum_{i=1}^{p} z_i^2 \sum_{k=1}^{n}(\hat{X}_{ik}^2 - 1)\right| \\
&\quad + \frac{1}{2\sqrt{np}}\frac{2|EX_{11}|}{\sigma^2}\sup_{\|\mathbf{z}\|=1}\left|\sum_{i=1}^{p}\sum_{j=1}^{p} z_i z_j \sum_{k=1}^{n}\hat{X}_{ik}\right| \\
&\quad + \frac{1}{2\sqrt{np}}\frac{n|EX_{11}|^2}{\sigma^2}\sup_{\|\mathbf{z}\|=1}\left|\sum_{i=1}^{p}\sum_{j=1}^{p} z_i z_j\right| + \frac{n}{2\sqrt{np}}\left|1-\frac{1}{\sigma^2}\right| \\
&= A_1 + A_2 + A_3 + A_4.
\end{aligned}
$$

By (7) and the strong law of large numbers, we have

$$
\begin{aligned}
A_1 &= \frac{|\sigma^2-1|}{2\sqrt{np}\sigma^2}\sup_{\|\mathbf{z}\|=1}\left|\sum_{k=1}^{n}\left(\left(\sum_{i=1}^{p} z_i\hat{X}_{ik}\right)^2 - \sum_{i=1}^{p} z_i^2\hat{X}_{ik}^2\right) + \sum_{i=1}^{p} z_i^2\sum_{k=1}^{n}(\hat{X}_{ik}^2-1)\right| \\
&\le \frac{|\sigma^2-1|\sqrt{np}}{2\sigma^2}\cdot\frac{1}{np}\left(2\left|\sum_{k=1}^{n}\sum_{i=1}^{p}\hat{X}_{ik}^2\right| + \left|\sum_{i=1}^{p}\sum_{k=1}^{n}(\hat{X}_{ik}^2-1)\right|\right) \\
&\le \frac{|\sigma^2-1|\sqrt{np}}{2\sigma^2}\cdot\frac{1}{np}\left(3\left|\sum_{k=1}^{n}\sum_{i=1}^{p} X_{ik}^2\right| + np\right) \\
&\to 0 \qquad \text{a.s.}
\end{aligned}
$$

Similarly, (6), Hölder's inequality and the strong law of large numbers yield

$$
\begin{aligned}
A_2 &\le \frac{1}{2\sqrt{np}}\cdot\frac{2|E\hat{X}_{11}|}{\sigma^2}\sup_{\|\mathbf{z}\|=1}\left|\sum_{j=1}^{p} z_j\right|\left|\sum_{i=1}^{p} z_i\sum_{k=1}^{n}\hat{X}_{ik}\right| \\
&\le \frac{1}{2\sqrt{np}}\cdot\frac{C}{\sigma^2(np)^{3/4}}\cdot\sqrt{p}\cdot\left(\sum_{i=1}^{p}\left(\sum_{k=1}^{n}\hat{X}_{ik}\right)^2\right)^{1/2} \\
&\le \frac{1}{2\sqrt{np}}\cdot\frac{C}{\sigma^2(np)^{3/4}}\cdot\sqrt{p}\cdot\left(n\sum_{i=1}^{p}\sum_{k=1}^{n}\hat{X}_{ik}^2\right)^{1/2} \\
&\le \frac{C}{\sigma^2(np)^{1/4}}\left|\frac{1}{np}\sum_{i=1}^{p}\sum_{k=1}^{n}\hat{X}_{ik}^2\right|^{1/2} \\
&\le \frac{C}{\sigma^2(np)^{1/4}}\left|\frac{1}{np}\sum_{i=1}^{p}\sum_{k=1}^{n} X_{ik}^2\right|^{1/2} \to 0 \qquad \text{a.s.}
\end{aligned}
$$

It is straightforward to conclude from (6) and (7) that

$$A_3 \to 0 \qquad \text{a.s.}, \qquad A_4 \to 0 \qquad \text{a.s.}$$

Thus, we have $\lambda_{\max}(\hat{\mathbf{A}}_p) - \lambda_{\max}(\tilde{\mathbf{A}}_p) \to 0$ a.s. By the above results, to prove (2), it is sufficient to show that $\limsup_{p\to\infty} \lambda_{\max}(\tilde{\mathbf{A}}_p) \le 1$ a.s. To this end, we note that the matrix $\tilde{\mathbf{A}}_p$ satisfies all the assumptions in Theorem 2. Therefore, we obtain (2) by Theorem 2 (whose argument is given in the next section). Together with (1), we finishes the proof of Theorem 1.

3. Proof of Theorem 2

Suppose that $\mathbf{z} = (z_1, \ldots, z_p)$ is a unit vector. By the Rayleigh–Ritz theorem, we then have

$$\begin{aligned} \lambda_{\max}(\mathbf{A}_p) &= \max_{\|z\|=1} \left(\sum_{i,j} z_i z_j A_{ij} \right) \\ &= \max_{\|z\|=1} \left(\sum_{i \ne j} z_i z_j A_{ij} + \sum_{i=1}^{p} z_i^2 A_{ii} \right) \\ &\le \lambda_{\max}(\mathbf{B}_p) + \max_{i \le p} |A_{ii}|, \end{aligned}$$

where $\mathbf{B}_p = (B_{ij})_{p\times p}$ with

$$B_{ij} = \begin{cases} 0, & \text{if } i = j, \\ \dfrac{1}{2\sqrt{np}} \displaystyle\sum_{k=1}^{n} X_{ik} X_{jk}, & \text{if } i \ne j. \end{cases}$$

To prove Theorem 2, it is sufficient to prove, for any $\epsilon > 0$, $\ell > 0$

$$P\big(\lambda_{\max}(\mathbf{B}_p) > 1 + \epsilon\big) = \mathrm{o}(p^{-l}) \tag{8}$$

and

$$P\left(\max_{i \le p} \frac{1}{\sqrt{np}} \left| \sum_{j=1}^{n} (X_{ij}^2 - 1) \right| > \epsilon \right) = \mathrm{o}(p^{-l}). \tag{9}$$

We first prove (9). To simplify notation, let $Y_j = X_{1j}^2 - 1$ and $C_1 = E|Y_1|^2$. Then $EY_j = 0$. Choose an appropriate sequence $h = h_p$ such that it satisfies, as $p \to \infty$

$$\begin{cases} h/\log p \to \infty, \\ \delta^2 h/\log p \to 0, \\ \dfrac{\delta^4 p}{C_1} \ge \sqrt{p}. \end{cases} \tag{10}$$

We then have

$$
\begin{aligned}
&P\left(\max_{i\le p}\frac{1}{\sqrt{np}}\left|\sum_{j=1}^{n}(X_{ij}^2-1)\right|>\epsilon\right)\\
&\quad\le p\cdot P\left(\left|\sum_{j=1}^{n}(X_{1j}^2-1)\right|>\epsilon\sqrt{np}\right)\\
&\quad\le \epsilon^{-h}p\left(\sqrt{np}\right)^{-h}E\left|\sum_{j=1}^{n}Y_j\right|^h\\
&\quad\le \epsilon^{-h}p\left(\sqrt{np}\right)^{-h}\sum_{m=1}^{h/2}\sum_{1\le j_1<j_2<j_m\le n}\sum_{\substack{i_1+i_2+\cdots+i_m=h\\ i_1\ge 2,\ldots,i_1\ge 2}}\frac{h!}{i_1!i_2!\cdots i_m!}E|Y_{j_1}|^{i_1}E|Y_{j_2}|^{i_2}\cdots E|Y_{j_m}|^{i_m}\\
&\quad\le \epsilon^{-h}p\left(\sqrt{np}\right)^{-h}\sum_{m=1}^{h/2}\sum_{\substack{i_1+i_2+\cdots+i_m=h\\ i_1\ge 2,\ldots,i_1\ge 2}}\frac{n!}{m!(n-m)!}\frac{h!}{i_1!i_2!\cdots i_m!}E|Y_1|^{i_1}E|Y_1|^{i_2}\cdots E|Y_1|^{i_m}\\
&\quad\le \epsilon^{-h}p\left(\sqrt{np}\right)^{-h}\sum_{m=1}^{h/2}\sum_{\substack{i_1+i_2+\cdots+i_m=h\\ i_1\ge 2,\ldots,i_1\ge 2}}n^m\frac{h!}{i_1!i_2!\cdots i_m!}C_1^m\left(\delta^2\sqrt{np}\right)^{h-2m}\\
&\quad\le \epsilon^{-h}p\sum_{m=1}^{h/2}m^h\left(\frac{\delta^4 p}{C_1}\right)^{-m}\delta^{2h}\le \epsilon^{-h}p\frac{h}{2}\cdot\left(\frac{\delta^2 h}{\log(\delta^4 p/C_1)}\right)^h\\
&\quad\le\left(\left(\frac{ph}{2}\right)^{1/h}\cdot\frac{2\delta^2 h}{\log p}\cdot\epsilon^{-1}\right)^h\le\left(\frac{\xi}{\epsilon}\right)^h=\mathrm{o}(p^{-\ell}),
\end{aligned}
$$

where ξ is a constant satisfying $0<\xi<\epsilon$. Below are some interpretations of the above inequalities:

(a) The fifth inequality is because, $\frac{n!}{m!(n-m)!}<n^m$, $|Y_1|<\delta^2\sqrt{np}$.
(b) We use the fact $\sum_{i_1+i_2+\cdots+i_m=h_{i_1\ge 2,\ldots,i_1\ge 2}}\frac{h!}{i_1!i_2!\cdots i_m!}<m^h$ in the sixth inequality.
(c) The seventh inequality uses the elementary inequality

$$
a^{-t}t^b\le\left(\frac{b}{\log a}\right)^b\qquad\text{for all } a>1, b>0, t\ge 1 \text{ and } \frac{b}{\log a}>1.
$$

(d) The last two inequalities are due to (10).
(e) With the facts that $\frac{\xi}{\epsilon}<1$, $h/\log p\to\infty$, the last equality is true.

Thus, (9) follows.

Next, consider (8). For any $\varsigma > 0$, we have

$$P\big(\lambda_{\max}(\mathbf{B}_p) \geq 1+\varsigma\big) \leq \frac{E\lambda^k_{\max}(\mathbf{B}_p)}{(1+\varsigma)^k} \leq \frac{Etr(\mathbf{B}^k_p)}{(1+\varsigma)^k}$$
$$\leq \frac{1}{(1+\varsigma)^k}\cdot\frac{1}{(2\sqrt{np})^k}\sum E(X_{i_1j_1}X_{i_2j_1}X_{i_2j_2}X_{i_3j_2}\cdots X_{i_kj_k}X_{i_1j_k}),$$

where $k = k_p$ satisfies, as $p \to \infty$

$$\begin{cases} k/\log p \to \infty, \\ \delta^{1/3}k/\log p \to 0, \\ \dfrac{\delta^2\sqrt[4]{p}}{k^3} \geq 1, \end{cases}$$

and the summation is taken with respect to $j_1, j_2, \ldots, j_k$ running over all integers in $\{1, 2, \ldots, n\}$ and $i_1, i_2, \ldots, i_k$ running over all integers in $\{1, 2, \ldots, p\}$ subject to the condition that $i_1 \neq i_2, i_2 \neq i_3, \ldots, i_k \neq i_1$.

In order to get an up bound for $|\sum EX_{i_1j_1}X_{i_2j_1}\cdots X_{i_kj_k}X_{i_1j_k}|$, we need to construct a graph for given $i_1, \ldots, i_k$ and $j_1, \ldots, j_k$, as in [7,11] and [3]. We follow the presentation in [3] and [11] to introduce some fundamental concepts associated with the graph.

For the sequence $(i_1, i_2, \ldots, i_k)$ from $\{1, 2, \ldots, p\}$ and the sequence $(j_1, j_2, \ldots, j_k)$ from $\{1, 2, \ldots, n\}$, we define a directed graph as follows. Plot two parallel real lines, referred to as *I-line* and *J-line*, respectively. Draw $\{i_1, i_2, \ldots, i_k\}$ on the *I-line*, called *I-vertices* and draw $\{j_1, j_2, \ldots, j_k\}$ on the *J-line*, known as *J-vertices*. The vertices of the graph consist of the *I-vertices* and *J-vertices*. The edges of the graph are $\{e_1, e_2, \ldots, e_{2k}\}$, where for $a = 1, \ldots, k$, $e_{2a-1} = i_aj_a$ are called the column edges and $e_{2a} = j_ai_{a+1}$ are called row edges with the convention that $i_{2k+1} = i_1$. For each column edge e_{2a-1}, the vertices i_a and j_a are called the ends of the edge i_aj_a and moreover i_a and j_a are, respectively, the initial and the terminal of the edge i_aj_a. Each row edge e_{2a} starts from the vertex j_b and ends with the vertex i_{b+1}.

Two vertices are said to coincide if they are both in the *I-line* or both in the *J-line* and they are identical. That is $i_a = i_b$ or $j_a = j_b$. Readers are also reminded that the vertices i_a and j_b are not coincident even if they have the same value because they are in different lines. We say that two edges are coincident if two edges have the same set of ends.

The graph constructed above is said to be a *W-graph* if each edge in the graph coincides with at least one other edge. See Figure 1 for an example of a *W-graph*.

Two graphs are said to be isomorphic if one becomes another by an appropriate permutation on $\{1, 2, \ldots, p\}$ of *I-vertices* and an appropriate permutation on $\{1, 2, \ldots, n\}$ of *J-vertices*. A *W-graph* is called a canonical graph if $i_a \leq \max\{i_1, i_2, \ldots, i_{a-1}\} + 1$ and $j_a \leq \max\{j_1, j_2, \ldots, j_{a-1}\} + 1$ with $i_1 = j_1 = 1$, where $a = 1, 2, \ldots, k$.

In the canonical graph, if $i_{a+1} = \max\{i_1, i_2, \ldots, i_a\} + 1$, then the edge j_ai_{a+1} is called a row innovation and if $j_a = \max\{j_1, j_2, \ldots, j_{a-1}\} + 1$, then the edge i_aj_a is called a column innovation. Apparently, a row innovation and a column innovation, respectively, lead to a new I-vertex and a new J-vertex except the first column innovation i_1j_1 leading to a new I-vertex i_1 and a new J-vertex j_1.

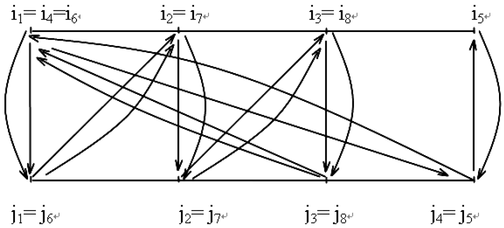

Figure 1. An example of *W-graph*.

We now classify all edges into three types, T_1, T_3 and T_4. Let T_1 denote the set of all innovations including row innovations and column innovations. We further distinguish the column innovations as follows. An edge $i_a j_a$ is called a T_{11} edge if it is a column innovation and the edge $j_a i_{a+1}$ is a row innovation; An edge $i_b j_b$ is referred to as a T_{12} edge if it is a column innovation but $j_b i_{b+1}$ is not a row innovation. An edge e_j is said to be a T_3 edge if there is an innovation edge $e_i, i < j$ so that e_j is the first one to coincide with e_i. An edge is called a T_4 edge if it does not belong to a T_1 edge or T_3 edge. The first appearance of a T_4 edge is referred to as a T_2 edge. There are two kinds of T_2 edges: (a) the first appearance of an edge that coincides with a T_3 edge, denoted by T_{21} edge; (b) the first appearance of an edge that is not an innovation, denoted by T_{22} edge.

We say that an edge e_i is single up to the edge e_j, $j \geq i$, if it does not coincide with any other edges among $e_1, \ldots, e_j$ except itself. A T_3 edge e_i is said to be regular if there are more than one innovations with a vertex equal to the initial vertex of e_i and single up to e_{i-1}, among the edges $\{e_1, \ldots, e_{i-1}\}$. All other T_3 edges are called irregular T_3 edges.

Corresponding to the above classification of the edges, we introduce the following notation and list some useful facts.

1. Denote by l the total number of innovations.
2. Let r be the number of the row innovations. Moreover, let c denote the column innovations. We then have $r + c = l$.
3. Define r_1 to be the number of the T_{11} edges. Then $r_1 \leq r$ by the definition of a T_{11} edge. Also, the number of the T_{12} edges is $l - r - r_1$.
4. Let t be the number of the T_2 edges. Note that the number of the T_3 edges is the same as the number of the innovations and there are a total of $2k$ edges in the graph. It follows that the number of the T_4 edges is $2k - 2l$. On the other hand, each T_2 edge is also a T_4 edge. Therefore, $t \leq 2k - 2l$.
5. Define μ to be the number of T_{21} edges. Obviously, $\mu \leq t$. The number of T_{22} edge is then $t - \mu$. Since each T_{21} edge coincides with one innovation, we let $n_i, i = 1, 2, \ldots, \mu$, denote the number of T_4 edges which coincide with the ith such innovation, $n_i \geq 0$.
6. Let μ_1 be the number of T_{21} edges which do not coincide with the other T_4 edges. That is $\mu_1 = \#\{i\colon\ n_i = 1, i = 1, 2, \ldots, \mu\}$, where $\#\{\cdot\}$ denotes the cardinality of the set $\{\cdot\}$.
7. Let m_j, $j = 1, 2, \ldots, t - \mu$, denote the number of T_4 edges which coincide with and include the jth T_{22} edge. Note that $m_j \geq 2$.

We now claim that

$$
\begin{aligned}
Etr(B_p^k) &\le \left(2\sqrt{np}\right)^{-k} \sum E(X_{i_1j_1}X_{i_2j_1}\cdots X_{i_kj_k}X_{i_1j_k}) \\
&= \left(2\sqrt{np}\right)^{-k} \sum{}' \sum{}'' \sum{}''' \sum_{*} E(X_{i_1j_1}X_{i_2j_1}\cdots X_{i_kj_k}X_{i_1j_k}) \\
&\le \left(2\sqrt{np}\right)^{-k} \sum_{l=1}^{k}\sum_{r=1}^{l}\sum_{r_1=0}^{r}\sum_{t=0}^{2k-2l}\sum_{\mu=0}^{t}\sum_{\mu_1=0}^{\mu}\sum_{*}\binom{k}{r}\binom{r}{r_1}\binom{k-r_1}{l-r-r_1}\binom{2k-l}{l} \\
&\quad \times k^{3t}(t+1)^{6k-6l}\left(\delta\sqrt[4]{np}\right)^{2k-2l-2t+\mu_1} p^{r+1}n^{l-r},
\end{aligned}
\tag{11}
$$

where the summation $\sum'$ is with respect to different arrangements of three types of edges at the $2k$ different positions, the summation $\sum''$ over different canonical graphs with a given arrangement of the three types of edges for $2k$ positions, the third summation $\sum'''$ with respect to all isomorphic graphs for a given canonical graph and the last notation $\sum_*$ denotes the constraint that $i_1 \ne i_2, i_2 \ne i_3, \ldots, i_k \ne i_1$.

Now, we explain why the above estimate is true:

(i) The factor $(2\sqrt{np})^{-k}$ is obvious.

(ii) If the graph is not a *W-graph*, which means there is a single edge in the graph, then the mean of the product of X_{ij} corresponding to this graph is zero (since $EX_{11}=0$). Thus, we have $l \le k$. Moreover, the facts that $r \le l$, $r_1 \le r$, $t \le 2k-2l$, $\mu \le t$ and $\mu_1 \le \mu$ are easily obtained from the fact 1 to the fact 7 listed before.

(iii) There are at most $\binom{k}{r}$ ways to choose r edges out of the k row edges to be the r row innovations. Subsequently, we consider how to select the column innovations. Observe that the definition of T_{11} edges, there are $\binom{r}{r_1}$ ways to select r_1 row innovations out of the total r row innovations so that the edge before each such r_1 row innovations is a T_{11} edge, column innovation. Moreover, there are at most $\binom{k-r_1}{l-r-r_1}$ ways to choose $l-r-r_1$ edges out of the remaining $k-r_1$ column edges to be the $l-r-r_1$ T_{12} edges, the remaining column innovations.

(iv) Given the position of the l innovations, there are at most $\binom{2k-l}{l}$ ways to select l edges out of the $2k-l$ edges to be T_3 edges. And the rest positions are for the T_4 edges. Therefore, the first summation $\sum'$ is bounded by $\sum_{l=1}^{k}\sum_{r=1}^{l}\sum_{r_1=0}^{r}\binom{k}{r}\binom{r}{r_1}\binom{k-r_1}{l-r-r_1}\binom{2k-l}{l}$.

(v) By definition, each innovation (or each irregular T_3 edges) is uniquely determined by the subgraph prior to the innovation (or the irregular T_3). Moreover, by Lemma 3.2 in [11] for each regular T_3 edge, there are at most $t+1$ innovations so that the regular T_3 edge coincides with one of them and by Lemma 3.3 in [11] there are at most $2t$ regular T_3 edges. Therefore, there are at most $(t+1)^{2t} \le (t+1)^{2(2k-2l)}$ ways to draw the regular T_3 edges.

(vi) Once the positions of the innovations and the T_3 edges are fixed there are at most $\binom{(r+1)c}{t} \le \binom{k^2}{t} \le k^{2t}$ ways to arrange the t T_2 edges, as there are $r+1$ *I-vertices* and c *J-vertices*. After t positions of T_2 edges are determined there are at most t^{2k-2l} ways to distribute $2k-2l$ T_4 edges among the t positions. So there are at most $k^{2t}\cdot t^{2k-2l}$ ways to arrange T_4 edges. It follows that $\sum''$ is bounded by $\sum_{t=0}^{2k-2l}(t+1)^{2(2k-2l)}k^{2t}\cdot t^{2k-2l}$.

(vii) The third summation $\sum'''$ is bounded by $n^c p^{r+1}$ because the number of graphs in the isomorphic class for a given graph is $p(p-1)\cdots(p-r)n(n-1)\cdots(n-c+1)$.

(viii) Recalling the definitions of $l, r, t, \mu, \mu_1, n_i, m_i$, we have

$$EX_{i_1j_1}X_{i_2j_1}\cdots X_{i_kj_k}X_{i_1j_k} = (EX_{11}^2)^{l-\mu}\left(\prod_{i=1}^{\mu} EX_{11}^{n_i+2}\right)\left(\prod_{i=1}^{t-\mu} EX_{11}^{m_i}\right), \tag{12}$$

where $\sum_{i=1}^{\mu} n_i + \sum_{i=1}^{t-\mu} m_i = 2k-2l$. Without loss of generality, we suppose $n_1 = n_2 = \cdots = n_{\mu_1} = 1$ and $n_{\mu_1+1}, \ldots, n_\mu \geq 2$ for convenience. It is easy to check that

$$E|X_{11}^s| \leq \begin{cases} M\left(\delta\sqrt[4]{np}\right)^{s-4}, & \text{if } s \geq 4, M = \max\{EX_{11}^4, |EX_{11}^3|\}, \\ \left(\delta\sqrt[4]{np}\right)^{s-2}, & \text{if } s \geq 2. \end{cases}$$

Thus, (12) becomes

$$\begin{aligned} &|EX_{i_1j_1}X_{i_2j_1}\cdots X_{i_kj_k}X_{i_1j_k}| \\ &\quad\leq \sum_{\mu=0}^{t}\sum_{\mu_1=0}^{\mu} |EX_{11}^3|^{\mu_1}|EX_{11}^4|^{t-\mu_1}\left(\delta\sqrt[4]{np}\right)^{\sum_{i=\mu_1+1}^{\mu} n_i - 2(\mu-\mu_1)}\left(\delta\sqrt[4]{np}\right)^{\sum_{i=1}^{t-\mu} m_i - 2(t-\mu)} \\ &\quad\leq \sum_{\mu=0}^{t}\sum_{\mu_1=0}^{\mu} M^t\left(\delta\sqrt[4]{np}\right)^{2k-2l-2t+\mu_1} \\ &\quad\leq \sum_{\mu=0}^{t}\sum_{\mu_1=0}^{\mu} k^t\left(\delta\sqrt[4]{np}\right)^{2k-2l-2t+\mu_1}, \qquad \text{when } k \text{ is large enough.} \end{aligned} \tag{13}$$

The above points regarding the T_2 edges are discussed for $t > 0$, but they are still valid when $t = 0$ with the convention that $0^0 = 1$ in the term t^{2k-2l}, because in this case there are only T_1 edges and T_3 edges in the graph and thus $l = k$.

Consider the constraint $\sum_*$ now. Note that for each T_{12} edge, say i_aj_a, it is a column innovation, but the next row edge j_ai_{a+1} is not a row innovation. Since $i_{a+1} \neq i_a$, the edge j_ai_{a+1} cannot coincide with the edge i_aj_a. Moreover, it also doesn't coincide with any edges before the edge i_aj_a since j_a is a new vertex. So j_ai_{a+1} must be a T_{22} edge. Thus, the number of the T_{12} edges cannot exceed the number of the T_{22} edges. This implies $l - r - r_1 \leq t - \mu$. Moreover, note that $\mu_1 \leq \mu$. We then have

$$\begin{aligned} &n^{-k/2}p^{-k/2}n^{l-r}p^{r+1}(np)^{k/2-l/2-t/2+\mu_1/4} \\ &\quad= (n/p)^{l/2}\cdot n^{-r-t/2+\mu_1/4}p^{r+1-t/2+\mu_1/4} \\ &\quad\leq \left(\sqrt{\frac{p}{n}}\right)^{r-r_1}\cdot p^{-t/2}p. \end{aligned} \tag{14}$$

We thus conclude from (11) and (14) that

$$
\begin{aligned}
Etr(B_p^k) \le 2^{-k} \sum_{l=1}^{k} \sum_{r=1}^{l} \sum_{r_1=0}^{r} \sum_{t=0}^{2k-2l} \sum_{\mu=0}^{t} \sum_{\mu_1=0}^{\mu} \binom{k}{r} \binom{r}{r_1} \binom{k-r_1}{l-r-r_1} \binom{2k-l}{l} \\
\times \left(\sqrt{\frac{p}{n}}\right)^{r-r_1} p^{-t/2} p k^{3t} (t+1)^{6k-6l} \delta^{2k-2l-2t+\mu_1}.
\end{aligned}
\tag{15}
$$

Moreover, we claim that

$$
\begin{aligned}
& p\left[2^{-k}\binom{k}{r}\right]\left[\binom{r}{r_1}\left(\sqrt{\frac{p}{n}}\right)^{r-r_1}\right]\left[\binom{k-r_1}{l-r-r_1}\delta^{l-r-r_1}\right] \\
& \quad \times \left[\binom{2k-l}{l}\left(\frac{\sqrt{p}\delta^3}{k^3}\right)^{-t}(t+1)^{6k-6l}\delta^{2k-2l}\right]\delta^{-(l-r-r_1)+3t-(2k-2l)} \cdot \delta^{2k-2l-2t+\mu_1} \\
& \le p^2\left(1+\sqrt{\frac{p}{n}}\right)^k (1+\delta)^k \left(1+\frac{24^3 k^3 \delta}{\log^3 p}\right)^{2k}.
\end{aligned}
\tag{16}
$$

Indeed, the above claim is based on the following five facts.

(1) $2^{-k}\binom{k}{r} \le 2^{-k}\sum_{r=0}^{k}\binom{k}{r} = 1$.
(2) $\binom{r}{r_1}(\sqrt{\frac{p}{n}})^{r-r_1} = \binom{r}{r-r_1}(\sqrt{\frac{p}{n}})^{r-r_1} \le \sum_{s=0}^{r}\binom{r}{s}(\sqrt{\frac{p}{n}})^s = (1+\sqrt{\frac{p}{n}})^r \le (1+\sqrt{\frac{p}{n}})^k$.
(3) $\binom{k-r_1}{l-r-r_1}\delta^{l-r-r_1} \le \sum_{s=0}^{k-r_1}\binom{k-r_1}{s}\delta^s = (1+\delta)^{k-r_1} \le (1+\delta)^k$.
(4) By the fact that $\binom{2k-l}{l} \le \binom{2k}{2l}$, and the inequality $a^{-t}(t+1)^b \le a(\frac{b}{\log a})^b$, for $a > 1$, $b > 0$, $t > 0$ and $\frac{\delta^2\sqrt{p}}{k^3} \ge \sqrt[4]{p}$, we have

$$
\begin{aligned}
\binom{2k-l}{l}\left(\frac{\sqrt{p}\delta^3}{k^3}\right)^{-t}(t+1)^{6k-6l}\delta^{2k-2l} &\le \binom{2k}{2l}\frac{\sqrt{p}\delta^3}{k^3}\left(\frac{6k-6l}{\log(\sqrt{p}\delta^3/k^3)}\right)^{6k-6l}\delta^{2k-2l} \\
&\le p\binom{2k}{2l}\left(\frac{24k}{\log p}\right)^{6k-6l}\delta^{2k-2l} \\
&\le p\binom{2k}{2l}\left(\frac{24^3k^3\delta}{\log^3 p}\right)^{2k-2l} \\
&\le p\sum_{s=0}^{2k}\binom{2k}{s}\left(\frac{24^3k^3\delta}{\log^3 p}\right)^{2k-s} \\
&= p\left(1+\frac{24^3k^3\delta}{\log^3 p}\right)^{2k}.
\end{aligned}
$$

(5) When p is large enough, $\delta^{-(l-r-r_1)+3t-(2k-2l)} \cdot \delta^{2k-2l-2t+\mu_1} = \delta^{t-(l-r-r_1)} \cdot \delta^{\mu_1} \le 1$, since $\delta \to 0$ and $l - r - r_1 \le t - \mu$.

Summarizing (15) and (16), we obtain that

$$\begin{aligned} Etr(B_p^k) &\le \sum_{l=1}^{k} \sum_{r=1}^{l} \sum_{r_1=0}^{r} \sum_{t=0}^{2k-2l} \sum_{\mu=0}^{t} \sum_{\mu_1=0}^{\mu} p^2 \left(1+\sqrt{\frac{p}{n}}\right)^k (1+\delta)^k \left(1+\frac{24^3 l^3 \delta}{\log^3 p}\right)^{2k} \\ &\le 8k^6 p^2 \left(1+\sqrt{\frac{p}{n}}\right)^k (1+\delta)^k \left(1+\frac{24^3 l^3 \delta}{\log^3 p}\right)^{2k} \\ &= \left((8k^6)^{1/k} p^{2/k} \left(1+\sqrt{\frac{p}{n}}\right)(1+\delta)\left(1+\frac{24^3 k^3 \delta}{\log^3 p}\right)^2\right)^k \\ &\le \eta^k, \end{aligned}$$

where η is a constant satisfying $1 < \eta < 1 + \epsilon$. Here the last inequality uses the facts below:

(i) $(p^2)^{1/k} \to 1$, because $k/\log p \to \infty$,
(ii) $(8k^6)^{1/k} \to 1$, because $k \to \infty$,
(iii) $(1 + \sqrt{\frac{p}{n}}) \to 1$, because $p/n \to 0$,
(iv) $(1 + \delta) \to 1$, because $\delta \to 0$,
(v) $\frac{24^3 \cdot k^3 \delta}{\log^3 p} \to 0$, because $\frac{\delta^{1/3} k}{\log p} \to 0$.

It follows that

$$P\left(\lambda_{\max}(\mathbf{B}_p) > 1 + \epsilon\right) \le \left(\frac{\eta}{1+\epsilon}\right)^k = \mathrm{o}(p^{-\ell})$$

since $k/\log p \to \infty$ and $\frac{\eta}{1+\epsilon} < 1$. The proof is complete.

4. Proof of Theorem 3

Note that

$$\mathbf{S}_1 = \mathbf{S} - \bar{\mathbf{s}}\bar{\mathbf{s}}'. \tag{17}$$

By the Fan inequality [6],

$$\sup_x |F^{\mathbf{A}_{p1}}(x) - F^{\mathbf{A}_p}(x)| \le \frac{1}{p}.$$

Thus from theorem in [2], we see that

$$F^{\mathbf{A}_{p1}}(x) \xrightarrow{\text{a.s.}} F(x),$$

specified in the introduction. It follows that

$$\liminf_{p\to\infty} \lambda_{\max}(\mathbf{A}_{p1}) \geq 1.$$

Let $\mathbf{z}$ be a unit vector. In view of (17), we obtain

$$\mathbf{z}'\mathbf{A}_{p1}\mathbf{z} = \mathbf{z}'\mathbf{A}_p\mathbf{z} - \frac{1}{2}\sqrt{\frac{n}{p}}\mathbf{z}'\bar{\mathbf{s}}\bar{\mathbf{s}}'\mathbf{z} \leq \mathbf{z}'\mathbf{A}_p\mathbf{z},$$

which implies that

$$\lambda_{\max}(\mathbf{A}_{p1}) \leq \lambda_{\max}(\mathbf{A}_p).$$

This, together with Theorem 1, finishes the proof of Theorem 3.

5. Proof of Theorem 4

Theorem 4 follows from Theorem 3 and the fact that

$$\|\mathbf{S}_2 - \Sigma\| = \|\Sigma^{1/2}(\mathbf{S}_1 - \mathbf{I}_p)\Sigma^{1/2}\| \leq \|\mathbf{S}_1 - \mathbf{I}_p\|\|\Sigma\|.$$

Acknowledgement

G.M. Pan was partially supported by a Grant M58110052 at the Nanyang Technological University, and by a grant # ARC 14/11 from Ministry of Education, Singapore.

References

[1] Adamczak, R., Litvak, A.E., Pajor, A. and Tomczak-Jaegermann, N. (2010). Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.* **23** 535–561. MR2601042
[2] Bai, Z.D. and Yin, Y.Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. MR0929083
[3] Bai, Z.D. and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416
[4] El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. MR2485011
[5] El Karoui, N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity. Preprint.
[6] Fan, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Natl. Acad. Sci. USA* **37** 760–766. MR0045952
[7] Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261. MR0566592
[8] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** 1–38. MR0650926

[9] Marčenko, V.A. and Pastur, L.A. (1967). Distribution for some sets of random matrices. *Math. USSR-Sb.* **1** 457–483.

[10] Wigner, E.P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* (2) **67** 325–327. MR0095527

[11] 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. MR0950344

Received May 2010 and revised December 2010