CONVERGENCE RATES OF SPECTRAL DISTRIBUTIONS OF LARGE SAMPLE COVARIANCE MATRICES[*]

Z. D. BAI[†], BAIQI MIAO[‡], AND JIAN-FENG YAO[§]

Abstract. In this paper, we improve known results on the convergence rates of spectral distributions of large-dimensional sample covariance matrices of size $p \times n$. Using the Stieltjes transform, we first prove that the expected spectral distribution converges to the limiting Marčenko–Pastur distribution with the dimension sample size ratio $y = y_n = p/n$ at a rate of $O(n^{-1/2})$ if y keeps away from 0 and 1, under the assumption that the entries have a finite eighth moment. Furthermore, the rates for both the convergence in probability and the almost sure convergence are shown to be $O_p(n^{-2/5})$ and $o_{a.s.}(n^{-2/5+\eta})$, respectively, when y is away from 1. It is interesting that the rate in all senses is $O(n^{-1/8})$ when y is close to 1.

1. Introduction. The spectral analysis of large-dimensional random matrices has been actively developed in the last decades since the initial contributions of Wigner (1955, 1958); also see the recent review by Bai (1999) and the book by Mehta (1991). Various limiting distributions were discovered including the Wigner semicircular law (Wigner, 1955), the Marčenko–Pastur law (Marčenko and Pastur, 1967), the limiting law for multivariate F matrices (Bai, Yin, and Krishnaiah (1987) and Silverstein (1985)) and the circular law (Bai and Yin (1986), Bai (1997)). The spectrum separation problem for large-dimensional sample covariance matrices was investigated in Bai and Silverstein (1998, 1999).

Let A be an $n \times n$ symmetric matrix, and $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of A. The spectral distribution F^A of A is defined as

$$F^A(x) = \frac{1}{n} \times \text{number of elements in } \{k : \lambda_k \leq x\}.$$

Let $\mathbf{X}_p = (x_{ij})_{p \times n}$ be a $p \times n$ observation matrix whose entries are mutually independent and have a common mean zero and variance 1. The entries of $\mathbf{X}_p$ may depend on n but we suppress the index n for simplicity. In this paper, we consider the sample covariance matrix $\mathbf{S} = n^{-1}\mathbf{X}_p\mathbf{X}_p^T$, where $\mathbf{X}^T$ denotes the transpose of the matrix $\mathbf{X}$. Assume that the ratio p/n of sizes tends to a positive limit y as $n \to \infty$. Under suitable moment conditions on the x_{ij} entries, it is known that the empirical spectral distribution (ESD) $F_p := F^{\mathbf{S}}$ converges to the Marčenko–Pastur distribution F_y with

[†]Department of Mathematics, Northeast Normal University, Changchun, 130024 China, and Department of Statistics and Applied Probability, National University of Singapore, Singapore, 117543 (stabaizd@leonis.nus.edu.sg). The research of this author was partially supported by NSFC grant 201471000 as well as by the NUS Research grant R-155-030-112.

[‡]Department of Statistics and Finance, University Science and Technology of China, Hefei, Anhui, China (bqmiao@ustc.edu.cn). The research of this author was supported in part by the National Foundation of Natural Science of China.

[§]IRMAR, Université de Rennes 1, Campus de Beaulieu, F-35042 Rennes Cedex, France (jian-feng. yao@univ-rennes1.fr).

index y with density

$$F_y'(x) = \begin{cases} \frac{1}{2\pi xy}\sqrt{(x-a)(b-x)} & \text{if } a < x < b, \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $a = (1 - \sqrt{y})^2$, $b = (1 + \sqrt{y})^2$.

An important question here concerns the problem of the convergence rates. However, no significant progress was made before the introduction of a novel and powerful tool, namely, the Berry–Esseen inequalities in terms of Stieltjes transforms, by Bai (1993a, 1993b). Using this methodology, Bai (1993b) proved that the expected ESD $\mathbb{E}F_p$ converges to F_{y_n} at a rate of $O(n^{-1/4})$ or $O(n^{-5/48})$ depending on whether y_n is far away or close to 1, respectively, where $y_n = p/n$. In another work by Bai, Miao, and Tsay (1997), these rates are also established for the convergence in probability of the ESD F_p itself. In later works of Bai, Miao, and Tsay (1999, 2002), the convergence rates for large Wigner matrices are significantly improved.

In this work, we further investigate the convergence rates for empirical spectral distributions for large sample covariance matrices and improve those results in the theorems to follow.

The following conditions will be used:

(C.1) $\mathbb{E}x_{ij} = 0$, $\mathbb{E}x_{ij}^2 = 1$, $1 \le i \le p$, $1 \le j \le n$.

(C.2) $\sup_{i,j,n} \mathbb{E}|x_{ij}|^8 < \infty$.

(C.3)   For any positive constant δ,

$$\sum_{ij} \mathbb{E}x_{ij}^8 I_{(|x_{ij}| \ge \delta\sqrt{n})} = o(n^2).$$

It is easy to see that condition (C.3) guarantees that there is a sequence $\{\delta = \delta_n \to 0\}$ such that

$$(1.1) \qquad\qquad \sum_{ij} \mathbb{E}x_{ij}^8 I_{(|x_{ij}| \ge \delta\sqrt{n})} = o(n^2\delta^8).$$

(C.2′) $\sup_{i,j,n} \mathbb{E}|x_{ij}|^k < \infty$ for any integer $k \ge 1$.

Throughout the paper, we use the notation $Z_n = O_p(a_n)$ if the sequence $(a_n^{-1}Z_n)$ is tight and use $Z_n = o_p(a_n)$ when $a_n^{-1}Z_n$ tends to 0 in probability. We shall also set $\|f\| = \sup_x |f(x)|$.

For simplicity, from now on we drop the index n from y and use the notation $y = y_n = p/n$. Finally, let us define

$$(1.2) \qquad \theta = \theta(n,y) = \begin{cases} \frac{-2\log_n(1-\sqrt{y})}{1+4\log_n(1-\sqrt{y})} & \text{if } y \le (1-n^{-1/8})^2, \\ \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

We now introduce the main results of the paper.

THEOREM 1.1. *Assume that the conditions* (C.1)–(C.3) *are satisfied. Then,*

$$\|\mathbb{E}F_p - F_y\| = O\left(\frac{n^{-1/[4\theta+2]}}{[1 - \sqrt{y} + n^{-1/[8\theta+4]}]}\right),$$

THEOREM 1.2. *Assume that the conditions* (C.1)–(C.3) *are satisfied. Then,*

$$\|F_p - F_y\| = O_p\left(\max\left\{\frac{n^{-(2/(5+\theta))}}{[1 - \sqrt{y} + n^{-(1/(5+\theta))}]}, \frac{n^{-1/[4\theta+2]}}{[1 - \sqrt{y} + n^{-1/[8\theta+4]}]}\right\}\right).$$

THEOREM 1.3. *Assume that the conditions* (C.1)–(C.3) *are satisfied. Then, with probability* 1,

$$\|F_p - F_y\| = o\left(\max\left\{\frac{n^{-(2/(5+\theta))+\eta}}{[1 - \sqrt{y} + n^{-(1/(5+\theta))}]}, \frac{n^{-1/[4\theta+2]}}{[1 - \sqrt{y} + n^{-1/[8\theta+4]}]}\right\}\right).$$

Remark on the convergence rates. If y is not close to 1, then $\theta \sim c/\log n$, and hence the convergence rates in the above three theorems are $O(n^{-1/2})$, $O_p(n^{-2/5})$, and $o_{a.s.}(n^{-2/5+\eta})$, respectively. When $y > 1 - O(n^{-1/8})$, $\theta = 1/2$, and hence the rates of the three theorems are $O(n^{-1/8})$, $O_p(n^{-1/8})$, and $O_{a.s.}(n^{-1/8})$, respectively. When y goes to 1 with intermediate rates, we may have intermediate convergence rates.

It is worth noticing that the convergence rates given above for the case $0 < y \le 1$ also apply to the case $y > 1$, since the last case can be reduced to the first case by interchanging the roles of row and column sizes p and n.

The proofs of these main results will be given in section 3. For convenience, we first introduce some necessary notation and preliminary consequences in section 2. Some necessary lemmas are postponed to section 4.

2. Definitions and easy consequences. Throughout the paper, the transpose of a possibly complex matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{\mathrm{T}}$ and its conjugate by $\overline{\mathbf{A}}$. For each fixed p, n, and $k = 1, \ldots, p$, let us denote by $\mathbf{x}_k = (x_{k1}, \ldots, x_{kn})^T$ the kth row of $\mathbf{X}_p$ arranged as a column vector, and let $\mathbf{X}_p(k)$ be the $(p-1) \times n$ submatrix obtained from $\mathbf{X}_p$ by deleting its kth row. Let us define

$$\alpha_k := \frac{1}{n}\mathbf{X}_p(k)\mathbf{x}_k, \quad \mathbf{S}_k := \frac{1}{n}\mathbf{X}_p(k)\mathbf{X}_p^T(k), \quad \mathbf{B}_k := \frac{1}{n}\mathbf{X}_p^T(k)\mathbf{D}_k\mathbf{X}_p(k),$$

(2.1)
$$\mathbf{B} := \frac{1}{n}\mathbf{X}_p^T\mathbf{D}\mathbf{X}_p, \quad \mathbf{D}_k := (\mathbf{S}_k - z\mathbf{I}_{p-1})^{-1}, \quad \mathbf{D} := (\mathbf{S} - z\mathbf{I}_p)^{-1},$$

$$\mathbf{\Gamma}_k := \mathbf{D}_k\overline{\mathbf{D}}_k, \qquad \mathbf{\Lambda}_k := \mathbf{D}_k\mathbf{S}_k\overline{\mathbf{D}}_k.$$

Here $\mathbf{I}_m$ is the m-dimensional identity matrix and z a complex number with a positive imaginary part.

Following Bai (1993b), the Stieltjes transform of the spectral distribution F_p of the sample covariance matrix $\mathbf{S}$ is defined for $z = u + iv$ with $v > 0$ by

$$m_p(z) = \int_{-\infty}^{\infty} \frac{1}{x - z}dF_p(x),$$

and it is well known that

$$m_p(z) = \frac{1}{p}tr(\mathbf{S} - z\mathbf{I}_p)^{-1}.$$

Similarly, the Stieltjes transform of the spectral distribution $F_p^{(k)}$ of the submatrix $\mathbf{S}_k$ satisfies

$$m_p^{(k)}(z) = \int_{-\infty}^{\infty} \frac{1}{x - z}dF_p^{(k)}(x) = \frac{1}{p-1}tr(\mathbf{S}_k - z\mathbf{I}_{p-1})^{-1}.$$

Finally, the Stieltjes transform of the "limiting" (by noting that $y = y_n$) Marčenko–Pastur distribution F_y is

(2.2)
$$m(z) = \int_{-\infty}^{\infty} \frac{1}{x - z}dF_y(x) = -\frac{y + z - 1 - \sqrt{(1 - y - z)^2 - 4yz}}{2yz}$$

for $0 < y \leq 1$. Here the square root $\sqrt{z}$ is the one with a positive imaginary part. Note that $m(z)$ is a root of the quadratic equation

$$yzm^2 + (y + z - 1)m + 1 = 0,$$

which implies that $m(z)m^*(z) = \frac{1}{yz}$, where

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

is the other root of the equation. We claim that

(2.3) $$|m(z)| < |m^*(z)| \quad \text{for all} \ \ z = u + iv, \ v > 0.$$

To see this, set

$$\alpha + i\beta = \sqrt{(1 + y - z)^2 - 4y} \quad \text{with} \quad \beta \geq 0.$$

We have

(2.4) $$\alpha\beta = v(u - y - 1),$$

(2.5) $$\beta^2 - \alpha^2 = (b - u)(u - a) + v^2.$$

First, note that $\beta = 0$ is impossible; otherwise we should have $u = 1 + y$ and (2.5) would be violated. Hence, $\beta > 0$ and $\alpha \geq 0$ if and only if $u \geq 1 + y$.

It is easy to see that

$$|m(z)| < |m^*(z)| \Leftrightarrow |y + z - 1 - (\alpha + i\beta)| < |y + z - 1 + (\alpha + i\beta)|$$

$$\Leftrightarrow \alpha(y - 1 + u) + \beta v > 0.$$

The last inequality clearly holds if $u \geq (1 + y)$ or $u \leq 1 - y$ (in this case the result was proved in Bai (1993b, p. 651)). Now assume for a $u \in (1 - y, 1 + y) \subset [a, b]$ that the inequality does not hold, i.e., $\alpha(y - 1 + u) + \beta v \leq 0$. This implies that $\beta v \leq |\alpha|[u - (1 - y)]$ (noting that $\alpha < 0$). Multiplying both sides by β and using (2.4), we get

$$\beta^2 \leq [u - (1 - y)][1 + y - u] \leq (b - u)(u - a),$$

which contradicts (2.5). The claim (2.3) is then proved.

This claim implies that $|m(z)| \leq 1/\sqrt{y|z|}$ for any z. On the other hand, when $u < a - v$, both real and imaginary parts of $m(z)$ are positive and increasing (a consequence of the integral formula (2.2) of $m(z)$). Thus, $|m(z)|$ can only reach its maximum when $u > a - v$. When $a < 2v$, we have $|m(z)| \leq 1/\sqrt{yv} \leq \frac{2\sqrt{2}}{\sqrt{y}(\sqrt{a}+\sqrt{v})} = \frac{2\sqrt{2}v_y}{\sqrt{yv}}$. When $a \geq 2v$, by noticing that $\sqrt{|z|} \geq \sqrt[4]{a^2/4 + v^2} \geq \frac{1}{2\sqrt{2}}(\sqrt{a} + \sqrt{v})$, we obtain the same bound as in the first case. Therefore, we obtain

(2.6) $$|m(z)| \leq \frac{2\sqrt{2}v_y}{\sqrt{yv}},$$

where

(2.7) $$v_y := v/[\sqrt{a} + \sqrt{v}] = v/[1 - \sqrt{y} + \sqrt{v}].$$

LEMMA 2.1. *Let* $\mathbf{x} = (x_1, \ldots, x_n)^T$ *and* $\mathbf{y} = (y_1, \ldots, y_n)^T$ *be independent real random vectors with independent elements. Suppose that for all* $1 \leq j \leq n$, $\mathbb{E}x_j = \mathbb{E}y_j = 0$, $\mathbb{E}|x_j|^2 = \mathbb{E}|y_j|^2 = 1$, $\mathbb{E}|x_j|^4 \leq L < \infty$ *and that* $\mathbf{A}$ *is an* $n \times n$ *complex symmetric matrix. Let* $\mu_k = \max_{j \leq n}(\mathbb{E}|x_j|^k, \mathbb{E}|y_j|^k)$. *Then*

(i) $\mathbb{E}|\mathbf{x}^T \mathbf{A} \mathbf{y}|^2 = tr(\mathbf{A}\bar{\mathbf{A}})$;

(ii) $\mathbb{E}|\mathbf{x}^T \mathbf{A} \mathbf{x}|^2 \leq Ltr(\mathbf{A}\bar{\mathbf{A}}) + |tr\mathbf{A}|^2$;

(iii) $\mathbb{E}|\mathbf{x}^T \mathbf{A} \mathbf{x} - tr\mathbf{A}|^2 \leq L(tr\mathbf{A}\bar{\mathbf{A}})$;

(iv) $\mathbb{E}|\mathbf{x}^T \mathbf{A} \mathbf{x} - tr\mathbf{A}|^{2k} \leq d_k \left[\mu_{4k} tr(\mathbf{A}\bar{\mathbf{A}})^k + (Ltr(\mathbf{A}\bar{\mathbf{A}}))^k \right]$ *for* $k \geq 2$ *and some positive constant* d_k *depending on* k *only.*

The proofs of (i)–(iii) are elementary and therefore omitted. The statement (iv) follows from Lemma 2.7 of Bai and Silverstein (1998).

LEMMA 2.2. *Let* G_1 *and* G_2 *be probability distribution functions and* $z = u + iv$, $v > 0$. *Then for each positive integer* m,

$$\left| \int_{-\infty}^{\infty} \frac{1}{|x-z|^m} d(G_1(x) - G_2(x)) \right| \leq \frac{2}{v^m} \|G_1 - G_2\|.$$

Proof. Let be $G^* := G_1 - G_2$. We have, by integration by parts,

$$\left| \int_{-\infty}^{\infty} \frac{1}{|x-z|^m} d(x) G^* \right|$$

$$= \left| -\int_{-\infty}^{\infty} G^*(x) d\left[\frac{1}{|x-z|^m} \right] \right|$$

$$= \left| -\int_{-\infty}^{Re(z)} G^*(x) d\left[\frac{1}{|x-z|^m} \right] + \int_{Re(z)}^{\infty} G^*(x) d\left[-\frac{1}{|x-z|^m} \right] \right|$$

$$\leq \|G^*\| \left\{ \int_{-\infty}^{Re(z)} d\left[\frac{1}{|x-z|^m} \right] + \int_{Re(z)}^{\infty} d\left[-\frac{1}{|x-z|^m} \right] \right\}$$

$$= \|G^*\| \left\{ \frac{1}{|x-z|^m} \bigg|_{-\infty}^{Re(z)} + \left(-\frac{1}{|x-z|^m} \bigg|_{Re(z)}^{\infty} \right) \right\} = \|G^*\| \frac{2}{v^m}. \qquad \square$$

We will need the following auxiliary variables:

$$\varepsilon_k = -\frac{1}{n} \sum_{j=1}^{n} (x_{kj}^2 - 1) + \frac{1}{n}(\mathbf{x}_k^T \mathbf{B}_k \mathbf{x}_k - \mathbb{E}tr\mathbf{B}),$$

$$\varepsilon_k^* = -\frac{1}{n} \sum_{j=1}^{n} (x_{kj}^2 - 1) + \frac{1}{n}(\mathbf{x}_k^T \mathbf{B}_k \mathbf{x}_k - tr\mathbf{B}_k),$$

$$\widetilde{\varepsilon}_k = \frac{1}{n}(tr\mathbf{B}_k - \mathbb{E}tr\mathbf{B}_k) = \frac{z}{n}(tr\mathbf{D}_k - \mathbb{E}tr\mathbf{D}_k),$$

$$\pi_k = \frac{1}{n}\mathbb{E}(tr\mathbf{B}_k - tr\mathbf{B}) = \frac{z}{n}\mathbb{E}(tr\mathbf{D}_k - tr\mathbf{D}),$$

$$\beta_k = -\frac{1}{n} \sum_{j=1}^{n} (x_{kj}^2 - 1) + z - 1 + \frac{1}{n}\mathbf{x}_k^T \mathbf{B}_k \mathbf{x}_k,$$

$$\beta_k^* = z - 1 + \frac{1}{n}tr\mathbf{B}_k,$$

$$\beta = z - 1 + \frac{1}{n}tr\mathbf{B}.$$

We summarize below some inequalities which will be used in the derivations. Let $\Delta = \| \mathbb{E}F_p - F_y \|$ and $M := \sup_{i,j,n} \mathbb{E}|x_{ij}|^4$. For fixed (n,p) and $1 \le k \le p$, we define the σ-algebra

$$\mathcal{F}^{(k)} = \sigma(\mathbf{x}_i \; : \; 1 \le i \le p, \; i \ne k), \quad \mathcal{F}_k = \sigma(\mathbf{x}_i \; : \; k < i \le p).$$

1. (from Lemma 3.3 of Bai (1993a)):

(2.8) $$|(p-1)F_p^{(k)}(x) - pF_p(x)| \le 1.$$

2. (from Lemma 2.2 and (2.8)):

(2.9) $$|tr\mathbf{D} - tr\mathbf{D}_k| = \left| \int_{-\infty}^{\infty} \frac{d[pF_p(x) - (p-1)F_p^{(k)}(x)]}{x - z} \right| \le 2v^{-1}.$$

3. (from equation (3.14) of Bai (1993b)):

(2.10) $$m_p(z) = \int_0^{\infty} \frac{1}{x - z} \, dF_p(x) = \frac{1}{p}tr\mathbf{D} = -\frac{1}{p}\sum_{k=1}^{p} \frac{1}{\beta_k}.$$

4. (from Lemma 2.2 of Bai, Miao, and Tsay (1997)):

(2.11) $$\mathbb{E}|m_p(z) - \mathbb{E}(m_p(z))|^2 \le p^{-1}v^{-2}.$$

5. (from $|\beta_k^*| \ge Im(\beta_k^*) = v(1 + n^{-1}tr\mathbf{\Lambda}_k)$):

(2.12) $$|\beta_k^*|^{-1}(1 + n^{-1}tr\mathbf{\Lambda}_k) \le v^{-1}.$$

6.

(2.13) $$|\beta_k| \ge Im(\beta_k) = v\left(1 + \frac{1}{n}\alpha_k^T \mathbf{D}_k \overline{\mathbf{D}_k}\alpha_k\right).$$

7.

(2.14) $$\left|1 + \frac{1}{n}\alpha_k^T \mathbf{D}_k{}^2\alpha_k\right| \le 1 + \frac{1}{n}\alpha_k^T \mathbf{D}_k \overline{\mathbf{D}_k}\alpha_k.$$

Let λ_{kj}, $j = 1, 2, \dots, p-1$, be the eigenvalues of $\mathbf{S}_k$ which can be decomposed into a diagonal form on the basis of orthonormal and real eigenvectors. Let $\mathbf{L}$ be a complex matrix having the product form $\mathbf{L} = \mathbf{M}^\ell \mathbf{N}^{\ell'}$ for some integers ℓ, ℓ' and factors $\mathbf{M}$, $\mathbf{N}$ equal to one of the matrices $\{\mathbf{D}_k, \overline{\mathbf{D}}_k, \mathbf{S}_k\}$. An important feature that we will frequently use in what follows is that such a matrix $\mathbf{L}$ can be decomposed into a diagonal form *on the same basis as the eigenvectors of* $\mathbf{S}_k$. Moreover, the eigenvalues of $\mathbf{L}$ can be straightforwardly expressed in terms of the λ_{kj}'s. In particular, we have the following.

LEMMA 2.3. *Assume that $|z| \le T$, where $T \ge 1$. Then for all integers $\ell \ge 1$,*

(2.15) $$tr(\mathbf{\Gamma}_k)^\ell \le \left(\frac{1}{v^2}\right)^{\ell-1} tr\mathbf{\Gamma}_k,$$

(2.16) $$tr(\mathbf{\Lambda}_k)^\ell \le \left(\frac{T}{v^2}\right)^{\ell-1} tr\mathbf{\Lambda}_k.$$

Proof. (i) The inequality (2.15) follows from

$$tr(\mathbf{\Gamma}_k)^\ell = \sum_{j=1}^{p-1} \frac{1}{|\lambda_{kj} - z|^{2\ell}} \leq v^{-2(\ell-1)} \sum_{j=1}^{p-1} \frac{1}{|\lambda_{kj} - z|^2} = v^{-2(\ell-1)} tr\mathbf{\Gamma}_k.$$

(ii) For the inequality (2.16), we have

$$tr(\mathbf{\Lambda}_k)^\ell = \sum_{j=1}^{p-1} \frac{\lambda_{kj}^\ell}{|\lambda_{kj} - z|^{2\ell}}.$$

The conclusion follows from the fact that the function $\varphi(\lambda) := \lambda^{-1}|\lambda - z|^2$ defined on $(0, \infty)$ is convex and has a unique minimum value φ^* satisfying

$$\varphi^* = 2(\sqrt{u^2 + v^2} - u) = 2\frac{v^2}{|z| + u} \geq \frac{v^2}{T}. \qquad \square$$

LEMMA 2.4. *For the Marčenko–Pastur distribution F_y, we have*

(2.17) $$\int_a^b \frac{1}{|x - z|^2} dF_y(x) \leq \frac{2}{y^{3/4}} v_y v^{-2}.$$

Proof. Since for $x \in [a, b]$, $(x - a)(b - x) \leq x(b - a) = 4x\sqrt{y}$, we have for any z,

$$\int_a^b \frac{1}{|x - z|^2} dF_y(x) = \int_a^b \frac{1}{|x - z|^2} \frac{1}{2\pi xy} \sqrt{(x - a)(b - x)} \, dx$$

$$\leq \frac{1}{\pi y^{3/4}} \int_a^b \frac{1}{\sqrt{x}|x - z|^2} \, dx.$$

The maximum of $\int_a^b 1/(\sqrt{x}|x - z|^2) \, dx$ can only be attained when $u \geq a$. We then have

$$\int_a^b \frac{1}{|x - z|^2} dF_y(x) \leq \frac{1}{\pi y^{3/4}} \int_0^\infty \frac{1}{\sqrt{x}|x - z|^2} \, dx$$

$$= \frac{1}{y^{3/4}} \frac{1}{|z|\sqrt{2|z| - 2u}} \leq \frac{1}{y^{3/4}} |z|^{-1/2} v^{-1}.$$

From this and by noticing that $|z|^{1/2} \geq \frac{1}{2}(\sqrt{a} + \sqrt{v})$ when $u > a$, the lemma is proved. $\square$

3. Proofs. We first truncate and centralize the random variables so that all random variables could be further considered as bounded (up to some order of n). In the subsection 3.2, we introduce a Bai inequality for the proofs of the main theorems. These proofs are then given in subsequent sections.

3.1. Truncation and centralization. Define $\hat{x}_{ij} = x_{ij}I(|x_{ij}| \leq \delta\sqrt{n})$ and $\tilde{x}_{ij} = (\hat{x}_{ij} - \mathbb{E}(\hat{x}_{ij}))/\sigma_{ij}$, where $\sigma_{ij}^2 = \mathbb{E}(\hat{x}_{ij} - \mathbb{E}(\hat{x}_{ij}))^2$. Here $\delta = \delta_n$ is chosen such that $\delta_n \to 0$ with a slow rate and such that (1.1) holds. We remind the reader that all the above variables depend on n, but the index is suppressed.

Define $p \times n$ matrices $\hat{X} = (\hat{x}_{ij})$ and $\tilde{X} = (\tilde{x}_{ij})$ and define $p \times p$ matrices $\hat{S} = \frac{1}{n}\hat{X}\hat{X}^T$ and $\tilde{S} = \frac{1}{n}\tilde{X}\tilde{X}^T$. Denote the ESDs of $\hat{S}$ and $\tilde{S}$ by $\hat{F}_p$ and $\tilde{F}_p$, respectively.

Z. D. BAI, BAIQI MIAO, AND JIAN-FENG YAO

We first estimate the truncation error $\|F_p - \hat{F}_p\|$. By (1.1),

$$\sum_{i,j} P\left(|x_{ij}| \geq \delta_n \sqrt{n}\right)$$

$$= \sum_{i,j} \mathbb{E}I\left(|x_{ij}| \geq \delta_n \sqrt{n}\right) \leq (\delta_n \sqrt{n})^{-8} \sum_{i,j} \mathbb{E}\left[x_{ij}^8 I\left(|x_{ij}| \geq \delta_n \sqrt{n}\right)\right]$$

$$(3.1) \qquad \leq cn^{-2}.$$

Let $\alpha \in (0,1)$. By the Markov inequality,

$$P\left(\sum_{i,j} I(|x_{ij}| \geq \delta_n \sqrt{n}) \geq n^{-\alpha}\right) \leq cn^{-2+\alpha},$$

which, together with the Borel–Cantelli lemma, implies that

$$(3.2) \qquad \mathrm{rank}(X - \hat{X}) \leq \sum_{i,j} I(|x_{ij}| \geq \delta_n \sqrt{n}) = O(n^{-\alpha}) \quad \text{a.s.}$$

By Lemma 2.6 of Bai (1999), we have

$$(3.3) \qquad \|F_p - \hat{F}_p\| = O(1/n^{1+\alpha}) \quad \text{a.s.}$$

The estimation (3.3) reduces the proofs to show that the three theorems remain true when F_p is replaced with $\hat{F}_p$.

Furthermore, recalling the proof of Lemma 2.7 of Bai (1999), we find that

$$\int |\hat{F}_p(x) - \tilde{F}_p(x)| dx = \frac{1}{p} \sum_{k=1}^{p} |\hat{\lambda}_k - \tilde{\lambda}_k|$$

$$(3.4) \qquad \leq \left(\frac{1}{np} tr(\hat{X} - \tilde{X})(\hat{X} - \tilde{X})^T \frac{2}{np} tr(\hat{X}\hat{X}^T + \tilde{X}\tilde{X}^T)\right)^{1/2},$$

where $\hat{\lambda}_k$ and $\tilde{\lambda}_k$, arranged in increasing order, are the eigenvalues of $\hat{S}$ and $\tilde{S}$, respectively.

Under the uniform boundedness of the fourth moments of the entries, it is easy to show that

$$(3.5) \qquad \frac{1}{np} tr\hat{X}\hat{X}^T \leq \frac{1}{np} \sum_{ij} |x_{ij}|^2 \to 1 \quad \text{a.s.}$$

Also,

$$(3.6) \qquad 1 \geq \max_{ij} \sigma_{ij}^2 \geq \min_{ij} \sigma_{ij}^2 \to 1.$$

Furthermore, by (3.5) and (3.6),

$$\left(\frac{1}{np} tr\tilde{X}\tilde{X}^T\right)^{1/2} = \frac{1}{\sqrt{np}} \left(\sum_{ij} |\tilde{x}_{ij}|^2\right)^{1/2}$$

$$\leq \frac{1}{\sqrt{np}(\min_{ij} \sigma_{ij})} \left(\left(\sum_{ij} |x_{ij}|^2\right)^{1/2} + \left(\sum_{ij} \mathbb{E}|x_{ij}|^2 I(|x_{ij}| \geq \delta \sqrt{n})\right)^{1/2}\right)$$

$$(3.7) \qquad \to 1 \quad \text{a.s.}$$

Note that

$$\frac{1}{np} tr(\hat{X} - \tilde{X})(\hat{X} - \tilde{X})^T$$

$$\leq \frac{2}{np} \sum_{ij} \left[|x_{ij}^2| \max_{ij} |1 - 1/\sigma_{ij}|^2 \right.$$

$$(3.8) \qquad\qquad \left. + \mathbb{E}^2\{|x_{ij}|I(|x_{ij}| \geq \delta\sqrt{n})\} \right] = O(n^{-6}) \text{ a.s.}$$

Here, the convergence rates follow from the facts that

(i) $\dfrac{1}{np} \sum_{ij} |x_{ij}^2| \to 1$ a.s.

(ii) $\max_{ij} |1 - 1/\sigma_{ij}|^2 \leq \max_{ij} \mathbb{E}^2\{|x_{ij}|^2 I(|x_{ij}| \geq \delta\sqrt{n})\} = O(\delta^{-12} n^{-6})$.

(iii) $\dfrac{1}{np} \sum_{ij} \mathbb{E}^2\{|x_{ij}|I(|x_{ij}| \geq \delta\sqrt{n})\} = O(\delta^6 n^{-7})$.

It follows from (3.4)–(3.8) that under conditions (C.1)–(C.3),

$$(3.9) \qquad\qquad \int |\hat{F}_p(x) - \tilde{F}_p(x)| dx = O(n^{-3}) \text{ a.s.}$$

Using Lemma 2.5 of Bai (1993b), the proofs of the three theorems reduce to show that the main theorem remains true when $\hat{F}_p$ is replaced with $\tilde{F}_p$. Note that the random variables $\tilde{x}_{ij}$ still satisfy the conditions (C.1)–(C.3). They also satisfy the additional condition

$$|x_{ij}| \leq \delta\sqrt{n}$$

(here, the constant δ should be 3δ if δ is the one we previously selected. For brevity, we still use δ). Also, for simplicity, we shall drop the tilde sign from various variables.

3.2. The Bai inequality. Suppose that G is a function of bounded variation. The Stieltjes transform g of G is defined as

$$g(z) = \int_{-\infty}^{\infty} \frac{1}{x - z} dG(x),$$

where $z = u + iv$ and $v > 0$. Our main tool is the following inequality (Bai (1993a)).

PROPOSITION 3.1. *Let G be a distribution function and H be a function of bounded variation satisfying $\int |G(x) - H(x)| dx < \infty$. Denote their Stieltjes transforms by $g(z)$ and $h(z)$, respectively. Then*

$$\|G - H\| \leq \frac{1}{\pi(1-\kappa)(2\gamma-1)} \left[\int_{-A}^{A} |g(z) - h(z)| du + \frac{2\pi}{v} \int_{|x|>B} |G(x) - H(x)| dx \right.$$

$$\left. + \frac{1}{v} \sup_x \int_{|y| \leq 2va_*} |H(x+y) - H(x)| dy \right],$$

where the constants $A > B$, γ, and a_* are restricted by

$$\gamma = \frac{1}{\pi} \int_{|u| \le a_*} \frac{1}{u^2 + 1} du > \frac{1}{2}, \quad and \quad \kappa = \frac{4B}{\pi(A - B)(2\gamma - 1)} \in (0, 1).$$

Denote the Stieltjes transform of F_p and F_y (recall our convention that $y = y_n = p/n$) by $m_p(z)$ and $m(z)$, respectively. Application of the Bai inequality with $(G, H) = (F_p, F_y)$, $A = 25$, and $B = 5$ gives, for some constant $c > 0$,

$$\|F_p - F_y\| \le c \left[\int_{-A}^{A} |m_p(z) - m(z)| \, du + \frac{1}{v} \int_{|x| > 5} |F_p(x) - e(x)| \, dx \right.$$

(3.10)
$$\left. + \frac{1}{v} \sup_x \int_{|u| \le 2va_*} |F_y(x + u) - F_y(x)| \, du \right],$$

where $e(x) = 1$ for $x > 0$ or $e(x) = 0$ otherwise. We shall estimate these three terms in the above bound successively and start with the last one.

(a) Estimate for $\sup_x \int_{|u| \le 2va_*} |F_y(x + u) - F_y(x)| \, du$.
LEMMA 3.1. *We have, for any $0 < v < 4\sqrt{y}$,*

$$\sup_x \int_{|u| \le v} |F_y(x + u) - F_y(x)| \, du \le \frac{11\sqrt{2(1 + y)}}{3\pi y} v v_y,$$

where $v_y = v/(\sqrt{a} + \sqrt{v})$ is defined as in (2.7).

Proof. It is enough to consider the part $0 \le u \le v$ in the integral only since the remaining part for $-v \le u \le 0$ can be handled in a similar way. Set $\Phi(\lambda) := \int_0^v [F_y(x + u) - F_y(x)] du$ with $x = a + \lambda$; we are estimating the maximum of $\Phi(\lambda)$. Without loss of generality, we need only consider the case that $\lambda \ge 0$ because $\int_0^v [F_y(x + u) - F_y(x)] du$ increases when $x \le a$. Then

$$\Phi(\lambda) = \int_0^v du \int_x^{x+u} F_y{}'(t) \, dt$$

$$= \int_{a+\lambda}^{a+\lambda+v} \frac{a + \lambda + v - t}{2\pi y t} \sqrt{(t - a)(b - t)} I_{[a,b]}(t) dt$$

(3.11)
$$= \int_{\lambda}^{\lambda+v} \frac{\lambda + v - u}{2\pi y (u + a)} \sqrt{u(4\sqrt{y} - u)} I_{[0, b-a]}(u) \, du.$$

Let $\phi(u) := (u + a)^{-1} \sqrt{u(4\sqrt{y} - u)}$. The derivative of $\log(\phi(u))^2$ is

$$\frac{1}{u} - \frac{1}{4\sqrt{y} - u} - \frac{2}{u + a} = \frac{2(2\sqrt{y}a - (1 + y)u)}{u(4\sqrt{y} - u)(u + a)}.$$

Note that the above equality holds also for $y = 1$ for which $a = 0$. Let $\rho := (1 + y)^{-1}(2a\sqrt{y})$. Thus $\phi(u)$ is decreasing when $u > \rho$ and increasing when $u < \rho$. Since

$$\frac{d\Phi(\lambda)}{d\lambda} = \frac{1}{2\pi y} \left(\int_{\lambda}^{\lambda+v} [\phi(u) - \phi(\lambda)] \, du \right),$$

it follows that for $\lambda > \rho$, $\Phi(\lambda)$ is decreasing and then $\Phi(\lambda) \leq \Phi(\rho)$; for $\lambda < \rho - v$, $\Phi(\lambda)$ is increasing and then $\Phi(\lambda) \leq \Phi(\rho - v)$. Hence, $\Phi(\lambda)$ reaches its maximum only for some $\lambda \in (\max(\rho - v, 0), \rho)$. Considering such a λ yields by (3.11)

$$\Phi(\lambda) \leq \frac{2y^{1/4}}{2\pi y} \int_\lambda^{\lambda+v} \frac{\lambda + v - u}{u + a} \sqrt{u} \, du$$

$$= 2(\pi y^{3/4})^{-1} \left\{ (\lambda + v + a) \left[(\sqrt{\lambda + v} - \sqrt{\lambda}) \right. \right.$$

$$\left. \left. - \sqrt{a} \left(\arctan \sqrt{\frac{\lambda + v}{a}} - \arctan \sqrt{\frac{\lambda}{a}} \right) \right] - \frac{1}{3} \left[(\lambda + v)^{3/2} - \lambda^{3/2} \right] \right\}.$$

Since

$$\sqrt{a} \left(\arctan \sqrt{\frac{\lambda + v}{a}} - \arctan \sqrt{\frac{\lambda}{a}} \right) \geq \frac{a}{\lambda + v + a} \left(\sqrt{\lambda + v} - \sqrt{\lambda} \right),$$

we get, by setting $\lambda^* = \sqrt{\lambda + v} - \sqrt{\lambda}$,

$$\Phi(\lambda) \leq \frac{2}{\pi y^{3/4}} \left\{ (a + \lambda + v) \left(\lambda^* - \frac{a}{a + \lambda + v} \lambda^* \right) - \lambda^* \left(\lambda + \sqrt{\lambda} \lambda^* + \frac{1}{3} \lambda^{*2} \right) \right\}$$

$$(3.12) \quad = \frac{2}{\pi y^{3/4}} \left[\sqrt{\lambda} \lambda^{*2} + \frac{2}{3} \lambda^{*3} \right].$$

Let $c^2 = \frac{1+y}{2\sqrt{y}}$. Since $\lambda + v \geq c^{-2} a$ and

$$(\sqrt{\lambda + v} + \sqrt{\lambda})^2 \geq \lambda + v + 2\sqrt{\lambda(\lambda + v)} \geq 2\sqrt{\lambda v} + 2\sqrt{\lambda c^{-2} a},$$

we have

$$\frac{\sqrt{\lambda}}{(\sqrt{\lambda + v} + \sqrt{\lambda})^2} \leq \frac{c}{2\sqrt{a} + 2c\sqrt{v}} \leq \frac{c}{2\sqrt{a} + 2\sqrt{v}},$$

$$\frac{1}{(\sqrt{\lambda + v} + \sqrt{\lambda})^3} \leq \frac{2c}{(\sqrt{a} + \sqrt{v})v},$$

where the last inequality follows from

$$(\sqrt{\lambda + v} + \sqrt{\lambda})^3 \geq \sqrt{\lambda + v} \, v$$

$$\geq \frac{1}{2} \left[\sqrt{c^{-2} a} + \sqrt{v} v \right]$$

$$\geq \frac{1}{2c} \left[\sqrt{v} + \sqrt{a} \right] v.$$

Hence

$$\Phi(\lambda) \leq \frac{2}{\pi y^{3/4}} \cdot \frac{11c}{6(\sqrt{a} + \sqrt{v})} v^2 = \frac{11\sqrt{2(1+y)}}{6\pi y} \frac{1}{\sqrt{v} + (1 - \sqrt{y})} v^2.$$

This completes the proof of the lemma. $\square$

(b) Estimate for $\frac{1}{v} \int_{|x|>5} |F_p(x) - e(x)| \, dx$. Let λ_p denote the largest eigenvalue of **S**. By Yin, Bai, and Krishnaiah (1988), for any positive constant (ε) and integer (ℓ_n) such that $\ell_n / \log n \to \infty$ and $\ell_n \delta_n^{1/4} / \log n \to 0$, we have

$$(3.13) \qquad E(\lambda_p)^{\ell_n} \leq c(b + \varepsilon)^{\ell_n}.$$

Therefore, for $x \geq 5$ and any fixed $t > 0$,

$$(3.14) \qquad P(\lambda_p > x) \leq c \left(\frac{b + \varepsilon}{x} \right)^{\ell_n} \leq c \left(\frac{b + \varepsilon}{x} \right)^2 \left(\frac{b + \varepsilon}{5} \right)^{\ell_n - 2} = o(x^{-2} n^{-t}).$$

Since $F_p(x) = e(x) = 0$ for $x \leq 0$, we have

$$\int_{|x|>5} |F_p(x) - e(x)| \, dx = \int_5^\infty [1 - F_p(x)] \, dx$$

$$= \int_5^\infty \frac{1}{p} \sum_{k=1}^p P(\lambda_k > x) \, dx \leq \int_5^\infty P(\lambda_p > x) \, dx$$

$$(3.15) \qquad \leq \int_5^\infty o(x^{-2} n^{-t}) \, dx = o(n^{-t}) \,.$$

By (3.15) we finally get

$$\int_{|x|>5} |F_p(x) - e(x)| \, dx = O(n^{-2}) \text{ a.s.}$$

Thus, for $v > cn^{-1}$, we have

$$v^{-1} \int_{|x|>5} |F_p(x) - e(x)| \, dx = O_{a.s.}(n^{-1}) \text{ a.s.}$$

(c) Conclusion. Summarizing previous steps gives

$$(3.16) \qquad \|F_p - F_y\| \leq c \left[\int_{-A}^A |m_p(z) - m(z)| \, du + O_{a.s.}(n^{-1}) + v_y \right].$$

To prove the main theorems, we need only estimate $|m_p(z) - m(z)|$. Bai (1993b) has proved that $\Delta = \Delta_{n,y} = \| \mathbb{E}F_p - F_y \| = O(n^{-5/48})$. In what follows, we shall treat Δ as at least of the order $O(n^{-5/48})$.

3.3. Proof of Theorem 1.1. We begin by estimating $| \mathbb{E}m_p(z) - m(z)|$, with various choices of v, subject to $cn^{-1/2} \leq v \leq 1$ for some $c > 0$. With the formula of m_p given in (2.10), let us define δ_p such that

$$(3.17) \qquad m_p(z) = - \frac{1}{z + y - 1 + yz\mathbb{E}m_p(z)} + \delta_p = - \frac{1}{\mathbb{E}\beta} + \delta_p.$$

Since

$$\frac{1}{\beta_k} = \frac{1}{\mathbb{E}\beta} \left(1 - \frac{\varepsilon_k}{\beta_k} \right),$$

it is easy to see that

$$\delta_p = \frac{1}{p}\sum_{k=1}^{p}\frac{1}{\mathbb{E}\beta}\frac{\varepsilon_k}{\beta_k} = \frac{1}{(\mathbb{E}\beta)^2}\left(\frac{1}{p}\sum_{k=1}^{p}\varepsilon_k - \frac{1}{p}\sum_{k=1}^{p}\frac{\varepsilon_k^2}{\beta_k}\right).$$

Now

$$|\mathbb{E}\delta_p|$$

$$\leq \frac{1}{p|\mathbb{E}\beta|^2}\sum_{k=1}^{p}\left(|\mathbb{E}\varepsilon_k| + \left|\mathbb{E}\frac{\varepsilon_k^2}{\beta_k}\right|\right)$$

$$= \frac{1}{p|\mathbb{E}\beta|^2}\sum_{k=1}^{p}\left[|\mathbb{E}(\varepsilon_k^* + \tilde{\varepsilon}_k) + \pi_k| + \left|\frac{1}{\mathbb{E}\beta}\mathbb{E}\varepsilon_k^2 - \frac{1}{(\mathbb{E}\beta)^2}\mathbb{E}\varepsilon_k^3 + \frac{1}{(\mathbb{E}\beta)^2}\mathbb{E}\left(\frac{\varepsilon_k^4}{\beta_k}\right)\right|\right]$$

$$\leq \frac{1}{p|\mathbb{E}\beta|^2}\left[\sum_{k=1}^{p}|\mathbb{E}(\varepsilon_k^* + \tilde{\varepsilon}_k) + \pi_k| + \sum_{k=1}^{p}\left|\frac{1}{\mathbb{E}\beta}\mathbb{E}\varepsilon_k^2\right| + \sum_{k=1}^{p}\left|\frac{1}{(\mathbb{E}\beta)^2}\mathbb{E}\varepsilon_k^3\right|\right.$$

$$\left. + \sum_{k=1}^{p}\left|\frac{1}{(\mathbb{E}\beta)^2}\mathbb{E}\left(\frac{\varepsilon_k^4}{\beta_k}\right)\right|\right]$$

$$= |\mathbb{E}\beta|^{-2}\left[I_0 + I_1 + I_2 + I_3\right].$$

We will estimate each I_i to obtain a bound on $|\mathbb{E}\delta_p|$ (cf. (3.19) below). Since $\mathbb{E}(\varepsilon_k^* + \tilde{\varepsilon}_k) = 0$, by (2.9), we have

$$I_0 = \frac{|z|}{p}\sum_{k=1}^{p}|\pi_k| = \frac{|z|}{pn}\sum_{k=1}^{p}|\mathbb{E}tr\mathbf{D}_k - \mathbb{E}tr\mathbf{D}| \leq |z|/(nv).$$

From Lemma 4.2, Remark 4.1, and noticing that $v \leq v_y$, we have

$$I_1 \leq \frac{1}{p|\mathbb{E}\beta|}\sum_{k=1}^{p}\mathbb{E}|\varepsilon_k|^2 = \frac{1}{p|\mathbb{E}\beta|}\sum_{k=1}^{p}(\mathbb{E}|\varepsilon_k^*|^2 + \mathbb{E}|\tilde{\varepsilon}_k|^2 + |\pi_k|^2)$$

$$\leq \frac{c}{|\mathbb{E}\beta|}\left(\left[\frac{1}{n} + \frac{\Delta + v_y}{nv^2}\right] + \frac{\Delta + v_y}{n^2v^4} + \frac{1}{n^2v^2}\right) \leq \frac{c(\Delta + v_y)}{|\mathbb{E}\beta|nv^2},$$

$$I_2 = \frac{1}{p|\mathbb{E}\beta|^2}\sum_{k=1}^{p}|\mathbb{E}\varepsilon_k^3| \leq \sum_{k=1}^{p}\left(\frac{1}{p|\mathbb{E}\beta|}\mathbb{E}|\varepsilon_k|^2 + \frac{1}{p|\mathbb{E}\beta|^3}\mathbb{E}|\varepsilon_k|^4\right).$$

Now

$$\frac{1}{p}\sum_{k=1}^{p}\mathbb{E}|\varepsilon_k|^4 \leq \frac{27}{p}\sum_{k=1}^{p}(\mathbb{E}|\varepsilon_k^*|^4 + \mathbb{E}|\tilde{\varepsilon}_k|^4 + |\pi_k|^4) \stackrel{\triangle}{=} c(I_{21} + I_{22} + I_{23}).$$

Since

$$tr\mathbf{B}_k\bar{\mathbf{B}}_k = tr(\mathbf{I}_{p-1} + z\mathbf{D}_k)(\mathbf{I}_{p-1} + \bar{z}\bar{\mathbf{D}}_k) \leq 2(p + |z|^2tr\mathbf{D}_k\bar{\mathbf{D}}_k),$$

We have from the proof of Lemma 4.1,

$$\mathbb{E}|\varepsilon_k^*|^4 \leq cn^{-2}\left\{1 + n^{-2}\mathbb{E}(tr\mathbf{B}_k\bar{\mathbf{B}}_k)^2\right\}$$

$$\leq cn^{-2}\left\{1 + n^{-2}\mathbb{E}(tr\mathbf{D}_k\bar{\mathbf{D}}_k)^2\right\}.$$

Now

$$
\begin{aligned}
\mathbb{E}(tr(\mathbf{D}_k\overline{\mathbf{D}_k}))^2 &= v^{-2}\,\mathbb{E}(\operatorname{Im}(tr(\mathbf{D}_k)))^2 \\
&\leq 2v^{-2}[v^{-2} + \mathbb{E}(\operatorname{Im}(tr(\mathbf{D})))^2] \\
&= 2v^{-4} + 2p^2v^{-2}\,\mathbb{E}(\operatorname{Im}(m_p(z)))^2 \\
&\leq 2v^{-4} + 4p^2v^{-2}|\,\mathbb{E}m_p(z)|^2 + 4p^2v^{-2}\,\mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^2 \\
&\leq cp^2v^{-4}(\Delta + v_y)^2 + cv^{-6}(\Delta + v_y) \leq cp^2v^{-4}(\Delta + v_y)^2,
\end{aligned}
$$

where the second inequality follows from (2.9) and the last steps follow from Proposition 4.1 and

$$
(3.18) \qquad |\,\mathbb{E}m_p(z)| \leq |\,\mathbb{E}m_p(z) - m(z)| + |m(z)| \leq v^{-1}(2\Delta + \alpha_y v_y),
$$

with $\alpha_y := 2\sqrt{2}/\sqrt{y}$ (see (2.6)). Thus

$$
\begin{aligned}
I_{21} &\leq c\left\{n^{-2} + n^{-2}v^{-4}(\Delta + v_y)^2\right\} \\
&\leq cn^{-2}v^{-4}(\Delta + v_y)^2.
\end{aligned}
$$

Also, considering $\mathbf{D}_k$ instead of $\mathbf{D}$ as in Proposition 4.1 and applying (2.8), one can show that for some L_0 such that for all $L_0 n^{-1/2} \leq v < 1$,

$$
I_{22} \leq c(\Delta + v_y)^2 n^{-4}v^{-8}.
$$

Since $|\pi_k| \leq |z|(nv)^{-1}$, we have $I_{23} \leq |z|^4(nv)^{-4}$, and hence,

$$
\begin{aligned}
p^{-1}\sum_{k=1}^p \mathbb{E}|\varepsilon_k|^4 &\leq c(I_{21} + I_{22} + I_{23}) \\
&\leq c[n^{-2}v^{-4}(\Delta + v_y)^2 + (\Delta + v_y)^2 n^{-4}v^{-8} + (nv)^{-4}] \\
&\leq cn^{-2}v^{-4}(\Delta + v_y)^2.
\end{aligned}
$$

Consequently, for some constant $c > 0$,

$$
I_2 \leq \frac{c(\Delta + v_y)}{|\,\mathbb{E}\beta|nv^2} + \frac{c(\Delta + v_y)^2}{|\,\mathbb{E}\beta|^3 n^2 v^4}
$$

and

$$
I_3 \leq \frac{1}{pv|\,\mathbb{E}\beta|^2}\sum_{k=1}^p \mathbb{E}|\varepsilon_k|^4 \leq \frac{c}{n^2 v^5|\,\mathbb{E}\beta|^2}(\Delta + v_y)^2.
$$

Summing up the above results, we obtain

$$
\begin{aligned}
|\,\mathbb{E}\delta_p| &\leq \frac{1}{|\,\mathbb{E}\beta|^2}[I_0 + I_1 + I_2 + I_3] \\
&\leq \frac{c}{|\,\mathbb{E}\beta|^2}\left[\frac{1}{nv} + \frac{\Delta + v_y}{nv^2|\,\mathbb{E}\beta|} + \frac{(\Delta + v_y)^2}{n^2 v^5|\,\mathbb{E}\beta|^2}\right] \\
(3.19) \qquad &\leq \frac{c}{|\,\mathbb{E}\beta|^2}\left[\frac{\Delta + v_y}{nv^2|\,\mathbb{E}\beta|} + \frac{(\Delta + v_y)^2}{n^2 v^5|\,\mathbb{E}\beta|^2}\right].
\end{aligned}
$$

For the positive constant L_0 (required by Proposition 4.1) and all $v \in [L_0 n^{-1/2}, 1)$, define

$$\varphi_n(v) = \sup_{|u| \leq A} |\mathbb{E}\delta_p| - \gamma v \quad \text{with} \quad \gamma = 1/[10(A+1)^2].$$

Checking the proofs of (3.39)–(3.40) of Bai (1993b), we find that there is a constant c such that

$$\varphi_n(v) \leq 0 \Longrightarrow \int_{-A}^{A} |\mathbb{E}m_p(z) - m(z)|\, du < cv.$$

In view of (3.16), we can then find a positive constant c_1 such that

(3.20) $$\varphi_n(v) \leq 0 \Longrightarrow \Delta < c_1 v_y.$$

The proof of the theorem will be complete once we have shown that for all large n and all $v \in [Ln^{-1/[4\theta+2]}, 1)$, we have $\varphi_n(v) \leq 0$, where L is a constant such that $L \geq L_0$ and

$$cM_0^2 \left[\frac{(c_1+1)M_0}{L^2} + \frac{(1+c_1)^2 M_0^2}{L^4} \right] < \gamma$$

and $M_0 = \gamma + 2c_1 + \alpha_y$.

Assume the contrary; i.e., there exists a $v_1 \in [Ln^{-1/[4\theta+2]}, 1)$ for which $\varphi_n(v_1) > 0$. By continuity of φ_n, there exists a $v_0 \in [Ln^{-1/[4\theta+2]}, 1)$ for which $\varphi_n(v_0) = 0$. As $[-A, A]$ is compact, there exists a $u_0 \in [-A, A]$ such that $0 = \varphi_n(v_0) = |\mathbb{E}\delta_p(u_0, v_0)| - \gamma v_0$. Let $z_0 = u_0 + iv_0$. By (3.17), (3.18), and Lemma 2.2, with $z = z_0$,

(3.21)
$$\frac{1}{|\mathbb{E}\beta|} = |-\mathbb{E}\delta_p(z_0) + \mathbb{E}[m_p(z_0) - m(z_0)] + m(z_0)| \leq |\mathbb{E}\delta_p(z_0)| + \frac{2\Delta + \alpha_y v_{y,0}}{v_0}$$

$$\leq (\gamma + 2c_1 + \alpha_y) \frac{v_{y,0}}{v_0} = M_0 \frac{v_{y,0}}{v_0}.$$

On the other hand, by definition (1.2) of θ, we have

$$\frac{v}{v_y} = \sqrt{a} + \sqrt{v} \geq \sqrt{a} \vee \sqrt{v} \geq n^{-\frac{\theta}{4\theta+2}}.$$

Therefore, for any $v \in [Ln^{-1/[4\theta+2]}, 1)$ we have

$$\frac{1}{nv^2}\left(\frac{v_y}{v}\right)^4 \leq 1/L^2, \qquad \frac{1}{n^2 v^4}\left(\frac{v_y}{v}\right)^6 \leq 1/L^4.$$

Thus, from (3.19) we have for $z = z_0$ (so $v = v_0$ and $v_y = v_{y,0}$)

$$|\mathbb{E}\delta_p(z_0)| \leq \frac{cM_0^2 v_y^2}{v^2}\left[\frac{(c_1+1)M_0 v_y^2}{nv^3} + \frac{(1+c_1)^2 M_0^2 v_y^4}{n^2 v^7}\right]$$

$$\leq vcM_0^2\left[\frac{(c_1+1)M_0}{L^2} + \frac{(1+c_1)^2 M_0^2}{L^4}\right] < \gamma v$$

by noticing the selection of L.

This leads to a contradiction of $\varphi_n(z_0) = 0$. The proof of Theorem 1.1 is complete.

3.4. Proof of Theorem 1.2. As the proof of (3.16), one can show that

$$
\mathbb{E}||F_p - F_y||
$$
$$
\leq c \left[\int_{-A}^{A} \mathbb{E}|m_p(z) - m(z)|du + v_y \right]
$$
$$
\leq c \left[\int_{-A}^{A} \mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|du + \int_{-A}^{A} |\mathbb{E}m_p(z) - m(z)|du + v_y \right].
$$

In the proof of Theorem 1.1, we have shown that $\int_{-A}^{A} |\mathbb{E}m_p(z) - m(z)|du = O(v)$ if $Ln^{-1/[2+4\theta]} < v < 1$.

Applying the Cauchy–Schwarz inequality, Remark 4.1, and the result $\Delta = O(n^{-1/[4\theta+2]})$ proved in Theorem 1.1, we conclude that

$$
\int_{-A}^{A} \mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|du \leq \int_{-A}^{A} (\mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^2)^{1/2}du
$$
$$
\leq cn^{-1/2}v_y^{1/2}v^{-2} \leq v
$$

for some positive constant c and all $cn^{-2/[5+\theta]} \leq v < 1$. Recall that we need a condition of $v > Ln^{-1/[4\theta+2]}$ to guarantee $\int_{-A}^{A} |\mathbb{E}m_p(z) - m(z)|du = O(v)$. The convergence rate we can guarantee is

$$
O_p \left(\max \left\{ \frac{n^{-(2/(5+\theta))}}{[1 - \sqrt{y} + n^{-(1/(5+\theta))}]}, \frac{n^{-1/[4\theta+2]}}{[1 - \sqrt{y} + n^{-1/[8\theta+4]}]} \right\} \right).
$$

The proof of Theorem 1.2 in this case is complete.   □

3.5. Proof of Theorem 1.3. Similarly, we have

$$
(3.22) \qquad ||F_p - F_y|| \leq c \left[\int_{-A}^{A} |m_p(z) - \mathbb{E}m_p(z)|du + v_y \right].
$$

Thus, to complete the proof of Theorem 1.3, setting $v = n^{-2/[5+\theta]+\eta}$ it suffices to show that

$$
(3.23) \qquad v^{-1} \int_{-A}^{A} |m_p(z) - \mathbb{E}m_p(z)| du \to 0 \qquad \text{a.s.}
$$

Now, applying Proposition 4.1, we obtain for each $\xi > 0$,

$$
P \left(\int_{-A}^{A} |m_p(z) - \mathbb{E}m_p(z)| du \geq \xi v \right)
$$
$$
\leq (v\xi)^{-2k}(2A)^{2k-1} \int_{-A}^{A} \mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^{2k} du
$$
$$
\leq \xi^{-2k}(2A)^{2k} \left[c_k \left(n^{-2}v^{-6}v_y \right)^k \right]
$$
$$
\leq c_k'(\varepsilon\xi)^{-2k}n^{-(5+\theta)\eta k}.
$$

The right-hand side of the above inequality is summable by choosing k such that $5\eta k > 1$. Recalling the condition used in the proof of Theorem 1.1, the convergence rate is

$$O_{a.s.}\left(\max\left\{\frac{n^{-(2/(5+\theta))+\eta}}{[1-\sqrt{y}+n^{-(1/(5+\theta))}]}, \frac{n^{-1/[4\theta+2]}}{[1-\sqrt{y}+n^{-1/[8\theta+4]}]}\right\}\right)$$

Thus, (3.23) is proved and the proof of Theorem 1.3 is complete. $\square$

4. Intermediate lemmas. In this section, we establish a few more technical lemmas. Let $\nu_\ell = \sup_{i,j,n}\{E|x_{ij}|^\ell\}$.

LEMMA 4.1. *For each $\ell \geq 1$ with $\nu_{4\ell} < \infty$, there exist positive constants c_ℓ independent of n and v such that for all n, v satisfying $nv \geq T$, we have*

$$(4.1) \qquad \mathbb{E}\left(\left.|\varepsilon_k^*|^{2\ell}\right|\mathcal{F}^{(k)}\right) \leq c_\ell(1+\lambda_p)^{\ell/2}n^{-\ell}\left(1+\frac{1}{n}tr\mathbf{\Lambda}_k\right)^\ell$$

and

$$(4.2) \qquad \mathbb{E}\left(\left.\frac{(\varepsilon_k^*)^{2\ell}}{|\beta_k^*|^\ell}\right|\mathcal{F}^{(k)}\right) \leq c_\ell(1+\lambda_p)^{\ell/2}n^{-\ell}v^{-\ell}.$$

Proof. We have

$$\mathbb{E}\left(|\varepsilon_k^*|^{2\ell}|\mathcal{F}^{(k)}\right) = \mathbb{E}\left(\left.\left|-\frac{1}{n}\sum_{j=1}^n(x_{kj}^2-1)+\frac{1}{n}(\mathbf{x}_k'\mathbf{B}_k\mathbf{x}_k-tr\mathbf{B}_k)\right|^{2\ell}\right|\mathcal{F}^{(k)}\right)$$

$$\leq 2^{2\ell-1}n^{-2\ell}\left\{\mathbb{E}\left|\sum_{j=1}^n(x_{kj}^2-1)\right|^{2\ell} + \mathbb{E}\left(\left.|\mathbf{x}_k'\mathbf{B}_k\mathbf{x}_k-tr\mathbf{B}_k|^{2\ell}\right|\mathcal{F}^{(k)}\right)\right\}$$

$$:= A + B.$$

For the first term A, by the Burkholder inequality (Burkholder (1973, p. 22)), we get

$$\mathbb{E}\left|\sum_{j=1}^n(x_{kj}^2-1)\right|^{2\ell}$$

$$\leq c_\ell\,\mathbb{E}\left[\sum_{j=1}^n(x_{kj}^2-1)^2\right]^\ell \leq c_\ell n^{\ell-1}\,\mathbb{E}\left[\sum_{j=1}^n(x_{kj}^2-1)^{2\ell}\right] \leq c_\ell\nu_{4\ell}n^\ell.$$

For the second term B, denoting the eigenvalues of $\mathbf{S}_k$ by λ_{kj} and noticing that their maximum is less than the largest eigenvalue λ_p of $\mathbf{S}$, we then have

$$tr\left(\mathbf{B}_k\overline{\mathbf{B}}_k\right) = tr\mathbf{B}_k + \overline{z}\,tr\mathbf{\Lambda}_k,$$

$$= \sum_{j=1}^{p-1}\frac{\lambda_{kj}}{\lambda_{kj}-z} + \overline{z}\sum_{j=1}^{p-1}\frac{\lambda_{kj}}{|\lambda_{kj}-z|^2} \leq \lambda_p^{1/2}\sum_{j=1}^{p-1}\frac{\lambda_{kj}^{1/2}}{|\lambda_{kj}-z|} + T\sum_{j=1}^{p-1}\frac{\lambda_{kj}}{|\lambda_{kj}-z|^2}$$

$$\leq (\lambda_p^{1/2}+T)(p-1+tr\mathbf{\Lambda}_k).$$

Therefore by Lemma 2.1,

$$E\left(\,|\mathbf{x}_k'\mathbf{B}_k\mathbf{x}_k - tr\mathbf{B}_k|^{2\ell}\,\big|\,\mathcal{F}^{(k)}\right)$$

$$\leq c_\ell(\nu_{4\ell} + M^\ell)(tr\mathbf{B}_k\overline{\mathbf{B}}_k)^\ell \leq c_\ell(\lambda_p^{1/2} + T)^\ell n^\ell \left(1 + \frac{1}{n}tr\mathbf{\Lambda}_k\right)^\ell .$$

Combining the bounds for A and B proves the first conclusion. The second conclusion immediately follows by taking into account inequality (2.12). $\square$

LEMMA 4.2. *If $n^{-1/2} \leq v < 1$, then there are positive constants C_1, C_2 such that for large n and each $1 \leq k \leq p$,*

$$\text{(i)} \quad |\mathbb{E}tr(\mathbf{D}_k\overline{\mathbf{D}}_k)| \leq C_1 p\frac{\Delta + v_y}{v^2}.$$

$$\text{(ii)} \quad \mathbb{E}|\varepsilon_k^*|^2 \leq C_2\frac{1}{n}\left(1 + |z|^2\frac{\Delta + v_y}{v^2}\right).$$

Proof. (i) Recall that $\Delta = \|\,\mathbb{E}F_p - F_y\|$. By Lemma 2.2,

$$\left|\int_{-\infty}^\infty \frac{1}{|x - z|^2}d(\,\mathbb{E}F_p(x) - F_y(x))\right| \leq \frac{2\Delta}{v^2}.$$

Application of Lemma 2.1 and inequality (2.8) yields that

$$|\,\mathbb{E}tr(\mathbf{D}_k\overline{\mathbf{D}}_k)| = \left|(p - 1)\int_{-\infty}^\infty \frac{1}{|x - z|^2}d[\,\mathbb{E}F_p^{(k)}(x)]\right|$$

$$\leq \left|\int_{-\infty}^\infty \frac{1}{|x - z|^2}d[(p - 1)\,\mathbb{E}F_p^{(k)}(x) - p\,\mathbb{E}F_p(x)]\right|$$

$$+ p\left|\int_{-\infty}^\infty \frac{1}{|x - z|^2}d[\,\mathbb{E}F_p(x) - F_y(x)]\right| + p\left|\int_{-\infty}^\infty \frac{1}{|x - z|^2}dF_y(x)\right|$$

$$\leq \frac{2}{v^2} + p\frac{2\Delta}{v^2} + p\left|\int_{-\infty}^\infty \frac{1}{|x - z|^2}dF_y(x)\right|.$$

Here, the bound of the last term follows from Lemma 2.4. The proof of conclusion (i) is complete.

(ii) This conclusion follows from (i), (4.1), and the fact

$$tr\mathbf{B}_k\overline{\mathbf{B}}_k = tr(\mathbf{I}_{p-1} + z\mathbf{D}_k)(\mathbf{I}_{p-1} + \overline{z}\overline{\mathbf{D}}_k) \leq 2(p + |z|^2 tr\mathbf{D}_k\overline{\mathbf{D}}_k).   \square$$

LEMMA 4.3. *Assume $|z| \leq T$ with $T \geq 2$. Then there are constants C_0, C_1 such that for all $v \geq C_0 n^{-1/2}$ and large n, we have*

$$(4.3) \qquad\qquad \sum_{k=1}^p \mathbb{E}(|\beta_k^*|^{-1}) \leq C_1 n(\Delta + v_y)v^{-1} .$$

Proof. From the definition of ε_k^*, we notice that $(\beta_k^*)^{-1} = \beta_k^{-1}(1 + \beta_k^{-1}\varepsilon_k^*)$. By (2.9),

$$|\beta_k^* - \beta| = \frac{1}{n}|-1 + z(tr\mathbf{D}_k - tr\mathbf{D})| \leq \frac{1}{n}\left(1 + \frac{|z|}{v}\right) \leq \frac{2T}{nv}.$$

By (3.13), it is easy to see that for any fixed $t > 0$ and all large n,

$$
\mathbb{E}\left[(1+\lambda_p)^{\ell}|W|\right] \leq 6^{\ell}\,\mathbb{E}|W| + o(\|W\|n^{-t}), \tag{4.4}
$$

where W is a bounded random variable with a nonrandom bound $\|W\|$. By this and taking into account (2.10), (4.2), and (2.6), we obtain

$$
\sum_{k=1}^{p} \mathbb{E}(|\beta_k^*|^{-1})
$$

$$
\leq \sum_{k=1}^{p}\mathbb{E}\left|\frac{1}{|\beta_k^*|}-\frac{1}{|\beta|}\right| + \mathbb{E}\left|\sum_{k=1}^{p}\left(\frac{1}{\beta}-\frac{1}{\beta_k^*}\right)\right| + \mathbb{E}\left|\sum_{k=1}^{p}\left(\frac{1}{\beta_k^*}-\frac{1}{\beta_k}\right)\right| + \mathbb{E}\left|\sum_{k=1}^{p}\beta_k^{-1}\right|
$$

$$
\leq 2\sum_{k=1}^{p}\mathbb{E}\frac{|\beta_k^*-\beta|}{|\beta||\beta_k^*|} + \sum_{k=1}^{p}\mathbb{E}\frac{|\varepsilon_k^*|}{|\beta_k^*|^2} + \sum_{k=1}^{p}\mathbb{E}\frac{|\varepsilon_k^*|^2}{|\beta_k||\beta_k^*|^2} + p\,\mathbb{E}|m_p(z)|
$$

$$
\leq \frac{c}{nv^2}\sum_{k=1}^{p}\mathbb{E}(|\beta_k^*|^{-1}) + \sum_{k=1}^{p}\mathbb{E}\frac{\left(\,\mathbb{E}(|\varepsilon_k^*|^2|\mathcal{F}^{(k)})\right)^{1/2}}{|\beta_k^*|^2} + \sum_{k=1}^{p}\mathbb{E}\frac{\mathbb{E}(|\varepsilon_k^*|^2|\mathcal{F}^{(k)})}{v|\beta_k^*|^2} + p\,\mathbb{E}|m_p(z)|
$$

$$
\leq c_*\left(\frac{1}{nv^2}+n^{-1/2}v^{-1}+\frac{1}{nv^2}\right)\sum_{k=1}^{p}\mathbb{E}(|\beta_k^*|^{-1}) + p\,\mathbb{E}|m_p(z)| + o(n^{-t})
$$

$$
\leq 3c_*(nv^2)^{-1/2}\sum_{k=1}^{p}\mathbb{E}(|\beta_k^*|^{-1}) + p\,\mathbb{E}|m_p(z)| + o(n^{-t})
$$

for all $v \geq 1/\sqrt{n}$. Let $C_0 = 1 \vee (6c_*)^2$. We have for all $v \geq \sqrt{C_0/n}$

$$
\sum_{k=1}^{p}\mathbb{E}(|\beta_k^*|^{-1}) \leq \frac{1}{2}\sum_{k=1}^{p}\mathbb{E}(|\beta_k^*|^{-1}) + p\,\mathbb{E}|m_p(z)| + o(n^{-t})
$$

$$
\leq 2p\,\mathbb{E}|m_p(z)| + o(n^{-t})
$$

$$
\leq 2p\,\mathbb{E}|m_p(z)-\mathbb{E}(m_p(z))| + 2p|\mathbb{E}(m_p(z))-m(z)| + 2p|m(z)| + o(n^{-t})
$$

$$
\leq 2\sqrt{p}v^{-1} + 4p\Delta/v + 2pv_y/v.
$$

The proof is now complete. $\square$

LEMMA 4.4. *Let* $z_k = \mathbb{E}(tr\mathbf{D}|\mathcal{F}_{k-1}) - \mathbb{E}(tr\mathbf{D}|\mathcal{F}_k)$. *Then* $tr\mathbf{D} - \mathbb{E}tr\mathbf{D} = \sum_{k=1}^{p} z_k$ *and* (z_k) *is a martingale difference with respect to* $(\mathcal{F}_k)$, $k = p, p-1, \ldots, 0$. *Moreover, we have the following formula for* z_k:

$$
z_k = \left\{\,\mathbb{E}\left(a_k|\mathcal{F}_{k-1}\right) - \mathbb{E}\left(a_k|\mathcal{F}_k\right)\right\} - \mathbb{E}\left(b_k|\mathcal{F}_{k-1}\right),
$$

with

$$
a_k = \frac{\varepsilon_k^*(1+\alpha_k^T\mathbf{D}_k^{\,2}\alpha_k)}{\beta_k^*\beta_k}, \quad b_k = \frac{\alpha_k^T\mathbf{D}_k^{\,2}\alpha_k - \frac{1}{n}tr[(\mathbf{I}+z\mathbf{D}_k)\mathbf{D}_k]}{\beta_k^*}. \tag{4.5}
$$

Proof. Since $\mathbb{E}(tr\mathbf{D}_k|\mathcal{F}_{k-1}) = \mathbb{E}(tr\mathbf{D}_k|\mathcal{F}_k)$, we have

$$
z_k = \mathbb{E}[(tr\mathbf{D}-tr\mathbf{D}_k)|\mathcal{F}_{k-1}] - \mathbb{E}[(tr\mathbf{D}-tr\mathbf{D}_k)|\mathcal{F}_k].
$$

On the other hand,

$$tr\mathbf{D} - tr\mathbf{D}_k = -\frac{1 + \frac{1}{n}\alpha_k^T\mathbf{D}_k{}^2\alpha_k}{\beta_k}$$

$$= -\frac{1 + \frac{1}{n}tr[(\mathbf{I} + z\mathbf{D}_k)\mathbf{D}_k]}{\beta_k^*} + \frac{\varepsilon_k^*(1 + \alpha_k^T\mathbf{D}_k{}^2\alpha_k)}{\beta_k^*\beta_k} - \frac{\alpha_k^T\mathbf{D}_k{}^2\alpha_k - \frac{1}{n}tr[(\mathbf{I} + z\mathbf{D}_k)\mathbf{D}_k]}{\beta_k^*}$$

$$= -\frac{1 + \frac{1}{n}tr[(\mathbf{I} + z\mathbf{D}_k)\mathbf{D}_k]}{\beta_k^*} + a_k - b_k.$$

The conclusion follows from

$$\mathbb{E}\left(\frac{1 + \frac{1}{n}tr[(\mathbf{I} + z\mathbf{D}_k)\mathbf{D}_k]}{\beta_k^*}\,\middle|\,\mathcal{F}_{k-1}\right) = \mathbb{E}\left(\frac{1 + \frac{1}{n}tr[(\mathbf{I} + z\mathbf{D}_k)\mathbf{D}_k]}{\beta_k^*}\,\middle|\,\mathcal{F}_k\right)$$

and

$$\mathbb{E}\left(\alpha_k^T\mathbf{D}_k{}^2\alpha_k\,\middle|\,\mathcal{F}^{(k)}\right) = \frac{1}{n}tr[(\mathbf{I} + z\mathbf{D}_k)\mathbf{D}_k]. \qquad \square$$

PROPOSITION 4.1. *For each $\ell > 1/2$ with $\nu_{4\ell} < \infty$, there exist positive constants c_ℓ and L_0 independent of n and v such that for all n, v satisfying $L_0 n^{-1/2} \leq v < 1$,*

$$\mathbb{E}|m_p(z) - Em_p(z)|^{2\ell} \leq c_\ell n^{-2\ell} v^{-4\ell}(\Delta + v_y)^\ell.$$

Proof. In the proof of the proposition, c_ℓ and $c_{\ell,0}$ will be used to denote universal positive constants which may depend on the moments up to order ℓ of underlying variables and may represent different values at different appearances, even in one expression. Recall that we have

$$m_p(z) - \mathbb{E}m_p(z) = \frac{1}{p}\left[tr\mathbf{D} - \mathbb{E}tr\mathbf{D}\right] = \sum_{k=1}^p z_k,$$

where the $\{z_k\}$ are defined as in Lemma 4.4. We have

$$\mathbb{E}\left(|z_k|^{2\ell}\,\middle|\,\mathcal{F}_k\right) = \mathbb{E}\left\{\left|\left[\mathbb{E}\left(a_k|\mathcal{F}_{k-1}\right) - \mathbb{E}\left(a_k|\mathcal{F}_k\right)\right] - \mathbb{E}\left(b_k|\mathcal{F}_{k-1}\right)\right|^{2\ell}\,\middle|\,\mathcal{F}_k\right\}$$

$$\leq 2^{2\ell-1}\,\mathbb{E}\left\{\left[\mathbb{E}\left(a_k|\mathcal{F}_{k-1}\right) - \mathbb{E}\left(a_k|\mathcal{F}_k\right)\right]^{2\ell} + \left[\mathbb{E}\left(b_k|\mathcal{F}_{k-1}\right)\right]^{2\ell}\,\middle|\,\mathcal{F}_k\right\}$$

$$\leq 2^{2\ell-1}\,\mathbb{E}\left\{\left[\mathbb{E}\left(a_k|\mathcal{F}_{k-1}\right)\right]^{2\ell} + \left[\mathbb{E}\left(b_k|\mathcal{F}_{k-1}\right)\right]^{2\ell}\,\middle|\,\mathcal{F}_k\right\}$$

$$\leq 2^{2\ell-1}\left\{\mathbb{E}\left((a_k)^{2\ell}\,\middle|\,\mathcal{F}_k\right) + \mathbb{E}\left((b_k)^{2\ell}\,\middle|\,\mathcal{F}_k\right)\right\}.$$

Note that by (2.13) and (2.14), $|a_k| \leq v^{-1}\,|\varepsilon_k^*/\beta_k^*|$. Hence by Lemma 4.1,

$$\mathbb{E}\left(|a_k|^{2\ell}\,\middle|\,\mathcal{F}^{(k)}\right) \leq \frac{1}{v^{2\ell}}\,\mathbb{E}\left(\left|\frac{\varepsilon_k^*}{\beta_k^*}\right|^{2\ell}\,\middle|\,\mathcal{F}^{(k)}\right) \leq c_{\ell,0}(1 + \lambda_p)^{\ell/2}n^{-\ell}v^{-3\ell}|\beta_k^*|^{-\ell}.$$

On the other hand, by Lemma 2.1 and assuming $\ell \geq 1$,

$$\mathbb{E}\left(|b_k|^{2\ell}\,\middle|\,\mathcal{F}^{(k)}\right) \leq c_{\ell,0}(n\beta_k^*)^{-2\ell}(\nu_{4\ell} + M^\ell)\left[tr(\mathbf{I} + z\mathbf{D}_k)(\mathbf{I} + \bar{z}\overline{\mathbf{D}}_k)\mathbf{D}_k\overline{\mathbf{D}}_k\right]^\ell.$$

By (2.12) and (2.16),

$$|\beta_k^*|^{-1} tr(\mathbf{I} + z\mathbf{D}_k)(\mathbf{I} + \overline{z}\overline{\mathbf{D}}_k)\mathbf{D}_k\overline{\mathbf{D}}_k \leq |\beta_k^*|^{-1} tr\mathbf{\Lambda}_k^2 \leq nTv^{-3},$$

which implies

$$\mathbb{E}\left\{|b_k|^{2\ell}\,\Big|\,\mathcal{F}_k\right\} \leq c_{\ell,0} n^{-\ell} v^{-3\ell}\,\mathbb{E}\left[|\beta_k^*|^{-\ell}\,\big|\,\mathcal{F}_k\right].$$

Therefore, for all $\ell \geq 1$,

$$\mathbb{E}\left(|z_k|^{2\ell}\,\big|\,\mathcal{F}_k\right) \leq c_{\ell,0}(1+\lambda_p)^{\ell/2} n^{-\ell} v^{-3\ell}\,\mathbb{E}\left[|\beta_k^*|^{-\ell}\,\big|\,\mathcal{F}_k\right]$$

$$(4.6) \qquad\qquad\qquad \leq c_{\ell,0}(1+\lambda_p)^{\ell/2} n^{-\ell} v^{-4\ell+1}\,\mathbb{E}\left[|\beta_k^*|^{-1}\,\big|\,\mathcal{F}_k\right].$$

Applying Lemma 4.3 and (4.4), it follows that, for $\ell \geq 1$,

$$(4.7) \qquad\qquad \sum_{k=1}^{p} \mathbb{E}|z_k|^{2\ell} \leq c_{\ell,0} n^{-\ell+1}(\Delta + v_y)v^{-4\ell}.$$

Case $\ell = 1$. Since $\{z_k\}$ is a martingale difference sequence, the above inequality yields

$$(4.8) \qquad \mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^2 = n^{-2}\sum_{k=1}^{p} \mathbb{E}|z_k|^2 \leq c_{1,0} n^{-2}(\Delta + v_y)v^{-4}.$$

The proposition is proved in this case.

Case $\frac{1}{2} < \ell < 1$. By applying the Burkholder inequality for the martingale and using the concavity of the function x^ℓ, we find

$$\mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^{2\ell}$$

$$\leq c_\ell p^{-2\ell}\,\mathbb{E}\left(\sum_{k=1}^{p}|z_k|^2\right)^\ell \leq c_\ell n^{-2\ell}\left[\mathbb{E}\left(\sum_{k=1}^{p}|z_k|^2\right)\right]^\ell \leq c_\ell n^{-2\ell}\left[(\Delta + v_y)v^{-4}\right]^\ell,$$

where the last step follows from the previous case $\ell = 1$. The lemma is then proved in this case.

Case $\ell > 1$. We proceed by induction in this general case. First, by another Burkholder inequality for the martingale (Burkholder (1973), p. 39), we have

$$\mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^{2\ell} \leq c_\ell p^{-2\ell}\left\{\sum_{k=1}^{p}\mathbb{E}|z_k|^{2\ell} + \mathbb{E}\left(\sum_{k=1}^{p}\mathbb{E}(|z_k|^2|\mathcal{F}_k)\right)^\ell\right\}$$

$$(4.9) \qquad\qquad\qquad \hat{=} I_1 + I_2.$$

By (4.7)

$$(4.10) \qquad I_1 \leq c_{\ell,0}(\Delta + v_y)n^{-3\ell+1}v^{-4\ell} \leq c_{\ell,0}(\Delta + v_y)^\ell n^{-2\ell}v^{-4\ell}.$$

The proposition already has been proved for the case $\frac{1}{2} < \ell \leq 1$. Suppose that the lemma is true for $\ell \leq 2^t$. Now, we consider the case where $2^t < \ell \leq 2^{t+1}$. Application of (4.6) with $\ell = 1$ gives

$$\sum_{k=1}^{n} \mathbb{E}\left(|z_k|^2\,\big|\,\mathcal{F}_k\right) \leq c_{1,0} n^{-1} v^{-3}(1+\lambda_p)^{1/2}\sum_{k=1}^{p} \mathbb{E}\left(|\beta_k^*|^{-1}\,\big|\,\mathcal{F}_k\right).$$

Hence, by taking into account (4.4) we get

$$I_2 \le c_{\ell,0}(nv)^{-3\ell}\, \mathbb{E}(1 + \lambda_p)^{\ell/2} \left(\sum_{k=1}^p \mathbb{E}(|\beta_k^*|^{-1}|\mathcal{F}_k) \right)^\ell$$

(4.11) $$\le c_{\ell,0} n^{-2\ell-1} v^{-3\ell} \sum_{k=1}^p \mathbb{E}|\beta_k^*|^{-\ell} + o(n^{-4\ell-1}v^{-4\ell}),$$

since $P(\lambda_p > 5) = o(n^{-4\ell})$. Notice that if $L_0 > \sqrt{2}$, then $nv^2 > 2$ and that

$$\left| |\beta|^{-1} - |\beta_k^*|^{-1} \right| \le |\beta^{-1} - (\beta_k^*)^{-1}| = \frac{|tr\mathbf{D} - tr\mathbf{D}_k|}{p|\beta||\beta_k^*|} \le \frac{1}{pv^2} \min(|\beta|^{-1}, |\beta_k^*|^{-1})$$

(this comes from (2.9) and $|\beta\beta_k^*|^{-1} \le v^{-1} \min(|\beta|^{-1}, |\beta_k^*|^{-1})$). This yields

$$|\beta_k^*|^{-1} \le |\beta|^{-1} + p^{-1}v^{-2}|\beta_k^*|^{-1} \le 2|\beta|^{-1}$$

and

$$|p\beta^{-1}| \le \left| \sum_{k=1}^p (\beta_k^*)^{-1} \right| + \sum_{k=1}^p |(\beta_k^*)^{-1} - \beta^{-1}| \le \left| \sum_{k=1}^p (\beta_k^*)^{-1} \right| + v^{-2}|\beta|^{-1}$$

$$\le 2 \left| \sum_{k=1}^p (\beta_k^*)^{-1} \right| \le 2 \left| \sum_{k=1}^p ((\beta_k^*)^{-1} - \beta_k^{-1}) \right| + 2 \left| \sum_{k=1}^p \beta_k^{-1} \right|$$

$$\le 2 \sum_{k=1}^p \frac{|\varepsilon_k^*|^2}{|\beta_k||\beta_k^*|^2} + 2p|m_p(z)|.$$

Therefore, by applying Lemma 4.1, and if we choose, $L_0 > (2c_{\ell,0})^{1/\ell}$ so that $c_{\ell,0}n^{-\ell}v^{-2\ell} < 1/2$, we have

$$\sum_{k=1}^p \mathbb{E}|\beta_k^*|^{-\ell} \le c_{\ell,0} \left(v^{-\ell} \sum_{k=1}^p \mathbb{E}\frac{|\varepsilon_k^*|^{2\ell}}{|\beta_k^*|^{2\ell}} + p\,\mathbb{E}|m_p(z)|^\ell \right)$$

$$\le c_{\ell,0} \left(n^{-\ell}v^{-2\ell} \sum_{k=1}^p \mathbb{E}|\beta_k^*|^{-\ell} + p\,\mathbb{E}|m_p(z)|^\ell \right)$$

$$\le 2c_{\ell,0}p\,\mathbb{E}|m_p(z)|^\ell.$$

From the above inequality and (4.11), we get by induction,

$$I_2 \le c_\ell n^{-2\ell}v^{-3\ell}\,\mathbb{E}|m_p(z)|^\ell + o(n^{-4\ell-1}v^{-4\ell})$$

$$\le c_\ell n^{-2\ell}v^{-3\ell} \left[\mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^\ell + |\mathbb{E}m_p(z) - m(z)|^\ell + |m(z)|^\ell \right]$$

$$\quad + o(n^{-4\ell-1}v^{-4\ell})$$

$$\le c_\ell n^{-2\ell}v^{-3\ell} \left[\mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^\ell + (\Delta + v_y)^\ell v^{-\ell} \right] + o(n^{-4\ell-1}v^{-4\ell})$$

$$\le c_\ell n^{-2\ell}v^{-4\ell}(\Delta + v_y)^\ell \left[\left(n^2v^2(\Delta + v_y) \right)^{-\ell/2} + 1 \right] + o(n^{-4\ell-1}v^{-4\ell})$$

(4.12) $$\le c_\ell n^{-2\ell}v^{-4\ell}(\Delta + v_y)^\ell.$$

Therefore by (4.9) and (4.12), it follows that

$$(4.13) \qquad \mathbb{E}|m_p(z) - \mathbb{E}m_p(z)|^{2\ell} \leq c_\ell n^{-2\ell}(\Delta + v_y)^\ell v^{-4\ell}.$$

The proof of Proposition 4.1 is complete. $\qquad \square$

Remark 4.1. Application of Proposition 4.1 to the case $\ell = 1$ gives that there is some constant $c_1 > 0$ such that

$$(4.14) \qquad \mathbb{E}|tr\mathbf{D}_k - \mathbb{E}tr\mathbf{D}_k|^2 \leq c_1(\Delta + v_y)v^{-4}.$$

It is also worth noticing that if we substitute $\mathbf{D}$ for any $\mathbf{D}_k$ with $k \leq n$, Proposition 4.1 as well as the above consequence (4.14) are still valid, with slightly different constants c_ℓ.

REFERENCES

Z. D. BAI (1993a), *Convergence rate of expected spectral distributions of large random matrices, Part* I: *Wigner matrices*, Ann. Probab., 21, pp. 625–648.

Z. D. BAI (1993b), *Convergence rate of expected spectral distributions of large random matrices, Part* II: *Sample covariance matrices*, Ann. Probab., 21, pp. 649–672.

Z. D. BAI (1997), *Circular law*, Ann. Probab., 25, pp. 494–529.

Z. D. BAI (1999), *Methodologies in spectral analysis of large dimensional random matrices. A review*, Statist. Sinica, 9, pp. 611–677.

Z. D. BAI, B. MIAO, AND J. TSAY (1997), *A note on the convergence rate of the spectral distribution of large random matrices*, Statist. Probab. Lett., 34, pp. 95–101.

Z. D. BAI, B. Q. MIAO, AND J. TSAY (1999), *Remarks on the convergence rate of the spectral distributions of Wigner matrices*, J. Theoret. Probab., 12, pp. 301–311.

Z. D. BAI, B. MIAO, AND J. TSAY (2002), *Convergence rates of the spectral distributions of large Wigner matrices*, Int. Math. J., 1, pp. 65–90.

Z. D. BAI AND J. W. SILVERSTEIN (1998), *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab., 26, pp. 316–345.

Z. D. BAI AND J. W. SILVERSTEIN (1999), *Exact separation of eigenvalues of large dimensional sample covariance matrices*, Ann. Probab., 27, pp. 1536–1555.

Z. D. BAI AND Y. Q. YIN (1986), *Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang*, Probab. Theory Related Fields, 73, pp. 555–569.

Z. D. BAI, Y. Q. YIN, AND P. K. KRISHNAIAH (1987), *On the limiting empirical distribution function of the eigenvalues of a multivariate F-matrix*, Probab. Theory Appl., 32, pp. 490–500.

D. L. BURKHOLDER (1973), *Distribution function inequalities for martingales*, Ann. Probab., 1, pp. 19–42.

V. A. MARČENKO AND L. A. PASTUR (1967), *Distribution of eigenvalues for some sets of random matrices*, Mat. Sb., 72, pp. 507–536.

M. L. MEHTA (1991), *Random Matrices*, Academic Press, New York.

J. W. SILVERSTEIN (1985), *The limiting eigenvalue distribution of a multivariate F matrix*, SIAM J. Math. Anal., 16, pp. 641–646.

E. P. WIGNER (1955), *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2), 62, pp. 548–564.

E. P. WIGNER (1958), *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2), 67, pp. 325–327.

Y. Q. YIN, Z. D. BAI, AND P. R. KRISHNAIAH (1988), *On the limit of the largest eigenvalue of the large dimensional sample covariance matrix*, Probab. Theory Related Fields, 78, pp. 509–531.