Convergence of a class of Toeplitz type matrices

Arup Bose * Sreela Gangopadhyay Koushik Saha

February 13, 2013

Abstract

We use the method of moments to study the spectral properties in the bulk for finite diagonal large dimensional random and non-random Toeplitz type matrices via the joint convergence of matrices in an appropriate sense. As a consequence we revisit the famous limit theorem of Szegö for non-random symmetric Toeplitz matrices.

Keywords: Eigenvalues, finite diagonal matrix, joint convergence, large dimensional random matrix, limiting spectral distribution, semi-circular law, Toeplitz matrix, weak convergence, method of moments.

AMS 2000 Subject Classification: 60B20, 60G70, 60G55.

1 Introduction

Suppose A_n is an $n \times n$ real symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. The *Empirical Spectral Distribution Function* (ESD) of A_n is defined as

$$F_{A_n}(x) = n^{-1} \sum_{i=1}^{n} \mathbb{I}\{\lambda_i \leq x\}.$$

Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of square matrices with the corresponding ESD $\{F_{A_n}\}_{n=1}^{\infty}$. The *Limiting Spectral Distribution* (LSD) of $\{A_n\}$ is defined as the weak limit F of the sequence $\{F_{A_n}\}$, if it exists. We also identify F with any random variable X with distribution F.

When the elements of $\{A_n\}$ are defined on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$, that is, $\{A_n\}$ are random, then $\{F_{A_n}(\cdot)\}$ are random and are functions of $\omega \in \Omega$ but we suppress this dependence. Let F be a non-random distribution function. We say the LSD of $\{A_n\}$ exists (and equals F) *almost surely* if for almost every $\omega \in \Omega$ and at all continuity points x of F

$$F_{A_n}(x) \to F(x) \text{ as } n \to \infty.$$

The *expected spectral distribution function* of A_n is defined as $\mathrm{E}(F_n(\cdot))$. This expectation is always a distribution function.

*Research supported by J.C.Bose Fellowship, Dept. of Science and Technology, Govt. of India.

We shall be dealing with more than one sequence of non-cummutative matrices and shall need an appropriate notion of joint convergence. We present this in brief. For a comprehensive description of this notion, see Anderson et al. (2010).

A *non-commutative probability space* is a pair $(\mathcal{A}, \phi)$ where $\mathcal{A}$ is a unital algebra[1] (with unity $\mathbf{1}$) and $\phi : \mathcal{A} \to \mathbb{C}$ is a linear functional satisfying $\phi(\mathbf{1}) = 1$. Elements of a non-commutative probability space will also be called (non-commutative) *random variables*.

Let $\mathcal{A}_n$ be the space of $n \times n$ symmetric random matrices with elements which are random variables with all moments finite. Then ϕ_n equal to $\frac{1}{n}\,\mathrm{E}_\mu[\mathrm{Tr}(\cdot)]$ or $\frac{1}{n}[\mathrm{Tr}(\cdot)]$ both yield non-commutative probability space.

For random variables $\{b_i\}_{i\in J} \subset \mathcal{A}$, their *joint moments* is the collection $\{\phi(b_{i_1} b_{i_2} \dots b_{i_k}),\ k \geq 1\}$, where each $b_{i_j} \in \{b_i\}_{i\in J}$.

Random variables $\{b_{i,n}\}_{i\in J} \in (\mathcal{A}_n, \phi_n)$ are said to *converge in law* to $\{b_i\}_{i\in J} \in (\mathcal{A}, \phi)$ (as $n \to \infty$) if each joint moment of $\{b_{i,n}\}_{i\in J}$ converges to the corresponding joint moment of $\{b_i\}_{i\in J}$. That is if

$$\phi_n[P(b_{1,n}, b_{2,n}, \dots, b_{J,n})] \to \phi[P(b_1, b_2, \dots, b_J)]$$

for all polynomials P. If this happens, we write

$$(b_{1,n}, b_{2,n}, \dots, b_{J,n}) \xrightarrow{\phi_n} (b_1, b_2, \dots, b_J).$$

If the random variables $b_{1,n}, \dots, b_{j,n}$ are $n \times n$ (non-random) matrices, then the above convergence is assumed to be with respect to $\phi_n = \frac{1}{n}\,\mathrm{Tr}$. If instead they are random matrices, then the above convergence is in one of the following two senses:

(i) We say that $(b_{1,n}, b_{2,n}, \dots, b_{j,n})$ converges to $(b_1, b_2, \dots, b_j)$ if the convergence holds with respect to $\phi_n = \frac{1}{n}\,\mathrm{E}\,\mathrm{Tr}$.

(ii) We say $(b_{1,n}, b_{2,n}, \dots, b_{j,n})$ converges almost surely to $(b_1, b_2, \dots, b_j)$ if the convergence holds with respect to $\phi_n = \frac{1}{n}\,\mathrm{Tr}$.

Consider a sequence of $n \times n$ symmetric matrices $A_n = ((a_{n,i,j}))_{1\leq i,j\leq n}$. For simplicity, we write $a_{i,j}$ for $a_{n,i,j}$. For any $k > 0$, $k < n$, the k-th upper and lower diagonals of A_n are defined as the finite sequence $\{a_{i,i+k},\ i = 1, \dots, n-k\}$ and $\{a_{i+k,i},\ i = 1, \dots, n-k\}$. A sequence of matrices A_n with increasing dimension is said to be *finite diagonal* if there is a k such that $a_{i,j} = 0$ whenever $|i-j| \geq k$.

One particular case of the famous Szegö's theorem says that if $\{a_i\}$ is square summable and $T_n = ((a_{|i-j|}))_{1\leq i,j\leq n}$, then the LSD of T_n exists as $n \to \infty$ and is the distribution of $f(U)$ where f is the Fourier function of $\{a_i\}$ defined on the unit interval and U is uniformly distributed on this interval. There are well-known generalisations of this result. In particular, the results of Kac et al. (1953) gives the limiting spectral distribution of A_n when for each k the elements on the k-th diagonal come from a given function h_k and $\{h_k\}$ satisfy some suitable condition. As suggested in Kac et al. (1953), if h_k are continuous then the (i,j)-th entry of A_n can be defined equal to $h_{|j-i|}((i+j)/2n+2)$, or $h_{|j-i|}(\min(i,j)/n+1)$, or $h_{|j-i|}(\max(i,j)/n+1)$. Later Trotter (1984) generalized this result for normal matrices with some suitable assumptions on the entries. In Section 2, we briefly describe the essence of their results.

[1]Since we shall be dealing with only real symmetric matrices, all our algebras are real.

In this article we look at the spectral properties in the bulk for finite diagonal large dimensional Toeplitz type matrices, which may or may not be random. We study this using the joint convergence and method of moments. In the process we revisit the above results, but restricting ourselves only to symmetric matrices.

Let $D_{r,n}$ be the $n \times n$ matrix whose (i,j)-th entry is given by

$$D_{r,n}(i,j) = g_r\left(\min\{i,j\}/n\right)\mathbb{I}(|i-j| = r)$$

where g_r is a real valued function on $[0,1]$ and for every $r < n$, let

$$T_{r,n} := \sum_{j=0}^{r-1} D_{j,n}$$

be a finite diagonal matrix.

Let U, U_1, U_2 denote i.i.d. random variables distributed uniformly on the interval $[0,\ 1]$. Assuming that $\{g_r\}$ are continuous, we first prove by using the moment method that for any fixed $r \geq 0$, $(D_{0,n}, D_{1,n}, \ldots, D_{r,n})$ converge in distribution to

$$\left(g_0(U_1), 2g_1(U_1)\cos(2\pi U_2), \ldots, 2g_r(U_1)\cos(2\pi r U_2)\right). \tag{1.1}$$

The convergence of the ESD of $T_{r,n}$ follows as a consequence. See Theorem 3.1. As a special interesting case, if $g_1(x) = x^{\frac{1}{2}}$, then ESD of $D_{1,n}$ converges to the well-known semi circle law. See Corollary 3.1. This raises the natural question: for which symmetric random matrices is the LSD equal to the LSD of a tri-diagonal non-random matrix $D_{1,n}$ for an appropriate choice of g_1? For symmetric circulant random matrices with i.i.d. entries having finite variance, it is known that the LSD is the normal distribution. We show that this is also the LSD of $D_{1,n}$ for a suitable choice of g_1. The problem for other suitable patterned random matrices (such as the Toeplitz or Hankel) leads to a Hamburger moment problem which does not seem to yield a quick resolution. See Section 3.4.

Then we move from finite diagonal matrices to the full matrix $T_n = \sum_{r=0}^{n-1} D_{r,n}$ by an application of the Mallows metric. Under a mild uniformity condition on the sequence $\{g_k\}$, we show that the ESD of T_n converges to the distribution of

$$G_\infty = g_0(U_1) + 2\sum_{j=1}^{\infty} g_j(U_1)\cos(2\pi j U_2). \tag{1.2}$$

See Corollary 3.2. Thus we get the result of Kac et al. (1953).

Finally in Section 3.3, we deal with finite diagonal matrices but with possibly unbounded $\{g_r\}$. Suppose $\{F_r\}$ are distribution functions such that

$$\int_{-\infty}^{\infty} |x|^k dF_r(x) < \infty \text{ for all } r \geq 0 \text{ and } k \geq 0. \tag{1.3}$$

Now let $g_r(x) = F_r^{-1}(x)$ for $0 < x < 1$ and $g_r(x) = 0$ for $x = 0, 1$. Then for fixed $r > 0$,

$$(D_{0,n}, D_{1,n}, \ldots, D_{r,n}) \xrightarrow{\frac{1}{n}\operatorname{Tr}} (F_0^{-1}(U_1), F_1^{-1}(U_1)2\cos(2\pi U_2), \ldots, F_r^{-1}(U_1)2\cos(2\pi r U_2)),$$

and as a consequence we get the LSD of $T_{r,n}$. See Theorem 3.2. As a corollary of this theorem, we derive the LSD of $\bar{T}_{r,n}$, where $\bar{T}_{r,n}$ is a random version of $T_{r,n}$. See Corollary 3.4. The monotonicity of F_r is crucial.

All our proofs are based on the elementary moment method and the use of Mallows metric.

2 Kac-Murdock-Szegö's and Trotter's results

Suppose $\{h_j\}$ is a sequence of real-valued continuous functions on $[0,1]$. Consider the $n \times n$ matrix $A_n = ((a_{i,j}))$ whose k-th upper and lower diagonals are given by

$$a_{i,i+k} = a_{i+k,i} = h_k\left(i/n\right), \; i = 1, \ldots, n-k, \; 0 \le k < n.$$

The results of Kac et al. (1953) imply that if

$$\sum_{j=0}^{\infty} \int_0^1 h_j^2(x)dx < \infty \text{ and } \sum_{j=0}^{\infty} \max_x |h_j(x)| < \infty$$

then the ESD of A_n converges to (the distribution of) the random variable

$$\sum_{j=-\infty}^{\infty} h_j(U_1) \exp(2\pi i j U_2) \tag{2.1}$$

where $h_{-j} = h_j$. As suggested in Kac et al. (1953), if the (i,j)-th entry of A_n is equal to $h_{|j-i|}((i+j)/2n+2)$, or $h_{|j-i|}(\min(i,j)/n+1)$, or $h_{|j-i|}(\max(i,j)/n+1)$, then also the LSD of A_n will be given by (2.1). For detail on this see Chapter II of Kac et al. (1953).

Trotter (1984) extended this result as follows. Suppose that $\{A_n, \; n \ge 1\}$ is a sequence of *normal* matrices. For each n, define a sequence of functions on $[0,1]$ as follows

$$\eta(A_n)_j(x) = 0 \;\; \forall \; x \in [0,1], \text{ if } |j| \ge n,$$

and for $0 \le j < n$,

$$\eta(A_n)_j(x) = \begin{cases} a_{i,i+j} & \forall \; x \in \left[\frac{i-1}{n}, \frac{i}{n}\right], \; i = 1, \ldots, n-j, \\ \\ 0 & \forall \; x \in \left(\frac{n-j}{n}, \; 1\right], \end{cases}$$

and for $-n < j \le 0$

$$\eta(A_n)_j(x) = \begin{cases} a_{i-j,i} & \forall \; x \in \left[\frac{i-1}{n}, \frac{i}{n}\right], \; i = 1, \ldots, n-j, \\ \\ 0 & \forall \; x \in \left(\frac{n-j}{n}, \; 1\right]. \end{cases}$$

Suppose $\{h_j, \; -\infty < j < \infty\}$ is a sequence of functions defined on $[0,1]$ such that

$$\sum_{j=-\infty}^{\infty} \int_0^1 h_j^2(x)dx < \infty \text{ and } \sum_{j=-\infty}^{\infty} \int_0^1 \left[h_j(x) - \eta(A_n)_j(x)\right]^2 dx \to 0.$$

Then again the ESD of A_n converges to the random variable (2.1). But here h_{-j} is not necessarily h_j.

Note that we can write A_n as a sum of matrices where only the kth diagonal is possibly non-zero, $1 \le k < n$. We may then consider the joint convergence of such matrices and establish a connection with Kac-Murdoch-Szegö's result.

3 Results

Recall that $D_{r,n}$ is given by

$$D_{r,n} = \begin{bmatrix} 0 & \dots & 0 & g_r\left(\frac{1}{n}\right) & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & g_r\left(\frac{2}{n}\right) & \dots & 0 \\ \vdots & \ddots & & & & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & g_r\left(\frac{n-r}{n}\right) \\ \vdots & & & \ddots & & & \vdots \\ g_r\left(\frac{1}{n}\right) & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & & & & \ddots & \vdots \\ 0 & \dots & g_r\left(\frac{n-r-1}{n}\right) & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & g_r\left(\frac{n-r}{n}\right) & 0 & \dots & 0 \end{bmatrix}.$$

3.1 Finite diagonal matrices

For every r, define the random variable

$$G_r = g_0(U_1) + 2\sum_{j=1}^{r-1} g_j(U_1)\cos(2\pi j U_2), \tag{3.1}$$

and let $\tilde{F}_r$ be its distribution function.

Theorem 3.1. *Suppose $\{g_r(\cdot)\}$ are continuous functions on $[0,1]$. Then for any fixed $r \ge 0$, $(D_{0,n}, D_{1,n}, \dots, D_{r,n})$ converge in distribution to*

$$\left(g_0(U_1), 2g_1(U_1)\cos(2\pi U_2), \dots, 2g_r(U_1)\cos(2\pi r U_2)\right). \tag{3.2}$$

As a consequence the ESD of $T_{r,n} := \sum_{j=0}^{r-1} D_{j,n}$ converges weakly to $\tilde{F}_r$.

Note that the limit in (3.2) is a commutative object.

Proof of Theorem 3.1. We first prove the theorem for $r = 2$. Consider first the particular monomials $D_{0,n}^{k_0} D_{1,n}^{k_1} D_{2,n}^{k_2}$ where $k_0, k_1, k_2 \ge 1$. Then

$$\frac{1}{n}\operatorname{Tr}\left(D_{0,n}^{k_0} D_{1,n}^{k_1} D_{2,n}^{k_2}\right) = \frac{1}{n}\sum_{1 \le i_1, i_2, \dots, i_{k_0+k_1+k_2} \le n} a_{i_1 i_2} a_{i_2 i_3} \dots a_{i_{k_0+1} i_{k_0+2}} \dots a_{i_{k_0+k_1+k_2} i_1}, \tag{3.3}$$

where first k_0 many factors in each product are from $D_{0,n}$, second k_1 factors are from $D_{1,n}$ and last k_2 factors are from $D_{2,n}$. So for example $a_{i_1 i_2}, \ldots, a_{i_{k_0} i_{k_0+1}}$ are elements of $D_{0,n}$ and they will be non-zero if $i_1 = i_2 = \cdots = i_{k_0+1}$.

Similarly, $a_{i_{k_0+1}, i_{k_0+2}}, \ldots, a_{i_{k_0+k_1}, i_{k_0+k_1+1}}$ are elements of D_1 and they will be non- zero if they satisfy the condition

$$i_{k_0+2} = i_{k_0+1} \pm 1,\ i_{k_0+3} = i_{k_0+2} \pm 1, \ldots, i_{k_0+k_1+1} = i_{k_0+k_1} \pm 1.$$

Similarly, for non-zero contribution from the last k_2 many $a_{i,j}$, we must have

$$i_{k_0+k_1+1} = i_{k_0+k_1} \pm 2, \ldots, i_{k_0+k_1+k_2} = i_{k_0+k_1+k_2-1} \pm 2 \text{ and } i_1 = i_{k_0+k_1+k_2} \pm 2.$$

Hence (3.3) reduces to

$$\begin{aligned} &\frac{1}{n} \operatorname{Tr}\left(D_{0,n}^{k_0} D_{1,n}^{k_1} D_{2,n}^{k_2}\right) \\ = \;& \frac{1}{n} \sum_{i_1=1}^{n} g_0^{k_0}\left(\frac{i_1}{n}\right) \left[\prod_{\substack{j=k_0+1 \\ |i_j - i_{j+1}|=1,\ i_{k_0+1}=i_1}}^{k_0+k_1} g_1\left(\frac{i_j}{n}\right) \right] \left[\prod_{\substack{j=k_0+k_1+1 \\ |i_j - i_{j-1}|=2, i_{k_0+k_1+k_2+1}=i_1}}^{k_0+k_1+k_2} g_2\left(\frac{i_j}{n}\right) \right] \end{aligned} \qquad (3.4)$$

Since $|i_{k_0+l} - i_{k_0+l+1}| = 1$ for $l = 1, 2, \ldots, k_1$, suppose the difference is $+1$ in j many places and it is -1 in $(k_1 - j)$ many places. So we have $\binom{k_1}{j}$ many arrangements for j many $+1$ in k_1 many places and possible values of j are $0, 1, 2, \ldots, k_1$.

Similarly, $|i_{k_0+k_1+j+1} - i_{k_0+k_1+j}| = 2$ for $j = 1, 2, \ldots, k_2$. Suppose the difference is $+2$ in j' many places and it is -2 in $(k_2 - j')$ many places. So for each $j' = 0, 1, 2, \ldots, k_2$, we have $\binom{k_2}{j'}$ many arrangements.

Since $i_1 = i_{k_0+k_1+k_2+1}$ we must have

$$j - (k_1 - j) + 2j' - 2(k_2 - j') = 0. \qquad (3.5)$$

Since $g_r, r \geq 0$ are assumed to be continuous on $[0,1]$, for any $\varepsilon > 0$, for large n,

$$|g_1(i/n) - g_1(j/n)| < \varepsilon \text{ if } |i - j| \leq k_1 \text{ and } |g_2(i/n) - g_2(j/n)| < \varepsilon \text{ if } |i - j| \leq k_2 \qquad (3.6)$$

Now using (3.4)–(3.6), we have

$$\begin{aligned} &\frac{1}{n} \operatorname{Tr}\left(D_{0,n}^{k_0} D_{1,n}^{k_1} D_{2,n}^{k_2}\right) \\ \cong\;& \frac{1}{n} \sum_{i_1=1}^{n} \left[g_0^{k_0}\left(\frac{i_1}{n}\right) \sum_{j=0}^{k_1} \binom{k_1}{j} g_1^{k_1}\left(\frac{i_1}{n}\right) \sum_{j'=0}^{k_2} \binom{k_2}{j'} g_2^{k_2}\left(\frac{i_1}{n}\right) \mathbb{I}(2j - k_1 + 4j' - 2k_2 = 0) \right] \\ =\;& \sum_{j=0}^{k_1} \sum_{j'=0}^{k_2} \binom{k_1}{j} \binom{k_2}{j'} \mathbb{I}(2j - k_1 + 4j' - 2k_2 = 0) \frac{1}{n} \sum_{i=1}^{n} g_0^{k_0}\left(\frac{i}{n}\right) g_1^{k_1}\left(\frac{i}{n}\right) g_2^{k_2}\left(\frac{i}{n}\right). \end{aligned} \qquad (3.7)$$

Now consider,

$$\mathrm{E}\left[g_0(U_1)^{k_0} g_1^{k_1}(U_1) \left(2\cos 2\pi U_2\right)^{k_1} g_2^{k_2}(U_1) \left(2\cos 4\pi U_2\right)^{k_2} \right]$$

$$
\begin{aligned}
&= \int_0^1 g_0^{k_0}(x)g_1^{k_1}(x)g_2^{k_2}(x)dx \int_0^1 (2\cos 2\pi y)^{k_1}(2\cos 4\pi y)^{k_2}\,dy \\
&= \int_0^1 g_0^{k_0}(x)g_1^{k_1}(x)g_2^{k_2}(x)dx \int_0^1 \left(e^{i2\pi y}+e^{-i2\pi y}\right)^{k_1}\left(e^{i4\pi y}+e^{-i4\pi y}\right)^{k_2}dy \\
&= \int_0^1 g_0^{k_0}(x)g_1^{k_1}(x)g_2^{k_2}(x)dx \int_0^1 \sum_{j=0}^{k_1}\sum_{j'=0}^{k_2}\binom{k_1}{j}\binom{k_2}{j'}e^{i2\pi y(j-(k_1-j)+2j'-2(k_2-j'))}dy \\
&= \int_0^1 g_0^{k_0}(x)g_1^{k_1}(x)g_2^{k_2}(x)dx \sum_{j=0}^{k_1}\sum_{j'=0}^{k_2}\binom{k_1}{j}\binom{k_2}{j'}\mathbb{I}(2j-k_1+4j'-2k_2=0) \qquad (3.8)
\end{aligned}
$$

since,

$$
\int_0^1 e^{i2\pi xk}dx = \begin{cases} 1 & \text{if } k=0 \\ 0 & \text{otherwise.} \end{cases}
$$

The expression in (3.7) is a Riemann sum which converges to that in (3.8). This proves the moment convergence for the monomial $D_{0,n}^{k_0}D_{1,n}^{k_1}D_{2,n}^{k_2}$.

Any other monomial in $\{D_{0,n}, D_{1,n}, D_{2,n}\}$ may be considered and convergence can be shown. For instance, to see how commutativity arises in the limit, consider the monomial $D_{1,n}^{k_1}D_{0,n}^{k_0}D_{2,n}^{k_2}$. Then, for sufficiently large n,

$$
\begin{aligned}
&\frac{1}{n}\operatorname{Tr}\left(D_{1,n}^{k_1}D_{0,n}^{k_0}D_{2,n}^{k_2}\right) \\
\cong\ &\frac{1}{n}\sum_{i=1}^{n}\left[g_0^{k_0}\left(\frac{i}{n}\right)\sum_{j=0}^{k_1}\binom{k_1}{j}g_1^{k_1}\left(\frac{i}{n}\right)\sum_{j'=0}^{k_2}\binom{k_2}{j'}g_2^{k_2}\left(\frac{i}{n}\right)\mathbb{I}(2j-k_1+4j'-2k_2=0)\right],
\end{aligned}
$$

and the last expression also converges to that in (3.8). This completes the proof of the theorem for $r=2$.

For $r>2$, instead of $D_0^{k_0}D_1^{k_1}D_2^{k_2}$ one can consider a monomial f in $\{D_{0,n}, D_{1,n}, \ldots, D_{r,n}\}$ and following similar arguments as above it can be shown that $\frac{1}{n}\operatorname{Tr} f(D_{0,n}, ..., D_{r,n})$ converges to $\mathrm{E}(f(g_0(U_1), g_1(U_1)2\cos 2\pi U_2, \ldots, g_r(U_1)2\cos 2\pi rU_2)$. Hence $(D_{0,n}, ..., D_{r,n})$ converges in law to $(g_0(U_1), g_1(U_1)2\cos 2\pi U_2, \ldots, g_r(U_1)2\cos 2\pi rU_2)$.

It remains to establish the LSD of $T_{r,n}$. For this, we again use the method of moments and use the following well known fact.

Suppose $\{Y_n\}$ is a sequence of real valued random variables with distributions $\{F_n\}$. Suppose that there exists some sequence $\{m_h\}$ such that as $n\to\infty$, $\mathrm{E}[Y_n^h] = \int x^h dF_n(x) \to m_h$ for every positive integer h. Suppose that there is a unique distribution F whose moments are $\{m_h\}$. Then Y_n (or equivalently F_n) converges to F in distribution.

Now let F_n be the ESD of $T_{r,n}$ and $\beta_{k,n}$ denote the k-th moment of F_n. Then

$$
\begin{aligned}
\beta_{k,n} &= \frac{1}{n}\operatorname{Tr}\Big(\sum_{j=0}^{r-1} D_{j,n}\Big)^k \\
&= \sum_{i_j\in\{0,1,2,\ldots,r-1\}} \frac{1}{n}\operatorname{Tr}\left(D_{i_1,n}D_{i_2,n}\cdots D_{i_k,n}\right)
\end{aligned}
$$

$$\begin{aligned} &\to \sum_{i_j \in \{0,1,2,\ldots,r-1\}} \mathrm{E}(f_{i_1} f_{i_2} \cdots f_{i_k}) \quad \text{as} \quad n \to \infty \quad \text{(by Theorem 3.1)} \\ &= \mathrm{E}(f_0 + f_1 + \cdots + f_{r-1})^k = m_k, \text{ say,} \end{aligned}$$

where $(f_0, f_1, \ldots, f_{r-1}) = (g_0(U_1), 2g_1(U_1)\cos(2\pi U_2), \ldots, 2g_{r-1}(U_1)\cos(2\pi r U_2))$ and E is expectation with respect to the distribution of U_1 and U_2. Since $\{|g_j|;\ 0 \le j < r\}$ are bounded (by say M), it follows that $m_k \le [(r+1)M]^k$ for all $k \ge 1$. Therefore they determine a distribution (namely $\tilde{F}_r$) uniquely. Hence F_n converges to $\tilde{F}_r$ weakly. □

Corollary 3.1. *If $g_1(x) = x^{\frac{1}{2}}$, then LSD of $D_{1,n}$ equals $2\sqrt{U_1}\cos(2\pi U_2)$ which has the semi-circular distribution with density*

$$f(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\,\mathbb{I}(-2 < x < 2).$$

Proof. This is easily proved by showing that for all non-negative integers k,

$$\mathrm{E}[2\sqrt{U_1}\cos(2\pi U_2)]^{2k} = \frac{(2k)!}{k!(k+1)!} = \frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4 - x^2}\,dx$$

and

$$\mathrm{E}[2\sqrt{U_1}\cos(2\pi U_2)]^{2k+1} = 0 = \frac{1}{2\pi}\int_{-2}^{2} x^{2k+1}\sqrt{4 - x^2}\,dx.$$

Since these moments determine the distribution uniquely, the results follows. □

Remark 3.1. Let for $r \ge 0$, $\tilde{D}_{r,n}$ be an $n \times n$ matrix with (i,j)-th element

$$\tilde{D}_{r,n}(i,j) = g_r(U)\mathbb{I}(|i - j| = r). \tag{3.9}$$

Suppose $g_r(\cdot)$ are continuous functions on $[0,1]$. Following arguments similar to that given in the proof of Theorem 3.1, one can show that

(i) For fixed $r > 0$,

$$\left(\tilde{D}_{o,n}, \tilde{D}_{1,n}, \ldots, \tilde{D}_{r,n}\right) \xrightarrow{\frac{1}{n}\mathrm{E}\,\mathrm{Tr}} (g_0(U_1), g_1(U_1)2\cos 2\pi U_2, \ldots, g_r(U_1)2\cos 2\pi r U_2)$$

and as a consequence the expected ESD of $\tilde{T}_{r,n} := \sum_{i=0}^{r} \tilde{D}_{i,n}$ converges weakly to $g_0(U_1) + \sum_{j=1}^{r} g_j(U_1)2\cos(2\pi j U_2)$.

(ii) For almost every value of U,

$$\left(\tilde{D}_{o,n}, \tilde{D}_{1,n}, \ldots, \tilde{D}_{r,n}\right) \xrightarrow{\frac{1}{n}\mathrm{Tr}} (g_0(U), g_1(U)2\cos 2\pi U_2, \ldots, g_r(U)2\cos 2\pi r U_2)$$

and hence for fixed $r > 0$, the ESD of $\tilde{T}_{r,n}$ converges weakly almost surely to $g_0(U) + \sum_{j=1}^{r} g_j(U)2\cos(2\pi j U_2)$. Note that in this limit U is a bounded constant.

3.2 When all diagonals are present

We now deal with matrices which may have all diagonals non-zero. To do this we need an appropriate metric which will allow such matrices to be approximated by k-diagonal matrices with finite but large k.

The Mallows metric (also known as Wasserstein metric) is defined on the space of all probability distributions with finite second moment. Let F and G be two distribution functions with finite second moment. Then the Mallows distance between F and G is defined as

$$d^2(F,G) = \inf_{X\sim F, Y\sim G} \mathrm{E}\,|X-Y|^2. \tag{3.10}$$

It is known that $d(F_n, F) \to 0$ if and only if F_n converges to F weakly and $\int x^2 dF_n(x) \to \int x^2 dF(x)$.

We shall use this metric in the proof of our next results. For this, we need the following upper bound of the Mallows metric distance between the ESD of two matrices: let A, B be two $n \times n$ real symmetric matrices with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ and $\beta_1 \le \beta_2 \le \cdots \le \beta_n$, respectively. Define a bivariate distribution function F as follows

$$F(x,y) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(\lambda_j \le x, \beta_j \le y).$$

Then the marginal distributions of F are F_A and F_B. Let (X,Y) be a random vector whose distribution function is F. Then

$$d^2(F_A, F_B) \le \mathrm{E}\,|X-Y|^2 = \frac{1}{n}\sum_{j=1}^{n}(\lambda_j - \beta_j)^2 \le \frac{1}{n}\,\mathrm{Tr}(A-B)^2. \tag{3.11}$$

The last inequality above is a standard result in matrix algebra; one can see a proof of this in Lemma 2.3 of Bai (1999).

Assume condition (i) of Corollary 3.2 holds. Then G_∞ is well defined. We can also consider G_r, G_∞ as random variables defined on the (same) probability space $([0,1]^2, \mathcal{B}, \mathrm{P})$ where P is the product measure on $[0,1]\times[0,1]$ and G_∞ is the limit of $\{G_r\}$ in $L^2([0,1]^2)$, that is, $\|G_r - G_\infty\|_2 \to 0$ as $r \to \infty$. Let $\tilde{F}_\infty$ be the distribution function of G_∞.

Corollary 3.2. *Suppose* $\{g_r\}$ *are continuous functions on* $[0,1]$. *Suppose*

(i) $\displaystyle\sum_{j=0}^{\infty}\int_0^1 g_j^2(x)dx < \infty$ *and*

(ii) $\displaystyle\sum_{j=0}^{n-1}\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} \left(g_j^2\left(\frac{k}{n}\right) - g_j^2(x)\right)dx \;\to 0$ *as* $n \to \infty$.

Then LSD of $T_n = \displaystyle\sum_{r=0}^{n-1} D_{r,n}$ *equals* $\tilde{F}_\infty$.

Proof. Let F_n denote the ESD of T_n and $F_{r,n}$ denote the ESD of $T_{r,n}$. Then we have

$$d(F_n, \tilde{F}_\infty) \le d(F_n, F_{r,n}) + d(F_{r,n}, \tilde{F}_r) + d(\tilde{F}_r, \tilde{F}_\infty).$$

Now using Mallows metric (3.10) and condition (i) of the Corollary, we have

$$d^2(\tilde{F}_r, \tilde{F}_\infty) \le \|G_r - G_\infty\|_2^2 \to 0.$$

Hence $\tilde{F}_r$ converges weakly to $\tilde{F}_\infty$. So for a fixed $\varepsilon > 0$, there exist an $R \in \mathbb{N}$ such that

$$d(\tilde{F}_r, \tilde{F}_\infty) \le \varepsilon \quad \text{for all} \quad r \ge R.$$

Now for a fixed $r \ge R$, by Theorem 3.1 $d(F_{r,n}, \tilde{F}_r) \to 0$ as $n \to \infty$. Finally, using (3.11) we have

$$\begin{aligned} d^2(F_n, F_{r,n}) &\le \frac{1}{n} \operatorname{Tr}(T_n - T_{r,n})^2 \\ &\le \frac{1}{n} \sum_{j=r}^{n-1} \sum_{k=1}^{n-j} g_j^2\left(\frac{k}{n}\right) \\ &\le \sum_{j=r}^{n-1} \left[\frac{1}{n} \sum_{k=1}^{n} g_j^2\left(\frac{k}{n}\right) - \int_0^1 g_j^2(x)dx \right] + \sum_{j=r+1}^{n-1} \int_0^1 g_j^2(x)dx, \end{aligned}$$

and due to conditions $(i), (ii)$, right side goes to zero as $n \to \infty$. Hence $d(F_n, \tilde{F}_\infty) \to 0$ as $n \to \infty$. This completes the proof of the corollary. □

Remark 3.2. Consider the symmetric Toeplitz matrix $T_n = ((a_{|i-j|}))_{n \times n}$ with $\sum_{i=0}^{\infty} a_i^2 < \infty$, then the LSD of T_n equals $a_0 + 2\sum_{j=0}^{\infty} a_j \cos(2\pi j U_1)$. This follows by taking $g_j(x) \equiv a_j$ for all $x \in [0, 1]$ and for all j. This is the famous Szegö's limit for non-random symmetric Toeplitz matrices.

The next result is similar to Corollary 3.2 for the matrices $\tilde{D}_{r,n}$. Recall from (3.9) that $\tilde{D}_{r,n}(i, j) = g_r(U)\mathbb{I}(|i - j| = r)$.

Corollary 3.3. *Suppose* $\{g_r\}$ *are continuous functions on* $[0, 1]$. *If* $\sum_{j=0}^{\infty} \int_0^1 g_j^2(x)dx < \infty$, *then expected ESD of* $\tilde{T}_n = \sum_{j=0}^{n-1} \tilde{D}_{j,n}$ *converges to* $\tilde{F}_\infty$.

Proof. Idea of the proof of this corollary is similar to the proof of Corollary 3.2 with a simple modification. Here we use notation similar to that in Corollary 3.2. Let F_n and $F_{r,n}$ denote the ESD of $\tilde{T}_n$ and $\tilde{T}_{r,n}$, and $\mathrm{E}\, F_n$, $\mathrm{E}\, F_{r,n}$ denote the expected ESD of $\tilde{T}_n$ and $\tilde{T}_{r,n}$, respectively. Now, using Mallows metric we have

$$d(\mathrm{E}\, F_n, \tilde{F}_\infty) \le d(\mathrm{E}\, F_n, \mathrm{E}\, F_{r,n}) + d(\mathrm{E}\, F_{r,n}, \tilde{F}_r) + d(\tilde{F}_r, \tilde{F}_\infty).$$

Now following similar arguments as given in the proof of Corollary 3.2 and using Remark 3.1, it can be shown that the second and the third term in the last expression goes to zero as $n, r \to \infty$. To show that the first term goes to zero, we first prove that

$$d^2(\mathrm{E}\, F_n, \mathrm{E}\, F_{r,n}) \leq \frac{1}{n} \mathrm{E}\, \mathrm{Tr}(\tilde{T}_n - \tilde{T}_{r,n})^2.$$

Suppose $\lambda_1(U) \leq \lambda_2(U) \leq \cdots \leq \lambda_n(U)$ and $\beta_1(U) \leq \beta_2(U) \leq \cdots \leq \beta_n(U)$ are the eigenvalues of $\tilde{T}_n$ and $\tilde{T}_{r,n}$ respectively. Then

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\lambda_i(U) \leq x) \quad \text{and} \quad F_{r,n}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\beta_i(U) \leq x).$$

Define

$$F_0(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\lambda_i(U) \leq x, \beta_i(U) \leq y).$$

Then the marginals of F_0 are F_n and $F_{r,n}$. Let

$$\tilde{F}_0(x, y) = \mathrm{E}_U(F_0(x, y)) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{P}(\lambda_i(U) \leq x, \beta_i(U) \leq y).$$

Then the marginals of $\tilde{F}_0$ are $\mathrm{E}\, F_n$ and $\mathrm{E}\, F_{r,n}$.

Define a transition probability measure on $\Omega \times \mathbb{R} \times \mathbb{R}$ by

$$\mu(\omega, dx, dy) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left(\lambda_i(U(\omega)) = x, \beta_i(U(\omega)) = y\right).$$

Let

$$\nu(dx, dy) = \int_{\Omega} \mu(\omega, dx, dy) dP(\omega).$$

Let (Z_1, Z_2) be a random vector whose cumulative distribution function is $\tilde{F}_0(x, y)$. Now

$$\begin{aligned}
\mathrm{E}(Z_1 - Z_2)^2 &= \int_{z_1, z_2} (z_1 - z_2)^2 d\tilde{F}_0(z_1, z_2) \\
&= \int_{z_1, z_2} (z_1 - z_2)^2 \nu(dz_1, dz_2) \\
&= \int_{z_1, z_2} (z_1 - z_2)^2 \int_{\Omega} \mu(\omega, dz_1, dz_2) dP(\omega) \\
&= \int_{\Omega} \int_{z_1, z_2} (z_1 - z_2)^2 \mu(\omega, dz_1, dz_2) dP(\omega) \\
&= \int_{\Omega} \frac{1}{n} \sum_{i=1}^{n} \left(\lambda_i(U(\omega)) - \beta_i(U(\omega))\right)^2 dP(\omega) \\
&= \frac{1}{n} \mathrm{E}\left[\sum_{i=1}^{n} \left(\lambda_i(U) - \beta_i(U)\right)^2\right].
\end{aligned}$$

Now using Mallows distance and (3.11)

$$d^2(\mathrm{E}\,F_n, \mathrm{E}\,F_{r,n}) \le E(Z_1 - Z_2)^2 = \frac{1}{n}\,\mathrm{E}\left[\sum_{i=1}^{n} (\lambda_i(U) - \beta_i(U))^2\right] \le \frac{1}{n}\,\mathrm{E}\,\mathrm{Tr}(\tilde{T}_n - \tilde{T}_{r,n})^2.$$

Under the given condition, the last expression goes to zero as $n \to \infty$. □

3.3 Unbounded entries

One crucial aspect of the results established so far is that the entries are bounded and the limit is also bounded. While general unbounded entries may be difficult to handle, we can handle some specific unbounded cases.

Theorem 3.2. *Suppose $\{F_r\}$ are distribution functions such that*

$$\int_{-\infty}^{\infty} |x|^k dF_r(x) < \infty \quad \text{for all } r \ge 0 \text{ and } k \ge 0. \tag{3.12}$$

Now let $g_r(x) = F_r^{-1}(x)$ for $0 < x < 1$ and $g_r(x) = 0$ for $x = 0, 1$. Then for fixed $r > 0$,

$$(D_{0,n}, D_{1,n}, \ldots, D_{r,n}) \xrightarrow{\frac{1}{n}\mathrm{Tr}} (F_0^{-1}(U_1), F_1^{-1}(U_1) 2\cos(2\pi U_2), \ldots, F_r^{-1}(U_1) 2\cos(2\pi r U_2)).$$

As a consequence the LSD of $T_{r,n} := \sum_{j=0}^{r} D_{j,n}$ equals $F_0^{-1}(U_1) + 2\sum_{j=1}^{r} F_j^{-1}(U_1)\cos(2\pi j U_2)$.

Remark 3.3. In the proof of this theorem we will use the monotonicity of F_r crucially. One can prove similar type of result with other appropriate, monotone $\{g_r\}$.

Let for each $r \ge 0$, $\{x_{r,n}; n \ge 1\}$ be an i.i.d. sequence of random variables with distribution function F_r. For fixed n, let $\{x_{r,(1)}, x_{r,(2)}, \ldots, x_{r,(n)}\}$ be the order statistic of $\{x_{r,1}, x_{r,2}, \ldots, x_{r,n}\}$. Let

$$\bar{D}_{r,n} = \begin{bmatrix} 0 & \cdots & 0 & x_{r,(1)} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & x_{r,(2)} & \cdots & 0 \\ \vdots & \ddots & & & & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & x_{r,(n-r)} \\ \vdots & & & \ddots & & & \vdots \\ x_{r,(1)} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & & & & \ddots & \vdots \\ 0 & \cdots & x_{r,(n-r-1)} & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & x_{r,(n-r)} & 0 & \cdots & 0 \end{bmatrix}.$$

If $F_{r,n}$ denote the empirical distribution function of $\{x_{r,1}, x_{r,2}, \ldots, x_{r,n}\}$ then (i,j)-th entry of $\bar{D}_{r,n}$ is $F_{r,n}^{-1}\left(\frac{\min\{i,j\}}{n}\right)$. So for fixed $r \ge 0$, we have a triangular sequence of symmetric random matrices $\{\bar{D}_{r,n}; n \ge 1\}$ and we have the following Corollary of the previous Theorem.

Corollary 3.4. *If $\{F_r; r \geq 0\}$ satisfy condition of Theorem 3.2 then for fixed r, the LSD of $\bar{T}_{r,n} := \sum_{j=0}^{r} \bar{D}_{j,n}$ equals $F_0^{-1}(U_1) + 2\sum_{j=1}^{r} F_j^{-1}(U_1)\cos(2\pi j U_2)$ almost surely.*

Proof. Let $F_{T_{r,n}}, F_{\bar{T}_{r,n}}$ be the ESDs of $T_{r,n}$ and $\bar{T}_{r,n}$ respectively. If we can show that

$$d(F_{T_{r,n}}, F_{\bar{T}_{r,n}}) \to 0 \ \ a.s.,$$

then from Theorem 3.2 the result follows. Now

$$\begin{aligned} d^2(F_{T_{r,n}}, F_{\bar{T}_{r,n}}) &\leq \frac{1}{n}\sum_{i=0}^{r}\sum_{j=1}^{n-i} 2\left(F_i^{-1}(j/n) - F_{i,n}^{-1}(j/n)\right)^2 \\ &\leq 2\sum_{i=0}^{r}\left[\frac{1}{n}\sum_{j=1}^{n-1}\left(F_i^{-1}(j/n) - F_{i,n}^{-1}(j/n)\right)^2\right]. \end{aligned}$$

Now fix $0 \leq i \leq r$ and consider

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n}\left(F_i^{-1}(j/n) - F_{i,n}^{-1}(j/n)\right)^2 \\ = \ &\frac{1}{n}\sum_{j=1}^{n}(F_i^{-1}(j/n))^2 + \frac{1}{n}\sum_{j=1}^{n}(F_{i,n}^{-1}(j/n))^2 - 2\frac{1}{n}\sum_{j=1}^{n}F_i^{-1}(j/n)F_{i,n}^{-1}(j/n) \\ = \ &I_1 + I_2 - 2I_3, \text{ say.} \end{aligned}$$

By condition (3.12) and DCT, I_1 converges to $\int_0^1 (F_i^{-1}(x))^2 dx$. Similarly

$$I_2 = \frac{1}{n}\sum_{j=1}^{n}(F_{i,n}^{-1}(j/n))^2 = \frac{1}{n}\sum_{j=1}^{n} x_{i,j}^2 \to \int_0^1 (F_i^{-1}(x))^2 dx$$

almost surely (by SLLN). We shall show that $I_3 \to \int_0^1 (F_i^{-1}(x))^2 dx$ a.s. Note that

$$I_3 = \frac{1}{n}\sum_{j=1}^{n} F_i^{-1}(j/n)F_{i,n}^{-1}(j/n) = \int_0^1 J_{i,n}(t)F_{i,n}^{-1}(t)dt$$

where $J_{i,n}(t) = F_i^{-1}(j/n)$ if $\frac{j}{n} \leq t < \frac{j+1}{n}$. Now

$$\int_0^1 J_{i,n}(t)F_{i,n}^{-1}(t)dt = \int_0^1 J_{i,n}(t)F_i^{-1}(t)dt + \left(\int_0^1 J_{i,n}(t)F_{i,n}^{-1}(t)dt - \int_0^1 J_{i,n}(t)F_i^{-1}(t)dt\right)$$

and again using condition (3.12) and DCT

$$\int_0^1 J_{i,n}(t)F_i^{-1}(t)dt \to \int_0^1 (F_i^{-1}(x))^2 dx.$$

Using Cauchy-Schwartz inequality we have

$$\left|\int_0^1 J_{i,n}(t)F_{i,n}^{-1}(t)dt - \int_0^1 J_{i,n}(t)F_i^{-1}(t)dt\right| \ = \ \left|\int_0^1 J_{i,n}(t)\left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)dt\right|$$

$$\leq \left[\int_0^1 J_{i,n}^2(t)dt\right]^{1/2} \left[\int_0^1 \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2\right]^{1/2}.$$

Again using condition (3.12) and DCT

$$\int_0^1 J_{i,n}^2(t)dt = \frac{1}{n}\sum_{j=1}^n (F_i^{-1}(j/n))^2 \to \int_0^1 (F_i^{-1}(x))^2 dx \text{ as } n \to \infty.$$

Now it remains to show that almost surely

$$\int_0^1 \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2 dt \to 0 \text{ as } n \to \infty.$$

Let $\varepsilon > 0$. First observe that

$$\int_{-\infty}^{\infty} x^2 dF_{i,n}(x) = \frac{1}{n}\sum_{j=1}^n x_{i,j}^2 \to \int_{-\infty}^{\infty} x^2 dF_i(x) \ \text{ a.s. (by SLLN)}$$

and hence we can choose an $\tilde{n} \in \mathbb{N}$ a.s. such that

$$\left|\int_0^1 \left[\left(F_{i,n}^{-1}(t)\right)^2 - \left(F_i^{-1}(t)\right)^2\right] dt\right| < \varepsilon \ \text{ for all } n \geq \tilde{n}. \tag{3.13}$$

Due to (3.12), we can have $\delta = \delta_\varepsilon$ such that

$$\int_{[\delta,1-\delta]^c} \left(F_i^{-1}(t)\right)^2 dt < \varepsilon. \tag{3.14}$$

$F_{i,n}^{-1}(\cdot)$ and $F_i^{-1}(\cdot)$ are bounded on $[\delta, 1-\delta]$ and $F_{i,n}^{-1}(t) \to F_i^{-1}(t)$ a.s., hence by DCT

$$\int_\delta^{1-\delta} \left[\left(F_{i,n}^{-1}(t)\right)^2 - \left(F_i^{-1}(t)\right)^2\right] dt < \varepsilon, \tag{3.15}$$

and

$$\int_\delta^{1-\delta} \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2 dt < \varepsilon. \tag{3.16}$$

for all $n \geq n'$. Now

$$\begin{aligned}
& \left|\int_{[\delta,1-\delta]^c} \left(F_{i,n}^{-1}(t)\right)^2 dt\right| \\
\leq\ & \left|\int_{[\delta,1-\delta]^c} \left[\left(F_{i,n}^{-1}(t)\right)^2 - \left(F_i^{-1}(t)\right)^2\right] dt\right| + \left|\int_{[\delta,1-\delta]^c} \left(F_i^{-1}(t)\right)^2 dt\right| \\
\leq\ & \left|\int_0^1 \left[\left(F_{i,n}^{-1}(t)\right)^2 - \left(F_i^{-1}(t)\right)^2\right] dt\right| + \left|\int_\delta^{1-\delta} \left[\left(F_{i,n}^{-1}(t)\right)^2 - \left(F_i^{-1}(t)\right)^2\right] dt\right| + \left|\int_{[\delta,1-\delta]^c} \left(F_i^{-1}(t)\right)^2 dt\right| \\
\leq\ & 3\varepsilon \qquad \text{using(3.13), (3.14) and (3.15)}
\end{aligned}$$

for all $n \geq N = \max\{n', \tilde{n}\}$. Finally using (3.17)

$$\begin{aligned}\int_0^1 \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2 dt &\leq \int_\delta^{1-\delta} \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2 dt + \int_{[\delta,1-\delta]^c} \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2 dt \\ &\leq \varepsilon + \int_{[\delta,1-\delta]^c} \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2 dt \qquad (3.17)\end{aligned}$$

for all $n \geq n'$ and since

$$\begin{aligned}\left[\int_{[\delta,1-\delta]^c} \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2 dt\right]^{1/2} &\leq \left[\int_{[\delta,1-\delta]^c} \left(F_{i,n}^{-1}(t)\right)^2 dt\right]^{1/2} + \left[\int_{[\delta,1-\delta]^c} \left(F_i^{-1}(t)\right)^2 dt\right]^{1/2} \\ &\leq \sqrt{3\varepsilon} + \sqrt{\varepsilon}, \qquad (3.18)\end{aligned}$$

for all $n \geq N$, we have from (3.17), (3.18),

$$\int_0^1 \left(F_{i,n}^{-1}(t) - F_i^{-1}(t)\right)^2 dt < 9\varepsilon \quad \text{for all } n \geq N.$$

This completes the proof of the corollary. □

Proof of Theorem 3.2. For simplicity we shall prove the theorem for $r = 1$. For $r > 1$, similar argument works.

We will consider three cases separately depending on the nature of the supports of F_0 and F_1.

Case 1. Both F_0 and F_1 are supported on $[0, \infty)$. Consider the monomial $D_{0,n}^{k_0} D_{1,n}^{k_1}$ where $k_0, k_1 \geq 1$. Then

$$\frac{1}{n} \operatorname{Tr}\left(D_{0,n}^{k_0} D_{1,n}^{k_1}\right) = \frac{1}{n} \sum_{1 \leq i_1, i_2, \ldots, i_{k_0+k_1} \leq n} a_{i_1 i_2} a_{i_2 i_3} \ldots a_{i_{k_0+1} i_{k_0+2}} \ldots a_{i_{k_0+k_1} i_1}, \qquad (3.19)$$

where first k_0 many factors in each product are from $D_{0,n}$ and second k_1 entries are from $D_{1,n}$. So

$$\frac{1}{n} \operatorname{Tr}\left(D_{0,n}^{k_0} D_{1,n}^{k_1}\right) = \frac{1}{n} \sum_{1 \leq i_1, i_2, \ldots, i_{k_0+k_1} \leq n} a_{i_1 i_2} a_{i_2 i_3} \ldots a_{i_{k_0+1} i_{k_0+2}} \ldots a_{i_{k_0+k_1} i_1} = I_1 + I_2, \text{ say}, \qquad (3.20)$$

where

$$I_1 = \frac{1}{n} \sum_{i_1=1}^{n-k_1-1} g_0^{k_0}\left(\frac{i_1}{n}\right) \left[\prod_{\substack{j=k_0+1 \\ i_{k_0+1}=i_1=i_{k_0+k_1+1}, |i_j - i_{j+1}|=1}}^{k_0+k_1} g_1\left(\frac{i_j \wedge i_{j+1}}{n}\right)\right],$$

$$I_2 = \frac{1}{n} \sum_{i_1=n-k_1}^{n} g_0^{k_0}\left(\frac{i_1}{n}\right) \left[\prod_{\substack{j=k_0+1 \\ i_{k_0+1}=i_1=i_{k_0+k_1+1}, |i_j - i_{j+1}|=1}}^{k_0+k_1} g_1\left(\frac{i_j \wedge i_{j+1}}{n}\right)\right].$$

Now due to condition (3.12), $I_2 \to 0$, as $n \to \infty$ and

$$\begin{aligned}
I_1 &= \frac{1}{n}\sum_{i_1=1}^{n-k_1-1} g_0^{k_0}\left(\frac{i_1}{n}\right)\left[\prod_{\substack{j=k_0+1\\ i_{k_0+1}=i_1=i_{k_0+k_1+1},|i_j-i_{j+1}|=1}}^{k_0+k_1} g_1\left(\frac{i_j\wedge i_{j+1}}{n}\right)\right]\\
&= \sum_{i_1=1}^{n-k_1-1}\int_{\frac{i_1+k_1}{n}}^{\frac{i_1+k_1+1}{n}} g_0^{k_0}\left(\frac{i_1}{n}\right)\left[\prod_{\substack{j=k_0+1\\ i_{k_0+1}=i_1=i_{k_0+k_1+1},|i_j-i_{j+1}|=1}}^{k_0+k_1} g_1\left(\frac{i_j\wedge i_{j+1}}{n}\right)\right]dx\\
&= \int_0^1\sum_{i_1=1}^{n-k_1-1} g_0^{k_0}\left(\frac{i_1}{n}\right)\mathbb{I}_{[\frac{i_1+k_1}{n},\frac{i_1+k_1+1}{n})}(x)\left[\prod_{\substack{j=k_0+1\\ i_{k_0+1}=i_1=i_{k_0+k_1+1},|i_j-i_{j+1}|=1}}^{k_0+k_1} g_1\left(\frac{i_j\wedge i_{j+1}}{n}\right)\right]dx\\
&=: \int_0^1 f_n(x)dx, \text{ say.}
\end{aligned}$$

Since g_0 and g_1 are continuous at almost all $x \in (0,1)$, we have as $n \to \infty$,

$$f_n(x) \xrightarrow{a.s.} g_0^{k_0}(x)\sum_{\substack{j=0\\ 2j-k_1=0}}^{k_1}\binom{k_1}{j}g_1^{k_1}(x) = g_0^{k_0}(x)\binom{k_1}{k_1/2}g_1^{k_1}(x), \text{ provided } k_1 \text{ is even}$$

and $f_n(x) = 0$, otherwise. Note that since in this case $|g_0|, |g_1|$ are non-decreasing in $x \in (0,1)$, $|f_n(x)| \le c_1|g_0(x)|^{k_0}|g_1(x)|^{k_1}$, where c_1 is a constant and also $\int_0^1 |g_0(x)|^{k_0}|g_1(x)|^{k_1}dx < \infty$, due to condition (3.12). So using the DCT, as $n \to \infty$

$$\frac{1}{n}\operatorname{Tr}\left(D_{0,n}^{k_0}D_{1,n}^{k_1}\right) = \int_0^1 f_n(x)dx \to \int_0^1 g_0^{k_0}(x)\binom{k_1}{k_1/2}g_1^{k_1}(x)dx = \mathrm{E}\left[g_0^{k_0}(U_1)\left(g_1(U_1)2\cos(2\pi U_2)\right)^{k_1}\right].$$

Similarly we can consider a polynomial in $D_{0,n}, D_{1,n}$ and similar convergence can be shown. This completes the proof of Case 1.

Case 2. F_0 and F_1 both are supported on $(-\infty, 0]$. In this case, similar argument as in case 1 will work but one has to be more careful since $|g_0|$ and $|g_1|$ are now non-increasing. As in (3.20) $\frac{1}{n}\operatorname{Tr}\left(D_{0,n}^{k_0}D_{1,n}^{k_1}\right)$ can be written as

$$\frac{1}{n}\operatorname{Tr}\left(D_{0,n}^{k_0}D_{1,n}^{k_1}\right) = \frac{1}{n}\sum_{i_1=1}^{k_1}\cdots + \frac{1}{n}\sum_{i_1=k_1+1}^{n}\cdots =: I_1 + I_2.$$

Here $I_1 \to 0$ and $I_2 \to \mathrm{E}\left[g_0^{k_0}(U_1)\left(g_1(U_1)2\cos(2\pi U_2)\right)^{k_1}\right]$ as $n \to \infty$.

Case 3. One of F_0 and F_1 is supported on $[0, \infty)$ and the other on $(-\infty, 0]$. Without loss of generality we assume that F_0 is supported on $[0, \infty)$. We will just sketch the idea behind the proof in this case.

We divide $[0,1]$ into two parts $J_1 := [0,1/2]$ and $J_2 := (1/2,1]$. First we restrict the sum in (3.19) to J_1 i.e i_1 from 1 to m_n where m_n is the largest positive integer so that $m_n/n \leq 1/2$. We call this part S_1 and the remaining part S_2. Observe that in S_1, g_0 is bounded and $g_0 = |g_0|$ is non-decreasing and hence the factor involving g_0 can be bounded by $|g_0(3/4)|^{k_0}$. But g_1 is unbounded near 0 and $|g_1|$ is non-increasing. In order to apply the DCT we argue as in case 2 and make necessary adjustments for g_1 near 0 so that S_1 is dominated by $c_2|g_0(3/4)|^{k_0}|g_1(x)|^{k_1}$, where c_2 is a constant. For S_2 observe that now g_1 is bounded and $|g_1|$ is non-increasing so that factors involving g_1 can be bounded by $|g_1(1/4)|^{k_1}$. However g_0 is unbounded near 1 and $g_0 = |g_0|$ is non-decreasing. So for g_0 we argue as in case 1. Using the DCT we get the result.

The general case can be handled using appropriate combinations of Cases 1–3.

□

3.4 Tridiagonal matrices and LSD of random matrices

We have seen in Corollary 3.1 that the semi-circle law arises as the LSD of a tri-diagonal non-random symmetric matrix. It is interesting to explore if other LSDs of random symmetric matrices arise as the LSD of non-random tri-diagonal matrices.

Let $\{x_j\}$ be i.i.d. random variables with mean zero and variance one. Let Z denote the random variable $Z = 2\cos(2\pi U)$ where U is uniformly distributed on $[0,1]$. Then observe that $\mathrm{E}(Z^{2k+1}) = 0$, $\mathrm{E}(Z^{2k}) = \frac{(2k)!}{(k!)^2}$ and Z has the following density

$$f_Z(x) = \frac{1}{\pi}\frac{1}{\sqrt{4-x^2}}, \ -2 \leq x \leq 2.$$

(i) Symmetric Circulant matrix. Consider the $n \times n$ random Symmetric Circulant matrix SC_n whose (i,j)-th entry is given by $SC_n(i,j) = x_{n/2-|n/2-|i-j||}$ for $1 \leq i,j \leq n$. Then it is well known that almost surely, the LSD of $n^{-1/2}SC_n$ is the standard normal distribution.

Let F_1 be the distribution function of standard symmetrized Rayleigh random variable R with density $f(x) = |x|\exp(-x^2)$ where $-\infty < x < \infty$. Then $E(R^{2k+1}) = 0$ and $\mathrm{E}(R^{2k}) = k!$. Now consider $D_{1,n}$ with $\frac{1}{\sqrt{2}}F_1^{-1}$ and then by Theorem 3.2, the ESD of $D_{1,n}$ converges to $\frac{1}{\sqrt{2}}F_1^{-1}(U_1)Z$. It is easy to see that

$$\mathrm{E}(\frac{1}{\sqrt{2}}F_1^{-1}(U_1)Z)^{2k+1} = 0 \text{ and } \mathrm{E}(\frac{1}{\sqrt{2}}F_1^{-1}(U_1)Z)^{2k} = \frac{(2k)!}{2^k k!}.$$

Thus LSD of $D_{1,n}$ is the same as that of $n^{-1/2}SC_n$.

(ii) Reverse Circulant matrix. Now consider the random Reverse Circulant matrix RC_n whose (i,j)-th entry is given by $RC_n(i,j) = x_{(i+j-2) \bmod n}$ for $1 \leq i,j \leq n$. Then almost surely the LSD of $n^{-1/2}RC_n$ is the symmetrized Rayleigh distribution given above.

Now let $R = Y.Z$ where Y and Z are independent random variables and R is a symmetrized Rayleigh variable and $Z = 2\cos 2\pi U$. Then

$$E(Y^{2k}) = \frac{E(R^{2k})}{E(Z^{2k})} = \frac{(k!)^3}{(2k)!} = m_{2k}, \text{ say}.$$

Also $m_{2k+1} = E(Y^{2k+1}) = 0$. It is easy to check that $\{m_k\}$ satisfies Carleman's condition. However we have been unable to resolve if it is a moment sequence. **If** it is indeed a moment sequence then there is a unique distribution F with this moment sequence. In that case, if we take $g_1 = F^{-1}$, then the LSD of $D_{1,n}$ is the same as that of $n^{-1/2}RC_n$.

(iii) In Bose et al. (2008), it is shown that under a mild condition on the pattern, any possible LSD of patterned random matrices is symmetric about zero and its even moments $\{\beta_{2k}\}$ are bounded above by the moments of a Gaussian variable. Then to identify an appropriate tri-diagonal matrix with the same limit, we may proceed as above. Now we need to solve $\mathrm{E}\, X^{2k}\, \mathrm{E}\, Z^{2k} = \beta_{2k}$ or

$$m_{2k} = \mathrm{E}\, X_{2k} = \frac{\beta_{2k}}{\mathrm{E}\, Z^{2k}} = \beta_{2k}\frac{(k!)^2}{(2k)!} \quad \text{and} \quad m_{2k+1} = 0.$$

It is easy to see that

$$m_{2k} \leq \frac{c^{2k}(2k)!}{2^k k!}\frac{(k!)^2}{(2k)!} = \frac{c^{2k}}{2^k}k!$$

and hence satisfies Carleman's condition. Hence **if** $\{m_k\}$ is a moment sequence, then there is a unique distribution F with these moments. In that case by choosing $g_1 = F^{-1}$, one can get the desired limit. However, it is not clear when it will form a moment sequence.

In particular, the question remains open if the LSD of random Toeplitz matrix and random Hankel matrix (which are not known in closed form) are the LSD of such non-random tri-diagonal matrices.

References

G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-19452-5.

Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9(3):611–677, 1999. ISSN 1017-0405. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.

A. Bose, R. Subhra Hazra, and K. Saha. Patterned random matrices and method of moments. In *Proceedings of the International Congress of Mathematicians. Volume IV*, pages 2203–2231, New Delhi, 2010. Hindustan Book Agency.

A. Bose, R. S. Hazra, and K. Saha. Convergence of joint moments for independent random patterned matrices. *Ann. Probab.*, 39(4):1607–1620, 2011. ISSN 0091-1798. doi: 10.1214/10-AOP597. URL `http://dx.doi.org/10.1214/10-AOP597`.

A. Bose and A. Sen, Another lok at the moment method for large dimensional random matrices. In *Elec. Journal. of Probab.*, Vol. 13, Paper no. 21, 588-628, 2008.

M. Kac, W. L. Murdock, and G. Szegö. On the eigenvalues of certain Hermitian forms. *J. Rational Mech. Anal.*, 2:767–800, 1953.

H. F. Trotter. Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock, and Szegő. *Adv. in Math.*, 54(1):67–82, 1984. ISSN 0001-8708. doi: 10.1016/0001-8708(84)90037-9. URL `http://dx.doi.org/10.1016/0001-8708(84)90037-9`.

Arup Bose, Sreela Gangopadhyay: Statistics and Mathematics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India.
email: bosearu@gmail.com.,gangopadhyay7@gmail.com.
Koushik Saha: Department of Mathematics, Indian Institute of Science, Bangalore 560012, India. email: koushik12@math.iisc.ernet.in.