Analysis and Mathematical Physics (2022) 12:12
https://doi.org/10.1007/s13324-021-00621-1

Macroscale behavior of random lower triangular matrices

J. E. Pascoe[1] · Tapesh Yadav[1]

Received: 15 April 2021 / Revised: 14 August 2021 / Accepted: 5 November 2021 /
Published online: 28 November 2021

Abstract

We analyze the macroscale behavior of random lower (and therefore upper) triangular matrices with entries drawn i.i.d. from a distribution with nonzero mean and finite variance. We show that such a matrix behaves like a probabilistic version of a Riemann sum and therefore in the limit behaves like the Volterra operator. Specifically, we analyze certain SOT-like and WOT-like modes of convergence for random lower triangular matrices to a scaled Volterra operator. We close with a brief discussion of moments.

Keyword Random lower triangular matrices

Mathematics Subject Classification 60B20

1 Introduction

The Wigner semicircle law states that a class of self adjoint random matrices called Wigner matrices go to semicircular element a.s. and in distribution asymptotically. Specifically, if one considers a large random Hermitian matrix with entries drawn i.i.d. from a suitably nice distribution, when we look at the histogram of the eigenvalues we see a semicircular shape, with perhaps one large exceptional eigenvalue. The theory of free probability and random matrix theory give various ways in which we can make this convergence formal [6].

In their breakthrough paper, [2], Dykema and Haagerup studied distribution limits of upper triangular random matrices with i.i.d. complex Gaussian entries having mean

Research supported by National Science Foundation DMS Grant #1953963.

✉ J. E. Pascoe
pascoej@ufl.edu

Tapesh Yadav
tapeshtadav@ufl.edu

1 Department of Mathematics, University of Florida, 1400 Stadium Rd, Gainesville, FL 32611, USA

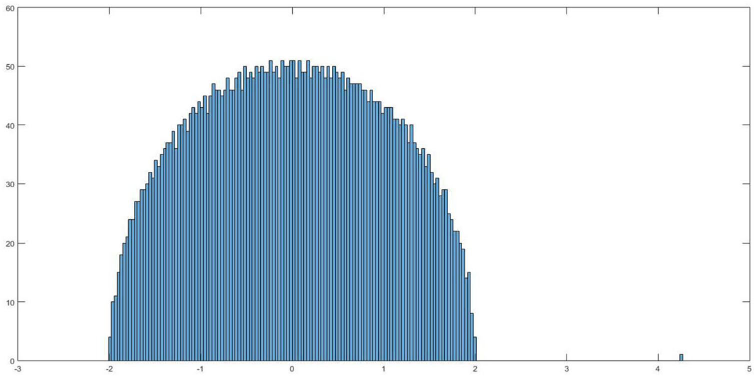

Fig. 1 A histogram of 5000 by 5000 Gaussian self adjoint (real) random matrix (GOE) with non zero mean. Each entry has mean 4/N and standard deviation $\frac{1}{\sqrt{N}}$, N = 5000. Note that there is one exceptional large eigenvalue, while the rest follow a semi-circular distribution

zero and variance c^2/n in the strictly upper triangular part, and i.i.d. random variables distributed according to a compactly supported measure μ on the main diagonal. For these matrices, a limiting non-commuting random variable exists and is called a DT-element or DT-operator. The DT-operators include Voiculescu's circular operator and elliptic deformations of it, as well as the circular free Poisson operators. Star moments of these operators show interesting combinatorial properties, as is explored in [7] by Sniady. Dykema and Haagerup [3] later proved that every DT-operator has a nontrivial, closed, hyperinvariant subspace. Furthermore, every DT-operator generates the von Neumann algebra $L(\mathbb{F}_2)$ of the free group on two generators. Some properties of such matrices, including some properties of the distribution of the singular values, were also independently investigated by Chelotis [1].

Let $X_N = \frac{1}{N}(X_{i,j}^N)_{i,j=1}^N$ denote an $N \times N$ random lower triangular matrix where $X_{i,j}^N$'s are i.i.d. random variables with finite mean μ_N and finite variance σ_N^2. Let T_N be the deterministic lower triangular matrix with each entry being $1/N$, i.e.

$$X_N = \frac{1}{N}\begin{bmatrix} X_{1,1}^N & 0 & \dots & 0 \\ X_{2,1}^N & X_{2,2}^N & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ X_{N,1}^N & X_{N,2}^N & \dots & X_{N,N}^N \end{bmatrix}, T_N = \frac{1}{N}\begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

Numerical experiments give useful information about the singular value distribution for large random matrices. For example, Fig. 2 shows the singular value distribution of $X_N = \frac{\pi}{N}(X_{i,j}^N)_{i,j=1}^N$ for $N = 5000$, where $X_{i,j}^N$ are i.i.d. Bernoulli (0,1) random variables for $1 \leq j \leq i \leq N$ and 0 otherwise. Singular values of X_N for large N behave like the singular value distribution for DT-operators near 0 and like the

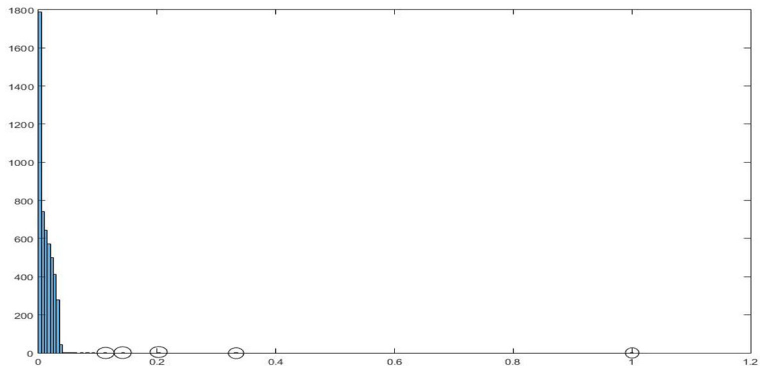

Fig. 2 singular value distribution of $X_N = \frac{\pi}{N}(X_{i,j}^N)_{i,j=1}^N$ for $N = 5000$, where $X_{i,j}^N$ is i.i.d. Bernoulli(0,1) random variable for all $1 \le j \le i \le N$ and 0 otherwise

Volterra operator away from 0. Our current investigation only concerns the asymptotic description of large singular values.

Let V be the **Volterra operator** on $L^2[0, 1]$ defined by

$$V(f)(x) = \int_0^x f(t)dt$$

for all $f \in L^2[0, 1]$. Let $W_N : \mathbb{C}^N \to L^2[0, 1]$ be the isometry taking a vector to a piecewise constant function, as formally defined in Eq. (1).

Theorem 1.1 (SOT-like convergence) Let $\{a_N\}_{N=0}^{\infty} \subset \mathbb{Z}^+$ be a non-negative increasing sequence. Let $\{k(N)\}$ be a sequence of non-negative real numbers such that $\frac{k(N)\sigma_N}{\sqrt{a_N}} \to 0$ and $\sum_{N=1}^{\infty} \frac{1}{k(N)^2} < \infty$. Let $\mu_{a_N} \to \mu$.

Then, for all $u \in L^2[0, 1]$,

$$W_{a_N} X_{a_N} W_{a_N}^*(u) \to \mu V(u) \text{ a.s..}$$

(Here, $W_{a_N}^*$ denotes the adjoint of W_{a_N}).

For instance, we have that $W_{2^N} X_{2^N} W_{2^N}^*(f) \to \mu V(f)$ a.s. for $f \in L^2[0, 1]$ whenever $\mu_N \to \mu$ and the standard deviations are uniformly bounded. We discuss important properties of SOT-like convergence and prove Theorem 1.1 in Sect. 2. The idea is that the T_N act on vector in $\mathbb{C}^N$ consisting of function values taken from equally distanced points in the interval [0, 1] and output the partial sums for that function, which converge to integral of the function as in a Riemann sum. The matrix T_N has singular values similar to Volterra operator, which are $\frac{2}{\pi(2n+1)}$ (see e.g. [4,5]).

We also have a WOT version of this theorem, which requires considerably weaker conditions for convergence.

Theorem 1.2 (WOT-like convergence) *Let* $\{a_N\}_{N=0}^{\infty} \subset \mathbb{Z}^+$ *be a non-negative increasing sequence. Let* $\{k(N)\}$ *be a sequence of non-negative real numbers such that* $\frac{k(N)\sigma_N}{a_N} \to 0$ *and* $\sum_{N=1}^{\infty} \frac{1}{k(N)^2} < \infty$*. Let* $\mu_{a_N} \to \mu$.

Then, for all $u, v \in L^2[0, 1]$,

$$\langle W_{a_N} X_{a_N} W_{a_N}^*(u), v\rangle \to \mu\langle V(u), v\rangle \text{ a.s.}$$

For instance, we can conclude WOT-like convergence along the sequence X_N whenever $\mu_N \to \mu$ and the standard deviations are uniformly bounded (specifically, when there is not enough variance to neccesitate taking a subsequence as in Theorem 1.1). We discuss WOT-like convergence and prove Theorem 1.1 in Sect. 2.

Also, one can remove the '*like*' from the above results to get WOT and SOT convergence if the random matrices X_N under consideration are uniformly bounded in operator norm a.s.. For example, the X_N will be uniformly bounded for Bernoulli $0-1$ random variables with fixed mean and variance.

In the last section, we give moment results for $X_N^* X_N$ for any random matrix with finite moments for each entry and of $N X_N^* X_N$ in the case of non-zero mean. The zero mean case was studied by Dykema and Haagerup [2] where each entry of X_N was Gaussian. We do not see any direct way to generalize their method to matrices X_N with non-Gaussian random variables. Also, in non-zero mean case, as Fig. 2 suggests, we do not get a mean zero spectrum with an exceptional eigenvalue, as is the case in non-zero mean Wigner matrices. Our empirical observations show a superimposition of singular values from the Volterra operator and DT operator.

2 SOT-like convergence

Let

$$B_N = \overline{\text{span}}\{e_1^N, \ldots, e_N^N\} \subseteq L^2[0, 1],$$

where for $1 \le i \le N$, the function $e_i^N = \sqrt{N}\mathbf{1}_{[(i-1)/N, i/N]}$ and $\mathbf{1}_{[(i-1)/N, i/N]}$ is the indicator function of the interval $[(i-1)/N, i/N]$. Note that $\{e_i^N\}_{i=1}^N$ form an orthonormal basis for B_N. We define $W_N : \mathbb{C}^N \to L^2[0, 1]$ by

$$W_N(a_1, \ldots, a_N) = \sum_{i=1}^{n} a_i e_i^N. \tag{1}$$

The map W_N takes $\mathbb{C}^N$ onto B_N isometrically. Note that W_N^* is a partial isometry which sends f to $(\langle f, e_1^N\rangle, \ldots, \langle f, e_N^N\rangle)$.

Let us begin with the following useful lemma.

Lemma 2.1 *Let V be the Volterra operator on $L^2[0, 1]$. Then, $W_N T_N W_N^* \to V$ in SOT.*

Proof Let

$$g_N(x) = W_N T_N W_N^*(f)(x).$$

We first show $g_N \to V(f)$ pointwise for each $f \in C([0,1])$. Without loss of generality, consider a non-negative continuous function $f \in C([0,1])$. There exists $x_i^N \in [(i-1)/N, i/N]$ such that $f(x_i^N) = \sqrt{N}\langle f, e_i^N \rangle$ by the intermediate value theorem. Define $a_N(x) = \min\{i \in \mathbb{N} : x \leq i/N\}$. For fixed $x \in [0,1]$,

$$(1/\sqrt{N}) \sum_{i=1}^{a_N(x)} \langle f, e_i^N \rangle \to \int_0^x f(t)dt$$

as $N \to \infty$. Thus $g_N(x) \to \int_0^x f = V(f)(x)$ pointwise. Therefore

$$g_N(x) \to \int_0^x f(t)dt$$

in $L^2[0,1]$ by the bounded convergence theorem (every function g_N is bounded by the sup norm of f). Hence,

$$\lim_{N\to\infty} ||(W_N T_N W_N^* - V)(f)|| \to 0.$$

Since we obtain convergence for all continuous functions on [0, 1], which are dense in $L^2[0,1]$, and the norms of $\{W_N T_N W_N^*\}$ and V are uniformly bounded by 2, (the Fröbenius norm of T_N, which dominates the operator norm, is given by $\sqrt{\frac{n(n+1)}{2n^2}}$, which is less than 2.), we conclude that

$$||(W_N T_N W_N^* - V)(f)|| \to 0$$

for all $f \in L^2[0,1]$. □

For $u = (u_1, u_2, \ldots, u_N) \in \mathbb{C}^N$, let u^2 denote the vector

$$u^2 = (|u_1|^2, |u_2|^2, \ldots, |u_N|^2) \in \mathbb{C}^N.$$

Lemma 2.2 *Let u be a unit vector in* $\mathbb{C}^N$. *Then*

$$E((\mu_N T_N - X_N)u) = 0,$$

and

$$E((\|(\mu_N T_N - X_N)u\|)^2) \leq \|\frac{\sigma_N^2}{N} T_N u^2\|_1.$$

Proof The first equality is direct. For the second inequality, observe that

$$E((\|(\mu_N T_N - X_N)u\|)^2) = \frac{1}{N^2} E\left(\sum_{i=1}^{N} |\sum_{j=1}^{i} (X_{i,j} - \mu)u_j|^2\right)$$

$$\leq \frac{1}{N^2} E\left(\sum_{i=1}^{N}\sum_{j=1}^{i} |(X_{i,j}-\mu)u_j|^2\right)$$
$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{i} E(|X_{i,j}-\mu|^2)|u_j|^2$$
$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{i} \sigma_N^2 |u_j|^2$$
$$= ||\frac{\sigma_N^2}{N} T_N u^2||_1$$

□

For a non-negative sequence $k(N)$, Chebychev's inequality implies that

$$P\left(\|(\mu_N T_N - X_N)u\| \geq k(N)\sqrt{\left\|\frac{\sigma_N^2}{N} T_N u^2\right\|_1}\right) \leq \frac{1}{k(N)^2}. \tag{2}$$

Therefore, we can finesse our estimate for the standard deviation into a statement about almost sure convergence.

Lemma 2.3 *Let $\{a_N\}_{N=0}^{\infty} \subset \mathbb{Z}^+$ be a non-negative increasing sequence. If there exists positive sequence $\{k(N)\}$ such that $\frac{k(N)\sigma_N}{\sqrt{a_N}} \to 0$ and $\sum_{N=1}^{\infty} \frac{1}{k(N)^2} < \infty$. Then $\|W_{a_N}(\mu_{a_N} T_{a_N} - X_{a_N})W_{a_N}^* u\| \to 0$ a.s. for all $u \in L^2[0, 1]$.*

Proof From Eq. (2), we get that,

$$P\left(\left\|(\mu_{a_N} T_{a_N} - X_{a_N})u\right\| \geq k(N)\sqrt{\left\|\frac{\sigma_N^2}{a_N} T_{a_N} u^2\right\|_1}\right) \leq \frac{1}{k(N)^2}.$$

The right hand side is summable. So, by the first Borel–Cantelli lemma, the probability that the events $\{\|(\mu_{a_N} T_{a_N} - X_{a_N})u\| \geq k(N)\sqrt{\|\frac{\sigma_N^2}{a_N} T_{a_N} u^2\|_1}\}$, occur infinitely often is 0. Observe that for a unit vector u, if u_{a_N} denotes $W_{a_N}^* u$, then u_{a_N} has norm less than or equal to 1 (as W^* is projection). Then $||T_{a_N} u_{a_N}^2||_1 \leq 1$ and hence

$$k(N)\sqrt{\|\frac{\sigma_N^2}{a_N} T_{a_N} u^2\|_1} \leq \frac{k(N)\sigma_N}{\sqrt{a_N}}.$$

So,

$$\|(\mu_{a_N} T_{a_N} - X_{a_N})u_{a_N}\| \leq \frac{k(N)\sigma_N}{\sqrt{a_N}} \text{ eventually a.s..}$$

This gives that

$$\begin{aligned}
&\|(\mu_{a_N} T_{a_N} - X_{a_N}) u_{a_N}\| \to 0 \text{ a.s.} \\
&\implies \|(\mu_{a_N} T_{a_N} - X_{a_N}) W^*_{a_N} u\| \to 0 \text{ a.s.} \\
&\implies \|W_{a_N}(\mu_{a_N} T_{a_N} - X_{a_N}) W^*_{a_N} u\| \to 0 \text{ a.s..}
\end{aligned}$$

This is true for any unit vector u, and hence for any vector in general. □

Proof of Theorem 1.1 Lemma 2.1 along with triangle inequality gives that $\|(\mu V - \mu_{a_N} W_{a_N} T_{a_N} W^*_{a_N})u\| \to 0$ for all $u \in L^2[0, 1]$. Hence,

$$\begin{aligned}
&\|(\mu V - W_{a_N} X_{a_N} W^*_{a_N})u\| \\
&\quad \le \|(\mu V - \mu_{a_N} W_{a_N} T_{a_N} W^*_{a_N})u\| + \|W_{a_N}(\mu_{a_N} T_{a_N} - X_{a_N}) W^*_{a_N} u\| \\
&\quad \to 0 \text{ a.s..}
\end{aligned}$$

□

2.1 Remarks on SOT convergence

(1) The above theorem is rather powerful. For example, if the variance goes to 0 at a rate faster than $\frac{1}{N^\epsilon}$ for some $\epsilon > 0$, then we have guaranteed convergence for any sequence a_N. In particular, for $a_N = N$ which gives $W_N X_N W^*_N(u) \to \mu V(u)$ a.s. (choose $k(N) = N^{\frac{1+\epsilon}{2}}$.).

(2) The sequence $\{k(N)\}$ may not exist in some cases. For example, let $\sigma_N = \sigma$ be constant. Then if $a_N = N$, we do not have any sequence which achieves the goal. This implies that if all the random variables come from the same distribution independent of size of matrix N, then the above theorem cannot guarantee convergence to the Volterra operator for the random matrices.

(3) If norm of the random matrices X_N can be bounded uniformly a.s., then we can conclude true SOT convergence.

An important case for convergence (for $a_N = 2^N$) can be seen in the corollary below.

Corollary 2.4 $\forall u \in L^2[0, 1]$ $\|(\mu V - W_{2^N} X_{2^N} W^*_{2^N})u\| \to 0$ *a.s. whenever for* $j \le i \le N$, $X^N_{i,j}$ *are i.i.d. random variables (independent of N) with mean* μ *and finite variance* σ.

Proof Choose $k(N) = 2^{N/4}$. □

3 WOT-like convergence

Let X_N be as earlier. We have the following variance bound.

Lemma 3.1 *Let u, v be vectors in $\mathbb{C}^N$, then*

$$E(\langle (X_N - \mu_N T_N)u, v\rangle) = 0$$

and

$$E(|\langle (X_N - \mu_N T_N)u, v\rangle|^2) \leq \frac{\sigma_N^2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{i}|v_i|^2|u_j|^2 \leq \frac{\sigma_N^2}{N^2}\|u\|^2\|v\|^2.$$

Proof The first equality is direct. The second inequality is also direct after expanding and using triangle inequality. □

As

$$P\left(|\langle (X_N - \mu_N T_N)u, v\rangle| > k(N)\frac{\sigma_N}{N}\|u\|\|v\|\right) \leq \frac{1}{k(N)^2}, \tag{3}$$

we obtain Theorem 1.2 via a similar argument to the proof of Theorem 1.1.

Equation (3) gives us that unlike the SOT-like case, whenever $\mu_N \to \mu$, we do have $\langle W_N X_N W_N^* u, v\rangle \to \langle \mu V u, v\rangle$ a.s. whenever $\{\sigma_N\}$ is uniformly bounded.

Let X_N be a random lower triangular matrix such that an entry is $\frac{1}{\delta(N)N}$ with probability $\delta(N)$ and 0 otherwise. This gives mean $\mu_N = 1$ and variance $\sigma_N^2 = \frac{1-\delta(N)}{\delta(N)}$. Then,

$$P\left(|\langle (X_N - T_N)u, v\rangle| > k(N)\sqrt{\frac{1-\delta(N)}{\delta(N)N^2}}\|u\|\|v\|\right) \leq \frac{1}{k(N)^2}. \tag{4}$$

3.1 Remarks on WOT convergence

(1) If $\delta(N)$ is bounded below uniformly, then $\langle W_N X_N W_N^* u, v\rangle \to \langle V u, v\rangle$ a.s. (Choose $k(N) = N^{1/2+\varepsilon}$).
(2) If $\delta(N) = N^{-d}$, and $d < 1$, we can show that we still have WOT-like convergence (choose $k(N) = N^{\frac{(3-d)}{4}}$). If $d \geq 1$, Theorem 1.2 cannot guarantee WOT-like convergence.

4 Asymptotic distribution of $X_N^* X_N$ and $N X_N^* X_N$

We will begin with the following observation about the deterministic matrix T_N. For fixed $1 \leq k \leq N$, let $\mathbb{1}_k^N$ be N by N deterministic matrix with entry $(\mathbb{1}_k^N)_{ij} = 1$ if

$i, j \le k$ and 0 otherwise.

$$\mathbb{1}_k^N = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}$$

Lemma 4.1

$$\lfloor N/2 \rfloor^{2n} \le Tr((N^2 T_N^* T_N)^n) \le N^{2n}$$

for all $n \ge 1$.

Proof Basic computations show that

$$N^2(T_N^* T_N) = \sum_{k=1}^{N} \mathbb{1}_k^N. \tag{5}$$

It follows that

$$\begin{aligned} Tr(N^2(T_N^* T_N))^n &= Tr(\sum_{k=1}^{N} \mathbb{1}_k^N)^n \\ &= \sum_{i_1,\dots,i_n=1}^{N} Tr(\mathbb{1}_{i_1}^N \mathbb{1}_{i_2}^N \dots \mathbb{1}_{i_n}^N) \\ &\le \sum_{i_1,\dots,i_n=1}^{N} Tr(\mathbb{1}_N^N)^n \\ &= \sum_{i_1,\dots,i_n=1}^{N} N^n \\ &= N^n(\sum_{i_1,\dots,i_n=1}^{N} 1) \\ &= N^n N^n = N^{2n}. \end{aligned}$$

This gives the upper bound. For lower bound, we observe that we can restrict indices $\lfloor N/2 \rfloor \le i_l \le N$ for all $l = 1, \dots, n$. Under this restriction,

$$Tr(\mathbb{1}_{i_1}^N \mathbb{1}_{i_2}^N \dots \mathbb{1}_{i_n}^N) \ge Tr(\mathbb{1}_{\lfloor N/2 \rfloor}^N)^n = (\lfloor N/2 \rfloor)^n.$$

Then

$$\begin{aligned}\sum_{i_1,\ldots,i_n=1}^{N} Tr(\mathbb{1}_{i_1}^N \mathbb{1}_{i_2}^N \ldots \mathbb{1}_{i_n}^N) &\geq \sum_{i_1,\ldots,i_n=\lfloor N/2 \rfloor}^{N} Tr(\mathbb{1}_{i_1}^N \mathbb{1}_{i_2}^N \ldots \mathbb{1}_{i_n}^N) \\ &\geq (\lfloor N/2 \rfloor)^n (\sum_{i_1,\ldots,i_n=\lfloor N/2 \rfloor}^{N} 1) \\ &\geq (\lfloor N/2 \rfloor)^n (\lfloor N/2 \rfloor)^n \\ &= (\lfloor N/2 \rfloor)^{2n}.\end{aligned}$$

This proves the lower bound. □

Lemma 4.2 *Let $\{X_{ij}^N\}$ be uniformly bounded by constant K a.s.. Then, $tr((X_N^* X_N)^n) \to 0$ a.s as $N \to \infty$ for all $n \geq 1$. (Here, tr denotes the normalized trace. That is, $tr(A) = \frac{1}{N} Tr(A)$ where N is the size of the matrix A.)*

Proof We observe that,

$$tr((X_N^* X_N)^n) \leq tr(K^2 T_N^* T_N)^n \ a.s..$$

By Lemma 4.1, we have that

$$\begin{aligned} & Tr(K^2 T_N^* T_N)^n && \leq K^{2n} \\ \implies & tr(K^2 T_N^* T_N)^n && \leq (K^{2n})/N \\ \implies & tr((X_N^* X_N)^n) && \leq (K^{2n})/N \to 0 \ a.s \text{ as } N \to \infty. \end{aligned}$$

□

Lemma 4.3 *Let $\{X_{ij}^N\}$ be i.i.d. random variables with finite moments. Then, $E[tr((X_N^* X_N)^n)] \to 0$ as $N \to \infty$ for all $n \geq 1$.*

Proof We observe that, after expanding $Tr((N^2 X_N^* X_N)^n)$, there are at most N^{2n} terms of the form $X_{j_1 i_1}^N X_{j_1 i_2}^N \ldots X_{j_n i_1}^N$ for $i_l, j_k \in 1, \ldots, N$ and $l, k \in \{1, \ldots, n\}$. Since the X_{ij} are i.i.d., we get that, for fixed n, the expectation of each term can take values from a finite set of numbers independent of N. For example, it can be $E[X_{11}]^{2n}$, if the pairs (i_l, j_l) and (i_{l+1}, j_l) are all distinct, i.e., every random variable is independent of each other in the term. It can be $E[X_{11}^{2n}]$, if $(i_l, j_l) = (i_{l+1}, j_{l+1})$ for all $l = 1, \ldots, n-1$, i.e., we have the same random variable multiplied $2n$ times. This gives that there are finitely many values that each term in the trace expansion can take. Let M_n be the maximum absolute value in this set. Each term $E|X_{j_1 i_1}^N X_{j_1 i_2}^N \ldots X_{j_n i_1}^N| \leq M_n$ for all $i_l, j_k \in \{1, \ldots, N\}$ and $l, k \in \{1, \ldots, n\}$, independent of N. Since there are at most N^{2n} such terms, we have that $E[Tr((N^2 X_N^* X_N)^n)] \leq M_n N^{2n}$, which gives $E[tr(X_N^* X_N)^n] \leq M_n/N \to 0$. Hence the claim. □

Lemma 4.4 *Let* $\{X_{ij}^N\}$ *be collection of i.i.d. random variables with mean,* $\mu \neq 0$. *Then,* $E[tr((NX_N^* X_N)^n)] \to \infty$ *as* $N \to \infty$ *for all* $n \geq 2$. *For* $n = 1$, $E[tr(NX_N^* X_N)] \to (\sigma^2 + \mu^2)/2$ *as* $N \to \infty$.

Proof If we expand $Tr((NX_N^* X_N)^n)$, we get that each term is of the form, $N^{-n}(X_{j_1 i_1}^N X_{j_1 i_2}^N \ldots X_{j_n i_1}^N)$ for $i_l, j_k \in 1, \ldots, N$ and $l, k \in 1, \ldots, n$. Note that a term equals 0 if $i_l < j_l$ or $i_{l+1} < j_l$. While i_l is free to take any value from $\{1, \ldots, N\}$ and j_l, j_{l-1} are restricted due to that, we can restrict $i_l \geq \lfloor N/2 \rfloor$ for all $l = 1, .., n$. Total such possibilities are at least $(N/2)^n$. Moreover, each j_l is free to take values till $\lfloor N/2 \rfloor$. The number of terms following this constraint are of order $O(N^{2n})$. Also, the number of paths $i_1 \to j_1 \to i_2 \to j_2 ... \to j_n \to i_1$, under the restriction that at least a pair of numbers i_l, j_k is same, is of order $O(N^{2n-1})$. Hence, terms with all distinct random variables $X_{j_1 i_1}^N X_{j_1 i_2}^N \ldots X_{j_n i_1}^N$ grow as $O(N^{2n})$, while the remaining terms grow at $O(N^{2n-1})$. If all of the random variables are distinct, we get that $E[X_{j_1 i_1}^N X_{j_1 i_2}^N \ldots X_{j_n i_1}^N] = \mu^{2n}$. Summing over each such term (the number of such terms is bigger that $K_n N^{2n}$ for some positive K_n) gives that $E[tr((NX_N^* X_N)^n)] \to \infty$ as $N \to \infty$.

For $n = 1$, we know that for any matrix A, $Tr(A^* A)$ is equal to the square sum of its entries. So, $E[tr(NX_N^* X_N)] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^i E[X_{ij}^2] = \frac{N(N+1)/2}{N^2}(\sigma^2 + \mu^2) \to \frac{(\sigma^2+\mu^2)}{2}$ as $N \to \infty$.

□

Acknowledgements We thank the anonymous referee and Ryan Tully-Doyle for help in the revision of this manuscript. We also thank Ryan Tully-Doyle, Austin Jacobs, Roland Speicher and Jamie Mingo for helpful discussions and observations related to this project.

Data availability statement The project did not generate significant data.

Declarations

Conflict of interest We do not believe there are significant conflicts of interest to declare.

References

1. Cheliotis, D.: Triangular random matrices and biorthogonal ensembles. Stat. Probab. Lett. **134**, 36–44 (2018)
2. Dykema, K., Haagerup, U.: DT-operators and decomposability of Voiculescus circular operator Am. J. Math **126**(1), 121–189 (2004)
3. Dykema, K., Haagerup, U.: Invariant subspaces of the quasinilpotent dt-operator. J. Funct. Anal. **209**(2), 332–366 (2004)
4. Halmos, P.: A Hilbert Space Problem Book. Springer-Verlag, New York (1982)
5. Lyubich, Y., Dashdondog, T.: The norms and singular numbers of polynomials of the classical volterra operator in $L^2(0, 1)$. Stud. Math. **199**, 171–184 (2010)
6. Mingo, J.A., Speicher, R.: Free Probability and Random Matrices, volume 35. Springer, (2017)
7. Śniady, P.: Multinomial identities arising from free probability theory. J. Comb. Theor., Ser. A **101**(1), 1–19 (2003)