Singular values of triangular random matrices

Dimitris Cheliotis [1]

October 6, 2022

Abstract

We study the singular values of lower triangular random matrices whose entries on and below the diagonal are i.i.d. complex valued random variables with variance 1. We prove that the empirical distribution of the appropriately scaled squares of the singular values converges to a measure whose moments we identify and also that, with probability 1, the rescaled largest singular value converges to $\sqrt{e}$ under the additional assumption of mean zero and finite fourth moment for the law of the matrix elements. These results have been proved in the past under the assumption that the elements of the matrix are i.i.d. with standard complex Gaussian distribution. Finally, we show how this model is connected with the model of hermitian and triangular with respect to the antidiagonal random matrices, studied by Basu, Bose, Ganguly, and Hazra in 2012.

1 Introduction and statement of the results

1.1 Singular values of random matrices

Singular values of random matrices are of importance in numerical analysis, multivariate statistics, information theory, and the spectral theory of random non-symmetric matrices. See the survey paper [4].

We state in this subsection two of the very basic results concerning singular values of random matrices that are relevant to our work.

Let $\{X_{i,j} : i, j \in \mathbb{N}^+\}$ be i.i.d. complex valued random variables with variance 1, and for $n, m \in \mathbb{N}^+$ consider the $n \times m$ matrix $X(n, m) := (X_{i,j})_{1 \le i \le n, 1 \le j \le m}$. Call $\lambda_1^{n,m} \ge \lambda_2^{n,m} \ge \cdots \ge \lambda_n^{n,m} \ge 0$ the eigenvalues of the Hermitian, positive definite matrix

$$S_{n,m} = \frac{1}{m} X(n, m) X(n, m)^*,$$

and

$$L_{n,m} := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i^{n,m}}$$

[1]Department of Mathematics, University of Athens, Panepistimiopolis 15784, Athens Greece.

MSC2010 Subject Classifications: 60B20, 60F15, 62E15

Keywords: Triangular random matrices, singular values, alternating trees, DT-operators, Lambert function.

Funding: This work was supported by the ERC Advanced Grant 267356-VARIS of Frank den Hollander and also by the Hellenic Foundation for Research and Innovation (H. F. R. I.) [under the “First Call for H. F. R. I. Research Projects to support Facutly members and Researchers and the procurement of high-cost research equipment” grant, Project Number: 1034.]

their empirical distribution. It was shown in [11] that for $c > 0$, with probability 1, as $n, m \to \infty$ so that $n/m \to c$,

$$L_{n,m} \Rightarrow \mathbf{1}_{a \le x \le b} \frac{1}{2\pi x c} \sqrt{(b-x)(x-a)}\, dx + \mathbf{1}_{c>1} \left(1 - \frac{1}{c}\right) \delta_0 \tag{1}$$

(weak convergence of measures) where $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$.

Regarding the largest eigenvalue, $\lambda_1^{n,m}$, it was proved in [9] that if $X_{1,1}$ is real valued and there is $\delta > 0$ satisfying $\mathbf{E}|X_{1,1}|^k \le k^{\delta k}$ for all integers $k \ge 2$, then with probability 1,

$$\lim_{\substack{n\to\infty \\ m/n \to c}} \lambda_1^{n,m} = b. \tag{2}$$

Then Bai and Yin showed in [2] that this convergence takes place under the assumption $\mathbf{E}|X_{1,1}|^4 < \infty$ and that this assumption is necessary for (2) to hold.

The quantity $\sqrt{\lambda_1^{n,m}}$ is the operator norm of the matrix $X(n,m)/\sqrt{m}$ (considered as an operator from $\mathbb{R}^m$ to $\mathbb{R}^n$, both equipped with the Euclidean norm), and it is for this reason that it has been studied in the Analysis literature (see, e.g., [7], [13]).

1.2 Triangular Wigner matrices

In this work, we study the singular values of certain triangular random matrices. The motivation comes from the purely mathematical viewpoint as triangular matrices are ingredients in several matrix decompositions.

Assume as above that $\{X_{i,j} : i, j \in \mathbb{N}^+, i \ge j,\}$ are i.i.d. complex valued random variables with variance 1, and for $n \in \mathbb{N}^+$ let $X(n)$ be the lower triangular $n \times n$ matrix whose (i,j) element is $X_{i,j}$ for $1 \le j \le i \le n$. Call $\lambda_1^{(n)} \ge \lambda_2^{(n)} \ge \cdots \ge \lambda_n^{(n)} \ge 0$ the eigenvalues of the Hermitian matrix

$$S_n = \frac{1}{n} X(n) X(n)^*,$$

and

$$L_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i^{(n)}}$$

their empirical distribution.

Our main results are the following two theorems, which are the analogues of (1) and (2) for the matrices $\{X(n)\}_{n \in \mathbb{N}^+}$.

Theorem 1. *With probability 1, $(L_n)_{n\ge1}$ converges weakly to a deterministic measure μ_0 on $\mathbb{R}$ with moments*

$$\int_{\mathbb{R}} x^k d\mu_0(x) = \frac{k^k}{(k+1)!} \tag{3}$$

for all $k \in \mathbb{N}$.

Theorem 2. *Assume that $X_{1,1}$ has mean 0, variance 1, and finite fourth moment. Then with probabilty 1, $\lim_{n\to\infty} \lambda_1^{(n)} = e$.*

Theorem 1 implies that the measure μ_0 has support $[0, e]$ and comes from a density which can be expressed in terms of the Lambert function (see Theorem 8.9 in [8]). The information on the support of

μ_0 gives that $\underline{\lim}_{n\to\infty} \lambda_1^{(n)} \geq e$ with probability 1. The inequality $\overline{\lim}_{n\to\infty} \lambda_1^{(n)} \leq e$, which is the content of Theorem 2, is not automatic. There are cases where the top eigenvalue has limit striclty larger than the top of the support of the limiting empirical spectral distribution (it is easy to construct examples). The scenarios where the two quantities are equal are treated on a case by case basis, taking into account the peculiarities of the ensemble under study (e.g., see [5] for a general result). The case of triangular matrices is one more model whose combinatorics allow us to prove the equality.

1.3 The triangular hermitian model

Take $(X_{i,j})_{i,j\geq 1}$ i.i.d. complex valued random variables with variance 1, $(Y_i)_{i\geq 1}$ i.i.d. real valued and independent of $(X_{i,j})_{i,j\geq 1}$, and consider the $n \times n$ matrix

$$W_n^u := \begin{pmatrix} Y_1 & X_{1,2} & X_{1,3} & \cdots & X_{1,n-1} & X_{1,n} \\ \overline{X}_{1,2} & Y_2 & X_{2,3} & \cdots & X_{2,n-1} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \overline{X}_{1,n-1} & \overline{X}_{2,n-1} & 0 & \cdots & 0 & 0 \\ \overline{X}_{1,n} & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{4}$$

That is, W_n^u is upper triangular with respect to the antidiagonal and hermitian (thus it has real spectrum). Essentially, this model has been studied in reference [3]. More specifically, that work makes the additional assumptions that $X_{1,2}$ is real valued (thus, W_n^u is symmetric) and has the same distribution as Y_1. In our setting, we have the following.

Theorem 3. *([3]) With probability one, the empirical spectral distribution of the matrix $(1/\sqrt{n})W_n^u$ converges weakly to a measure μ^u symmetric around 0 and whose image under the map $x \mapsto x^2$ is the measure μ_0 of Theorem 1.*

We prove in Section 4 that Theorems 1 and 3 imply one another with the help of a simple argument.

Remark 1. (Related works) The conclusions of Theorems 1 and 2 have been proved in the works [8] and [7] respectively under the assumption that $X_{1,1}$ is standard complex Gaussian random variable. Additionally, the conclusion of Theorem 1 was proved in [10] under the assumption that $X_{1,1}$ is real valued with $\mathbf{E}|X_{1,1}|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$.

Our proofs of Theorems 1 and 2 follow the classical method of moments and path counting used for the analogous theorems for Wigner and sample covariance matrices (see e.g., Chapter 2 in [13]). The crucial ingredient in our analysis is the notion of rooted alternating plane tree, which appears because of the triangular structure of the matrix. Theorems 1, 2, and 3 are proved in Sections 2, 3, and 4 respectively.

2 The limiting empirical spectral distribution

In this and the next section, we will use some notions from graph theory. For us, a *graph* is an ordered triple (V, E, ϕ), where V, E are two sets (called the set of *vertices* and *edges* respectively), and

$$\phi : E \to \{\{x, y\} : x, y \in V\} \tag{5}$$

a map. The interpretation is that $\phi(v)$ gives the two vertices that the edge v connects, also called *ends* of v (see the Appendix of [12]). Such a graph is not directed, and can have several edges with the same ends (multiple edges) and edges with both ends coinciding (loops).

2.1 Proof of Theorem 1

We follow the proof of Theorem 3.7 in [1]. There, all matrix elements are i.i.d., so that everything in that proof transfers to our case (by just replacing all superdiagonal elements with zero) except the computation of the moments of the limiting measure. In particular, the first step of that proof shows that we can assume that $X_{1,1}$ has mean 0 and is bounded. With these additional assumptions, we prove that

$$\lim_{n\to\infty} \mathbf{E}\left\{\int x^k dL_n(x)\right\} = \frac{k^k}{(k+1)!} \tag{6}$$

for all positive integers k. And this will complete the proof. We abbreviate the matrix $X(n)$ to X.
We have

$$\begin{aligned}
\mathbf{E}\left\{\int x^k dL_n(x)\right\} &= \frac{1}{n}\mathbf{E}\,\mathrm{tr}(S_n^k) = \frac{1}{n^{k+1}}\mathbf{E}\,\mathrm{tr}\{(XX^*)^k\} \\
&= \frac{1}{n^{k+1}}\mathbf{E}\Big\{\sum_{1\le i_1,i_2,\dots,i_k\le n} (XX^*)_{i_1,i_2}(XX^*)_{i_2,i_3}\cdots(XX^*)_{i_k,i_1}\Big\} \\
&= \frac{1}{n^{k+1}}\mathbf{E}\Big\{\sum_{\substack{1\le i_1,i_2,\dots,i_k\le n\\ 1\le j_1,j_2,\dots,j_k\le n}} X_{i_1,j_1}X^*_{j_1,i_2}X_{i_2,j_2}X^*_{j_2,i_3}\cdots X_{i_k,j_k}X^*_{j_k,i_1}\Big\} \\
&= \frac{1}{n^{k+1}}\sum_{\substack{1\le i_1,i_2,\dots,i_k\le n\\ 1\le j_1,j_2,\dots,j_k\le n}} \mathbf{E}(X_{i_1,j_1}\overline{X}_{i_2,j_1}X_{i_2,j_2}\overline{X}_{i_3,j_2}\cdots X_{i_k,j_k}\overline{X}_{i_1,j_k}).
\end{aligned} \tag{7}$$

Now for a term with indices $i_1, i_2, \dots, i_k, j_1, j_2, \dots, j_k$, we let $i_{k+1} := i_1$, $\mathbf{i} := (i_1, i_2, \dots, i_k)$, $\mathbf{j} := (j_1, j_2, \dots, j_k)$ and consider the graph $G(\mathbf{i}, \mathbf{j})$ with vertex set

$$V(\mathbf{i}, \mathbf{j}) = \{(1, i_1), (1, i_2), \dots, (1, i_k), (2, j_1), (2, j_2), \dots, (2, j_k)\}$$

(its cardinality is not necessarily $2k$ because of repetitions), set of edges

$$\{(2r-1, \{(1, i_r), (2, j_r)\}), (2r, \{(2, j_r), (1, i_{r+1})\}) : r = 1, 2, \dots, k\},$$

which has cardinality $2k$, and the map ϕ maps $(x, \{y, z\})$ to $\{y, z\}$. We call a vertex of the form $(1, i)$ an I-vertex, and a vertex of the form $(2, i)$ a J-vertex. Note that this graph has no loops since all its edges connect a J-vertex with an I-vertex, which are always different.

From $G(\mathbf{i}, \mathbf{j})$ we generate a simple graph $G_1(\mathbf{i}, \mathbf{j})$ by identifying edges with equal ends (i.e, we remove multiple edges). Formally, $G_1(\mathbf{i}, \mathbf{j})$ has vertex set $V(\mathbf{i}, \mathbf{j})$, edge set

$$\{\{(1, i_r), (2, j_r)\}, \{(2, j_r), (1, i_{r+1})\} : r = 1, 2, \dots, k\},$$

and the map ϕ_1 is the identity map (see Figure 1).
As explained in [1] (in the proof of relation (3.1.6) there, pages 49, 50), the limit as $n \to \infty$ of the quantity in (7) remains the same if we keep only the summands whose indices $(\mathbf{i}, \mathbf{j})$ satisfy the following:

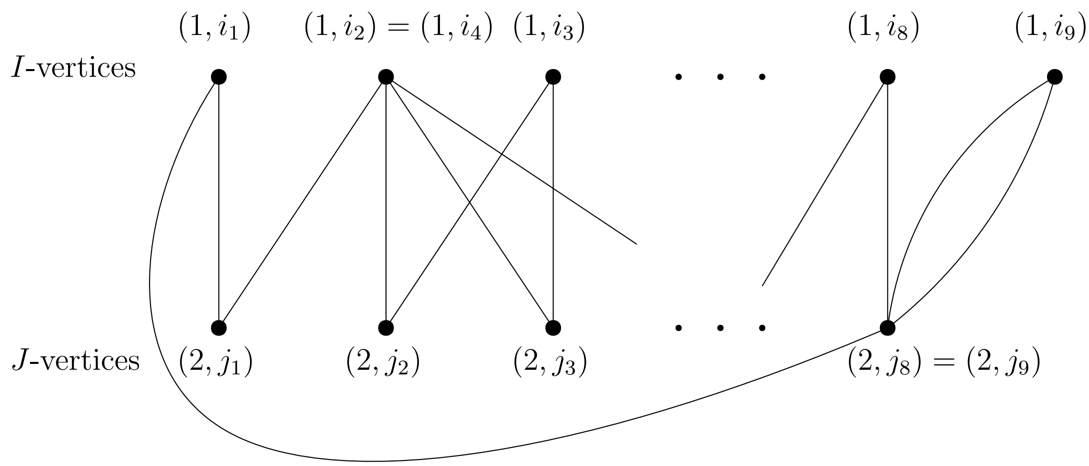

Figure 1: A possible graph $G_1(\mathbf{i}, \mathbf{j})$

(i). The graph $G_1(\mathbf{i}, \mathbf{j})$ is a tree with $k+1$ vertices.

(ii). The path $(1, i_1) \to (2, j_1) \to (1, i_2) \to (2, j_2) \to \cdots \to (1, i_k) \to (2, j_k) \to (1, i_1)$ traverses each edge of the tree exactly twice, in opposite directions of course.

In fact, the pair $(\mathbf{i}, \mathbf{j})$ defines an *ordered tree* (also called *plane tree*), that is, a rooted tree on which we have specified an order among the children of each vertex. The root of the tree is $(1, i_1)$, and among two vertices with common parent, we declare smaller the one whose label appears first in the sequence $(i_1, j_1, i_2, j_2, \ldots, i_k, j_k)$. This order is not related to the way the labels of the vertices compare as real numbers.
In our case, the fact that X is triangular implies that the term corresponding to $(\mathbf{i}, \mathbf{j})$ can be non zero only when the following additional restriction is satisfied:

(iii). $j_1 \le i_1, i_2$ and $j_2 \le i_2, i_3,,...,$ and $j_k \le i_k, i_1$.

That is, each j index is smaller than its two neighbors.

Call $\Delta(n, k)$ the set of pairs of indices $(\mathbf{i}, \mathbf{j})$ with elements from $\{1, 2, \ldots, n\}$ that satisfy (i), (ii), (iii) above, and $\hat{\Delta}(n, k)$ the subset of it for which

(iv). $\{i_1, i_2, \ldots, i_k\} \cap \{j_1, j_2, \ldots, j_k\} = \emptyset$.

Any pair $(\mathbf{i}, \mathbf{j}) \in \hat{\Delta}(n, k)$ satisfies the following strengthening of (iii) above.

$$j_1 < i_1, i_2 \text{ and } j_2 < i_2, i_3, , ..., \text{ and } j_k < i_k, i_1.$$

An example of a pair $(\mathbf{i}, \mathbf{j}) \in \hat{\Delta}(n, k)$ is $((5, 5, 3, 7, 5, 6), (4, 2, 2, 2, 1, 1))$. Figure 2 shows the tree that this defines. The path $i_1 \to j_1 \to i_2 \to j_2 \to \cdots i_k \to j_k \to i_1$ travels the tree from left to right.

Lemma 1. *For positive integers n, k with $n \ge k+1$, it holds*

$$|\hat{\Delta}(n, k)| = \binom{n}{k+1} k^k.$$

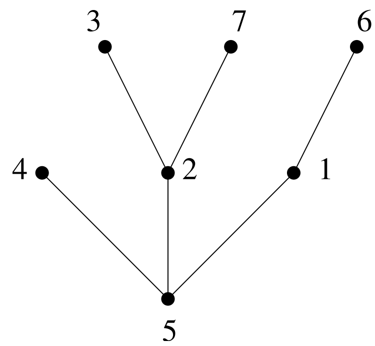

Figure 2: The graph $G_1(\mathbf{i},\mathbf{j})$ corresponding to the pair $(\mathbf{i},\mathbf{j}) = ((5,5,3,7,5,6),(4,2,2,2,1,1))$.

Proof. A tree with r vertices labeled $\{1,2,\dots,r\}$ is called *alternating* if for each path on it, with vertices in order of appearance having labels $v_1, v_2, \dots, v_s$, it holds

$$v_1 < v_2 > v_3 < v_4 > \dots \text{ or}$$
$$v_1 > v_2 < v_3 > v_4 < \dots$$

Denote by pr_2 the projection in the second coordinate, i.e., $\mathrm{pr}_2(x,y) = y$. When $(\mathbf{i},\mathbf{j}) \in \hat{\Delta}(n,k)$, the set $\mathrm{pr}_2(V(\mathbf{i},\mathbf{j}))$, which is $\{i_1, i_2, \dots, i_k, j_1, j_2, \dots, j_k\}$, has cardinality $k+1$ and can take $\binom{n}{k+1}$ values. Take one of them, say $\{1,2,\dots,k+1\}$. The elements $(\mathbf{i},\mathbf{j}) \in \hat{\Delta}(n,k)$ for which $\mathrm{pr}_2(V(\mathbf{i},\mathbf{j})) = \{1,2,\dots,k+1\}$ are in a one to one and onto correspondence with the rooted alternating plane trees with $k+1$ vertices labeled $1,2,\dots,k+1$ and such that the root has label larger than its children. Figure 2 shows the tree corresponding to the pair $(\mathbf{i},\mathbf{j}) = ((5,5,3,7,5,6),(4,2,2,2,1,1))$. The number of such trees equals k^k [It follows from Theorem 1.3 in [6]]. ∎

Lemma 2. *For each $k \in \mathbb{N}^+$,*

$$\lim_{n\to\infty} \frac{1}{n^{k+1}} |\Delta(n,k) \backslash \hat{\Delta}(n,k)| = 0.$$

Proof. The elements of $\Delta(n,k)\backslash\hat{\Delta}(n,k)$ map injectively to the labeled trees with $k+1$ vertices and at most k labels from $\{1,2,\dots,n\}$. The number of such trees is at most

$$\frac{1}{k+1}\binom{2k}{k}\sum_{j=1}^{k}(n)_j < 2^k k n^k.$$

Here, $(n)_j$ denotes the falling factorial. The lemma follows. ∎

As we mentioned above, in the sum in (7), we can ignore the indices $(\mathbf{i},\mathbf{j}) \notin \Delta(n,k)$. Then each $(\mathbf{i},\mathbf{j}) \in \Delta(n,k)$ appears in the sum and the term corresponding to it equals 1 due to the assumptions on the distribution of the X's and property (ii) above. Thus

$$\lim_{n\to\infty} \mathbf{E}\left\{\int x^k dL_n(x)\right\} = \lim_{n\to\infty} \frac{1}{n^{k+1}}|\Delta(n,k)| = \lim_{n\to\infty}\frac{1}{n^{k+1}}|\hat{\Delta}(n,k)| = \frac{k^k}{(k+1)!},$$

using the previous two lemmas, which concludes the proof of the theorem.

Remark 2. We compare this theorem with the special case of the Marchenko-Pastur law [relation (1)] when $m = n$. Then the density in (1) becomes $(2\pi)^{-1}\sqrt{(4-x)/x}\mathbf{1}_{(0,4)}(x)$, and its k-th moment equals $C_k := (k+1)^{-1}\binom{2k}{k}$, the k-th Catalan number, for each $k \in \mathbb{N}$.

The proof of Theorem 1 mimics the proof of that result, and the point that the proofs differ is in the number of terms in (7) that contribute to the limit and each equals 1.
In the Marchenko-Pastur case, that number equals the number of labeled rooted ordered trees with $k+1$ vertices and labels from $\{1, 2, \dots, n\}$, which is

$$(n)_{k+1}C_k \sim n^{k+1}C_k.$$

In the triangular matrix case, that number equals the number of labeled rooted alternating ordered trees with $k+1$ vertices and labels from $\{1, 2, \dots, n\}$, which is

$$\binom{n}{k+1}k^k \sim n^{k+1}\frac{k^k}{(k+1)!}$$

3 The largest eigenvalue. Proof of Theorem 2

Taking k^{th} root and limit as $k \to \infty$ in (3) we get that the supremum of the support of μ_0 is e. And then, the convergence of L_n to μ_0 implies that $\underline{\lim}\,\lambda_1^n \geq e$ with probability one. The aim of this section is to show that $\overline{\lim}\,\lambda_1^n \leq e$ with probability one. Our proof parallels the one of the Bai-Yin theorem as given in Section 2.3 of [13]. The idea is to control a high enough moment of the maximum eigenvalue, and this is accomplished in the next proposition.

Proposition 1. *Assume that $X_{1,1}$ has zero mean and variance 1. Fix $C_1, C_2 > 0$, $\varepsilon \in (0, 1/2)$. There exists positive integer n_0 with the following property. For $n \geq n_0$, if the support of $|X_{1,1}|$ is contained in $[0, C_1 n^{1/2-\varepsilon}]$ and k is an integer with $1 \leq k \leq C_2 \log^2 n$, then*

$$\mathbf{E}\,\mathrm{tr}\{(X(n)X(n)^*)^k\} \leq 2e^k n^{k+1}. \tag{8}$$

Proof. Let $d_n := C_2 \log^2 n$. Pick n_0 so that for all $n \geq n_0$ it holds

$$(1 + 2C_1^2)(4d_n^5)^7 < n^{2\varepsilon}. \tag{9}$$

Take n, k as in the statement of the proposition. As in (7), we write

$$\begin{aligned}
\mathbf{E}\,\mathrm{tr}\{(X(n)X(n)^*)^k\} &= \sum_{\substack{1\leq i_1,i_2,\dots,i_k\leq n \\ 1\leq j_1,j_2,\dots,j_k\leq n}} \mathbf{E}(X_{i_1,j_1}\overline{X}_{i_2,j_1}X_{i_2,j_2}\overline{X}_{i_3,j_2}\cdots X_{i_k,j_k}\overline{X}_{i_1,j_k}) \\
&= \sum_{\mathbf{i},\mathbf{j}} \mathbf{E}(X_{G(\mathbf{i},\,\mathbf{j})}) \\
&\leq \sum_{\mathbf{i},\mathbf{j}} \mathbf{E}(|X_{G(\mathbf{i},\,\mathbf{j})}|),
\end{aligned} \tag{10}$$

where to the pair $(\mathbf{i},\,\mathbf{j}) = ((i_1, i_2, \dots, i_k), (j_1, j_2, \dots, j_k))$ of indices, we correspond the graph $G(\mathbf{i}, \mathbf{j})$ as in Subsection 2.1 and the term $X_{G(\mathbf{i},\,\mathbf{j})} := X_{i_1,j_1}\overline{X}_{i_2,j_1}X_{i_2,j_2}\overline{X}_{i_3,j_2}\cdots X_{i_k,j_k}\overline{X}_{i_1,j_k}$.

In the sum (10), we isolate the pairs $(\mathbf{i},\,\mathbf{j}) \in \hat{\Delta}(n,k)$. We call these pairs *good*, and the rest, *bad*. The contribution of the good pairs to the sum is

$$\binom{n}{k+1} k^k = n^{k+1} \frac{k^k}{(k+1)!} < n^{k+1} e^k.$$

The inequality follows by the series expansion of e^k.

Now we need to bound the contribution of the bad pairs to (10). Take such a pair $(\mathbf{i},\,\mathbf{j})$. The path

$$(1, i_1) \to (2, j_1) \to (1, i_2) \to (2, j_2) \to \cdots (1, i_k) \to (2, j_k) \to (1, i_1) \tag{11}$$

is a cycle that traverses the graph $G_1(\mathbf{i}, \mathbf{j})$. List the edges $e_1, e_2, \dots, e_s$ of $G_1(\mathbf{i}, \mathbf{j})$ in order of appearance in the cycle, and call $a_1, a_2, \dots, a_s$ their multiplicities in the cycle. That is, a_q is the number of times the (undirected) edge e_q appears in the cycle. If any of these multiplicities is 1, we have $\mathbf{E}(X_{G(\mathbf{i},\,\mathbf{j})}) = 0$. We assume therefore that all are at least 2. Using the information about the mean, variance, and support of $|X_{1,1}|$, we get that for any integrer $a \geq 2$ it holds $\mathbf{E}(|X_{1,1}|^a) \leq (C_1 n^{1/2-\varepsilon})^{a-2}$. Thus

$$\mathbf{E}(|X_{G(\mathbf{i},\,\mathbf{j})}|) \leq \prod_{i=1}^{s} \mathbf{E}|X_{e_i}|^{a_i} \leq (C_1 n^{1/2-\varepsilon})^{a_1+\cdots+a_s-2s} = (C_1 n^{1/2-\varepsilon})^{2k-2s}. \tag{12}$$

Cycles that are generated by bad pairs we call them *bad cycles*. For integers $s \geq 1$ and $a_1, \dots, a_s \geq 2$, let

$$N_{a_1,a_2,\dots,a_s} = \text{the number of bad cycles whose edges have multiplicities } a_1, a_2, \dots, a_s. \tag{13}$$

The contribution of the bad pairs to (10) is at most

$$\sum_{s=1}^{k} (C_1 n^{1/2-\varepsilon})^{2k-2s} \sum_{a_1,a_2,\dots,a_s} N_{a_1,a_2,\dots,a_s}. \tag{14}$$

Now, using the bound on $N_{a_1,a_2,\dots,a_s}$, obtained in Lemma 3 below, we get that the expression in (14) is at most

$$e^k (4k^4)^5 \sum_{s=1}^{k} (C_1 n^{1/2-\varepsilon})^{2k-2s} (4k^4)^{2(k-s)} n^{\min\{s+1,k\}} \sum_{a_1,a_2,\dots,a_s} 1. \tag{15}$$

The inside sum is over all s-tuples of integers greater than or equal to 2 with sum $2k$. By subtracting 2 from each a_i, we get an s-tuple of non-negative integers with sum $2k-2s$. The number of such s-tuples is $\left(\!\binom{s}{2k-2s}\!\right)$ (combinations with repetition), which is at most $s^{2(k-s)} \leq k^{2(k-s)}$. Separating the $s = k$ term, and letting $w = k - s$, we get for (15) the bound

$$e^k (4k^4)^5 \left\{ n^k + n^{k+1} \sum_{w=1}^{k-1} \left(\frac{C_1^2 (4k^5)^2}{n^{2\varepsilon}} \right)^w \right\}.$$

By the choice of n_0 [see (9)] and since $k \leq d_n$, we have $C_1^2 (4k^5)^2 / n^{2\varepsilon} < 1/2$, and thus the sum in the last expression is bounded by $2C_1^2 (4k^5)^2 / n^{2\varepsilon}$. We conclude that the contribution of the bad pairs to the sum (10) is at most

$$e^k (4k^4)^5 \left(n^k + n^{k+1-2\varepsilon} 2C_1^2 (4k^5)^2 \right) = e^k n^{k+1} \left(\frac{(4k^4)^5}{n} + \frac{2C_1^2 (4k^5)^7}{n^{2\varepsilon}} \right) < e^k n^{k+1}.$$

In the last inequality, we used the inequalities $2\varepsilon < 1, k \leq d_n$, and (9). This finishes the proof of the proposition. ∎

Now Theorem 2 follows by adapting the arguments of Theorems 2.3.23, 2.3.24 (the Bai-Yin Theorem) in [13]. It is in those theorems that the finite fourth moment assumption is used. In the rest of the section, we prove the crucial estimate we invoked in the proof above in order to bound (14). We adopt and present the terminology of Section 5.1.1 of [1].

Lemma 3. $N_{a_1,a_2,\ldots,a_s} \le e^k (4k^4)^{2(k-s)+5} n^{\min\{s+1,k\}}$.

Proof. Take a cycle as in (11), assume that it has edge multiplicities $a_1, a_2, \ldots, a_s \ge 2$, and label the vertices as

$$v_1 \to v_2 \to \cdots \to v_{2k} \to v_{2k+1} = v_1. \tag{16}$$

Each step in the cycle we call a *leg*. More formally, legs are the elements of the set $\{(a,(v_a,v_{a+1})) : a = 1,2,\ldots,2k\}$, which are exactly the edges of $G(\mathbf{i},\mathbf{j})$ if we replace (v_a,v_{a+1}) with $\{v_a,v_{a+1}\}$.

For $1 \le a < b$, we say that the leg $(a,(v_a,v_{a+1}))$ is *single* up to b if $\{v_a,v_{a+1}\} \ne \{v_c,v_{c+1}\}$ for every $c \in \{1,2,\ldots,b-1\}, c \ne a$. We classify the $2k$ legs of the cycle into 4 sets T_1,T_2,T_3,T_4. The leg $(a,(v_a,v_{a+1}))$ belongs to

T_1: if $v_{a+1} \notin \{v_1,\ldots,v_a\}$. I. e., the leg leads to a new vertex.

T_3: if there is a T_1 leg $(b,(v_b,v_{b+1}))$ with $b < a$ so that $a = \min\{c > b : \{v_c,v_{c+1}\} = \{v_b,v_{b+1}\}\}$.
I. e., at the time of its appearence, it increases the multiplicity of a T_1 edge of $G(\mathbf{i},\mathbf{j})$ from 1 to 2.

T_4: if it is not T_1 or T_3.

T_2: if it is T_4 and there is no $b < a$ with $\{v_a,v_{a+1}\} = \{v_b,v_{b+1}\}$.
I.e., at the time of its appearence, it creates a new edge but leads to a vertex that has appeared already.

Moreover, a T_3 leg $(a,(v_a,v_{a+1}))$ is called *irregular* if there is exactly one T_1 leg $(b,(v_b,v_{b+1}))$ which has $b < a$, $v_a \in \{v_b,v_{b+1}\}$, and is single up to a. Otherwise the leg is called *regular*.
It is immediate that a T_4 leg is one of the following three kinds.

a) It is a T_2 leg.

b) Its appearance increases the multiplicity of a T_2 edge from 1 to 2.

c) Its edge marks the third or higher order appearance of an edge.

The number of edges of $G_1(\mathbf{i},\mathbf{j})$ is s. Let also

t: the number of vertices of $G_1(\mathbf{i},\mathbf{j})$.

ℓ: the number of edges of $G_1(\mathbf{i},\mathbf{j})$ that have multiplicity at least 3.

m: the number of T_2 legs.

r: the number of regular T_3 legs.

We have for r, t, and $|T_4|$ the following bounds

$$r \le 2m, \tag{17}$$

$$t = s + 1 - m \le k, \tag{18}$$

$$|T_4| = 2m + 2(k-s). \tag{19}$$

The first relation is Lemma 5.6 in [1]. The second is true because if we remove the m edges traveled by T_2 legs, we get a tree with $s - m$ edges and t vertices, and in any tree the number or vertices equals the number of edges plus one. Then the inequality is true because $s \le k$ (all edges of $G(\mathbf{i},\mathbf{j})$ have multiplicity at least 2) and if $s = k$ then $m \ge 1$ since the cycle is bad. For the last relation, note that $|T_3| = |T_1| = t-1$ and thus, using (18) too, we have $|T_4| = 2k - 2(t-1) = 2k - 2(s-m)$.

Now back to the task of bounding $N_{a_1,\dots,a_s}$. We fix a cycle as in the beginning of the proof and give each vertex an *index* in $\{1, 2, \dots, t\}$ which records the order of the first appearance of the vertex in the cycle. Then, we record

- for each T_4 leg, a) its order in the cycle, b) the index of its initial vertex, c) the index of its final vertex, and d) the index of the final vertex of the next leg in case that leg is T_1. This gives a $Q_1 \subset \{1, 2, \dots, 2k\} \times (\{1, 2, \dots t\}^2 \cup \{1, 2, \dots t\}^3)$ with $|T_4|$ elements.

- for each regular T_3 leg, a) its order in the cycle, b) the index of its initial vertex, and c) the index of its final vertex. This gives a $Q_2 \subset \{1, 2, \dots, 2k\} \times \{1, 2, \dots, t\}^2$ with r elements.

- the index of each J-vertex, say $(2, j)$, for which $j \in \{i_1, i_2, \dots, i_k\}$. This gives a $Q_3 \subset \{1, 2, \dots t\}$. Let $u := |Q_3|$.

- the set $Q_4 := \{i_1, i_2, \dots i_k, j_1, j_2, \dots, j_k\} \subset \{1, 2, \dots, n\}$, which has $t - u$ elements.

- $\hat{G}(\mathbf{i},\ \mathbf{j})$, a rooted, ordered, alternating tree with t vertices and labels $\{1, 2, \dots, t\}$ which we proceed to define.

We call U the set of all indices that appear as fourth coordinate in elements of Q_1. These are indices of final vertices of T_1 legs.

The graph $\hat{G}(\mathbf{i},\ \mathbf{j})$ is defined as follows.

A. Vertices, edges, root, ordering:
   In $G_1(\mathbf{i},\ \mathbf{j})$, we erase edges that were traveled by T_2 legs in $G(\mathbf{i},\ \mathbf{j})$. We thus get a simple graph $\hat{G}(\mathbf{i},\ \mathbf{j})$ which is a spanning tree of $G(\mathbf{i},\ \mathbf{j})$. Indeed, it has the same set of vertices as $G(\mathbf{i},\ \mathbf{j})$ and is connected because the edges we erased connect vertices that were already connected by a different route in $G(\mathbf{i},\ \mathbf{j})$. And it is a tree because if there were a simple cycle in it, we would be able to find in it an edge traveled by a T_2 leg of $G(\mathbf{i},\ \mathbf{j})$, which is false. We declare as root of the tree the vertex $(1, i_1)$ and we order the children of each vertex in accordance with the ordering of their indices.

B. Labels:
   Let $f : Q_4 \to \{1, 2, \dots, t - u\}$ be the unique strictly increasing map between these two sets. To each vertex (a, b) of the tree $\hat{G}(\mathbf{i},\ \mathbf{j})$ (recall that $a \in \{1, 2\}$) we assign the label $f(b)$. If $u > 0$, then we do a final relabeling. If for a J-vertex $v := (2, j)$ the I-vertex $(1, j)$ is present in the graph $\hat{G}(\mathbf{i},\ \mathbf{j})$, then we increase by one the label of every vertex (of the I or J kind) which has at the moment label $\ge f(j)$ except v (after that, $f(j)$ will be different from all labels of I-vertices). We do this procedure sequentially by checking whether the above scenario happens at a j with $f(j) = 1$, and continuing upward for the values $2, 3, \dots, t - u$. In the end, no two vertices will have the same label, and the set of labels will be $\{1, 2, \dots, t\}$.

Step A gives an ordered, rooted tree of t vertices, and Step B, together with the property

$$j_1 \le i_1, i_2 \text{ and } j_2 \le i_2, i_3, , \dots, \text{ and } j_k \le i_k, i_1$$

that the indices have, gives that the labeling is alternating (see definition in the proof of Lemma 1) with the root having label larger than its children. We denote by $\hat{G}(\mathbf{i},\mathbf{j})$ the resulting rooted, labeled tree.

We claim that, having $Q_1, Q_2, Q_3, Q_4, \hat{G}(\mathbf{i},\mathbf{j})$ we can reconstruct the cycle (16). We do this in three steps.

Step 1. We determine what kind each leg of the cycle is and what the index of its initial and its final vertex is. These data are known for the T_4 and T_3 regular legs. The remaining legs are T_1 or T_3 irregular. We discover the nature of each of them by traversing the cycle from the beginning as follows. The first leg is T_1 since the graph $G(\mathbf{i},\mathbf{j})$ does not have loops (each of its edges connects an I-vertex with a J-vertex). Assume that we have arrived at a vertex v_i in the cycle with the smallest i for which the nature of the leg $\ell_i := (i, (v_i, v_{i+1}))$ is not known yet. If the vertex v_i has no children in $\hat{G}(\mathbf{i},\mathbf{j})$ that we haven't encountered up to the leg ℓ_{i-1}, then ℓ_i is T_3 irregular. If the vertex v_i does have such children, call z the oldest among them (that is, the one that appears earlier in the cycle).

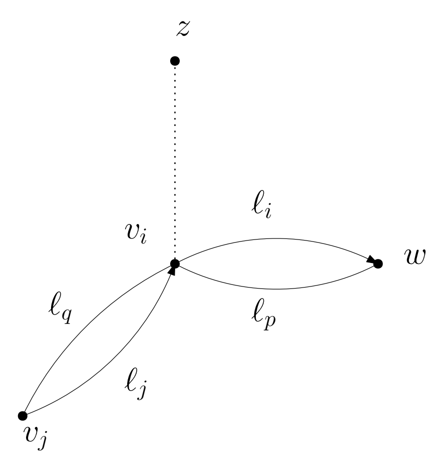

Figure 3: The case $z \notin U$.The legs $\ell_i, \ell_j (i < j)$ are T_3, while ℓ_p, ℓ_q are T_1.

• If $z \in U$, then in case it was included in Q_1 because of ℓ_{i-1} (and we have the date to check this), we have that ℓ_i is T_1 with $v_{i+1} = z$, while in case it was included with a leg $\ell_{i'}$ with index $i' \geq i$, we have that ℓ_i can't be T_1 (because then v_{i+1} would be a child of v_i appearing earlier than z, contradicting the choice of z), thus ℓ_i is T_3 irregular.

• If $z \notin U$, we will show that $\ell_i = (i, (v_i, w))$ is T_1. Assume on the contrary that it is T_3 irregular. Clearly $z \neq w$, and call ℓ_p $(p < i)$ the T_1 leg that has vertices v_i, w and is single up to $i-1$. The cycle will visit the vertex v_i at a later point, with a leg $\ell_j = (j, (v_j, v_i))$ with $j > i$ and $v_j \neq z, v_j \neq v_i$, in order to create the edge that connects v_i with z (that is, $\ell_{j+1} = (j+1, (v_i, z))$ will be T_1), see Figure 3. The leg ℓ_j is not T_1 because v_i has been visited by an earlier leg, and it is not T_4 because we assumed that $z \notin U$. It has then to be T_3. Thus, there is a leg ℓ_q connecting vertices v_i, v_j that is T_1.

If $q < i$, then we consider two cases. If $v_j = w$, then ℓ_j is T_4, because the edge v_i, w has been traveled already by ℓ_p, ℓ_i, and this would force $z \in U$, a contradiction. If $v_j \neq w$, then ℓ_i would have been T_3 regular as there are at least two T_1 legs (ℓ_p, ℓ_q) with order less than i with one vertex v_i, traveling different edges, and single up to $i-1$, again a contradiction because ℓ_i is T_1 or T_3 irregular.

If $q > i$, then $v_j (\neq z)$ is a child of v_i (that is, the T_1 leg ℓ_q goes from v_i to v_j) that appears after leg ℓ_i but earlier than z, which contradicts the definition of z. We conclude that ℓ_i is T_1.

Thus, having $\hat{G}(\mathbf{i},\mathbf{j}), Q_1, Q_2$ allows to determine the index of the initial and final vertex of all legs, and the only thing remaining for the recovery of all the data of the cycle (16) is the elements i_r, j_r in the legs. This is determined in the next two steps.

Step 2. In the tree $\hat{G}(\mathbf{i},\mathbf{j})$, we undo the second part of the label assignment procedure. I.e., we visit one after another the J vertices of $\hat{G}(\mathbf{i},\mathbf{j})$ that have indices in Q_3, and we decrease by one the index of every vertex (I or J) with index strictly bigger than the one of the vertex we visit. This gives a labeled tree with label set $\{1,2,\dots,t-u\}$, which is $\hat{G}(\mathbf{i},\mathbf{j})$ with its initial labeling.

Step 3. In the labeled tree obtained in the previous step, we change the label i to $g(i)$, where $g:\{1,2,\dots,t-u\}\to Q_4$ is the unique strictly increasing map between these two sets. This reveals the labels i_r, j_r in each element of the cycle (16).

The above imply that the number of bad cycles with given t,u,r is at most

$$(2kt^2(t+1))^{|T_4|}(2kt^2)^r t^u \binom{n}{t-u}(t-1)^{t-1} \le \frac{n^{t-u}}{(t-u)!}(t-1)!\, e^{t-1}(4k^4)^{r+|T_4|}t^u \tag{20}$$

$$\le e^{t-1}n^t\left(\frac{t^2}{n}\right)^u (4k^4)^{r+|T_4|}. \tag{21}$$

We used the bounds $t\le k, (t-1)^{t-1}\le (t-1)!e^{t-1}, (t-1)!/(t-u)!\le t^u$. By the choice of n_0 [see (9)], we have $t^2<n$. Moreover, using (18) and

$$r+|T_4|\le 4m+2(k-s), \tag{22}$$

which follows from (17) and (19), we bound the expression in (21) by

$$\begin{aligned} e^k n^{s+1-m}(4k^4)^{4m+2(k-s)} &= e^k(4k^4)^{2(k-s)}n^{s+1}\left(\frac{(4k^4)^4}{n}\right)^m \\ &\le e^k(4k^4)^{2(k-s)}n^{s+1}\left(\frac{(4k^4)^4}{n}\right)^{\mathbf{1}_{m\ge 1}} \le e^k(4k^4)^{2(k-s)+4}n^{\min\{s+1,k\}}. \end{aligned} \tag{23}$$

The last inequality is true if $s<k$ because $(4k^4)^4/n<1$ (by the choice of n_0), while if $s=k$, it is again valid because then we additionally have $m\ge 1$ as the cycle is bad. Recall that $s\le k$ always.

Summing the bound (23) over all possible values of t,r,u, which are no more than k for each (regarding r, note that $r\le |T_3|=t-1$), we get the claim of the lemma. ■

4 The triangular hermitian model

In this section, we prove that Theorem 1 implies Theorem 3. Similar arguments show the reverse implication. The crucial element is equality (26).

We repeat all the steps of the classical proof of the semicircle law through the moment method (see, .e.g., [1] Theorem 2.5) and the only point that differs is in the computation of the limit

$$\lim_{n\to\infty}\frac{1}{n}\mathbf{E}\,\mathrm{tr}\left\{\left(\frac{W_n^u}{\sqrt{n}}\right)^r\right\} \tag{24}$$

for each $r\in\mathbb{N}^+$ (by that point of the proof, one shows that we can assume that the distribution of $X_{1,2}$ has compact support and the distribution of Y_1 is concentrated at zero). For r odd, it is easy to see that this limit is zero, while for r even, say $r=2k$, we will show that

$$\lim_{n\to\infty}\frac{1}{n^{k+1}}\mathbf{E}\,\mathrm{tr}\{(W_n^u)^{2k}\} = \int x^k\, d\mu_0(x). \tag{25}$$

By the change of variables formula, the last integral equals $\int x^{2k}\, d\mu^u(x)$, and all the above suffice to prove that the limiting measure is μ^u.

To prove (25), we use the random variables $(X_{i,j})_{i,j\ge 1}$ that define W_n^u and define the matrix

$$U_n := (X_{i,j}\mathbf{1}_{i+j\le n+1})_{1\le i,j\le n}.$$

This is the non-hermitian analogue of W_n^u.
Claim: For each $k\in\mathbb{N}^+$, we have

$$\lim_{n\to\infty}\frac{1}{n^{k+1}}\mathbf{E}\,\mathrm{tr}\{(W_n^u)^{2k}\} = \lim_{n\to\infty}\frac{1}{n^{k+1}}\mathbf{E}\,\mathrm{tr}\{(U_nU_n^*)^k\}. \tag{26}$$

Proof of the claim: For simplicity, we denote W_n^u and U_n by W and U respectively. The expectations in (26) equal

$$\mathbf{E}\,\mathrm{tr}\{(WW)^k\} = \sum_{\substack{1\le i_1,i_2,\dots,i_k\le n\\ 1\le j_1,j_2,\dots,j_k\le n}} \mathbf{E}(W_{i_1,j_1}W_{j_1,i_2}W_{i_2,j_2}W_{j_2,i_3}\cdots W_{i_k,j_k}W_{j_k,i_1}), \tag{27}$$

$$\mathbf{E}\,\mathrm{tr}\{(UU^*)^k\} = \sum_{\substack{1\le i_1,i_2,\dots,i_k\le n\\ 1\le j_1,j_2,\dots,j_k\le n}} \mathbf{E}(U_{i_1,j_1}(U^*)_{j_1,i_2}U_{i_2,j_2}(U^*)_{j_2,i_3}\cdots U_{i_k,j_k}(U^*)_{j_k,i_1}). \tag{28}$$

Each index set of i's and j's defines a graph with vertices $\{i_1,j_1,i_2,j_2,\dots,i_k,j_k\}$ and edges $\{(i_r,j_r),(j_r,i_{r+1}) : r=1,2,\dots,k\}$, where $i_{k+1}:=i_1$. Let $\mathbf{i}:=(i_1,i_2,\dots,i_k), \mathbf{j}:=(j_1,j_2,\dots,j_k)$ and call $\Gamma(\mathbf{i},\mathbf{j})$ the aforementioned graph. As is well known, the limits in (26) stay unaffected if we keep in the sums (27), (28) only the terms for which $\Gamma(\mathbf{i},\mathbf{j})$ is a tree with $k+1$ vertices. Then the path $i_1\to j_1\to i_2\to j_2\to\cdots\to i_k\to j_k\to i_1$ traverses each edge of the tree exactly twice. For each $r\in\{1,\dots,k\}$ there is exactly one $s(r)\in\{1,\dots,k\}$ such that $(i_r,j_r)=(i_{s(r)+1},j_{s(r)})$ and $(i_r,j_r)\ne(i_{r'},j_{r'})$ for all $r'\ne r$, so that the contribution of the pair $(\mathbf{i},\mathbf{j})$ to both sums is the same because

$$\mathbf{E}(W_{i_1,j_1}W_{j_1,i_2}W_{i_2,j_2}W_{j_2,i_3}\cdots W_{i_k,j_k}W_{j_k,i_1}) = \prod_{r=1}^k \mathbf{E}(W_{i_r,j_r}W_{j_{s(r)},i_{s(r)+1}}) \tag{29}$$

$$= \prod_{r=1}^k \mathbf{E}(W_{i_r,j_r}W_{j_r,i_r}) = \prod_{r=1}^k \mathbf{E}(|W_{i_r,j_r}|^2) = 1 \tag{30}$$

$$= \prod_{r=1}^k \mathbf{E}\{U_{i_r,j_r}(U^*)_{j_{s(r)},i_{s(r)+1}}\} = \mathbf{E}\{U_{i_1,j_1}(U^*)_{j_1,i_2}U_{i_2,j_2}(U^*)_{j_2,i_3}\cdots U_{i_k,j_k}(U^*)_{j_k,i_1}\}. \tag{31}$$

This proves the claim.
Now, the claim implies (25) as follows. Call P_n the $n\times n$ permutation matrix that has ones in the antidiagonal, i.e., $(P_n)_{i,j}=\mathbf{1}_{i+j=n+1}$. Then $\tilde X(n) := P_nU_n$ is of the form of the matrix $X(n)$ of subsection 1.2 and $\tilde X(n)\tilde X(n)^*$ and $U_nU_n^*$ are similar. It follows (see (6), (7)) that the right hand side of (26) equals $k^k/(k+1)!$, which is $\int x^k\, d\mu_0(x)$.

References

[1] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer Series in Statistics. Springer, New York, second edition, 2010. URL http://dx.doi.org/10.1007/978-1-4419-0661-8.

[2] Z. Bai and Yin (1988). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix. *The Annals of Probability*, 16(4):1729–1741.

[3] R. Basu, A. Bose, S. Ganguly, and R. S. Hazra (2012). Spectral properties of random triangular matrices. *Random Matrices: Theory and Applications*, 1(03):1250003.

[4] D. Chafaï (2009). Singular values of random matrices. *Lecture Notes.* Available at author's website.

[5] D. Chafaï and K. Tikhomirov (2018). On the convergence of the extremal eigenvalues of empirical covariance matrices with dependence. *Probability Theory and Related Fields*, 170(3):847–889.

[6] C. Chauve, S. Dulucq, and A. Rechnitzer (2001). Enumerating alternating trees. *J. Combin. Theory Ser. A*, 94(1):142–151.

[7] B. Collins, P. Gawron, A. E. Litvak, and K. Życzkowski (2014). Numerical range for random matrices. *Journal of Mathematical Analysis and Applications*, 418(1):516–533.

[8] K. Dykema and U. Haagerup (2004). DT-operators and decomposability of Voiculescu's circular operator. *American journal of mathematics*, 126(1):121–189.

[9] S. Geman (1980). A limit theorem for the norm of random matrices. *The Annals of Probability*, 8(2): 252–261.

[10] A. Lytova and L. Pastur (2015). On a limiting distribution of singular values of random band matrices. *arXiv preprint arXiv:1503.04353*.

[11] V. A. Marchenko and L. A. Pastur (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536.

[12] R. P. Stanley. *Enumerative Combinatorics, Volume 1, second edition.* 2011.

[13] T. Tao. *Topics in random matrix theory*, volume 132. AMS Bookstore, 2012.