# Triangular random matrices and biorthogonal ensembles

Dimitris Cheliotis [1]

April 18, 2014

**Abstract**

We study the singular values of certain triangular random matrices. When their elements are i.i.d. standard complex Gaussian random variables, the squares of the singular values form a biorthogonal ensemble, and with an appropriate change in the distribution of the diagonal elements, they give the biorthogonal Laguerre ensemble. For triangular Wigner matrices, we give alternative proofs for the convergence of the empirical distribution of the appropriately scaled squares of the singular eigenvalues to a distribution with support $[0, e]$, as well as for the almost sure convergence of the rescaled largest singular eigenvalue to $\sqrt{e}$ under the additional assumption of mean zero and finite fourth moment for the law of the matrix elements.

## 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 Chafaï (2009).

We state in this subsection three 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 eigenvalues, alternating trees, determinantal processes, biorthogonal ensembles, DT-operators, Lambert function.

their empirical distribution. It was shown in Marchenko and Pastur (1967) that for $c > 0$, with probability 1, as $n, m \to \infty$ so that $n/m \to c$, $L_{n,m}$ converges weakly to the measure

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

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

Regarding the largest eigenvalue, it was proved in Geman (1980) under certain moment assumptions, that with probability 1, $\lambda_1^{n,m}$ converges to $b$ as $n, m \to \infty$. Then Bai and Yin (1988) showed that this convergence takes place under the assumption that $|X_{1,1}|$ has finite fourth moment and that this assumption is necessary for the validity of the conclusion.

When the $X_{i,j}$ follow the standard complex Gaussian distribution and $n \le m$, the vector $(\lambda_1^{n,m}, \lambda_2^{n,m}, \ldots, \lambda_n^{n,m})$ has density with respect to Lebesgue measure in $\mathbb{R}^n$ which is

$$\frac{1}{\prod_{k=1}^n \Gamma(m-n+k)\Gamma(k)} e^{-\sum_{k=1}^n x_k} \Big(\prod_{k=1}^n x_i\Big)^{m-n} \prod_{1 \le i \le j \le n} (x_i - x_j)^2 \mathbf{1}_{x_1 > x_2 > \cdots > x_n > 0}. \tag{2}$$

See, for example, relation (3.16) in Forrester (2010).

### 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. The results of this subsection have appeared before, and we offer alternative proofs.

Assume as above that $\{X_{i,j} : i, j \in \mathbb{N}^+, i \ge j,\}$ are i.i.d. complex valued 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.

The fact that $L_n$ converges weakly and description of the limit was given in Dykema and Haagerup (2004).

**Theorem 1.** *With probability 1, $(L_n)_{n \ge 1}$ converges weakly to a deterministic measure $\mu_0$ on $\mathbb{R}$ with moments*

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

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

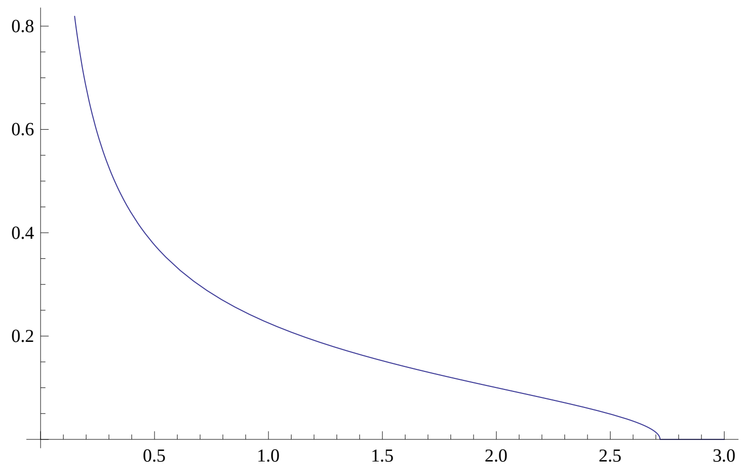


Figure 1: The graph of the density $f_0(x) = \pi^{-1}\,\mathrm{Im}\left\{-\frac{1}{xW(-x^{-1})}\right\}$. Its support is $[0, e]$.

The measure $\mu_0$ comes from a density which can be expressed in terms of the Lambert function. This is a multivalued function, it is the inverse of $w \mapsto we^w$. We will use the the principal branch, $W$, of this inverse, which is defined in $\mathbb{C}\backslash(-\infty, -e^{-1}]$. $W$ is analytic in $\mathbb{C}\backslash(-\infty, -e^{-1}]$ and can be extended to $\mathbb{C}$ so that it is continuous on the closed upper half plane (See Section 4 of Corless et al. (1996)). Below, $W$ will denote this extention.

**Corollary 1.** *The measure $\mu_0$ has*

*(i). continuous density $f_0$ with support* $[0, e]$*.*

*(ii). Stieltjes transform*

$$S(z) = -1 - \frac{1}{zW(-z^{-1})} = -1 + e^{W(-z^{-1})} \tag{4}$$

*for all $z \in \mathbb{C}$ with* $\mathrm{Im}(z) > 0$*.*

*(iii). R-transform*

$$R(z) = -\frac{1}{(1-z)\log(1-z)} - \frac{1}{z}$$

*for all $z \in \mathbb{C}$ with* $|z| < 1$*.*

The graph of the density $f_0$ is shown in Figure 1.

**Remark 1.** Near 0, the density $f_0$ behaves as

$$f_0(x) \sim \frac{1}{x(W(x^{-1}))^2} \sim \frac{1}{x(\log x)^2},$$

so that it explodes much faster than $1/\sqrt{x}$, which is the speed of explosion of the Marchenko-Pastur density in the case $m = n$. Near $e$, $f_0$ behaves as

$$f_0(x) \sim \frac{\sqrt{2}}{\pi e^{3/2}}\sqrt{e - x}. \tag{5}$$

We prove these statements in Subsection 2.2.

The next result, which is analogous to the result of Yin and Bai, appeared in the recent preprint Collins et al. (2014).

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

**Remark 2.** When this research begun, a few months ago, we were not aware that the result of Theorem 1 was known. In Dykema and Haagerup (2004), the theorem is phrased in the language of free probability, and its proof uses tools from that area together with combinatorial arguments.

The proof of Theorem 2 in Collins et al. (2014) uses probabilistic and operator theoretic arguments.

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 Tao (2012)). The crucial ingredient in our analysis is the notion of rooted alternating plane tree, which appears because of the triangular structure of the matrix.

### 1.3 The complex Gaussian case

In the case that the random variables $\{X_{i,j} : i, j \in \mathbb{N}^+, i \geq j\}$ in the previous subsection are complex standard normal, we can compute explicitly the joint distribution of the eigenvalues of $X(n)X(n)^*$.

**Theorem 3.** *For each positive integer $n$:*

*(i). The vector $\Lambda_n := (\lambda_1^{(n)}, \lambda_2^{(n)}, \dots, \lambda_n^{(n)})$ of the eigenvalues $\lambda_1^{(n)} \geq \lambda_2^{(n)} \geq \cdots \geq \lambda_n^{(n)}$ of $X(n)X^*(n)$ has density given by*

$$f_{\Lambda_n}(x_1, x_2, \dots, x_n) = \frac{1}{\prod_{j=1}^{n-1} j!}\, e^{-\sum_{j=1}^n x_j} \prod_{i<j} (x_i - x_j)(\log x_i - \log x_j) \mathbf{1}_{x_1 > x_2 > \cdots > x_n > 0} \tag{6}$$

*(ii). The point process $\{\lambda_1^{(n)}, \lambda_2^{(n)}, \dots, \lambda_n^{(n)}\}$ is determinantal.*

The theorem will be implied by the more general Theorems 4 and 5 of the next subsection.

### 1.4 Eigenvalue realization of the biorthogonal Laguerre ensemble

Fix a positive integer $n$. Consider the lower triangular matrix $(X_{i,j})_{1\leq i,j \leq n}$ with $\{X_{i,j} : 1 \leq j < i \leq n\}$ standard complex normal variables and $X_{k,k}$ having density

$$f_k(z) = \frac{1}{\pi \Gamma(c_k)}\, e^{-|z|^2} |z|^{2(c_k - 1)} \tag{7}$$

for all $z \in \mathbb{C}$, where for $(c_k)_{1 \leq k \leq n}$ we assume that they form an arithmetic progression with difference $\theta \in \mathbb{R}$ so that all its terms are positive. Thus $X_{k,k}$ can be written as

$$X_{k,k} = \frac{1}{\sqrt{2}}\, e^{i\phi_k} Y_k \tag{8}$$

where $Y_k$ follows the $\chi_{2c_k}$ distribution and $\phi_k$ is uniform on $[0, 2\pi)$ independent of $Y_k$.

It is enough to consider the case that $\theta \geq 0$, because otherwise taking $P$ the matrix with ones in the antidiagonal, we see that the matrix $(PXP^*)^*$ becomes of the form we consider with difference $\tilde{\theta} = -\theta$. Thus, we assume that there are $\theta \geq 0$, $b > 0$ so that

$$c_k = \theta(k-1) + b \tag{9}$$

for all $k \in \{1, 2, \ldots, n\}$. We denote the matrix constructed with this prescription by $X^{\theta,b}(n)$. For the distribution of the squares of its singular eigenvalues we have the following theorem.

**Theorem 4.** *The vector $\Lambda_n := (\lambda_1^{(n)}, \lambda_2^{(n)}, \ldots, \lambda_n^{(n)})$ of the eigenvalues $\lambda_1^{(n)} \geq \lambda_2^{(n)} \geq \cdots \geq \lambda_n^{(n)} \geq 0$ of $X^{\theta,b}(n) X^{\theta,b}(n)^*$ has density $f_{\Lambda_n}(x_1, x_2, \ldots, x_n)$ given by*

$$\frac{1}{\prod_{j=1}^{n-1} j!} \frac{\theta^{-n(n-1)/2}}{\prod_{k=1}^{n} \Gamma(c_k)} e^{-\sum_{j=1}^n x_j} \Big( \prod_{j=1}^n x_j^{b-1} \Big) \prod_{1 \leq i < j \leq n} (x_i - x_j)(x_i^\theta - x_j^\theta) \, \mathbf{1}_{x_1 > x_2 > \cdots > x_n > 0} \tag{10}$$

*when $\theta > 0$, and*

$$\frac{1}{\prod_{j=1}^{n-1} j!} \frac{1}{\Gamma(b)^n} e^{-\sum_{j=1}^n x_j} \Big( \prod_{j=1}^n x_j^{b-1} \Big) \prod_{1 \leq i < j \leq n} (x_i - x_j)(\log x_i - \log x_j) \, \mathbf{1}_{x_1 > x_2 > \cdots > x_n > 0} \tag{11}$$

*when $\theta = 0$.*

**Remark 3.** i) When $\theta = 0$ and $b = 1$, the matrix $X^{\theta,b}(n)$ is exactly $X(n)$ of the previous subsection. And thus we get Part (i) of Theorem 3.

ii) When $\theta = 1$ and $b = m - n$, with $m \geq n$ positive integers, (10) is the density (2). This is expected because there is a unitary matrix $U$ so that $X(m,n)U \stackrel{d}{=} [X^{1,m-n}(n), 0]$, where 0 is the $n \times (m-n)$ zero matrix.

iii) The density in (10) is the density of the $n$-point biorthogonal Laguerre ensemble, so termed and studied in Section 4 of Borodin (1999), with parameter pair $(\alpha, \theta)$ being $(b-1, \theta)$. Note that (11) is the $\theta \to 0$ limit of (10).

iv) Densities of the form (10) were introduced by Muttalib (1995) in the context of disordered conductors. A good approximation for the conductance of such a conductor is given by the sum $\sum_{k=1}^n (1 + x_k)^{-1}$, where $\{x_k : 1 \leq k \leq n\}$ are the eigenvalues of a certain positive definite random matrix. It is asserted in the above reference that the assumption that these eigenvalues come from a density of the form (10), with $\theta \neq 1$, which has two two-body interaction terms, namely $\prod_{i<j}(x_i - x_j)$ and $\prod_{i<j}(x_i^\theta - x_j^\theta)$, matches better experimental measurements from the conductor in the metalic regime than models with only one such term, as is (2). It is also shown that this density defines a determinantal point process. Later, Borodin (1999) gave an explicit formula for the kernel of the process, and using it determined the $n \to \infty$ limit at the hard edge (i.e., at 0) of an appropriate scaling of the process.

The formula for $f_{\Lambda_n}$ implies that $\{\lambda_1^{(n)}, \lambda_2^{(n)}, \ldots, \lambda_n^{(n)}\}$ is a biorthogonal ensemble (Borodin (1999), Forrester (2010) Section 5.8). And this allows to prove with little effort that the ensemble is a determinantal point process. In the case $\theta > 0$, this is already known. We cover next the $\theta = 0$ case. Define

$$g_{j,k} := \int_0^\infty x^j (\log x)^k e^{-x}\, dx$$

for $j, k \in \mathbb{N}$, and consider the matrix $G := (g_{i,j})_{i,j\in\mathbb{N}}$.

**Theorem 5.** *For each positive integer $n$:*

*(i). The matrix $G^{(n)} := (g_{j,k})_{0\le j,k\le n-1}$ is invertible.*

*(ii). The point process $\{\lambda_1^{(n)}, \lambda_2^{(n)}, \ldots, \lambda_n^{(n)}\}$ with law given by (11) is determinantal with kernel*

$$K_n(x,y) = e^{-\frac{x+y}{2}} (xy)^{\frac{b-1}{2}} \sum_{j,k=1}^{n} c_{j-1,k-1}^{(n)} (\log y)^{j-1} x^{k-1}.$$

*where $(c_{j,k}^{(n)})_{0\le j,k\le n-1}$ is the inverse of $G^{(n)}$.*

Finally, we come to the empirical spectral distribution $L_n^{\theta,b}$ of $X^{\theta,b}(n)X^{\theta,b}(n)^*/n$. The work in Dykema and Haagerup (2004) implies that this converges to a non trivial limit. To explain this connection, we need the notion of a $DT$-element.

Let $\nu$ a probability measure on $\mathbb{C}$ with compact support, and $c > 0$. For each $n$, let $T_n$ be an $n \times n$ matrix with $(T_n)_{i,j} = 0$ if $1 \le i \le j \le n$, and $\{(T_n)_{i,j} : 1 \le j < i \le n\}$ i.i.d. standard complex Gaussian. And let $D_n$ be a diagonal $n \times n$ matrix with i.i.d. diagonal elements, each having law $\nu$, and independent of $T_n$. Let $Z_n := D_n + cn^{-1/2}T_n$. It can be proved that for each $k \ge 1$ and $\varepsilon(1), \varepsilon(2), \ldots, \varepsilon(k) \in \{1, *\}$ the limit

$$\lim_{n\to\infty} \frac{1}{n} \mathbf{E}(\mathrm{tr}\{Z_n^{\varepsilon(1)} Z_n^{\varepsilon(2)} \cdots Z_n^{\varepsilon(k)}\}) \tag{12}$$

exists (Theorem 2.1 in Dykema and Haagerup (2004)).

**Definition 1.** An element $x$ of a $*$-noncommutative probability space $(\mathcal{A}, \phi)$ is called a $DT(\nu, c)$-element if for every $k \ge 1$ and $\varepsilon(1), \varepsilon(2), \ldots, \varepsilon(k) \in \{1, *\}$, we have that $\phi(x^{\varepsilon(1)} x^{\varepsilon(2)} \cdots x^{\varepsilon(k)})$ equals the value in (12).

And we are now ready to discuss the convergence of the sequence $(L_n^{\theta,b})_{n\ge1}$.

**Theorem 6.** *The empirical distribution of the eigenvalues of $X^{\theta,b}(n)X^{\theta,b}(n)^*/n$ converges to a measure $\mu_\theta$ whose moments are the moments of $xx^*$ where $x$ is a $DT(\nu_\theta, 1)$ element, and $\nu_\theta$ is the uniform measure on the disc $D(0, \sqrt{\theta}) := \{z \in \mathbb{C} : |z| \le \sqrt{\theta}\}$.*

Note that the limit does depend on $b$. In the case that $\theta > 1$ and $b = 1$, it is proven in Claeys and Romano (2013), see Paragraph 4.5.1, that the measure $\mu_\theta$ has density $f_\theta$ with support $I_\theta = [0, (1+\theta)^{1+1/\theta}]$. To describe it, let $J : \mathbb{C}\backslash[-1, 0] \to \mathbb{C}$ with

$$J(z) = \theta(z+1)\left(\frac{z+1}{z}\right)^{1/\theta}.$$

For each $x$ interior point of $I_\theta$, there are exactly two solutions of $J(z) = x$, which are conjugate complex numbers. Call them $I_-(x), I_+(x)$ so that $\mathrm{Im}(I_+(x)) > 0$. Then the density $f_\theta$ is given by

$$f_\theta(x) = \begin{cases} \frac{\theta}{2\pi x i}(I_+(x) - I_-(x)) & \text{if } x \in \big(0, (1+\theta)^{1+1/\theta}\big), \\ 0 & \text{if } x \in \mathbb{R}\backslash\big(0, (1+\theta)^{1+1/\theta}\big). \end{cases} \tag{13}$$

**Orientation:** Theorem 1 and its corollary are proved in Section 2, while Theorems 4, 5, 6 are proved in Sections 4, 5, 6 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 $(V, E, \phi)$, where $V, E$ are two sets (called the sets of vertices and edges respectively), and $\phi$ is a map from $E$ to $\{\{x, y\} : x, y \in V\}$. The interpretation is that $\phi(v)$ gives the two vertices that the edge $v$ connects, also called ends of $v$ (see Stanley Vol. 1, pg. 573). 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 Bai and Silverstein (2010). 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 this additional assumption, we prove that

$$\lim_{n\to\infty} \mathbf{E}\left\{\int x^k dL_n(x)\right\} = \frac{k^k}{(k+1)!} \tag{14}$$

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

We have

$$\begin{aligned} \mathbf{E}\left\{\int x^k dL_N(x)\right\} &= \frac{1}{n}\mathbf{E}\,\mathrm{tr}(S_n^k) = \frac{1}{n^{k+1}}\mathbf{E}\,\mathrm{tr}\{(XX^*)^k\} \\ &= \frac{1}{n^{k+1}}\mathbf{E}\Bigg\{\sum_{1\le i_1, i_2, \ldots, i_k \le n} (XX^*)_{i_1,i_2}(XX^*)_{i_2,i_3}\cdots(XX^*)_{i_k,i_1}\Bigg\} \\ &= \frac{1}{n^{k+1}}\mathbf{E}\Bigg\{\sum_{\substack{1\le i_1, i_2, \ldots, i_k \le n \\ 1\le j_1, j_2, \ldots, j_k \le n}} X_{i_1,j_1}X^*_{j_1,i_2}X_{i_2,j_2}X^*_{j_2,i_3}\cdots X_{i_k,j_k}X^*_{j_k,i_1}\Bigg\} \\ &= \frac{1}{n^{k+1}}\sum_{\substack{1\le i_1, i_2, \ldots, i_k \le n \\ 1\le j_1, j_2, \ldots, j_k \le n}} \mathbf{E}(X_{i_1,j_1}\overline{X}_{i_2,j_1}X_{i_2,j_2}\overline{X}_{i_3,j_2}\cdots X_{i_k,j_k}\overline{X}_{i_1,j_k}). \end{aligned} \tag{15}$$

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

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

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

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

which has cardinality $2k$, and the map $\phi$ 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.

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

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

and the maps $\phi_1$ is the identity map.
As explained in Bai and Silverstein (2010) (in the proof of relation (3.1.6) there, pages 49, 50), when we take $n \to \infty$ in (15), the only indices $(\mathbf{i}, \mathbf{j})$ contributing are those for which:

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

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

In fact, the pair $(\mathbf{i}, \mathbf{j})$ defines an *ordered* (also called *plane*) tree, that is, a tree on which we have specified an order among the children of each vertex. Among two vertices with common parent, we declare smaller the one that 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 labels of the vertices.
In our case, the fact that $X$ is triangular imposes the additional restrictions

(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 $\{i_1, i_2, \ldots, i_k\} \cap \{j_1, j_2, \ldots, j_k\} = \emptyset$. A pair $(\mathbf{i}, \mathbf{j}) \in \hat{\Delta}(n, k)$, instead of (iii) above, satisfies the stronger property

(iv). $j_1 < i_1, i_2$ and $j_2 < i_2, i_3$,,..., 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.

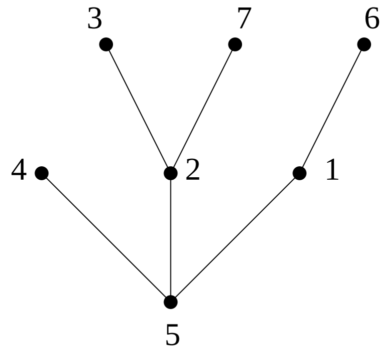


Figure 2: The tree corresponding to the pair $(\mathbf{i},\mathbf{j}) = ((5,5,3,7,5,6),(4,2,2,2,1,1))$

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

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

*Proof.* A tree with $r$ vertices labeled $\{1,2,\ldots,r\}$ is called *alternating* if for each path $v_1, v_2, \ldots, v_s$ on it we have

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

The set $V(\mathbf{i},\mathbf{j})$ can take $\binom{n}{k+1}$ values. Take one of them, say $\{1,2,\ldots,k+1\}$. The indices in $(\mathbf{i},\mathbf{j}) \in \Delta(n,k)$ for which $V(\mathbf{i},\mathbf{j}) = \{1,2,\ldots,k+1\}$ are in a one to one correspondence with the rooted alternating plane trees with $k+1$ vertices labeled $1,2,\ldots,k+1$ and such that the root is 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 $k^k$ (Theorem 3 in Chauve et al. (2001)). ∎

**Lemma 2.**

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

*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 $\{1,2,\ldots,n\}$. 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. ∎

The expectation in (15) corresponding to each $(\mathbf{i},\mathbf{j}) \in \Delta(n,k)$ equals 1 due to the assumptions on the distribution of the $X$'s and property (ii) above. Thus

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

which concludes the proof of the theorem.

## 2.2 The limiting measure. Proof of Corollary 1

(i) By Theorem 2.4.3 in Anderson et al. (2010) and the analytic properties of $W$, we get that the measure $\mu$ has support in $[0, e]$ and in $(0, e]$ has density

$$f(x) = \pi^{-1} \operatorname{Im} \left( \frac{-x^{-1}}{W(-x^{-1})} \right).$$

To exclude the possibility of mass at 0, we show that the integral of $f$ in $(0, 1]$ is 1. First, for $\delta > 0$, we compute

$$\begin{aligned} \int_\delta^e \frac{-x^{-1}}{W(-x^{-1})}\, dx &= \int_{-\delta^{-1}}^{-e^{-1}} \frac{1}{yW(y)}\, dy = \int_{-\delta^{-1}}^{-e^{-1}} \frac{1+W(y)}{W(y)^2} W'(y)\, dy \qquad (16) \\ &= \int_{-\delta^{-1}}^{-e^{-1}} \left( -\frac{1}{W(y)} + \log W(y) \right)' dy = 1 + \log(-1) + \frac{1}{W(-\delta^{-1})} - \log W(-\delta^{-1}). \end{aligned}$$

The second equality is true because the equation defining $W$ easily gives that $W(y) = y\{1 + W(y)\}W'(y)$. Now for $x \to -\infty$, we have $\operatorname{Re}(W(x)) \to \infty$ and $|\operatorname{Im}(W(x))| \le \pi$. Thus $\operatorname{Im}(\log W(x)) \to 0$, and the imaginary part of the integral in (16) converges to $\pi$ as $\delta \to 0^+$. So that the integral of $f$ in $(0, e]$ equals 1.

(ii) For $z \in \mathbb{C}, |z| > e$, we have

$$S(z) = -\sum_{k=0}^{\infty} \frac{1}{z^{k+1}} m_k = -\sum_{k=1}^{\infty} (k-1)^{(k-1)} \frac{(1/z)^k}{k!} = -L(1/z), \qquad (17)$$

where $L$ is the exponential generating function of the sequence $\{(k-1)^{(k-1)} : k \ge 1\}$. It is shown in Lemma 1 of Chauve et al. (2001) that $L$ satisfies

$$L(u) - 1 = -e^{\frac{u}{L(u)-1}}$$

for $u \in \mathbb{C}, |u| < e^{-1}$ so that $G(u) := u/(L(u)-1)$ satisfies $G(u)e^{G(u)} = -u$. And since $G(u) \in \mathbb{R}$ for $u \in (-e^{-1}, e^{-1})$, we have $G(u) = W(-u)$ for all $u \in \mathbb{C}, |u| < e^{-1}$. Combining $L(u) = 1 + (u/G(u))$ with (17), we get the claim for all $z \in \mathbb{C}, |z| > e$. The rest follows form the fact that both $S$ and $W$ are analytic in $\{z \in \mathbb{C} : \operatorname{Im}(z) > 0\}$.

(iii) By definition, $R(z) = K(z) - z^{-1}$ where $K$ satisfies $S(K(z)) = -z$. Omitting the analytic details, we note that this is written as $-1 + e^{W(-K(z)^{-1})} = -z$ so that

$$-K(z)^{-1} = W^{-1}(\log(1-z)) = \log(1-z)e^{\log(1-z)} = (1-z)\log(1-z),$$

proving the claim. ■

*Proof of Remark 1.* : For $x > 0$, let $a(x) := \operatorname{Re}(W(-x^{-1})), b(x) := \operatorname{Im}(W(-x^{-1}))$. Then

$$f_T(x) = \frac{1}{x\pi} \frac{b(x)}{\{a(x)^2 + b(x)^2\}}. \qquad (18)$$

1. Behavior near 0. By Corless et al. (1996), Section 4, $\lim_{x\to 0^+} b(x) = \pi$, while it is an easy exercise to see that $\lim_{x\to 0^+} a(x) = \infty$ and

$$\lim_{x\to 0^+} \{a(x) - W(x^{-1})\} = 0.$$

Thus the first $\sim$ has been proved. The second is elementary.
2. Behavior near $e$. Since $W(-e^{-1}) = -1$, we have $a(e) = 1, b(e) = 0$, and the denominator in (18) is $\sim 1$ as $x \to e^-$. If we let $p(z) = \sqrt{2(1 - ez^{-1})}$, relation 4.22 in Corless et al. (1996) implies that $b(x) \sim \operatorname{Im} p(x)$ as $x \to e$. But for $x \in (0, e)$, $p(x) = i\sqrt{2/x}\sqrt{e - x}$. This proves our claim. ■

## 3 The largest eigenvalue. Proof of Theorem 2

Theorem 1 and Corollary 1 give that $\underline{\lim}\, \lambda_1^n \geq e$. The aim of this section will be to show that $\overline{\lim}\, \lambda_1^n \leq e$. Our proof parallels the one of the Bai-Yin theorem as given in Section 2.3 of Tao (2012). The idea is to control a high enough moment of the maximum eigenvalue, and this is accomplished in the next proposition.

**Proposition 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 $[-C_1 n^{1/2-\varepsilon}, C_1 n^{1/2-\varepsilon}]$ and $k$ is an integer with $1 \leq k \leq C_2 \log^2 n$, then*

$$\mathbf{E}\, \mathrm{tr}\{(X(n)X(n)^*)^k\} \leq 2e^k n^{k+1}. \tag{19}$$

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

$$(2d_n)^6 < n, \tag{20}$$

$$(1 + 2C_1^2)(2d_n)^{48} < n^{2\varepsilon}. \tag{21}$$

Take $n, k$ as in the statement of the proposition. As in (15), we write

$$\begin{aligned}
\mathbf{E}\, \mathrm{tr}\{(X(n)X(n)^*)^k\} &= \sum_{\substack{1\leq i_1, i_2, \ldots, i_k \leq n \\ 1 \leq j_1, j_2, \ldots, j_k \leq n}} \mathbf{E}(X_{i_1,j_1} \overline{X}_{i_2,j_1} X_{i_2,j_2} \overline{X}_{i_3,j_2} \cdots X_{i_k,j_k} \overline{X}_{i_1,j_k}) \\
&= \sum_{\mathbf{i},\mathbf{j}} \mathbf{E}(X_{G(\mathbf{i},\, \mathbf{j})}) \\
&\leq \sum_{\mathbf{i},\mathbf{j}} \mathbf{E}(|X_{G(\mathbf{i},\, \mathbf{j})}|),
\end{aligned} \tag{22}$$

where to the pair $(\mathbf{i}, \mathbf{j}) = ((i_1, i_2, \ldots, i_k), (j_1, j_2, \ldots, j_k))$ 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 (22), we isolate the pairs $(\mathbf{i}, \mathbf{j})$ that satisfy (i), (ii), (iv) in the proof of Theorem 1. 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 (22). Take such a pair $(\mathbf{i},\ \mathbf{j})$. The path

$$(1, i_1) \to (2, j_1) \to (1, i_2) \to (2, j_2) \to \cdots (1, i_k) \to (2, j_k) \to (1, i_1) \tag{23}$$

is a cycle that traverses the graph $G_1(\mathbf{i}, \mathbf{j})$. List the edges $e_1, e_2, \ldots, e_s$ of $G_1(\mathbf{i}, \mathbf{j})$ in order of appearance in the cycle, and call $a_1, a_2, \ldots, a_s$ their multiplicities in the cycle. That is, $a_q$ is the number or 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 $a \geq 2$ integer it holds $\mathbf{E}(|X_{1,1}|^a) \leq (C_1 n^{1/2-\varepsilon})^{a-2}$. Thus

$$\mathbf{E}(|X_{G(\mathbf{i},\ \mathbf{j})}|) \leq \prod_{i=1}^{s} \mathbf{E}|X_{e_i}|^{a_i} \leq (C_1 n^{1/2-\varepsilon})^{a_1+\cdots+a_s-2s} = (C_1 n^{1/2-\varepsilon})^{2k-2s}. \tag{24}$$

Cycles that are generated by bad pairs we call them bad cycles. For integers $s \geq 1$ and $a_1, \ldots, a_s \geq 2$, let $N_{a_1,a_2,\ldots,a_s}$ be the number of bad cycles whose edges have multiplicities $a_1, a_2, \ldots, a_s$. The contribution of the bad pairs to (22) is at most

$$\sum_{s=1}^{k} (C_1 n^{1/2-\varepsilon})^{2k-2s} \sum_{a_1,a_2,\ldots,a_s} N_{a_1,a_2,\ldots,a_s}. \tag{25}$$

Using Lemma 3, we bound the last sum by

$$e^k (2k)^{10} \sum_{s=1}^{k} (C_1 n^{1/2-\varepsilon})^{2k-2s} (2k)^{36(k-s)} n^{\min\{s+1,k\}} \sum_{a_1,a_2,\ldots,a_s} 1. \tag{26}$$

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 $2k - 2k$. The number of such $s$-tuples is $\binom{2k-s-1}{2k-2s}$ (combinations with repetition) which is at most $(2k)^{2(k-s)}$. Separating the $s = k$ term, and letting $w = k - s$, we get for (26) the bound

$$e^k (2k)^{10} \left\{ n^k + n^{k+1} \sum_{w=1}^{k-1} \left( \frac{C_1^2 (2k)^{38}}{n^{2\varepsilon}} \right)^w \right\}.$$

By the choice of $n_0$, we have $C_1^2 (2k)^{38} / n^{2\varepsilon} < 1/2$, and the sum in the last expression is bounded by $2C_1^2 (2k)^{38} / n^{2\varepsilon}$. Thus, the contribution of the bad pairs to the sum (22) is at most

$$e^k (2k)^{10} \left( n^k + n^{k+1-2\varepsilon} 2C_1^2 (2k)^{38} \right) \leq e^k n^{k+1-2\varepsilon} (1 + 2C_1^2)(2k)^{48} < e^k n^{k+1}.$$

In the last equality, we used again the choice of $n_0$. 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 Tao (2012). In the rest of the section, we prove the crucial estimate we invoked in the proof above.

**Lemma 3.** $N_{a_1,a_2,\dots,a_s} \le e^k (2k)^{36(k-s)+10} n^{\min\{s+1,k\}}$.

*Proof.* Take a cycle as in (23), and label the vertices as $v_1 \to v_2 \to \cdots \to v_{2k} \to v_{2k+1} = v_1$. Each step in the cycle we call a leg. More formally, legs are the elements of the set $\{(a,(v_a,v_{a+1})) : a = 1,2,\dots,2k\}$. For $1 \le a < b$, we say that the leg $(a,(v_a,v_{a+1}))$ is single up to $b$ if $\{v_a,v_{a+1}\} \neq \{v_c,v_{c+1}\}$ for every $c \in \{1,2,\dots,b-1\}$. 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 $v_{a+1} \notin \{v_1,\dots,v_a\}$.

$T_3$: if there is $T_1$ leg $(b,(v_b,v_{b+1}))$ with $b < a$ so that $a = \min\{c > b : \{v_c,v_{c+1}\} = \{v_b,v_{b+1}\}\}$.

$T_4$: if it is not $T_1$ or $T_3$.

$T_2$: if it is $T_4$ and there is no $b < a$ with $\{v_a,v_{a+1}\} = \{v_b,v_{b+1}\}$.

Moreover, a $T_3$ leg $(a,(v_a,v_{a+1}))$ is called irregular if there is exactly one $T_1$ leg $(b,(v_b,v_{b+1}))$ with $b < a$ satisfying $\{v_c,v_{c+1}\} \neq \{v_b,v_{b+1}\}$ for all $c \in \{1,2,\dots,a-1\}\backslash\{b\}$ and $v_a \in \{v_b,v_{b+1}\}$. Otherwise the leg is called regular.

We already know that the number of edges of $G_1(\mathbf{i},\mathbf{j})$ is $s$. Let also

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

$\ell$: the number of edges that have multiplicity at least 3.

$m$: the number of $T_2$ legs.

$r$: the number of regular $T_3$ legs.

The number of edges with multiplicity 2 is $s-\ell$. Thus,

$$\sum_{i:a_i>3} a_i = 2k - 2(s-\ell). \tag{27}$$

On the other hand, the same sum is at least $3\ell$. Thus $\ell \le 2(k-s)$. Lemma 2.3 in Bai and Yin (1988) says that $r \le 2|T_4|$, while for $|T_4|$ we have the bound

$$|T_4| \le m + \sum_{i:a_i\ge 3} a_i \le m + 6(k-s). \tag{28}$$

Now back to the task of bounding $N_{a_1,\dots,a_s}$. Given a cycle, we 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

- the locations of regular $T_3$ legs in the cycle and the index of the vertex each has as a final vertex.
- the locations of $T_4$ legs in the cycle and the index of the vertex each has as a final vertex.
- the index of each J vertex, say $(2,j)$, for which $j \in \{i_1,i_2,\dots,i_k\}$.

Finally, a cycle defines a rooted, ordered, alternating tree with $t$ vertices and labels $\{1,2,\dots,t\}$. We get this tree as follows.

1. Graph: 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 $\hat{G}(\mathbf{i},\, \mathbf{j})$ (i.e., with no multiple edges) which is in fact 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. Root of the tree is the vertex $(1, i_1)$, and the children of each vertex are ordered according to their index.

2. Labels: Initially, to each vertex $(a, b)$ of this tree (recall that $a \in \{1, 2\}$) we assign the label $b$. Next, we assign new labels so that the labels of any two vertices have the same order as before, but the labels used form an initial segment of the positive integers. Assume that they are $\{1, 2, \ldots, t-u\}$ for some integer $u \geq 0$. It is $u > 0$ exactly when $u$ J-vertices have label that agrees with the label of one I-vertex. We do now a final relabeling. If a J vertex $v := (2, j)$ has $j \in \{i_1, i_2, \ldots, i_k\}$, we increase by one the label of every vertex (I or J) which has at the moment label $\geq j$ except $v$. We do this procedure sequentially by checking equality for the label 1 and continuing upward. In the end, no two vertices will have the same label, and the set of labels will be $\{1, 2, \ldots, t\}$.

Step 1 gives an ordered rooted tree of $t$ vertices, and Step 2 together with the property

$$j_1 \leq i_1, i_2 \text{ and } j_2 \leq i_2, i_3, , \ldots, \text{ and } j_k \leq i_k, i_1$$

that the indices have, gives us that the labeling is alternating (see definition in the proof of Lemma 1) with the root larger than its children.

Having these elements, we can reconstruct the cycle up to the names of the $t-u$ different labels (the locations of the labels are determined by the labeled tree). Because the locations of all legs in the cycle are either known or can be inferred. The same holds for the ending point of each leg. Since all legs start at the ending point of some other leg, the location of the legs in the cycle determines the starting points too.

Thus, the number of bad cycles with given $t, u, r, |T_4|$ is at most

$$\binom{n}{t-u}(t-1)^{t-1}(2k)^r t^r (2k)^{|T_4|} t^{|T_4|} t^u \leq n^{t-u}\frac{(t-1)!}{(t-u)!}\, e^{t-1}(2k)^{2(r+|T_4|)} t^u \tag{29}$$

$$\leq e^{t-1} n^t \left(\frac{t^2}{n}\right)^u (2k)^{2(r+|T_4|)}. \tag{30}$$

We used the rough bound $t \leq 2k$. By the choice of $n_0$, we have $t^2 < n$. Moreover, since $t \leq s+1-m \leq k+1$ and

$$r + |T_4| \leq 3|T_4| \leq 3m + 18(k-s), \tag{31}$$

the bound in (30) is less than

$$e^k n^{s+1-m}(2k)^{6m+36(k-s)} = e^k (2k)^{36(k-s)} n^{s+1}\left(\frac{(2k)^6}{n}\right)^m \leq e^k (2k)^{36(k-s)+6} n^{\min\{s+1,k\}}. \tag{32}$$

The last inequality is true because $s \le k$ always, $(2k)^6/n < 1$ by the choice on $n_0$, and because $s = k$ implies $m \ge 1$ as the cycle is bad. Summing the bound (32) over all possible values of $t, r, u, |T_4|$, which are at most $2k$ for each, we get the claim of the lemma. ∎

# 4 Distribution of singular values for $X^{\theta,b}(n)$

Define the following sets of matrices

$\mathcal{T}_n$: lower triangular $n \times n$ matrices with elements in $\mathbb{C}$ and diagonal elements in $\mathbb{C}\backslash\{0\}$.

$\mathcal{T}_n^+$: elements of $\mathcal{T}_n$ with diagonal elements in $(0, \infty)$,

$\mathcal{V}_n$: diagonal $n \times n$ matrices with diagonal elements complex of modulus 1.

$\mathcal{M}_n^+$: positive definite $n \times n$ matrices with elements in $\mathbb{C}$.

We identify the spaces $\mathcal{T}_n, \mathcal{T}_n^+, \mathcal{V}_n$ with $\mathbb{R}^{n(n-1)} \times (\mathbb{R}\backslash\{0\})^{2n}, \mathbb{R}^{n(n-1)} \times (0, \infty)^n, [0, 2\pi)^n$ respectively, and view $\mathcal{M}_n^+$ as a subset of $n \times n$ Hermitian matrices, which we identify with $\mathbb{R}^{n^2}$. Densities of random variables with values in these spaces are meant with respect to the corresponding Lebesgue measure.

Consider the maps $g : \mathcal{T}_n^+ \times \mathcal{V}_n \to \mathcal{T}_n, h : \mathcal{T}_n^+ \to \mathcal{M}_n^+$ with $g(T, V) = TV$, $h(Y) := YY^*$. They are both one to one and onto. Call $g^{-1} := (\gamma_1, \gamma_2)$, and $X := X^{\theta,b}(n)$. Then $XX^* = h(\gamma_1(X))$ provided that $X \in \mathcal{T}_n$, which holds with probability 1. We will use this relation in order to find the joint law of the elements of $XX^*$, and then, the law of its eigenvalues will follow from a well known formula.

**Lemma 4.** *The Jacobian of the map $g$ has absolute value*

$$\prod_{j=1}^{n} t_{j,j}.$$

*Proof.* Let $X = TV$ and call $x_{i,j}$ its $(i, j)$ element. The Jacobian matrix of $g$ has $n$ blocks, one for each column. The block corresponding to column $j$ is

$$\frac{\partial(x_{j,j}^R, x_{j,j}^I, x_{j+1,j}^R, x_{j+1,j}^I, \ldots, x_{n,j}^R, x_{n,j}^I)}{\partial(\theta_j, t_{j,j}, t_{j+1,j}^R, t_{j+1,j}^I, \ldots, t_{n,j}^R, t_{n,j}^I)},$$

and its determinant equals $-t_{j,j}$. ∎

**Lemma 5.** *The map $h$ has Jacobian*

$$2^n \prod_{i=1}^{n} t_{i,i}^{2(n-i)+1}.$$

*Proof.* This is Proposition 3.2.6 of Forrester (2010). ∎

In the following, we use the notation set in Subsection 1.4. Let $C(\theta,b) := \big(\prod_{k=1}^n \Gamma(c_k+1)\big)^{-1}$. The density of $X^{\theta,b}(n)$ is

$$f_{X^{\theta,b}(n)}(x) = \frac{1}{\pi^{n(n+1)/2}}\, C(\theta,b)\, e^{-\sum_{1\le j\le i\le n}|x_{i,j}|^2} \prod_{k=1}^n |x_{k,k}|^{2c_k} \tag{33}$$

$$= \frac{1}{\pi^{n(n+1)/2}}\, C(\theta,b)\, e^{-\operatorname{tr}(xx^*)} \prod_{k=1}^n |x_{k,k}|^{2c_k} \tag{34}$$

for all $x \in \mathbb{C}^{n(n+1)/2}$.

For an $n\times n$ matrix $a = (a_{i,j})_{1\le i,j\le n}$ and $k \in \{1,2,\dots,n\}$, we denote by $a_k$ its main $k\times k$ minor, that is, the matrix $(a_{i,j})_{1\le i,j\le k}$.

**Proposition 2.** *Let $X := X^{\theta,b}(n)$. The matrix $A := XX^*$ has density*

$$f_A(a) = \frac{1}{\pi^{n(n-1)/2}} \frac{1}{\prod_{k=1}^n \Gamma(c_k)}\, e^{-\operatorname{tr}(a)} \{\det(a)\}^{c_n-1} \{\det(a_1)\det(a_2)\cdots\det(a_{n-1})\}^{-\theta-1} \tag{35}$$

*for all $a \in \mathcal{M}_n^+$, and $f_A(a) = 0$ for every Hermitian matrix not an element of $\mathcal{M}_n^+$.*

*Proof.* Since $XX^* = h(\gamma_1(X))$, our first step is to find the distribution of $T := \gamma_1(X)$.

Claim: $T$ has density

$$f_T(t) = (2\pi)^n f_X(t) \prod_{j=1}^n t_{j,j}.$$

Proof of the claim: For any positive measurable function defined on $\mathcal{T}_n$, we have

$$\mathbf{E}\{s(T)\} = \mathbf{E}\{s(\gamma_1(X))\} = \int s(r_1(x)) f_X(x)\,dx = \int_{\mathcal{T}_n^+}\int_{[0,2\pi)^n} s(\gamma_1(g(t,\theta)) f_X(g(t,\theta)) |Jg(t,\theta)|\,d\theta\,dt$$

$$= (2\pi)^n \int_{\mathcal{T}_n^+} s(t) f_X(t) \prod_{j=1}^n t_{j,j}\,dt.$$

In the last equality we used Lemma 4, and the fact that $f_X(TV) = f_X(T)$ for all $V$ of the form $\operatorname{diag}(e^{i\theta_1}, e^{i\theta_2},\dots,e^{i\theta_n})$. Thus, the claim is proved.

Now, for given $a \in \mathcal{M}_n^+$, let $t := h^{-1}(a)$. Then

$$f_A(a) = f_T(h^{-1}(a))|Jh^{-1}(a)| = (2\pi)^n f_X(h^{-1}(a)) \Big(\prod_{j=1}^n t_{j,j}\Big) \frac{1}{|Jh(h^{-1}(a))|}$$

$$= (2\pi)^n \frac{1}{\pi^{n(n+1)/2}}\, C(\theta,b)\, e^{-\operatorname{tr}(a)} \prod_{j=1}^n |t_{j,j}|^{2(c_j-1)} \frac{1}{2^n \prod_{j=1}^n t_{j,j}^{2(n-j)+1}} \prod_{j=1}^n t_{j,j}$$

$$= \frac{1}{\pi^{n(n-1)/2}}\, C(\theta,b)\, e^{-\operatorname{tr}(a)} \Big(\prod_{i=1}^n t_{j,j}^{2(n-j-c_j+1)}\Big)^{-1}$$

$$= \frac{1}{\pi^{n(n-1)/2}}\, C(\theta,b)\, e^{-\operatorname{tr}(a)} \Big(\prod_{j=1}^n t_{j,j}^2\Big)^{c_n-1} \Big(\prod_{j=1}^n t_{j,j}^{2(n-j)}\Big)^{-(1+\theta)}.$$

In the third equality we used Lemma 5, and in the last equality the fact that $-c_j = \theta(n-j) - c_n$ for all $j \in \{1, 2, \dots, n\}$. Finally, we express the products involving the variables $t_{j,j}$ in terms of the variable $a$. Since $T$ is lower triangular, we have $a_i = T_i T_i^*$. Thus

$$\det(a_i) = |\det(T_i)|^2 = (t_{1,1} t_{2,2} \dots t_{i,i})^2.$$

Multiplying these equalities for all $1 \le i \le n-1$, we get

$$\det(a_1)\det(a_2)\cdots\det(a_{n-1}) = \prod_{i=1}^{n} t_{i,i}^{2(n-i)}.$$

This finishes the proof of the proposition. ∎

*Proof of Theorem 4.* From relations (4.1.17), (4.1.18) in Anderson et al. (2010), and the fact that $X^{\theta,b}(n)X^{\theta,b}(n)^*$ is positive definite, we have that the vector of the eigenvalues in decreasing order has density

$$f_{\Lambda_n}(\lambda) = C_n \prod_{i<j} (\lambda_i - \lambda_j)^2 \int_{U(n)} f_A(H D_\lambda H^*)(dH)\, \mathbf{1}_{\lambda_1 > \lambda_2 > \ldots > \lambda_n > 0}$$

where $\lambda := (\lambda_1, \lambda_2, \dots, \lambda_n)$, $D_\lambda$ is the diagonal matrix with diagonal $\lambda$, $(dH)$ is the normalized Haar measure on $U(n)$, and the constant $C_n$ is

$$C_n := \frac{\pi^{n(n-1)/2}}{\prod_{j=1}^{n-1} j!}.$$

Thus, writing $a := H D_\lambda H^*$ and taking into account Proposition 2, we get

$$f_{\Lambda_n}(\lambda) = \frac{C(\theta, b)}{\prod_{j=1}^{n-1} j!} \left\{ \prod_{i<j} (\lambda_i - \lambda_j)^2 \right\} e^{-\sum_{j=1}^n \lambda_j} \left( \prod_{j=1}^n \lambda_j \right)^{c_n - 1} K(\lambda)\, \mathbf{1}_{\lambda_1 > \lambda_2 > \ldots > \lambda_n > 0} \tag{36}$$

with

$$K(\lambda) := \int_{U(n)} \{\det(a_1)\det(a_2)\cdots\det(a_{n-1})\}^{-\theta-1} (dH). \tag{37}$$

The computation of the last integral is given in Lemma 6. Combining that computation with (36), we finish the proof. ∎

**Lemma 6.** *For $\theta \ge 0$, the integral in* (37) *equals*

$$K(\lambda) = \frac{\prod_{1 \le i < j \le n} \int_{\lambda_j}^{\lambda_i} x^{-\theta-1}\, dx}{\prod_{1 \le i < j \le n} (\lambda_i - \lambda_j)} = \left( \prod_{i=1}^n \lambda_i \right)^{-\theta(n-1)} \frac{\prod_{1 \le i < j \le n} \int_{\lambda_j}^{\lambda_i} x^{\theta-1}\, dx}{\prod_{1 \le i < j \le n} (\lambda_i - \lambda_j)}. \tag{38}$$

*Proof.* To simplify the exposition, we introduce a binary relation which we denote by $\succ$. $x \succ y$ means that there is $k \in \mathbb{N}^+$ so that $x = (x_1, x_2, \dots, x_{k+1}) \in \mathbb{R}^{k+1}$, $y = (y_1, y_2, \dots, y_k) \in \mathbb{R}^k$, and

$$x_1 \ge y_1 \ge x_2 \ge y_2 \cdots \ge x_n \ge y_n \ge x_{k+1}.$$

For $x = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, we let $Y(x)$ be the set of all elements $(y^{(k-1)}, y^{(k-2)}, \ldots, y^{(1)})$ of $\mathbb{R}^{k-1} \times \mathbb{R}^{k-2} \times \cdots \times \mathbb{R}^2 \times \mathbb{R}$ that satisfy

$$\lambda \succ y^{(k-1)} \succ y^{(k-2)} \succ \cdots \succ y^{(2)} \succ y^{(1)}.$$

One can easily verify that

$$\mathrm{Vol}(Y(x)) = \prod_{1 \le i < j \le k} \frac{x_i - x_j}{j - i}. \tag{39}$$

We can now start the proof of the lemma. For each $i = 1, 2, \ldots, n-1$, call $x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \cdots, x_i^{(i)})$ the vector of the eigenvalues of the symmetric matrix $a_i$ with $x_1^{(i)} \ge x_2^{(i)} \ge \cdots \ge x_i^{(i)}$. Then the integrand in (37) is simply

$$\prod_{i=1}^{n-1} \prod_{j=1}^{i} (x_j^{(i)})^{-\theta-1}.$$

Under $(dH)$, the law of $a$ is the one of an $n \times n$ GUE matrix conditioned to have eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, and according to Proposition 4.7 in Baryshnikov (2001), the law of $(x^{(n-1)}, x^{(n-2)}, \ldots, x^{(1)})$ is the uniform on $Y(\lambda)$ with respect to Lebesgue measure. Thus the integral equals

$$\frac{1}{\mathrm{Vol}(Y(\lambda))} \underbrace{\int_{\lambda_n}^{\lambda_{n-1}} \int_{\lambda_{n-1}}^{\lambda_{n-2}} \cdots \int_{\lambda_2}^{\lambda_1}}_{n-1 \text{ integrals}} \underbrace{\int_{x_{n-1}^{(n-1)}}^{x_{n-2}^{(n-1)}} \cdots \int_{x_2^{(n-1)}}^{x_1^{(n-1)}}}_{n-2 \text{ integrals}} \cdots \underbrace{\int_{x_2^{(2)}}^{x_1^{(2)}}}_{1 \text{ integral}} \prod_{1 \le j \le i \le n-1} (x_j^{(i)})^{-\theta-1} \prod_{1 \le j \le i \le n-1} dx_j^{(i)}. \tag{40}$$

Assume now that $\theta > 0$. Let $\tilde{\lambda}(\theta) := (\lambda_n^{-\theta}, \lambda_{n-1}^{-\theta}, \ldots, \lambda_1^{-\theta})$, and in the integral make the change of variables $y_j^{(i)} = (x_j^{(i)})^{-\theta}$ for all $1 \le j \le i \le n-1$. Then the previous expression becomes

$$\frac{1}{\mathrm{Vol}(Y(\lambda))} (-\theta)^{-n(n-1)/2} \underbrace{\int_{\lambda_n^{-\theta}}^{\lambda_{n-1}^{-\theta}} \int_{\lambda_{n-1}^{-\theta}}^{\lambda_{n-2}^{-\theta}} \cdots \int_{\lambda_2^{-\theta}}^{\lambda_1^{-\theta}}}_{n-1 \text{ integrals}} \underbrace{\int_{y_{n-1}^{(n-1)}}^{y_{n-2}^{(n-1)}} \cdots \int_{y_2^{(n-1)}}^{y_1^{(n-1)}}}_{n-2 \text{ integrals}} \cdots \underbrace{\int_{y_2^{(2)}}^{y_1^{(2)}}}_{1 \text{ integral}} \prod_{i=1}^{n-1} \prod_{j=1}^{i} dy_j^{(i)}$$

$$= \theta^{-n(n-1)/2} \frac{\mathrm{Vol}(Y(\tilde{\lambda}(\theta)))}{\mathrm{Vol}(Y(\lambda))} = \theta^{-n(n-1)/2} \frac{\prod_{1 \le i < j \le n} (\lambda_j^{-\theta} - \lambda_i^{-\theta})}{\prod_{1 \le i < j \le n} (\lambda_i - \lambda_j)} \tag{41}$$

$$= \left( \prod_{i=1}^{n} \lambda_i \right)^{-\theta(n-1)} \frac{\theta^{-n(n-1)/2} \prod_{1 \le i < j \le n} (\lambda_i^{\theta} - \lambda_j^{\theta})}{\prod_{1 \le i < j \le n} (\lambda_i - \lambda_j)}. \tag{42}$$

And the lemma is proved in this case. In the case $\theta = 0$, in the integral of (40), we let $y_j^{(i)} = \log x_j^{(i)}$ for all $1 \le j \le i \le n-1$ and proceed as above. Alternatively, we can take $\theta \to 0$ in the last expression. ∎

# 5 Determinantal process. Proof of Theorem 5

*Proof.* (i). For each positive integer $n$ and $y_1, y_2, \ldots, y_n \in \mathbb{R}$, we let

$$\Delta(y_1, y_2, \ldots, y_n) := \det(y_k^{j-1})_{1 \le j, k \le n} = \prod_{1 \le j < k \le n} (y_k - y_j).$$

Equation (3.3) of Deift and Gioev (2009) gives that the determinant of $G^{(n)}$ is

$$\det\left(\int_0^\infty x^j(\log x)^k e^{-x}\,dx\right)_{0\le j,k<n} = \frac{1}{n!}\int_0^\infty\cdots\int_0^\infty \det(x_k^{j-1})_{1\le j,k\le n}\det((\log x_k)^{j-1})_{1\le j,k\le n}\prod_{k=1}^n dx_k \tag{43}$$

$$= \frac{1}{n!}\int_0^\infty\cdots\int_0^\infty \Delta_n(x_1,\ldots,x_n)\Delta_n(\log x_1,\ldots,\log x_n)\prod_{k=1}^n dx_k \tag{44}$$

The integrand is positive, thus the determinant is not zero.

(ii) Follows from part (i) and Proposition 5.8.1 of Forrester (2010). ■

In the rest of the section, we discuss the structure of $G^{(n)}$ and compute explicitly the value of its determinant.

**Lemma 7.** *The matrix $G := (g_{i,j})_{i,j\in\mathbb{N}}$ has an LU factorization $G = LU$ with*

$$L_{i,j} = |s(i+1,j+1)| \qquad \text{for } i \ge j \ge 0, \tag{45}$$

$$U_{i,j} = (j)_i g_{0,j-i} \qquad \text{for } 0 \le i \le j. \tag{46}$$

*Here $s$ denotes the Stirling number of the first kind.*

We follow the convention $(0)_i = 1$ for all nonnegative integers $i$.

*Proof.* We compute the exponential generating function of the sequence $(g_{j,k})_{j,k\in\mathbb{N}}$.

$$\begin{aligned}\sum_{j,k=0}^\infty \frac{u^j}{j!}\frac{v^k}{k!}g_{j,k} &= \int_0^\infty e^{-x}e^{ux}e^{v\log x}\,dx = \int_0^\infty e^{-x}e^{-(1-u)x}x^v\,dx = (1-u)^{-v-1}\Gamma(1+v)\\ &= \sum_{j=0}^\infty (v+1)(v+2)\cdots(v+j)\frac{u^j}{j!}\sum_{s=0}^\infty \frac{\Gamma^{(s)}(1)}{s!}v^s\\ &= \sum_{j=0}^\infty \frac{u^j}{j!}\sum_{r=0}^{j}(-1)^{j-r}s(j+1,r+1)v^r\sum_{s=0}^\infty\frac{\Gamma^{(s)}(1)}{s!}v^s\\ &= \sum_{j,k=0}^\infty \frac{u^j}{j!}v^k\sum_{r=0}^{j\wedge k}\frac{\Gamma^{(k-r)}(1)}{(k-r)!}(-1)^{j-r}s(j+1,r+1).\end{aligned}$$

Since for $m,n\in\mathbb{N}$ the integer $s(m,n)$, if not zero, has sign $(-1)^{m-n}$ and $\Gamma^{(n)}(1) = g_{0,n}$, we get

$$g_{j,k} = \sum_{r=0}^{j\wedge k}|s(j+1,r+1)|(k)_r g_{0,k-r}$$

for all $j,k\in\mathbb{N}$. This proves the factorization $G = LU$. ■

Since $L_{k,k} = 1$ and $U_{k,k} = k!$ for all $k\in\mathbb{N}$, we obtain

$$\det(G^{(n)}) = 1!2!\cdots(n-1)!.$$

## 6 Limiting empirical distribution in the $\theta > 0$ case

*Proof of Theorem 6*: We first prove the following.

Claim: The sequence $\left(X^{\theta,b}(n)/\sqrt{n}\right)_{n\geq 1}$ converges in $*$-moments to a $DT(\nu_\theta, 1)$ element.

Recall the form of the diagonal elements of $X^{\theta,b}(n)$ in (8). Call $\eta_n$ the law of the vector $(X_{k,k}/\sqrt{n})_{1\leq k\leq n}$, $\tilde{\eta}_n$ its symmetrization, that is, the law of $(X_{\pi(k),\pi(k)}/\sqrt{n})_{1\leq k\leq n}$, where $\pi$ is a random permutation of $\{1, 2, \ldots, n\}$ independent of the matrix, and $\tilde{\eta}_n^{(p)}$ the projection onto the first $p$ coordinates of $\tilde{\eta}_n$ ($p \in \{1, 2, \ldots, n\}$). According to Theorem 2.13 in Dykema and Haagerup (2004), it is enough to show that $\tilde{\eta}_n^{(p)}$ converges in $*$-moments to $\times_1^p \nu_\theta$. Since the law $\eta_n$ is a product measure, it is enough to show this only for $p = 1$. Clearly the law of $\tilde{\eta}_n^{(1)}$ is radially symmetric. It suffices therefore to prove that for any function $h : \mathbb{R} \to \mathbb{R}$ of at most polynomial growth, it holds that

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{E}h\left(\frac{Y_k^2}{2n}\right) \to \frac{1}{\theta}\int_0^\theta h(r)\,dr \tag{47}$$

as $n \to \infty$. Take independent random variables $(W_k)_{k\geq 1}$ so that $W_1 \sim \Gamma(b, 1)$ and $W_k \sim \Gamma(\theta, 1)$ for all $k \geq 2$. Then $Y_k^2/2$ has the same law as $S_k := W_1 + W_2 + \cdots + W_k$. The left hand side of (47) is

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{E}h\left(\frac{S_k}{k}\frac{k}{n}\right). \tag{48}$$

Now, $S_k/k$ converges pointwise to $\theta$, and satisfies a large deviations principle with speed $n$ and good rate function having a unique zero at $\theta$. At the same time, $h$ has at most polynomial growth. It is easy then to show that (48) converges to $\int_0^1 h(\theta x)\,dx$, proving our claim.

Call $x$ the limit mentioned in the claim. We can assume that $x$ is an element of a von Neumann algebra (see Remark 2.3 in Dykema and Haagerup (2004)), and thus there is a unique compactly supported measure in $\mathbb{R}$ that has the same moments as $xx^*$ (Lemma 5.2.19 in Anderson et al. (2010)). The theorem follows by combining this with the above claim. ■

**Acknowledgments:** This research was carried out while I spent the fall semester of 2013 at Leiden University. I thank Frank den Hollander for the invitation, and the entire probability group for the stimulating atmosphere.

## 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.

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`.

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.

Y. Baryshnikov (2001). GUEs and queues. *Probab. Theory Related Fields*, 119(2):256–274.

A. Borodin (1999). Biorthogonal ensembles. *Nuclear Phys. B*, 536(3):704–732.

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

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

T. Claeys and S. Romano (2013). Biorthogonal ensembles with two-particle interactions. *arXiv preprint arXiv:1312.2892*.

B. Collins, P. Gawron, A. E. Litvak, and K. Życzkowski (2014). Numerical range for random matrices. *arXiv preprint arXiv:1309.6203*.

R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth (1996). On the Lambert $W$ function. *Adv. Comput. Math.*, 5(4):329–359.

P. Deift and D. Gioev. *Random matrix theory: invariant ensembles and universality*, volume 18. American Mathematical Soc., 2009.

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

P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

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

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

K. A. Muttalib (1995). Random matrix models with additional interactions. *Journal of Physics A: Mathematical and General*, 28(5):L159.

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