Strong Convergence of the Empirical Distribution of Eigenvalues of Large Dimensional Random Matrices

by

Jack W. Silverstein*
Department of Mathematics
Box 8205
North Carolina State University
Raleigh, North Carolina 27695-8205

Summary

Let X be $n \times N$ containing i.i.d. complex entries with $\mathsf{E}|X_{11} - \mathsf{E}X_{11}|^2 = 1$, and T an $n \times n$ random Hermitian non-negative definite, independent of X. Assume, almost surely, as $n \to \infty$, the empirical distribution function (e.d.f.) of the eigenvalues of T converges in distribution, and the ratio n/N tends to a positive number. Then it is shown that, almost surely, the e.d.f. of the eigenvalues of $(1/N)XX^*T$ converges in distribution. The limit is nonrandom and is characterized in terms of its Stieltjes transform, which satisfies a certain equation.

* Supported by the National Science Foundation under grant DMS-9404047

AMS 1991 subject classifications. Primary 60F15; Secondary 62H99.

Key Words and Phrases. Random matrix, empirical distribution function of eigenvalues, Stieltjes transform.

1. Introduction. For any square matrix A with only real eigenvalues, let F^A denote the empirical distribution function (e.d.f.) of the eigenvalues of A (that is, $F^A(x)$ is the proportion of eigenvalues of $A \leq x$). This paper continues the work on the e.d.f. of the eigenvalues of matrices of the form $(1/N)XX^*T$, where X is $n \times N$ containing i.i.d. complex entries with $\mathsf{E}|X_{11} - \mathsf{E}X_{11}|^2 = 1$, T is $n \times n$ random Hermitian non-negative definite, independent of X, and n, N are large but on the same order of magnitude. Assuming the entries of X and T to be real, it is shown in Yin [4] that, if n and N both converge to infinity while their ratio n/N converges to a positive quantity c, and the moments of F^T converge almost surely to those of a nonrandom probability distribution function (p.d.f.) H satisfying the Carleman sufficiency condition, then, with probability one, $F^{(1/N)XX^*T}$ converges in distribution to a nonrandom p.d.f. F. The aim of this paper is to extend the limit theorem to the complex case, with arbitrary H having mass on $[0, \infty)$, assuming convergence in distribution of F^T to H almost surely.

Obviously the method of moments, used in the proof of the limit theorem in Yin [4], cannot be further relied on. As will be seen, the key tool in understanding both the limiting behavior of $F^{(1/N)XX^*T}$ and analytic properties of F is the Stieltjes transform, defined for any p.d.f. G as the analytic function

$$m_G(z) \equiv \int \frac{1}{\lambda - z} dG(\lambda) \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : Im\, z > 0\}.$$

Due to the inversion formula

$$G\{[a, b]\} = \frac{1}{\pi} \lim_{\eta \to 0^+} \int_a^b Im\, m_G(\xi + i\eta) d\xi$$

(a, b continuity points of G), convergence in distribution of a tight sequence of p.d.f.'s is guaranteed once convergence of the corresponding Stieltjes transforms on a countable subset in $\mathbb{C}^+$ possessing at least one accumulation point is verified.

One reason for using the Stieltjes tranform to characterize spectral e.d.f's is the simple way it can be expressed in terms of the resolvent of the matrix. Indeed, for $p \times p$ A having real eigenvalues $\lambda_1, \ldots, \lambda_p$,

$$m_{F^A}(z) = (1/p) \sum_{i=1}^{p} \frac{1}{\lambda_i - z} = (1/p) \mathsf{tr}\, (A - zI)^{-1}$$

(tr denoting trace, and I the identity matrix).

Stieltjes transform methods are used in Marčenko and Pastur [1] on matrices of the form $A + (1/N)X^*TX$, where the entries of X have finite absolute fourth moments (independence is replaced by a mild dependency condition reflected in their mixed second and fourth moments), $T = \text{diag}(\tau_1, \ldots, \tau_n)$ with τ_i's i.i.d. having p.d.f. H where H is an arbitrary p.d.f. on $\mathbb{R}$, and A is $N \times N$ nonrandom Hermitian with F^A converging vaguely to a (possibly degenerate) distribution function. For ease of exposition, we confine our discussion only to the case $A = 0$. It is proven that $F^{(1/N)X^*TX}(x) \xrightarrow{i.p.} \underline{F}(x)$ for all $x \neq 0$ (it can be shown that $\underline{F}$ is absolutely continuous away from 0. See Silverstein and Choi [3] for results on the analytic behavior of $\underline{F}$ and F), where $m_{\underline{F}}$ is the solution to

$$m = -\left(z - c\int \frac{\tau dH(\tau)}{1+\tau m}\right)^{-1} \tag{1.1}$$

in the sense that, for every $z \in \mathbb{C}^+$, $m = m_{\underline{F}}(z)$ is the unique solution in $\mathbb{C}^+$ to (1.1).

Under the assumptions on X originally given in this paper, strong convergence for T diagonal is proven in Silverstein and Bai [2] with no restriction on $\{\tau_1, \ldots, \tau_n\}$ other than its e.d.f. converges in distribution to H almost surely. The proof takes an approach more direct than in Marčenko and Pastur [1] (which involves the construction of a certain partial differential equation), providing a clear understanding of why $m_{\underline{F}}$ satisfies (1.1), at the same time displaying where random behavior primarily comes into play (basically from Lemma 2.1 given below).

The difference between the spectra of $(1/N)XX^*T$ and $(1/N)X^*TX$ is $|n - N|$ zero eigenvalues, expressed via their e.d.f.'s by the relation

$$F^{(1/N)X^*TX} = (1 - \tfrac{n}{N})I_{[0,\infty)} + \tfrac{n}{N}F^{(1/N)XX^*T}$$

(I_B denoting the indicator function on the set B). It follows that their Stieltjes transforms satisfy

$$m_{F^{(1/N)X^*TX}}(z) = -\frac{(1-\frac{n}{N})}{z} + \tfrac{n}{N}m_{F^{(1/N)XX^*T}}(z) \quad z \in \mathbb{C}^+. \tag{1.2}$$

Therefore, in the limit (when F and $\underline{F}$ are known to exist)

$$\underline{F} = (1-c)1_{[0,\infty)} + cF,$$

and

$$m_{\underline{F}}(z) = -\frac{(1-c)}{z} + cm_F(z) \quad z \in \mathbb{C}^+. \tag{1.3}$$

From (1.1) and (1.3) it is straightfoward to conclude that for each $z \in \mathbb{C}^+$, $m = m_F(z)$ is a solution to

$$m = \int \frac{1}{\tau(1-c-czm) - z} dH(\tau). \tag{1.4}$$

It is unique in the set $\{m \in \mathbb{C} : -\frac{(1-c)}{z} + cm \in \mathbb{C}^+\}$.

It is remarked here that (1.1) reveals much of the analytic behavior of $\underline{F}$, and consequently F (Silverstein and Choi [3]), and should be viewed as another indication of the importance of Stieltjes transforms to these types of matrices. Even when H has all moments, it seems unlikely much information about F can be extracted from the explicit expressions for the moments of F given in Yin [4].

Using again the Stieltjes transform as the essential tool in analyzing convergence, this paper will establish strong convergence of $F^{(1/N)XX^*T}$ to F under the weakest assumptions on non-negative definite T. In order to keep notation to a minimum, the statement of the result and its proof will be expressed in terms of the Hermitian matrix $(1/N)T^{1/2}XX^*T^{1/2}$, $T^{1/2}$ denoting a Hermitian square root of T

The following theorem will be proven.

Theorem 1.1. Assume on a common probability space

a) For $n = 1, 2, \ldots$ $X_n = (X^n_{ij})$, $n \times N$, $X^n_{ij} \in \mathbb{C}$, i.d. for all n, i, j, independent across i, j for each n, $\mathsf{E}|X^1_{1\,1} - \mathsf{E}X^1_{1\,1}|^2 = 1$.

b) $N = N(n)$ with $n/N \to c > 0$ as $n \to \infty$.

c) T_n $n \times n$ random Hermitian non-negative definite, with F^{T_n} converging almost surely in distribution to a p.d.f. H on $[0, \infty)$ as $n \to \infty$.

d) X_n and T_n are independent.

Let $T_n^{1/2}$ be the Hermitian non-negative square root of T_n, and let $B_n = (1/N)T_n^{1/2}X_nX_n^*T_n^{1/2}$ (obviously $F^{B_n} = F^{(1/N)X_nX_n^*T_n}$). Then, almost surely, F^{B_n} converges in distribution, as $n \to \infty$, to a (nonrandom) p.d.f. F, whose Stieltjes transform $m(z)$ ($z \in \mathbb{C}^+$) satisfies (1.4), in the sense that, for each $z \in \mathbb{C}^+$, $m = m(z)$ is the unique solution to (1.4) in $D_c \equiv \{m \in \mathbb{C} : -\frac{(1-c)}{z} + cm \in \mathbb{C}^+\}$.

The proof will be given in the next section. Much of the groundwork has already been laid out in Silverstein and Bai [2], in particular the first step, which is to truncate and centralize the entries of X, and Lemma 2.1. Therefore we will, on occasion, refer the reader to the latter paper for further details.

Proof of Theorem 1.1. As in Silverstein and Bai [2], the dependency of the variables on n will occasionally be dropped.

Through two successive stages of truncations and centralizations (truncate at $\pm\sqrt{N}$, centralize, truncate at $\pm \ln N$, centralize) and a final scaling, the main part of section 3 in Silverstein and Bai [2] argues that the assumptions on the entries of X can be replaced by standardized variables bounded in absolute value by a fixed multiple of $\ln N$. Write T in its spectral decompositon: $T = U(\mathrm{diag}(\tau_1, \dots, \tau_n))U^*$. By replacing the diagonal matrix T_α in that paper with $U(\mathrm{diag}(\tau_1 I_{(\tau_1 \le \alpha)}, \dots, \tau_n I_{(\tau_n \le \alpha)}))U^*$, exactly the same argument applies in the present case. However, the following proof requires truncation of the eigenvalues of T. It is shown in Silverstein and Bai [2] (and used in section 3) that for any $N \times n$ matrix Q, if $\alpha = \alpha_n \to \infty$ then

$$\|F^{QTQ^*} - F^{QT_\alpha Q^*}\| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty,$$

where $\|\cdot\|$ here denotes the sup norm on functions.

Therefore, we may assume along with the conditions in Theorem 1.1

1) $|X_{1\,1}| \le \log n$, where $\log n$ denotes the logarithm of n with a certain base (defined in Silverstein and Bai [2]),
2) $\mathsf{E} X_{1\,1} = 0$, $\mathsf{E}|X_{1\,1}|^2 = 1$
3) $\|T\| \le \log n$,

where here and throughout the following $\|\cdot\|$ denotes the spectral norm on matrices.

The following two results are derived in Silverstein and Bai [2]. The first accounts for much of the truth of Theorem 1.1 due to random behavior. The second relies on the following fact (which contributes much to the form of equation (1.4)): For $n \times n$ B and $q \in \mathbb{C}^n$ for which B and $B + qq^*$ are invertible,

$$q^*(B + qq^*)^{-1} = \frac{1}{1 + q^* B^{-1} q} q^* B^{-1}, \tag{2.1}$$

(follows from $q^* B^{-1}(B + qq^*) = (1 + q^* B^{-1} q) q^*$).

Lemma 2.1 (Lemma 3.1 of Silverstein and Bai [2]). Let C be an $n \times n$ matrix with $\|C\| \le 1$, and $Y = (X_1, \dots, X_n)^T$, where the X_i's are i.i.d. satisfying conditions 1) and 2) above. Then

$$\mathsf{E}|Y^* C Y - \mathsf{tr}\, C|^6 \le K n^3 \log^{12} n$$

where the constant K does not depend on n, C, nor on the distribution of X_1.

Lemma 2.2 (Lemma 2.6 of Silverstein and Bai [2]). Let $z \in \mathbb{C}^+$ with $v = Im\, z$, A and B $n \times n$ with B Hermitian, and $q \in \mathbb{C}^n$. Then

$$\left|\mathsf{tr}\left((B - zI)^{-1} - (B + \tau qq^* - zI)^{-1}\right) A\right| \le \frac{\|A\|}{v}.$$

The next lemma contains some additional inequalities.

Lemma 2.3. For $z = u + iv \in \mathbb{C}^+$ let $m_1(z)$, $m_2(z)$ be Stieltjes transforms of any two p.d.f.'s, A and B $n \times n$ with A Hermitian non-negative definite, and $r \in \mathbb{C}^n$. Then

a) $$\|(m_1(z)A + I)^{-1}\| \le \max(4\|A\|/v, 2)$$

b) $$|\mathrm{tr}\, B((m_1(z)A+I)^{-1} - (m_2(z)A+I)^{-1})| \le |m_2(z) - m_1(z)| n \|B\|\, \|A\| (\max(4\|A\|/v, 2))^2$$

c)
$$|r^* B(m_1(z)A+I)^{-1} r - r^* B(m_2(z)A+I)^{-1} r| \le |m_2(z) - m_1(z)| \|r\|^2 \|B\| \|A\| (\max(4\|A\|/v, 2))^2$$

($\|r\|$ denoting Euclidean norm on r).

Proof: Notice b) and c) follow easily from a) using basic matrix properties. We have for any positive x $|m_1(z)x + 1|^2 = (Re\, m_1(z)x + 1)^2 + (Im\, m_1(z))^2 x^2$. Using the Cauchy-Schwarz inequality it is easy to show $|Re\, m_1(z)| \le (Im\, m_1(z)/v)^{1/2}$. This leads us to consider minimizing over m the expression $(vx)^2 m^4 + (mx + 1)^2$, or with $y = mx$ the function $f(y) = ay^4 + (y + 1)^2$ where $a = (v/x)^2$. Upon considering y on either side of $-1/2$ we find $f(y) \ge \min(a/16, 1/4)$, from which a) follows.

We proceed with the proof of Theorem 1.1

Fix $z = u + iv \in \mathbb{C}^+$. Let $\underline{B}_n = (1/N)X^*TX$, $m_n = m_{F^{B_n}}$, and $\underline{m}_n = m_{F^{\underline{B}_n}}$. Let $c_n = n/N$. In Silverstein and Bai [2] it is argued that, almost surely, the sequence $\{F^{\underline{B}_n}\}$, for diagonal T, is tight. The argument carries directly over to the present case. Thus the quantity

$$\delta = \inf_n Im\, \underline{m}_n(z) \ge \inf_n \int \frac{v\, dF^{\underline{B}_n}(\lambda)}{2(\lambda^2 + u^2) + v^2}$$

is positive almost surely.

For $j = 1, 2, \ldots, N$, let $q_j = (1/\sqrt{n})X_{\cdot\, j}$ ($X_{\cdot\, j}$ denoting the j^{th} column of X), $r_j = (1/\sqrt{N})T^{1/2}X_{\cdot\, j}$, and $B_{(j)} = B^n_{(j)} = B_n - r_j r_j^*$. Write

$$B_n - zI + zI = \sum_{j=1}^N r_j r_j^*.$$

Taking the inverse of $B_n - zI$ on the right on both sides and using (2.1) we find

$$I + z(B_n - zI)^{-1} = \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j} r_j r_j^* (B_{(j)} - zI)^{-1}.$$

Taking the trace on both sides and dividing by N we have

$$c_n + zc_n m_n = \frac{1}{N}\sum_{j=1}^{N} \frac{r_j^*(B_{(j)} - zI)^{-1} r_j}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j} = 1 - \frac{1}{N}\sum_{j=1}^{N} \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j}.$$

From (1.2) we see that

$$\underline{m}_n(z) = -\frac{1}{N}\sum_{j=1}^{N} \frac{1}{z(1 + r_j^*(B_{(j)} - zI)^{-1} r_j)}. \tag{2.2}$$

For each j we have

$$Im\, r_j^*((1/z)B_{(j)} - I)^{-1} r_j = \frac{1}{2i} r_j^*\big(((1/z)B_{(j)} - I)^{-1} - ((1/\overline{z})B_{(j)} - I)^{-1}\big) r_j$$

$$= \frac{v}{|z|^2} r_j^*((1/z)B_{(j)} - I)^{-1} B_{(j)}((1/\overline{z})B_{(j)} - I)^{-1} r_j \geq 0.$$

Therefore

$$\frac{1}{|z(1 + r_j^*(B_{(j)} - zI)^{-1} r_j)|} \leq \frac{1}{v}. \tag{2.3}$$

Write $B_n - zI - \big(-z\underline{m}_n(z)T_n - zI\big) = \sum_{j=1}^{N} r_j r_j^* - (-z\underline{m}_n(z))T_n$. Taking inverses and using (2.1), (2.2) we have

$$(-z\underline{m}_n(z)T_n - zI)^{-1} - (B_n - zI)^{-1} = (-z\underline{m}_n(z)T_n - zI)^{-1}\left[\sum_{j=1}^{N} r_j r_j^* - (-z\underline{m}_n(z))T_n\right](B_n - zI)^{-1}$$

$$= \sum_{j=1}^{N} \frac{-1}{z(1 + r_j^*(B_{(j)} - zI)^{-1} r_j)}\left[(\underline{m}_n(z)T_n + I)^{-1} r_j r_j^*(B_{(j)} - zI)^{-1} - \tfrac{1}{N}(\underline{m}_n(z)T_n + I)^{-1} T_n (B_n - zI)^{-1}\right].$$

Taking the trace and dividing by n we find

$$\tfrac{1}{n}\mathsf{tr}\,(-z\underline{m}_n(z)T_n - zI)^{-1} - m_n(z) = \frac{1}{N}\sum_{j=1}^{N} \frac{-1}{z(1 + r_j^*(B_{(j)} - zI)^{-1} r_j)} d_j \tag{2.4}$$

where

$$d_j = q_j^* T^{1/2}(B_{(j)} - zI)^{-1}(\underline{m}_n(z)T_n + I)^{-1} T^{1/2} q_j - \tfrac{1}{n}\mathsf{tr}\,(\underline{m}_n(z)T_n + I)^{-1} T_n (B_n - zI)^{-1}.$$

From Lemma 2.2 we see $\max_{j\leq N}|m_n(z) - m_{F^{B_{(j)}}}(z)| \leq \dfrac{1}{nv}$.

Let for each j $\underline{m}_{(j)}(z) = -\frac{(1-c_n)}{z} + c_n m_{F^{B_{(j)}}}(z)$. From (1.2) we have for any positive p

$$\max_{j \le N} \log^p n |\underline{m}_n(z) - \underline{m}_{(j)}(z)| \to 0 \quad \text{as } n \to \infty. \tag{2.5}$$

Also, by writing

$$\underline{m}_{(j)}(z) = -\frac{1}{Nz} + \left(\frac{N-1}{N}\right)\left(-\frac{(1-\frac{n}{N-1})}{z} + \frac{n}{N-1} m_{F^{B_{(j)}}}(z)\right),$$

from (1.2) we see that $\underline{m}_{(j)}$ is the Stieltjes transform of a p.d.f.

Using condition 3), Lemma 2.1, Lemma 2.3 a), the fact that q_j is independent of both $B_{(j)}$ and $\underline{m}_{(j)}(z)$, and $\|(A-zI)^{-1}\| \le 1/v$ for any Hermitian matrix A, we find

$$\mathsf{E}|\,\|q_j\|^2 - 1|^6 \le K\frac{\log^{12} n}{n^3}$$

and for n sufficiently large

$$\mathsf{E}|q_j^* T^{1/2}(B_{(j)} - zI)^{-1}(\underline{m}_{(j)}(z)T_n + I)^{-1}T^{1/2}q_j \\ - \tfrac{1}{n}\mathsf{tr}\, T^{1/2}(B_{(j)} - zI)^{-1}(\underline{m}_{(j)}(z)T_n + I)^{-1}T^{1/2}|^6 \le \frac{K4^6}{v^{12}}\frac{\log^{24} n}{n^3}.$$

Therefore we have almost surely as $n \to \infty$

$$\max_{j \le N} \max[|\,\|q_j\|^2 - 1|, |q_j^* T^{1/2}(B_{(j)} - zI)^{-1}(\underline{m}_{(j)}(z)T_n + I)^{-1}T^{1/2}q_j \\ - \tfrac{1}{n}\mathsf{tr}\, T^{1/2}(B_{(j)} - zI)^{-1}(\underline{m}_{(j)}(z)T_n + I)^{-1}T^{1/2}|] \to 0. \tag{2.6}$$

We concentrate on a realization for which (2.6) holds, $\{F^{B_n}\}$ is tight (implying $\delta > 0$), and F^{T_n} converges in distribution to H. From condition 3), Lemma 2.2, Lemma 2.3 b), c), (2.5), and (2.6) we find that $\max_{j \le N}|d_j| \to 0$ as $n \to \infty$. Therefore, from (2.3), (2.4)

$$\tfrac{1}{n}\mathsf{tr}\,(-z\underline{m}_n(z)T_n - zI)^{-1} - m_n(z) \to 0 \quad \text{as } n \to \infty.$$

Consider a subsequence $\{n_i\}$ on which $\{m_{n_i}(z)\}$ (bounded in absolute value by $1/v$) converges to a number m. Let $\underline{m} = -\frac{(1-c)}{z} + cm$ be the corresponding limit of $\underline{m}_{n_i}(z)$. We have $Im\,\underline{m} \ge \delta$ so that $m \in D_c$. We use the fact that for $m' \in \mathbb{C}^+$, $\tau \in \mathbb{R}$, $|1/(m'\tau + 1)| \le |m'|/Im\,m'$ and $|\tau/(m'\tau+1)| \le 1/Im\,m'$ to conclude that the function

$$f(\tau) = \frac{1}{\underline{m}\tau + 1}$$

is bounded and satisfies

$$\left|\frac{1}{\underline{m}_{n_i}(z)\tau+1}-f(\tau)\right|\leq\frac{|\underline{m}|}{\delta^2}|\underline{m}_{n_i}(z)-\underline{m}|.$$

Therefore

$$\frac{1}{n}\mathsf{tr}\,(-z\underline{m}_{n_i}(z)T_n-zI)^{-1}=-\frac{1}{z}\int\frac{1}{\underline{m}_{n_i}\tau+1}dF^{T_{n_i}}(\tau)\to-\frac{1}{z}\int\frac{1}{\underline{m}\tau+1}dH(\tau)\qquad\text{as } n\to\infty.$$

Thus m satisfies (1.4). Since m is unique we have $m_n(z)\to m$. Thus, with probability one, F^{B_n} converges in distribution to F having Stieltjes transform defined through (1.4). This completes the proof of Theorem 1.1.

REFERENCES

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

[2] J. W. Silverstein and Z. D. Bai, *On the empirical distribution of eigenvalues of a class of large dimensional random matrices,* J. Multivariate Anal. 54 (1995), pp. 175-192.

[3] J. W. Silverstein and S.I. Choi, *Analysis of the limiting spectral distribution of large dimensional random matrices,* J. Multivariate Anal. 54 (1995), pp. 295-309.

[4] Y. Q. Yin, *Limiting spectral distribution for a class of random matrices,* J. Multivariate Anal. 20 (1986), pp. 50-68.