On the largest eigenvalue of random matrices with general variance profile

Dimitris Cheliotis
Department of Mathematics
National and Kapodistrian University of Athens
Panepistimiopolis, Athens 15784, Greece.
dcheliotis@math.uoa.gr

Michail Louvaris
Department of Mathematics
National and Kapodistrian University of Athens
Panepistimiopolis, Athens 15784, Greece.
louvarismixalis@gmail.com

January 3, 2023

Abstract

In this paper we establish the convergence of the largest eigenvalue of general variance profile random matrices to the largest element of the support of the limiting measure under very general assumptions for the variance profile of the matrices. We also prove that it is sufficient for the entries of the matrix to have finite only the 4-th moment, instead of all the moments. This is a generalization of previously known results.

1 Introduction

The problem of understanding the operator norm of a random matrix with independent entries is in general multidisciplinary both from mathematical and non-mathematical point of view. From a mathematical point of view, tools from classical probability, geometric analysis, combinatorics, free probability and more have been used. That problem dates back to 1981, where in [18] the convergence of the largest eigenvalue of Wigner matrices (symmetric, i.i.d. entries) to the edge of the limiting distribution was established when the entries of the matrix are bounded. Next, in [2], the authors gave necessary and sufficient conditions for the entries of a Wigner matrix to converge. One of those conditions was that the entries should have finite 4-th moment. Similar bounds have been given to non-symmetric matrices with i.i.d. entries. Then, the difference of the largest eigenvalue and its limit, after re-normalization, was proven to converge to the Tracy-Widow law in [26]. Later, universality results were established for sparse random matrix models, for example in [21] for random graphs and in [24] for random banded matrices. Moreover, sharp non-asymptotic results for a general class of matrices were established in [4] and in [3].
All the models mentioned above can be generally considered as random matrices with general variance profile, i.e. random matrices whose entries' variances are not fixed and can depend on the dimension. These models have also drawn a lot of attention lately, see for example [12], [13], where non-Hermitian models are considered. Furthermore, in [27], the author characterized the limiting E.S.D. through the notion of graphons. The convergence of the largest eigenvalue to the largest element of the support of the limiting distribution was established in the recent works [22] and [16] for some class of random matrices with general variance profile under the assumption that the entries of the matrix have finite all moments. In this paper we generalize the previously mentioned results, i.e., we establish the convergence of the largest eigenvalue of general variance profile random matrices to the largest element of the support of the limiting measure under

very general assumptions for the variance profile of the matrices. We also prove that it is sufficient for the entries of the matrix to have finite only the 4-th moment, instead of all the moments.

2 Description of results

Let A_N be a sequence of symmetric random matrices with independent entries (up to symmetry) such that {Genikes ypo

Assumption 2.1.

- $\mathbf{E}a_{i,j}^{(N)} = 0, \mathbf{E}|a_{i,j}^{(N)}|^2 \leq 1$ for all i, j, N, and $\sup_N \max_{i,j\in[N]} \mathbf{E}|a_{i,j}^{(N)}|^4 < \infty$.
- For any constant $\varepsilon > 0$ it is true that

$$\sum_{i,j} \mathbf{P}(|a_{i,j}^{(N)}| \geq \varepsilon\sqrt{N}) \to 0 \tag{2.1}$$ {MaxToZero}

We let $s_{i,j}^{(N)} := \mathbf{E}\{|a_{i,j}^{(n)}|^2\}$. Note that these conditions imply the assumptions in the beginning of section 3 of [27]. Note that condition (2.1) always holds if we assume that the entries of A_N are i.i.d. with finite 4-th moment.

Notation 2.2. For any $N \times N$ symmetric matrix A with eigenvalues $\{\lambda_i(A)\}_{i\in[N]}$, the measure

$$\frac{1}{N}\sum_{i\in[N]} \delta_{\lambda_i(A)}$$

will be the Empirical Spectral Distribution (E.S.D.) of A. Moreover we will use the following notation for the operator norm of the matrix A,

$$\lambda_{\max}(A) := \max_{i\in[N]} |\lambda_i(A)| = |A|_{op}.$$

{Assumption

Assumption 2.3. There exists a probability measure μ such that for every $k \in \mathbb{N}$

$$\lim_{N\to\infty} \int x^k d\mu_N(x) = \lim_{N\to\infty} \frac{1}{N}\mathbf{E}\,\mathrm{tr}\left(\frac{A_N}{\sqrt{N}}\right)^k = \begin{cases} \int x^k d\mu(x) & k \in 2\mathbb{N}, \\ 0 & \text{else.} \end{cases} \tag{2.2}$$ {ESDMoments}

Moreover, set

$$\mu_\infty := \lim_{k\to\infty} \left(\int x^{2k} d\mu(x)\right)^{1/2k}. \tag{2.3}$$ {SupOfSuppt}

The measure μ has compact support as we will show below [see (3.6)], and since its odd moments are zero, μ is symmetric. Thus μ_∞ is finite and equals the maximum of the support of μ.
Assumptions 2.1 provide some sufficient condition for the entries of the matrix to be controllable. Assumptions 2.3 assumes the convergence of the empirical spectral distribution of sequence of the matrices. Both of them are more or less standard and can be found in the literature of Wigner-type matrices with general variance, see for example [27]. Next we give some sufficient conditions in order for the largest eigenvalue to converge. The main difficulty which the next conditions will try to address is how to compare high order moments of the matrix with μ_∞.

{Assumptions

Assumption 2.4. For every $N \in \mathbb{N}$ and $i, j \in [N]$ it is true that

$$s_{i,j}^{(N)} \leq \min\{s_{2i,2j}^{(2N)}, s_{2i-1,2j}^{(2N)}, s_{2i-1,2j-1}^{(2N)}\}$$

In order to give the next sufficient condition we first give some necessary definitions.

{orismos gra

Definition 2.5. We call graphon any measurable function $W : [0, 1] \times [0, 1] \to \mathbb{R}$ which is symmetric and integrable.

Note that any $N \times N$ matrix A_N which satisfies Assumptions 2.1 defines a graphon as follows

$$W_N(x, y) := s_{\lceil Nx\rceil,\lceil Ny\rceil}^{(N)}.$$

Definition 2.6. For any graphon W and any multigraph $G = (V, E)$, the isomorphism density from G to W is

$$t(G, W) = \int_{[0,1]^{|V|}} \prod_{i,j\in E} W(x_i, x_j) \prod_{i\in|V|} dx_i$$

Definition 2.7. Define $\mathbf{C}_k$ to be the set of all non-isomorphic plane rooted trees with $k+1$ vertices, i.e. all trees with $k+1$ vertices, a vertex distinguished as a root and an ordering amongst the children of any vertex. It is known that the number of such trees is the $k-$th Catalan number, i.e.,

$$|\mathbf{C}_k| = \frac{1}{k+1}\binom{2k}{k}, \tag{2.4}$$ {CatalNum}

and a trivial bound that we will use is $|\mathbf{C}_k| \leq 2^{2k}$.

Let A_N be a sequence of matrices that satisfy Assumptions 2.1 and call W_N the isomorphism density of A_N as is defined after Definition 2.5.

{Assumption

Assumption 2.8. There exists a graphon W such that

$$\lim_{N\to\infty} t(T, W_N) = t(T, W)$$

for any finite tree T. Moreover for the tree with two vertices and one edge, i.e. $T \in \mathbf{C}_1$, it is true that for any $D > 0$ there exists some $N_0 = N_0(D)$ such that for any $N \geq N_0$ it is true that

$$|t(T, W_N) - t(T, W)| \leq O(N^{-D}) \tag{2.5}$$ {fast conver

Remark 2.9. Note that in Assumption 2.8 we do not have to assume convergence of the E.S.D. of the matrix because this is ensured by the Assumption of the convergence of the graphon for every finite tree. This will be explained in subsection 4.2.

Definition 2.10. NA OXI DEFN For any $N \in \mathbb{N}$ and any two $N \times N$ matrices A, B we will denote $A \odot B$ their Hadamard product which is the $N \times N$ matrix with entries the entry-wise product of A, B, i.e.,

$$\{A \odot B\}_{i,j} = \{A\}_{i,j}\{B\}_{i,j}$$

The assumptions we made so far will lead to convergence in probability of the largest eigenvalue. Next we will give some extra condition, which will lead to the almost sure convergence of the largest eigenvalue.

{assumfora.s

Assumption 2.11. Suppose that A_N is a sequence of symmetric random matrices, with independent entries (up to symmetry), such that there exist a random variable X with mean 0, variance 1 and finite 4-th moment which stochastically dominates the entries of A_N in the following sense

$$\mathbf{P}(|\{A_N\}_{i,j}| \geq t) \leq \mathbf{P}(|X| \geq t), \text{ for all } t \in [0, \infty] \tag{2.6}$$ {stochdom}

and for any $N \in \mathbb{N}$ and any $i, j \in [N]$.

Instead of Assumption 2.11, an easier to check (but stronger) assumption for a model of random matrices is the following.

{a.s.remeasi

Remark 2.12. Note that if A_N can be written as the Hadamard product of two matrices Σ_N and A'_N, where A'_N is a sequence of symmetric random matrices with i.i.d. entries all following the same law, with 0 mean, unit variance and finite $4-$th moment and for each N the entries Σ_N belong to the set $[0, 1]$, then Assumption 2.11 will hold.

We are now ready to present our first main result.

{to theorima

Theorem 2.13. *Let A_N be a sequence of matrices satisfying Assumption 2.1. Then if either Assumptions 2.3 and 2.4 hold or Assumption 2.8 holds, it is true that*

$$\lim_{N\to\infty} \lambda_{\max}\left(\frac{A_N}{\sqrt{N}}\right) = \mu_\infty \text{ in probability} \tag{2.7}$$ {siglisi sto

Moreover if the sequence of matrices A_N satisfy Assumption 2.11 the convergence in (2.7) *improves from probability to almost surely.*

Note that Assumption (2.4) is restrictive and does not cover a lot of the well-known and studied models. So in what follows we try to take advantage of Theorem 2.13 and adjust it in order to prove the convergence of the largest eigenvalue for a a general class of random matrix models. We first give a definition.

{defngenvar}

Definition 2.14. Let A_N be a sequence of matrices for which Assumptions 2.1 hold. Moreover suppose that there exists an integer valued sequence d_N for which it is true that $\lim_{N\to\infty} \frac{d_N}{N} = 0$ and such that for each N there are d_N- orthogonally convex and closed $\{\mathcal{A}^{(N)}\}_{i\in[d_N]}$, subsets of $[0,N]^2$ with the following properties.

- It is true that
$$\mathcal{A}_1^{(N)} \subseteq \mathcal{A}_2^{(N)} \subseteq \cdots \subseteq \mathcal{A}_{d_N}^{(N)} = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le y \le N\} \tag{2.8}$$
- For any $N \in \mathbb{N}$, $m \in [d_N]$ there exists some there exists $f \in [d_{2N}]$ such that
$$\{2(x,y) : (x,y) \in \left(\mathcal{A}_m^{(N)} \setminus \cup_{l\le m-1}\mathcal{A}_l^{(N)}\right)^o \cap [N]^2\} \subseteq \left(\mathcal{A}_f^{(2N)} \setminus \cup_{l\le f-1}\mathcal{A}_l^{(2N)}\right)^o \tag{2.9}$$ {growth of i

  and

  Here the notation A^o stands for the interior of a set A.
- In what follows the notation bd(A) stands for the boundary of a set A. For every $m \in [d_N]$ the intersection of bd$(\mathcal{A}_m^{(N)})$ with any line segment which connects two elements $(x_1,y_1),(x_2,y_2) \in \{(x,y) \in [N]^2 : 0 < x \le y < N\}$ such that either $x_1 = x_2$ or $y_1 = y_2$, contain at most 2 elements.

For every $m \in [d_N]$ set
$$\mathcal{B}_m^{(N)} := \mathcal{A}_m^{(N)} \setminus \cup_{l\le m-1}\mathcal{A}_{l\le m-1}^{(N)}$$
Then if for all $(i,j) \in [N]^2$ the variance of the (i,j)-entry of A_N is given by
$$s_{i,j}^{(N)} := \sum_{m\in[d_N]} s_m^{(N)} 1_{(i\wedge j, i\vee j)\in\mathcal{B}_m^{(N)}} \tag{2.10}$$ {defnofvaria

for some set of numbers $\{s_i\}_{i\in[d_N]}$. We will call the sequence of matrices A_N, random matrix model whose variance profile is given by a generalized step function

The following Theorem is a corollary of Theorem 2.13 and covers the cases where the entries of the variance profile matrix are the values of a step function or a continuous function which is coordinate decreasing (or increasing).

{Theorima gi

Theorem 2.15. *Let A_N be a random matrix model whose variance profile is given by a generalized step function. Then if it also satisfies Assumptions 2.3 and for every $N \in \mathbb{N}$ and $\{i,j\}$ such that there exists some $m \in [d_N]$ such that if $(\min\{i,j\},\max\{i,j\}) \in (\mathcal{B}_m^{(N)})^o$ then*
$$s_{i,j}^{(N)} \le s_{2i,2j}^{(2N)} \tag{2.11}$$ {extrassucon

Then
$$\lim_{N\to\infty} \lambda_{\max}\left(\frac{A_N}{\sqrt{N}}\right) = \mu_\infty \quad \textit{in probability}$$
Moreover if the sequence of matrices A_N satisfy Assumption 2.11 the convergence in (2.7) *improves from probability to almost surely.*

Remark 2.16. The sequence d_N in Definition 2.14 can grow to infinity but each new set that emerges at some N_0 will not contribute to the entries of the matrix A_N, i.e., it will not contain a natural number until much later in the sequence. See for example the discussion in Remark 7.18.

Will we need a figure here as well here ? In the next corollary we prove that it is sufficient for a model to be well-approximated by sequence which satisfies one of the Assumptions above.

{symperasma

Corollary 2.17. *Let A_N be a sequence of matrices satisfying Assumptions 2.1 and 2.3. Write A_N as*
$$A_N = \Sigma_N \odot A'_N, \tag{2.12}$$
where Σ_N is the matrix with entries the standard deviation of A_N, i.e. $(\mathbf{E}\{A_N^2\}_{i,j})^{1/2} = \Sigma_{i,j}^{(N)}$ and A'_N is random symmetric matrix with independent entries all with zero mean, unit variance and finite $4-$th moment. Moreover, suppose that for every $n \in \mathbb{N}$ there exists a sequence of $\Sigma_N^{(n)}$ such that the sequence of matrices $A_N^{(n)} := \Sigma_N^{(n)} \odot A'_N$

satisfies Assumptions 2.1 and 2.3 and either Assumptions 2.4 either Assumptions 2.8 or the Assumptions of Theorem 2.15. Furthermore, for every $n \in \mathbb{N}$ denote by $\mu^{(n)}$ the limiting distribution of the E.S.D. of $A_N^{(n)}$ and by $\mu_\infty^{(n)}$ the largest element in the support of $\mu^{(n)}$. Suppose that

$$\lim_{n\to\infty} \mu_\infty^{(n)} = \mu_\infty, \quad \mu^{(n)} \Rightarrow \mu \quad \text{in distribution,}$$

$$\lim_n \limsup_N \max_{i,j} |\{\Sigma_N\}_{i,j} - \{\Sigma_N^{(n)}\}_{i,j}| = 0$$

. Then

$$\lim_{N\to\infty} \lambda_{\max}\left(\frac{A_N}{\sqrt{N}}\right) = \mu_\infty \quad \text{in probability}$$

Moreover again, if the sequence of matrices A_N' satisfy Assumption 2.11 the convergence in (2.7) improves from probability to almost surely.

Remark 2.18. Note that Theorem 2.15 covers the case that $s_{i,j}^N = \sigma(\frac{i}{N}, \frac{j}{N})$ for some symmetric step function $\sigma : [0, 1] \times [0, 1] \to [0, 1]$ and Corollary 2.17 covers the case that σ is symmetric and continuous. The last fact is true by Lemma 6.4 of [21], where the deviation matrices when σ is continuous is approximated by step functions.

3 Analysis of high order moments

{highordermo

We will relate the largest eigenvalue with a high moment of the measure μ_N and at the same time this moment will be controlled by μ_∞. We start by analysing the convergence in (2.2). In general, it is true that

$$\mathbf{E}\,\mathrm{tr}(A^{2k}) = \sum_{i_1, i_2, \dots\dots\dots, i_{2k} \in [N]} \mathbf{E} \prod_{l=1}^{2k} a_{i_l, i_{l+1}} \tag{3.1}$$ {geniko athr

with the convention that $i_{2k+1} = i_1$.

Now, for a term with indices $i_1, i_2, \dots, i_{2k}$, we let $\mathbf{i} := (i_1, i_2, \dots, i_{2k})$ and $X(\mathbf{i}) := \prod_{l=1}^{2k} a_{i_l, i_{l+1}}$. Then consider the graph $G(\mathbf{i})$ with vertex set

$$V(\mathbf{i}) = \{i_1, i_2, \dots, i_{2k}\},$$

and set of edges

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

For such an $\mathbf{i}$ we also use the term cycle.

As explained in [1] (in the proof of relation (3.1.6) there, pages 49, 50 or in Theorem 3.2 of [27]), the limit

$$\lim_{N\to\infty} \frac{1}{N^{k+1}} \mathbf{E}\,\mathrm{tr}(A^{2k})$$

remains the same if in (3.1) we keep only the summands whose indices $\mathbf{i}$ satisfy the following:

1. The graph $G(\mathbf{i})$ is a tree with $k + 1$ vertices.
2. The path $i_1 \to i_2 \to \cdots \to i_{2k} \to i_1$ traverses each edge of the tree exactly twice, in opposite directions of course.

$G(\mathbf{i})$ becomes an ordered rooted tree if we mark i_1 as the root and declare a child smaller than another if it appears earlier in the cycle.

Cycles $\mathbf{i}$ that don't satisfy either of 1 or 2 above we call bad cycles. So the sum in (3.1) can be written as

$$\mathbf{E}\,\mathrm{tr}(A^{2k}) = M_N(k) + B_N(k), \tag{3.2}$$ {xoristo ath

where

$$M_N(k) := \sum_{G \in \mathbf{C}_k} \sum_{\mathbf{i} \in [N]^{2k} : G_1(\mathbf{i}) \sim G} \prod_{\{i,j\} \in E(G_1(\mathbf{i}))} s_{i,j}^{(N)}, \tag{3.3}$$ {orismos kal

$$B_N(k) := \sum_{\mathbf{i} \in [N]^{2k} : \text{bad cycle}} \mathbf{E} X(\mathbf{i}). \tag{3.4}$$ {oi kakoi or

Here $\mathbf{C}_k$ are the ordered rooted trees with k edges and $G(\mathbf{i}) \sim G$ means that the graphs are isomorphic as ordered rooted trees.

Note that by the uniform bound on the variances of A_N it is true that

$$M_N(k) \le |\mathbf{C}_k| N^{k+1}, \tag{3.5}$$

which, with the use of (2.4), implies that

$$\limsup_{k\to\infty}\left(\limsup_{N\to\infty}\frac{M_N(k)}{N^{k+1}}\right)^{1/2k} \le 2. \tag{3.6}$$

{upboundcata

The plan is to control the expectation of the trace in (3.2) through an appropriate bound involving various $M_N(j)'s$. To control the term $B_N(k)$, we adopt the analysis of Section 2.3 of [25].

{protasi gia

Proposition 3.1. *Let A_N be a matrix that satisfies Assumption 2.1. Assume additionally that the absolute value of the entries of the matrix are all supported in $[0, CN^{\frac{1}{2}-\epsilon}]$ for some $\epsilon > 0$. Then for all N large enough and all $k < N$ it is true that*

$$|B_N(k)| \le \sum_{s=1}^{k}(4k^5)^{2k-2s}\left((CN)^{\frac{1}{2}-\epsilon}\right)^{2k-2s}\sum_{t=1}^{(s+1)\wedge k}(4k^4)^{4(s+1-t)}M_N(t-1). \tag{3.7}$$

{BadCycBnd}

Proof. We bound each term of the sum defining $B_N(k)$. Take a bad cycle $\mathbf{i}$ and let

- t: the number of vertices of $G(\mathbf{i})$,
- s: the number of the edges of $G(\mathbf{i})$,
- $e_1, e_2, \dots, e_s$: the edges of $G(\mathbf{i})$ in order of appearance in the cycle,
- $a_1, a_2, \dots, a_s$: the multiplicities of $e_1, e_2, \dots, e_s$ in the cycle.

That is, a_q is the number of times the (undirected) edge e_q appears in the cycle. Note that $t \le s+1$ (true for all graphs) and $t \le k$ because the cycle is bad.

Additionally, we let $T(\mathbf{i})$ be the rooted ordered tree obtained from $G(\mathbf{i})$ by keeping only edges that lead to a new vertex at the time of their appearance in the cycle. The root is i_1 and we declare a child of a vertex smaller than another if it appears earlier in the cycle.

To bound $|\mathbf{E}X(\mathbf{i})|$, notice that if any of $a_1, a_2, \dots, a_s$ is 1, we have $\mathbf{E}X(\mathbf{i}) = 0$ by the independence of the elements of A_N and the zero mean assumption. We assume therefore that all multiplicities are at least 2. Using the information about the mean, variance, and support of $|a_{i,j}^{(N)}|$, we get that for any integer $a \ge 2$ it holds $\mathbf{E}(|a_{i,j}^{(N)}|^a) \le (C_1 N^{1/2-\varepsilon})^{a-2} s_{i,j}^{(N)}$. Thus

$$\mathbf{E}|X(\mathbf{i})| = \prod_{i=1}^{s}\mathbf{E}|X_{e_i}|^{a_i} \le (C_1 N^{1/2-\varepsilon})^{a_1+\cdots+a_s-2s}\prod_{\{i,j\}\in E(G_1(\mathbf{i}))} s_{i,j}^{(N)} \le (C_1 N^{1/2-\varepsilon})^{2k-2s}\prod_{\{i,j\}\in E(T(\mathbf{i}))} s_{i,j}^{(N)}. \tag{3.8}$$

In the second inequality, we used the fact that $s_{i,j}^{(N)} \in [0,1]$ for all i, j, N. For integers $s, t \ge 1, a_1, \dots, a_s \ge 2$ and $T \in \mathbf{C}_{t-1}$ let

$$N_{T,a_1,a_2,\dots,a_s} = \begin{array}{l}\text{the number of bad cycles with } T(\mathbf{i}) \sim T, \text{ vertex set } \{1,2,\dots,t\},\\ \text{and edge multiplicities } a_1, a_2, \dots, a_s.\end{array} \tag{3.9}$$

Consequently,

$$\begin{aligned}
|B_N(k)| &\le \sum_{s=1}^{k}\sum_{t=1}^{k\wedge(s+1)}\sum_{a_1,a_2,\dots,a_s}(C_1 N^{1/2-\varepsilon})^{2k-2s}\sum_{T\in\mathbf{C}_{t-1}} N_{T,a_1,a_2,\dots,a_s}\sum_{\mathbf{i}\in[N]^{2k}:T(\mathbf{i})\sim T}\ \prod_{\{i,j\}\in E(T(\mathbf{i}))} s_{i,j}^{(N)} && (3.10)\\
&\le \sum_{s=1}^{k}\sum_{t=1}^{k\wedge(s+1)}\sum_{a_1,a_2,\dots,a_s}(C_1 N^{1/2-\varepsilon})^{2k-2s}\sum_{T\in\mathbf{C}_{t-1}} N_{T,a_1,a_2,\dots,a_s}\sum_{\mathbf{i}\in[N]^{2(t-1)}:T(\mathbf{i})\sim T}\ \prod_{\{i,j\}\in E(T(\mathbf{i}))} s_{i,j}^{(N)} && (3.11)\\
&\le \sum_{s=1}^{k}\sum_{t=1}^{k\wedge(s+1)}\sum_{a_1,a_2,\dots,a_s}(4k^4)^{4(s+1-t)+2(k-s)}(C_1 N^{1/2-\varepsilon})^{2k-2s}M_N(t-1). && (3.12)
\end{aligned}$$

{FirstGraphB

The inside sum is over all s-tuples of integers $a_1, a_2, \ldots, a_s$ 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 $s^{2(k-s)} \le k^{2(k-s)}$. Thus the above sum is bounded by

$$\sum_{s=1}^{k}(4k^5)^{2(k-s)}(C_1N^{1/2-\varepsilon})^{2k-2s}\sum_{t=1}^{k\wedge(s+1)}(4k^4)^{4(s+1-t)}M_N(t-1). \tag{3.13}$$

□

4 Proof of Theorem 2.13

By relation (3.2) and Proposition 3.1, it is clear that one needs to control the behaviour of the crucial part of high order traces, i.e., $M_N(k)$. More precisely, in order to give an upper bound on the largest eigenvalue we will study the behaviour of $M_N(k)$ when $k = O(\log^2(N))$. Firstly by assumption 2.3 one has that there is a probability measure μ which is symmetric and compactly supported such that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i(\frac{A_N}{\sqrt{N}})}(-\infty,x) = \mu(-\infty,x) \tag{4.1}$$

{liminf}

for all $x \in \mathbb{R}$ continuity points of the function $\mu(-\infty, x)$. This implies that

$$\liminf_N \lambda_{\max}(\frac{A_N}{\sqrt{N}}) \ge \mu_\infty \quad \text{a.s.}$$

where

$$\mu_\infty = |\mu|_{L_\infty} = \lim_{k\to\infty}\left(\int x^{2k}d\mu(x)\right)^{1/2k}$$

So in order to prove Theorem 2.13 one needs to prove that

$$\limsup_N \lambda_{\max}(\frac{A_N}{\sqrt{N}}) \le \mu_\infty \tag{4.2}$$

{H anisotita

We will prove (4.2) separately for each one of the assumptions 2.4 and 2.8.

4.1 Proof of Theorem 2.13 with Assumptions 2.4

In this subsection we will prove (4.2) under Assumptions 2.4. The following proposition is crucial.

{protasi gia

Lemma 4.1. *Let A_N be a sequence of matrices that satisfies Assumptions 2.1, 2.3 and 2.4. Then for every $k, N \in \mathbb{N}$ such that $k < N$ it is true that*

$$M_N(k) \le N^{k+1}\mu_\infty^{2k}$$

Proof. Fix $N, k \in \mathbb{N} : k < N$ and a tree $T \in \mathbf{C}_k$. Then, for each $\mathbf{d} := (d_1, d_2, \ldots, d_{k+1}) \in \{-1, 0\}^{k+1}$ consider the function

$$\phi_{\mathbf{d}} : [N]^{k+1} \to [2N]^{k+1}$$

with

$$\phi_{\mathbf{d}}(i_1, i_2, \cdots, i_{2k+1}) = 2(i_1, i_2, \cdots ., i_{2k+1}) + (d_1, d_2, \cdots, d_{k+1})$$

for all $i_1, i_2, \ldots, i_{k+1} \in [N]^{k+1}$. Note that each $\phi_{\mathbf{d}}$ is one to one and for different vectors $\mathbf{d}, \mathbf{d}' \in \{-1, 0\}^{k+1}$, the image of $\phi_{\mathbf{d}}$ is disjoint from that of $\phi_{\mathbf{d}'}$.
Lastly, by assumption 2.4 for any $\mathbf{d} \in \{-1, 0\}^{k+1}$ it is true that

$$\sum_{T\in\mathbf{C}_k}\sum_{i_1,i_2,\cdots i_{k+1}\in[N]^{k+1}}\prod_{i,j\in E(T)}s_{i,j}^{(N)} \le \sum_{T\in\mathbf{C}_k}\sum_{\mathbf{i}\in\phi_{\mathbf{d}}\left([N]^{k+1}\right)}\prod_{i,j\in E(T)}s_{i,j}^{(2N)} \tag{4.3}$$

{anisotita g

So if one sums over all possible trees in $\mathbf{C}_k$, (4.3) implies that

$$2^{k+1}M_N(k) = \sum_{d_1,d_2\cdots d_{k+1}}\sum_{T\in\mathbf{C}_k}\sum_{i_1,i_2,\cdots i_{k+1}\in[N]^{k+1}}\prod_{i,j\in E(T)}s_{i,j}^{(N)} \le \sum_{T\in\mathbf{C}_k}\sum_{\mathbf{i}\in\phi_{\mathbf{d}}\left([N]^{k+1}\right)}\prod_{i,j\in E(T)}s_{i,j}^{(2N)} \le M_{2N}(k). \tag{4.4}$$

{anisotita g

By applying (4.4) inductively, one can prove that for fixed $N, k \in \mathbb{N}$ the sequence

$$q_m := M_{2^m N}(k)/(2^m N)^{k+1}$$

is increasing in the variable m. So by (2.2) it is true that

$$\sup_m q_m = \lim_{m\to\infty} q_m = \mu_{2k}.$$

Lastly, by a trivial inequality on the p–norms of the limiting measure, one has that

$$\mu_{2k} \leq \mu_\infty^{2k}.$$

And the proof follows from the inequality $q_0 \leq \mu_{2k} \leq \mu_\infty^{2k}$. □

Next, we are going to give a more precise estimate on the high order moments of the matrices and hence we will give the necessary bound on the largest eigenvalue under an additional assumption on the support of the entries of the matrix. {anisotita m

Proposition 4.2. *Let A_N be a sequence of matrices that satisfy Assumptions 2.1, 2.3 and 2.4. Moreover assume that the entries of A_N are all supported in $[0, CN^{\frac{1}{2}-\delta}]$ for some constant C and some small $\delta > 0$. Then*

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = \mu_\infty \quad a.s.$$

Proof. First, we claim that for any $k = O(\log^2(N))$ it holds

$$\mathbf{E}\,\mathrm{tr}(A_N^{2k}) \leq N^{k+1}\mu_\infty^{2k}\{1 + o(1)\}^{2k}. \tag{4.5}$$ {teleytaia a

Indeed, combining Propositions 3.1 and 4.1, we get

$$\mathbf{E}\,\mathrm{tr}(A^{2k}) \leq \mu_\infty^{2k}N^{k+1} + \sum_{s=1}^{k}(4k^5)^{2k-2s}\left((CN)^{\frac{1}{2}-\epsilon}\right)^{2k-2s}\sum_{t=1}^{(s+1)\wedge k}(4k^4)^{4(s+1-t)}N^t\mu_\infty^{2(t-1)}. \tag{4.6}$$ {anisotita p

Note that if $\mu_\infty = 0$ then (4.5) is implied by NA. So we assume that $\mu_\infty > 0$. Next, we focus on the second summand in the right hand side of the previous inequality, for N large enough. By a trivial bound on the geometric series one has that the right hand side in (4.6) is bounded by

$$\frac{2k^2}{\mu_\infty^2}\sum_{s=1}^{k}(4k^5)^{2k-2s}\left((CN)^{\frac{1}{2}-\delta}\right)^{2k-2s}(4k^4)^{4(s+1)}\left(\frac{N\mu_\infty^2}{(4k^4)^4}\right)^{(s+1)\wedge k} \tag{4.7}$$

$$= 2^9k^{18}N^k(\mu_\infty^2)^{k-1} + \frac{2k^2}{\mu_\infty^2}(N\mu_\infty^2)^{k+1}\sum_{s=1}^{k-1}\left(\frac{(4k^5)^2C^{1-2\delta}}{N^{2\delta}\mu_\infty^2}\right)^{k-s} \tag{4.8}$$

$$\leq 2^9k^{18}N^k(\mu_\infty^2)^{k-1} + 32C^{1-2\delta}k^{13}(\mu_\infty^2)^{(k-1)}N^{k+1-2\delta}, \tag{4.9}$$

and (4.5) follows.

Now we can prove (4.2). Fix $\epsilon > 0$ and pick $k := [C_1 \log N]$, where $C_1 > 1 + (2/\epsilon)$ [obviously this k is $O((\log N)^2)$]. By Markov's inequality and (4.5), one has that

$$\mathbb{P}\left(\lambda_{\max}(\frac{A_N}{\sqrt{N}}) \geq \mu_\infty(1+\epsilon)\right) \leq \mathbb{P}\left(\lambda_{\max}^{2k}(\frac{A_N}{\sqrt{N}}) \geq \mu_\infty^{2k}(1+\epsilon)^{2k}\right) \leq \frac{1}{\mu_\infty^{2k}(1+\epsilon)^{2k}}\frac{1}{N^k}\mathbf{E}\lambda_{\max}^{2k}(A_N) \tag{4.10}$$

$$\leq \frac{1}{\mu_\infty^{2k}(1+\epsilon)^{2k}}\frac{1}{N^k}\mathbf{E}\,\mathrm{tr}(A_N^{2k}) \leq N\left(\frac{1+o(1)}{1+\epsilon}\right)^{2k} \tag{4.11}$$

Note that the expression $N\left(\frac{1+o(1)}{1+\epsilon}\right)^{C_1\log N}$ is summable bacause of the choice of C_1, which implies that

$$\limsup_N \lambda_{\max}\left(\frac{A_N}{\sqrt{N}}\right) \leq \mu_\infty \text{ a.s.},$$

since ϵ was arbitrary, and completes the proof. □

Proof of Theorem 2.13 with Assumptions 2.4. . Fix $\epsilon > 0$ and define the matrices $A_N^{\leq}, A_N^{>}$ as the ones whose (i,j) element is $a_{i,j}^N 1_{|a_{i,j}| \leq N^{\frac{1}{2}-\epsilon}}$ and $a_{i,j}^N 1_{|a_{i,j}| > N^{\frac{1}{2}-\epsilon}}$ respectively. For a random matrix $H := (h_{i,j})$, $\mathbf{E}H$ denotes the matrix whose (i,j) element is $\mathbf{E}h_{i,j}$ provided that the mean value of $h_{i,j}$ can be defined.

Weyl's inequality (Theorem 4.3.1 in [20]) gives that

$$\frac{1}{\sqrt{N}}\lambda_{\max}(A_N) \leq \frac{1}{\sqrt{N}}\left(\lambda_{\max}(A_N^{\leq} - \mathbf{E}A_N^{\leq}) + \lambda_{\max}(\mathbf{E}A_N^{\leq}) + \lambda_{\max}(A_N^{>})\right). \tag{4.12}$$ {anisotita w

The second and the third summands in the right hand side of (4.12) , due to Assumptions (2.1), can be proven that they are asymptotically negligible (in probability) completely analogously to the proof of Theorem 2.3.23 in [25]. In particular one can show that

- The (deterministic) sequence $N^{-\frac{1}{2}}\lambda_{\max}(\mathbf{E}(A_N^{\leq}))$ converges to 0. Write down why.
- The sequence $N^{-\frac{1}{2}}\lambda_{\max}((A_N^{\geq}))$ converges to 0 in probability.

To deal with the first summand, which refers to the matrix $A_N^{\leq} - \mathbf{E}A_N^{\leq}$, we will use Proposition 4.2. Notice that due to Lemma 3.6 of [27], the empirical spectral distributions of the matrices $(A_N^{\leq} - \mathbf{E}A_N^{\leq})/\sqrt{N}, A_N/\sqrt{N}$ converge weakly to the same probability measure almost surely. Moreover by Weyl's inequality one has that (**How exactly? Needs justification Because of the new definition of lamda(max)**) Next let

$$\tilde{A}_N := A_N^{\leq} - \mathbf{E}A_N^{\leq}, \tag{4.13}$$ {orismos tou

$$s_{i,j}^{(N),\leq} := \mathbf{E}\{(\tilde{A}_N)_{i,j}^2\}. \tag{4.14}$$

So by a direct application of Proposition 3.1, one has that for any $k < N$

$$\mathbf{E}\,\mathrm{tr}(\tilde{A}_N^{2k}) \leq \tilde{M}_N(k) + \sum_{s=1}^{k}(4k^5)^{2k-2s}\left((CN)^{\frac{1}{2}-\epsilon}\right)^{2k-2s}\sum_{t=1}^{(s+1)\wedge k}(4k^4)^{4(s+1-t)}\tilde{M}_N(t-1) \tag{4.15}$$

where $\tilde{M}_N(m)$ are the terms (3.3) for $m \in [N]$ and for the matrix $\tilde{A}_N$. Furthermore,

$$s_{i,j}^{(N)} = \mathbf{E}(A_N^2) \geq \mathbf{E}\{(A_N^{\leq})^2\} \geq s_{i,j}^{(N),\leq},$$

so

$$\mathbf{E}\,\mathrm{tr}(\tilde{A}_N^{2k}) \leq M_N(k) + \sum_{s=1}^{k}(4k^5)^{2k-2s}\left((CN)^{\frac{1}{2}-\epsilon}\right)^{2k-2s}\sum_{t=1}^{(s+1)\wedge k}(4k^4)^{4(s+1-t)}M_N(t-1) \tag{4.16}$$

where $M_N(m)$ are the (3.3) for the matrix A_N. Thus, by Proposition 4.1 and similarly to the proof of (4.5) one can show that for $k = O(\log^2(N))$,

$$\mathbf{E}\,\mathrm{tr}(\tilde{A}_N^{2k}) \leq N^{k+1}\mu_\infty^{2k}(1+o(1))^{2k}.$$

Then, as in the proof of Proposition 4.2, we get

$$\limsup_N \lambda_{\max}\left(\frac{\tilde{A}_N}{\sqrt{N}}\right) \leq \mu_\infty \quad \text{, a.s. ,} \tag{4.17}$$ {limsupoftru

which ends the proof. □

4.2 Proof of Theorem 2.13 under Assumption 2.8

{to theorima

Firstly, note that

$$\sum_{T\in\mathbf{C}_k} t(T, W_N) = \sum_{T\in\mathbf{C}_k}\frac{1}{N^{k+1}}M_N(k).$$

Thus, since asymptotically the only contributing parts in the k-th moment of the E.S.D. are the terms $M_N(k)$, as is proven in Theorem 3.2 of [27], Assumption 2.3 is implied by Assumption 2.8.

Next we need an analogue of (4.5), and this is proven in the next proposition.

{Protasi gia

Proposition 4.3. *Suppose A_N is a sequence of random matrices such that Assumptions 2.1,2.8 hold. Moreover assume that the entries of A_N are all supported in $(0, CN^{\frac{1}{2}-\delta}]$ for some $\delta, C > 0$. Then for $k = O(\log(N))$ it is true that*

$$\mathbf{E}\,\mathrm{tr}\,A_N^{2k} \leq 2N^{k+1}\mu_\infty^{2k}\{1+o(1)\}^{2k}.$$

Proof. Firstly note that for $k < N$

$$\mu_{2k} := \lim_{N\to\infty} \frac{1}{N^{k+1}} M_N(k) = \sum_{T\in\mathbf{C}_k} \int_{[0,1]^{k+1}} \prod_{(i,j)\in E(T)} W(x_i, x_j) dx_1 dx_2 \cdots dx_{k+1}. \tag{4.18}$$

{oriakh roph

and similarly

$$\frac{1}{N^{k+1}} M_N(k) = \sum_{T\in\mathbf{C}_k} \int_{[0,1]^{k+1}} \prod_{(i,j)\in E(T)} W_N(x_i, x_j) dx_1 dx_2 \cdots dx_{k+1}. \tag{4.19}$$

Fix $T \in \mathbf{C}_k$. Enumerate the edges of T in the order of first appearance during a depth first search algorithm. For $\{i,j\} \in E(T)$ denote $\{i,j\}_{\text{ord}}$ to be its enumeration. So for any $0 \le l \le k$ define the following quantities.

$$\mu_N^{(l)}(k,T) = \int_{[0.1]^{k+1}} \prod_{(i,j)\in E(T):\{i,j\}_{\text{ord}}\le l} W_N(x_i, x_j) \prod_{(i,j)\in E(T):\{i,j\}_{\text{ord}}\ge l+1} W(x_i, x_j) dx_1 dx_2 \cdots dx_{k+1}. \tag{4.20}$$

Note that

$$\sum_{T\in\mathbf{C}_k} \mu_N^{(0)}(k,T) = \mu_{2k}, \quad \sum_{T\in\mathbf{C}_k} \mu_N^{(k)}(k,T) = \frac{M_N(k)}{N^{k+1}}. \tag{4.21}$$

Fix $l \in [k]$. Then since all the variances are uniformly bounded by 1 and by Assumption (2.5) it is true that for any $D > 0$ there exists some $N_0 = N_0(D)$ such that for $N \ge N_0$ it is true that

$$|\mu_N^{(l)}(k,T) - \mu_N^{(l-1)}(k,T)| \le \left| \int_{[0,1]^2} W_N(x,y) - W(x,y) dy dx \right| \le C_1 \frac{1}{N^D} \tag{4.22}$$

{IncrBound}

for some absolute constant C_1. Note that

$$\left| \frac{M_N(k)}{N^{k+1}} - \mu_{2k} \right| \le \sum_{T\in\mathbf{C}} \left| \sum_{l=1}^{k} \{\mu_N^{(l)}(k,T) - \mu_N^{(l-1)}(k,T)\} \right| \le \sum_{T\in\mathbf{C}_k} \sum_{l=1}^{k} |\mu_N^{(l)}(k,T) - \mu_N^{(l-1)}(k,T)| \tag{4.23}$$

$$\le |\mathbf{C}_k| k C_1 \frac{1}{N^{\epsilon_N}} \le \frac{C_1 k 2^{2k}}{N^D}. \tag{4.24}$$

In the last inequality we used the trivial bound $|\mathbf{C}_k| \le 2^{2k}$. Thus, by the previous inequality and the trivial inequality $\mu_{2k} \le \mu_\infty^{2k}$ one has that

$$\frac{M_N(k)}{N^{k+1}} \le \mu_\infty^{2k} + k 2^{2k} C_1 \frac{1}{N^D}. \tag{4.25}$$

{anisotita g

Note that for any $C' \in (0,\infty)$ if $k = C' \log N$ then for D large enough one has that

$$\lim_{N\to\infty} k \left(\frac{2}{\mu_\infty} \right)^{2K} C_1 \frac{1}{N^D} = 0$$

and so any $k = O(\log(N))$ and for N large enough one has that

$$\frac{M_N(k)}{N^{k+1}} \le 2\mu_\infty^{2k}.$$

Thus, if one applies (4.25) to Proposition 3.1 and for $k = O(\log(N))$ one has that

$$M_N(k) \le 2\mu_\infty^{2k} N^{k+1} + 2 \sum_{s=1}^{k} (4k^5)^{2k-2s} \left((CN)^{\frac{1}{2}-\epsilon} \right)^{2k-2s} \sum_{t=1}^{(s+1)\wedge k} (4k^4)^{4(s+1-t)} N^t \mu_\infty^{t-1}.$$

Now the proof of Proposition 4.3 will proceed similarly to the proof of (4.5) so it is omitted. □

Proof of Theorem 2.13 with Assumptions 2.8. Given Proposition 4.25 the proof continues analogously to the Proof of Theorem 2.13 with Assumptions 2.4 and therefore it omitted. □

4.3 Proof of almost sure convergence, under the extra Assumption 2.11

Until this point we have proven that if a sequence of random matrices A_N satisfies Assumptions 2.1, 2.3 and 2.4 or 2.1 and 2.8 then

$$\lim_{N\to\infty} \lambda_{\max}\left(\frac{A_N}{\sqrt{N}}\right) = \mu_\infty \text{ in probability.} \tag{4.26}$$

For any $\epsilon > 0$ and $C > 0$, define the matrices $A_N^{\leq}$, $\mathbf{E}A_N^{\leq}$ the matrix with the truncated entries of A_N at scale $CN^{\frac{1}{2}-\epsilon}$ and the matrix with entries their expected value $\mathbf{E}A_N^{\leq}$. See the precise definitions of this matrices in the Proof of Theorem 2.13 with assumptions 2.4. For these matrices we have proven the following facts.

- As is mentioned before the empirical spectral distributions of the matrices $(A_N^{\leq} - \mathbf{E}A_N^{\leq})/\sqrt{N}, A_N/\sqrt{N}$ converge weakly a.s. the same probability measure. So due to that and (4.17) one has that

$$\mu_\infty \leq \liminf_N \lambda_{\max}\left(\frac{A_N^{\leq} - \mathbf{E}A_N^{\leq}}{\sqrt{N}}\right) \leq \limsup_N \lambda_{\max}\left(\frac{A_N^{\leq} - \mathbf{E}A_N^{\leq}}{\sqrt{N}}\right) \leq \mu_\infty \ \text{ a.s.} \tag{4.27}$$

{a.s.ina.s.}

- For the deterministic sequence $\mathbf{E}A_N^{\leq}$ it is true that

$$\lim_{N\to\infty} \lambda_{\max}\left(\frac{\mathbf{E}A_N^{\leq}}{\sqrt{N}}\right) = 0 \tag{4.28}$$

{determin a.

Assume that the sequence of matrices A_N satisfy also Assumption 2.11. Let X be the random variable that stochastically dominates the entries of A_N in the sense of (2.6). Let X_N be a sequence of symmetric random matrices after an appropriate coupling such that for all $N \in \mathbb{N}$ and $i, j \in [N]^2$ it is true that

$$|a_{i,j}^{(N)}| \leq |\{X_N\}_{i,j}| \tag{4.29}$$

{anisotita s

and the entries of X_N are independent up to symmetry and all following the same law as X. Furthermore note that since the random variable X has finite 4−th moment it is true that for any $\epsilon \in (0, 1)$

$$\sum_{m=1}^{\infty} 2^{2m}\mathbf{P}(|X| \geq 2^{\frac{m}{2}(1-\epsilon)}) < \infty \tag{4.30}$$

{sumfinitea.

From (4.1) it is sufficient to show that for any $\epsilon > 0$

$$\mathbf{P}\left(\limsup_N \left(\frac{A_N}{\sqrt{N}}\right) \leq \mu_\infty + \epsilon\right) = 0 \tag{4.31}$$

{a.s.probzer

Due to (4.27), (4.28) and the triangular inequality for the operator norm of matrices, the probability in (4.31) is bounded by

$$\mathbf{P}\left(\lambda_{\max}\left(\frac{A_N}{\sqrt{N}}\right) \leq \lambda_{\max}\left(\frac{A_N^{\leq}}{\sqrt{N}}\right) + \epsilon\right) \leq \mathbf{P}\left(\text{For all } N \in \mathbb{N},\ \text{there exists } k \geq N : A_N \neq A_N^{\leq}\right) \tag{4.32}$$

$$= \lim_{k\to\infty} \mathbf{P}\left(\cup_{N\geq 2^k} \cup_{1\leq i\leq j\leq N} \{|a_{i,j}^{(N)}| \geq CN^{\frac{1}{2}-\epsilon}\}\right) \tag{4.33}$$

{i.o for A}

Now note that the quantity in (4.33) is bounded by

$$\lim_{k\to\infty} \mathbf{P}\left(\cup_{N\geq 2^k} \cup_{1\leq i\leq j\leq N} \{|\{X_N\}_{i,j}| \geq CN^{\frac{1}{2}-\epsilon}\}\right) = \mathbf{P}\left(\text{For all } N \in \mathbb{N},\ \text{there exists } k \geq N : X_N \neq X_N^{\leq}\right) \tag{4.34}$$

{i.o for X}

due to (4.29). In (4.34) the matrix $X_N^{\leq}$ is the truncation of the matrix X_N when the absolute values of the entries of X_N are smaller or equal to $CN^{\frac{1}{2}-\epsilon}$. The probability in the right hand side of (4.34) is known to be 0, see for example [1] there, pages 94 and 95. This last fact is true due to (4.30).

5 Proof of Theorem 2.15

Set

$$\mathcal{B}'_{(N)} := \cup_{m\in d_N} (\mathcal{B}_m^{(N)})^o, \tag{5.1}$$

$$\mathcal{B}_N := \{(x, y) \in \mathbb{R}^2 : (x + d_x, y + d_y) \in \mathcal{B}'_N \text{ for any } \{d_x, d_y\} \in \{-1, 0, 1\}^2\}, \tag{5.2}$$

and define the matrices

$$\{A_N^{(1)}\}_{i,j} = 1_{(i,j)\in\mathcal{B}_N}\{A_N\}_{i,j}, \quad \{A_N^{(2)}\}_{i,j} = 1_{(i,j)\notin\mathcal{B}_N}\{A_N\}_{i,j}$$

. One can use Weyl's inequality similarly to (**??**) **NA BALOUME CITE** and Theorem 2.13 to show that it suffices to prove that the sequence of matrices $A_N^{(1)}$ satisfies the Assumptions of Theorem 2.13 and that

$$\lim_{N\to\infty} \lambda_{\max}\left(\frac{A_N^{(2)}}{\sqrt{N}}\right) = 0 \quad \text{in probability} \tag{5.3}$$

{oi ligoi or

In order to establish (5.3), we will need the following lemma.

{anisotita p

Lemma 5.1. *For any $m \in \mathbb{N}$ and any plane rooted tree with m edges, $T \in \mathbf{C}_m$ enumerate the vertices of T according to their order of appearance during a depth first algorithm. We will say that two numbers $l, n \in [m+1]$ are connected in T if the $l-$th vertex of T is connected with the $l-$th vertex through an edge. Moreover set*

$$acc(N, m, T) := \{(i_1, \cdots i_{m+1}) \in [N]^{m+1} : \text{ if } l,n \text{ are connected in } T \text{ for some } n, l \in [m+1] \text{ then } \{i_l, i_n\} \in \mathcal{A}_N\}.$$

Then

$$|acc(N, m, T)| \leq N6d_N^m$$

Proof. For the proof of this lemma we will use induction.

For $m = 2$ we will show that the number of non-identically 0 elements per row is at most $6d_N$ and the implication will follow. Fix $i \in [N]$. The number of non-identical 0 entries in the $i-$th row is bounded by the number of times the lines $y = i$ and $x = i$ intersect with the boundary of the set $\mathcal{A}_m^{(N)}$ for any $m \in [d_N]$ times the number of entries which are at a distance at most 1 from that intersection. By the 3−rd property in Definition 2.10 we have that their are at most two such elements in the intersection for each segment. Moreover, for any such intersection there are at most 3 elements in the axis $y = i$ and 3 at most in the axis $x = i$ with natural coordinates of distance at most 1 from that intersection. Thus, the number of non-identical 0 entries in the $i-$th row is the bounded by $12d_N$.

Suppose that the desired inequality holds for any $l : l \leq m-1$. Fix $T \in \mathbf{C}_m$. Enumerate the vertices of T according to their appearance during a depth first search algorithm. Let v_{m+1} denote the last appearing leaf of T during a depth first search algorithm and let v_d denote its only neighbour in T, for some $d \in [m-1]$. Then the tree $T \setminus \{v_{m+1}\}$ is a plane rooted tree with $m-1$ edges for which the induction hypothesis holds. So

$$|\text{acc}(N, m, T)| \leq |\text{acc}(N, m-1, T \setminus \{v_{m+1}\})| \max_{i\in[N]} |\{(i_{m+1}, i_d) \in \mathcal{A}_N, \ \text{ given that } \ i_d = i\}| \leq N(12d_N)^{m-1}12d_N$$

□

So by the uniform bound on the variance of A_N in Assumptions 2.1 and Lemma 5.3 one has that for any $k < N$

$$M_N^{(2)}(k) \leq N2^k(12d_N)^k \tag{5.4}$$

{anisotita g

where $M_N^{(2)}(k)$ are the terms in (3.3) for the matrix $A_N^{(2)}$. Now we are ready to prove the Theorem.

Proof of Theorem 2.15. One can show that similarly to the proof of Theorem 2.13, it is sufficient to prove (5.3) under the extra assumption that the absolute value of the entries is bounded by $CN^{\frac{1}{2}-\epsilon}$ for some constants $C, \epsilon > 0$. So by Proposition 3.1 and (5.4) one has that for $k = O(\log^2 N)$

$$\mathbf{E}\,\text{tr}\{A^{(2)_N}\}^{2k} \leq N^{k+1}\left(\frac{12d_N}{N}\right)^k + \sum_{s=1}^{k}(4k^5)^{2k-2s}\left((CN)^{\frac{1}{2}-\epsilon}\right)^{2k-2s}\sum_{t=1}^{(s+1)\wedge k}(4k^4)^{4(s+1-t)}N^t\left(\frac{12d_N}{N}\right)^{t-1} \tag{5.5}$$

{anisotita p

The term on the right-hand side (5.5) can be treated completely analogously to right hand-side term of (4.6) in the Proof of Proposition 4.2. Thus, one can conclude that for arbitrary small $\delta > 0$ and for any $k = O(\log(N))$ it is true that

$$\mathbf{P}\left(\lambda_{\max}\left(\frac{A_N^{(2)}}{\sqrt{N}}\right) \geq \delta\right) \leq \frac{1}{(\sqrt{N}\delta)^{2k}}\mathbf{E}\lambda_{\max}^{2k}(A_N) \leq \frac{1}{(\sqrt{N}\delta)^{2k}}\mathbf{E}\,\text{tr}\{A_N^{(2)}\}^{2k} \leq N\left(\frac{o(1)}{\delta}\right)^{2k}$$

which implies (5.3).

Moreover fix $(i,j) \in [N]^2$ such that $i < j$. Then if $(i,j)] \subseteq \mathcal{B}_N$ there exists some $m \in [d_N]$ such that

$$\{(k,x) \in [N]^2 : \mathrm{dist}((i,j),(k,x)) \leq 1\} \subseteq (\mathcal{B}_m^{(N)})^o$$

But from Assumption (2.9) there exists some $f \in [d_{2N}]$ such that

$$\{(k,x) \in [2N]^2 : \mathrm{dist}((2i,2j),(k,x)) = 0 \text{ or } 2\} \subseteq (\mathcal{B}_f^{(2N)})^o$$

but since $(\mathcal{B}_f^{(2N)})^o$ is an orthogonally convex set one can conclude that

$$\{(k,x) \in [2N]^2 : \mathrm{dist}((2i,2j),(k,x)) \leq 2\} \subseteq (\mathcal{B}_f^{(2N)})^o$$

and in particular due to (2.10),(2.11) one can conclude that the sequence of matrices $A_N^{(1)}$ satisfies Assumption 2.4.

□

Almost sure convergence under the extra Assumption 2.11. To improve the convergence in Theorem 2.15 one can notice that if the matrix sequence of matrices A_N satisfy Assumption 2.11 the same will be true for the sequences of matrices $A_N^{(1)}$ and $A_N^{(2)}$. Note that due to Lemma 3.6 in [27] and (5.5), the almost sure convergence of the E.S.D. of the matrices $A_N^{(2)}$ to the measure which is concentrated at 0 is implied. So with these facts one can prove that the convergence in (5.3) can be improved to almost surely, similarly to the proof of the almost sure convergence in Theorem 2.13. Lastly one can notice that the matrix $A_N^{(1)}$ satisfy the assumptions for the almost sure convergence in Theorem 2.13 which ends the proof. □

6 Proof of Corollary 2.17

Firstly note that for any $n \in \mathbb{N}$

$$\lambda_{\max}\left(\frac{A_N^{(n)}}{\sqrt{N}}\right) \to \mu_\infty^{(n)}, \quad \text{in probability} \tag{6.1}$$ {anisotita k

Also if we extra assume that A'_N satisfies Assumption 2.11 the convergence in (6.1) improves into almost surely. Next fix $\epsilon > 0$ and n_0 large enough such that for every $n \geq n_0$ it is true that $\mu_\infty + \epsilon \geq \mu_\infty^{(n)}$. Then, due to Weyl's inequality, one has that

$$\mathbf{P}\left(\lambda_{\max}\left(\frac{A_N}{\sqrt{N}}\right) \geq \mu_\infty + 2\epsilon\right) \leq \mathbf{P}\left(\lambda_{\max}\left(\frac{A_N}{\sqrt{N}}\right) \geq \mu_\infty^{(n)} + \epsilon\right) \leq \mathbf{P}\left(\lambda_{\max}\left(\frac{A_N - A_N^{(n)}}{\sqrt{N}}\right) + \lambda_{\max}\left(\frac{A_N^{(n)}}{\sqrt{N}}\right) \geq \mu_\infty^{(n)} + \epsilon\right) \tag{6.2}$$

$$\leq \mathbf{P}\left(\lambda_{\max}\left(\frac{A_N - A_N^{(n)}}{\sqrt{N}}\right) \geq \frac{\epsilon}{2}\right) + \mathbf{P}\left(\lambda_{\max}\left(\frac{A_N^{(n)}}{\sqrt{N}}\right) \geq \mu_\infty^{(n)} + \frac{\epsilon}{2}\right) \tag{6.3}$$ {pano fragma

The second term in (6.3) is asymptotically negligible due to (6.1). For the first term one can show, similarly to the proof of Theorem 2.13, that it suffices to prove that $\lambda_{\max}(\frac{A_N - A_N^{(n)}}{\sqrt{N}})$ is negligible, when for each N the absolute values of the entries of the matrix $A_N - A_N^n$ are concentrated in $(0, CN^{\frac{1}{2}-\epsilon})$ for some $C, \epsilon > 0$. . But one can easily show that **Why? Is this ok?**

$$M'_N(t) \leq N^{t+1} 2^t \left(\max_{i,j} \{\Sigma_N\}_{i,j} - \{\Sigma_N^{(n)}\}_{i,j}\right)^t$$

where $M'_N(t)$ are the terms defined in 3.3 for the matrices $A_N - A_N^{(n)}$. In particular due to Proposition 3.1, one can show that the largest eigenvalue of $A_N - A_N^{(n)}$ is asymptotically negligible, similarly to the proof of Theorem 2.13.

So after increasing n_0 if necessary, we conclude that the proof.

The improvement to almost sure convergence under the extra Assumption 2.11 for the sequence of matrices A'_N can be proven similarly to the proof of the analogous part of Theorem 2.15 and therefore it is omitted.

7 Examples

In this section we present examples on which our theorems can be applied.

7.1 Wigner and Wigner-type matrices

Our theory would not be complete if it couldn't be applied to the most widespread random matrix models the Wigner matrices, i.e., the case that the entries of A_N are i.i.d. In what follows we establish the convergence of the largest eigenvalue of Wigner matrices perturbed by an error term matrix.

{theorempert

Theorem 7.1. *Let A^{∞} be an infinite symmetric matrix with i.i.d. entries all with zero mean, unit variance and bounded 4-th moment. Then define a sequence of increasing dimension matrices such that*

$$A_N = A_N^{(1)} + \Sigma_N \odot A_N^{(1)}$$

where $A_N^{(1)}$ is an $N \times N$ symmetric sub-matrix of A^{∞} and the entries of the matrices $A_N^{(2)} := \Sigma_N \odot A_N^{(1)}$ satisfy

$$\lim_{N\to\infty} \max_{i,j} |\mathbf{E}\{A_N^{(2)}\}_{i,j}^2| = 0. \tag{7.1}$$

{mikres dias

Then

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = 2 \ \ a.s.$$

Proof. Firstly note that the sequence of matrices $A_N^{(1)}$ satisfy Assumptions 2.1, Assumptions 2.3, 2.8 and 4.27. Note that its limiting distribution is the semicircle law. So by Theorem 2.13 one has that

$$\lambda_{max}(\frac{A_N^{(1)}}{\sqrt{N}}) = 2 \ \text{ a.s.}$$

By Weyl's inequality one has that

$$\max_{i\in[N]} |\lambda_i(\frac{A_N}{\sqrt{N}}) - \lambda_i(\frac{A_N^{(1)}}{\sqrt{N}})| \leq \max_i |\lambda_i(\frac{A_N^{(2)}}{\sqrt{N}})| = \lambda_{\max}(\frac{A_N^{(2)}}{\sqrt{N}})$$

So it is sufficient to show that the largest eigenvalue of $N^{-\frac{1}{2}}A_N^{(2)}$ tends to 0. Note that the sequence of matrices $A_N^{(2)}$ satisfy Assumption 2.11. Thus similarly to the proof of the almost sure convergence in Theorem 2.15, it is sufficient to prove that the largest eigenvalue of $N^{-\frac{1}{2}}A_N^2$ tends to 0 under the extra assumption that the absolute value of the entries of $A_N^{(2)}$ are supported in $(0, CN^{\frac{1}{2}-\epsilon})$, for some $C, \epsilon > 0$. So by Proposition 3.1 one can show that for $k = O(\log^2 N)$ it is true that

$$\mathbf{E}\,\mathrm{tr}\{A_N^{(2)}\}^{2k} \leq N^{k+1}\left(\max_{i,j} \mathbf{E}\{A_N^{(2)}\}^2\right)^k$$

which implies that the largest eigenvalue of $N^{-\frac{1}{2}}A_N^2$ tends to 0. □

7.2 Random matrix with variance profile given by a step function

The next example we provide are Random Matrices whose variance profile is given by a step function. It is also a widespread random matrix model.

{theoremstep

Theorem 7.2. *Let $\{A_N\}_{N\in\mathbb{N}}$ be a sequence of symmetric random matrices each of them with dimension N. Suppose there exist two other sequence of matrices Σ_N, A'_N such that $A_N = \Sigma_N \odot A'_N$. Suppose that the entries of A'_N for all N are independent (up to symmetry) and identically distributed random variables with zero mean, unit variance and finite 4-th moment. Moreover suppose that for each $i, j \in [N]$*

- $\{\Sigma_N\}_{i,j} = \sigma(\frac{i}{N}, \frac{j}{N})$ *for some* $\sigma : [0,1]^2 \to [0,1]$.
- *The function σ is symmetric, i.e. $\sigma(x,y) = \sigma(y,x)$ for any $x.y \in [0,1]$. Furthermore suppose that there exist $m \in \mathbb{N}$, disjoint intervals $\{I_i\}_{i\in[m]}$ such that $[0,1] = \cup_{i\in[m]} I_i$ and $\{s_{i,j}\}_{1\leq i\leq j\leq m} \in [0,1]^{\frac{m(m+1)}{2}}$ such that*

$$\sigma(x,y) = \sum_{i,j\in[m]} \sigma_{\max\{i,j\},min\{i,j\}} 1_{I_i}(x) 1_{I_j}(y)$$

Then the sequence of matrices A_N satisfy Assumptions 2.1. It also satisfies 2.3 due to Theorem 1.2 of [22], So there exists a limiting measure μ of the E.S.D. of $N^{-\frac{1}{2}}A_N$ with bounded support. Set μ_∞^σ to be the largest element of the support of μ. Then

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = \mu_\infty^\sigma \quad a.s.$$

Proof. It is sufficient to show that the deviation profile matrix Σ_N satisfy the Assumptions of Theorem 2.15. Note that for the number of times that the values of $\sigma(\frac{i}{N}, \frac{j}{N})$ may differ for $i, j \in [N]^2$ are at most m^2. Moreover the underlying sets from Definition 2.14 are

$$\mathcal{B}_{i,j}^{(N)} = \{(x, y) \in \mathbb{R}^2 : (\frac{x}{N}, \frac{y}{N}) \in (I_i^o \times I_j^o)\} \cap \{(x.y) \in \mathbb{R}^2 : 0 \le x \le y \le N\} \text{for any } i, j \in [m] \tag{7.2}$$ {defnofsetsv

It is obvious that the sets in 7.2 satisfy the first two assumptions of Definition 2.14. For the third one, we may easily modify the sets $\{I_i\}_{i\in[m]}$ to some intervals $\{I_i^{(N)}\}_{i\in[m]}$ depending on N, but for which the edges of the sets are irrational numbers and such that the variances of the matrices A_N will remain the same. This implies the third assumption of Definition 2.14. Lastly note that the sequence of matrices satisfy the Assumption 2.11, see Remark 2.12. So we can apply Theorem 2.15 and complete the proof. □

{coraproxste

Corollary 7.3. *Suppose that A_N is a sequence of matrices such that it can decomposed in the following sense*

$$A_N = A_N^{(1)} + \Sigma_N \odot A_N^{(1)},$$

where $A_N^{(1)}$ satisfies the assumptions of Theorem 7.2 and the matrices $A_N^{(2)} := \Sigma_N \odot A_N^{(1)}$ satisfy $\lim_{N\to\infty} \max_{i,j} |\mathbf{E}\{A_N^{(2)}\}_{i,j}^2| = 0$. *Then*

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = \mu_\infty^\sigma \quad a.s.$$

Proof. The proof is completely analogous to the proof of Theorem 7.1 and therefore it is omitted. □

7.3 Random matrix with variance profile given by a continuous function

The next example we provide are Random Matrices whose variance profile is given by a the continuous functions. It is also a widespread random matrix model.

{theoremconf

Theorem 7.4. *Let $\{A_N\}_{N\in\mathbb{N}}$ be a sequence of symmetric random matrices each of them with dimension N. Suppose there exist two other sequence of matrices Σ_N, A_N' such that $A_N = \Sigma_N \odot A_N'$. Suppose that the entries of A_N' for all N are independent (up to symmetry) and identically distributed random variables with finite 4-th moment. Moreover suppose that for each $i, j \in [N]$*

- $\{\Sigma_N\}_{i,j} = \sigma(\frac{i}{N}, \frac{j}{N})$ *for some* $\sigma : [0, 1]^2 \to [0, 1]$.
- *The function σ is symmetric and continuous.*

Then the sequence of matrices A_N satisfy Assumptions 2.1. It also satisfy 2.3 and so there exists a limiting measure μ of the E.S.D. of $N^{-\frac{1}{2}}A_N$ with bounded support. Set μ_∞^σ to be the largest element of the support of μ. Then

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = \mu_\infty^\sigma \quad a.s.$$

Proof. It is sufficient to show that the deviation profile matrix Σ_N satisfy the Assumptions of Corollary 2.17. This is true by approximating the matrix $\Sigma_N \odot A_N'$ by matrices $\Sigma_N^{(n)} A_N'$ where the entries of $\Sigma_N^{(n)}$ are given by some appropriately chosen step functions which will depend on n, similarly to the proof of Lemma 6.4 of [22]. □

{coraproxcon

Corollary 7.5. *Suppose that A_N is a sequence of matrices such that it can decomposed in the following sense*

$$A_N = A_N^{(1)} + \Sigma_N \odot A_N^{(1)},$$

where $A_N^{(1)}$ satisfies the assumptions of Theorem 7.4 and the entries of the matrices $A_N^{(2)} := \Sigma_N \odot A_N^{(1)}$ satisfy the following

$$\lim_{N\to\infty} \max_{i,j} |\mathbf{E}\{A_N^{(2)}\}_{i,j}^2| = 0.$$

Then

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = \mu_\infty^\sigma \quad a.s.$$

Proof. The proof is completely analogous to the proof of Theorem 7.1 and therefore it is omitted. □

Remark 7.6. In Theorem 1.3 [22] the author proved the analogous results to Corollaries 7.3 and 7.5 under the extra assumption that the entries of the matrix A'_N have sub-Gaussian Laplace and in particular finite moments. transforms.

7.4 Generalized step functions, more examples

In the Random Matrix Theory literature what is commonly described as Random matrices with variance-profile given by a step function are more or less what we describe in Theorem 7.2. In this subsection we give some examples which are covered by the generalized version of this variance-profile matrices but not from the "standard" step functions.

Theorem 7.7 (Non-Periodic Band Matrices with Bandwidth proportional to the dimension)**.** *Let A'_N be a sequence of symmetric random matrices with i.i.d entries all following a law with* 0 *mean, unit variance and finite 4-th moment. Set A_N to be the matrix with entries*

$$\{A_N\}_{i,j} = 1_{|i-j|\le pN}\{A'_N\}_{i,j}, \quad i,j \in [N],$$

for some $p \in (0,1]$. Then for the sequence of matrices A_N Assumptions 2.1 hold. Moreover due to Theorem 4 of [7] Assumptions 2.3 also hold for the sequence of matrices A_N. So

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = \mu_\infty \quad a.s.$$

Proof. This is a straightforward application of Theorem 2.15, where the underlying sets described in Definition 2.14 are

$$\mathcal{A}_1^{(N)} := \{(x,y) \in [0,N]^2 : |x-y| \le p\} \cap \{(x,y) \in \mathbb{R}^2 : x \le y\}, \text{ and } \mathcal{A}_2^{(N)} := \{(x,y) \in \mathbb{R}^2 : 0 \le x \le y \le N\}$$

and the underlying numbers are $s_1 = 1$ and $s_2 = 0$. □

Remark 7.8. The random band matrix models have been extensively studied after the novel work in [7] and have tremendous application in various research areas. When the bandwidth of the matrices is periodic, i.e., the distance from the diagonal outside which the entries are 0 is periodic, the operator norm of such matrices has been extensively studied, see for example [24] or the survey [10]. Moreover when the bandwidth of such matrices is non-periodic but the bandwidth (the maximum non identically zero entries per row) is $o(N)$ but tends to infinity has also been examined in [5]. To the best of our knowledge the convergence of the largest eigenvalue of non-Periodic Band Matrices with Bandwidth proportional to the dimension was not established.

7.4.1 Random Gram Matrices

Lets X be an $N \times M$ matrix with independent, centered entries with unit variance, where $\frac{M}{N}$ converges to some positive constant as $N \to \infty$. It is known that the empirical spectral distribution of XX^T, after rescaling. converges to the Marˇcenko-Pastur law [23]. Moreover the convergence of largest eigenvalue to the largest element of its support has been established, see for example [1], However, some applications in wireless communication require understanding the spectrum of XX^T, where X has a variance profile, see for example [19] or [14]. Such matrices are called random Gram matrices. In this subsection we establish the convergence of the largest eigenvalue of random Gram matrices to the largest element of the support of its limiting distribution, assuming that there exists a limiting distribution for the E.S.D. Firstly we give some necessary definition.

{defnfewbadp

Definition 7.9. Let $A_{N,M}$ be a sequence of $N \times M$ matrices, where $M = \lceil cN \rceil$ for some constant $c \in [0, 1]^+$ and for N large enough. Suppose that the matrix

$$\begin{bmatrix} \mathbf{O}_{N,N} & \begin{bmatrix} A^T_{N,M} \\ \mathbf{0}_{N-M,N} \end{bmatrix} \\ \begin{bmatrix} A_{N,M} & \mathbf{0}_{N,N-M} \end{bmatrix} & \mathbf{O}_{N,N} \end{bmatrix}$$

are random matrices with a variance profile given by a generalized step function, see Definition 2.14, where $\mathbf{O}_{k,l}$ is an $k \times l$ matrices with all of its entries equal to 0. The we will say that the random matrices $A_{N,M}$ are non-symmetric random matrices with their variance profile given by a generalized step function.

{thmgrammatr

Theorem 7.10. *Let $A_{N,M}$ be a sequence of $N \times M$ matrices, where $M = \lceil cN \rceil$ for some constant $c \in [0, 1]^+$ and for N large enough. Suppose that the entries of $A_{N,M}$ are independent and satisfy Assumptions 2.1. Furthermore suppose that there exists a non-trivial probability measure μ with compact such that the E.S.D. of the matrices $N^{-1} A_{N,M} A^T_{N,M}$ converges to μ. Set μ_∞ to be the largest element in the support of μ. Then if one of the next Assumptions holds*

- *The matrices $A_{N,M}$ are non-symmetric random matrices with their variance profile given by a generalized step functions*
- *The matrix $A_{N,M}$ can be decomposed in the following sense NA BALO STHN ARXH HADAMARD PRODUCT $A_{N,M} = \Sigma_{N,M} \odot A'_{N.M}$, for two sequence of $N \times M$ matrices $\Sigma_{N,M}, A'_{N,M}$ such that the entries of $A'_{N,M}$ are centered, independent random variables with unit variance and bounded 4-th moment and $\Sigma_{N,M}$ is a sequence of matrices with values given by a continuous function similarly to Theorem 7.4.*

It is true that

$$\lambda_{\max}\left(\frac{A_{N,M} A^T_{N,M}}{N}\right) \to \mu_\infty \quad \text{in probability} \tag{7.3}$$ {convofgram

Moreover if there exists a random variable X with 0 mean, unit variance and finite 4-th moment that stochastically dominates, in the sense of 2.6, the entries of $A_{N,M}$ for all N, M then the convergence in (7.3) can be improved to almost surely.

Proof. Firstly set

$$\mathbf{A}_N := \begin{bmatrix} A_{N,M} & \mathbf{0}_{N,N-M} \end{bmatrix}$$

where $0_{N,N-M}$ denotes the $N \times N - M$ matrix with 0 one every entry. Note that the matrix $\mathbf{A}_N$ symmetric so that $A_{N,M} A^T_{N.M} = \mathbf{A}_N \mathbf{A}^T_N$. Next define the sequence of matrices

$$\tilde{A}_N = \begin{bmatrix} 0 & \mathbf{A}^T_N \\ \mathbf{A}_N & 0 \end{bmatrix}$$

Note that

$$\det(\lambda^2 \mathbb{I}_N - \mathbf{A}_N \mathbf{A}^T_N) = \det(\begin{bmatrix} \lambda \cdot \mathbb{I}_\mathbb{N} & \mathbf{A}^T_N \\ \mathbf{A}_N & \lambda \cdot \mathbb{I}_\mathbb{N} \end{bmatrix}) = \det(\lambda \cdot \mathbb{I}_{2N} + \tilde{A}_N)$$

where $\mathbb{I}_k$ is the $k \times k$ identity matrix for any $k \in \mathbb{N}$. Thus, the eigenvalues of $\tilde{A}_N$ are the square root of $\mathbf{A}_N \mathbf{A}^T_N$ and their negative ones. The matrix $\tilde{A}_N$ is called the symmetrization of $\mathbf{A}_N$.

Now the sequence of matrices $\tilde{A}_N$ satisfy Assumptions 2.1. Moreover the Assumptions 2.3 are also satisfied and if ν is the limiting probability measure of the E.S.D. of $\tilde{A}_N$, then if $X \sim \nu$ and $Y \sim \mu$ it is true that X^2 has the same law as Y.

Lastly note that Definition (7.9) implies the Assumptions of Theorem 2.15 for the sequence of Matrices $\tilde{A}_N$ and the second Assumption implies the Assumptions of Theorem 7.4. Thus by a direct application of one of the Theorems above the proof is complete. □

Remark 7.11. In [19] the authors showed that if the variances of the entries of $A_{N,M}$ are given by the values of a continuous function (and some extra assumptions such as bounded $4+\epsilon$ moments of the entries) the limiting distribution of the E.S.D. of $A_N A_N^T$ does exist. So in Theorem 7.10 we prove the convergence of the largest eigenvalue of these models as well. The authors in [19] also studied the non-centered version of these models, i.e. when the entries of the matrix do not have 0 mean, but we do not cover this case with our result.

Corollary 7.12 (Triangular matrices)**.** *Let $\{A_N\}_{N\in\mathbb{N}}$ be a sequence of $N\times N$ lower triangular matrices. with iid entries all following a law with* 0 *mean, unit variance and finite 4-th moment. Then*

$$\lim_{N\to\infty} \lambda_{max}\left(\frac{A_N A_N^T}{N}\right) = e \quad a.s.$$

Proof. As is also explained in the proof of Theorem 7.10 the asymptotic behavior of the eigenvalues a sample covariance random matrix model, i.e. $A_N A_N^T$, can be equivalently described as the asymptotic behavior of symmetric matrices with independent entries. More precisely set

$$\tilde{A}_N = \begin{bmatrix} 0 & A_N^T \\ A_N & 0 \end{bmatrix}$$

The eigenvalues of $\tilde{A}_N$ are the square roots of the eigenvalues of $A_N A_N^T$ and their negative ones. Moreover in [11] it is shown that the E.S.D. of $N^{-1}A_N A^T$ converges to a probability measure whose support's largest element is e. Equivalently the largest element of the support of limiting measure of the E.S.D. of $\tilde{A}_N$ is $\sqrt{e}$, see Remark 2.2 of [9] for a more detailed discussion on this phenomenon. Moreover the sequence of matrices $\tilde{A}_N$ satisfies the Assumptions of Theorem 2.15, where the underlying sets described in Definition 2.14 are

$$\mathcal{A}_1^{(N)} := \{(x,y) \in [0,N^2 : x + \frac{N}{2} \le y\} \text{ and } \mathcal{A}_2^{(N)} := \{(x,y) \in \mathbb{R}^2 : 0 \le x \le y \le N\}$$

and the underlying number are $s_1 = 1$ and $s_2 = 0$. So after an application of Theorem 2.15 for the sequence of matrices $\tilde{A}_N$, the proof is completed. □

7.5 Random Block Matrices

The Random Block Matrix models have application in Modeling and Optimization problems. Their spectral properties have been investigated in [17] and [15], under various assumptions. Next, in [27] the author proves the convergence of the E.S.D. under very general conditions. Next we present the Random Block Matrices.

Definition 7.13. Let A and B be two matrices of dimension $N\times M$ and $K\times L$ respectively. Let $a_{i,j}$ be the (i,j)−th entry of A. We define the Kronecker product of A and B to be the $NK\times ML$ matrix

$$A\otimes B := \begin{bmatrix} a_{1,1}B & a_{1,2}B & \cdots a_{1,M}B \\ a_{2,1}B & a_{2,2}B & \cdots a_{2,M}B \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ a_{N,1}B & a_{N,2}B & \cdots a_{N,M}B \end{bmatrix}$$

Definition 7.14. Fix $d\in\mathbb{N}$ and a set of numbers $\{a_i\}_{i\in[d]} \subseteq (0,1]^d$ such that $\sum_{i\in[d]} a_i = 1$. Let A_N be an $N\times N$ matrix such that $A_N = \sum_{k,l\in[d]^2} E_{k,l}\otimes A_N^{k,l}$, where $E_{k,l}$ is the matrix with 1 in the (k,l)−entry and 0 in every other entry and $A_N^{(k,l)}$ is an $a_k^{(N)} \times a_l^{(N)}$ matrix under the extra convention that $\left(A_N^{(k,l)}\right)^T = A_N^{(L,K)}$. Each of the matrices $A^{(k,l)}$ has i.i.d. entries and the sequences $a_k^{(N)}$ are sequences such that {denf random

$$\lim_{N\to\infty} \frac{a_k^N}{N} = a_k, \text{ for } k\in[d]$$

and

$$\sum_{i=1}^{d} a_i^{(N)} = N \text{ for any } N\in\mathbb{N}$$

Then the sequence of matrices A_N will be called random block matrix model.

Theorem 7.15. *Let A_N be a random block matrix model. Assume that the entries of A_N satisfy the Assumptions 2.1. Moreover assume that there are some numbers $\{s_{i,j}\}_{(i,j)\in[d]^2}$ such that for the variance of the entries of the $(k,l)-th$ block is equal to $s_{k,l}$ for any $N \in \mathbb{N}$.*

Then there exists a probability measure μ with compact support such that the E.S.D. of $\frac{A_N}{\sqrt{N}}$ tends to μ. Set μ_∞ to be the largest element in the support of μ. Then

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = \mu_\infty \ a.s.$$

Proof. The existence of the limiting distribution μ follows from Theorem 6.1 of [27]. The convergence of the largest eigenvalue is a consequence of Corollary 7.3 with the underlying sets being

$$I_i = \left(\sum_{j<i} a_i \sum_{j\leq i} a_i\right], \text{ for any } i \in [d].$$

and the underlying values of the step function being $\{s_{i,j}\}_{(i,j)\in[d]^2}$ □

In the previous theorem the number of blocks of the matrix A_N is fixed. Next we give present an analogue of these matrices with the number of blocks tending to infinity. Under some extra assumptions we prove the convergence of the largest eigenvalue. Until this point we have presented examples for which Theorem 2.15 is applied but the number of orthogonally convex sets in Definition 2.14 is fixed. So next we give an example where this does not hold. {denf random

Definition 7.16. Fix a non-decreasing sequence of natural numbers, d_N tending to infinity. For each $N \in [N]$ fix a set of numbers $\{a_i^{(d_N)}\}_{i\in[d_N]} \subseteq (1,N]^{d_N}$ such that $\sum_{i\in[d_N]}\lfloor a_i^{(d_N)}\rfloor = N$. Let A_N be an $N\times N$ matrix such that $A_N = \sum_{k,l\in[d_N]^2} E_{k,l}\otimes A_N^{k,l}$, where $E_{k,l}$ is the matrix with 1 in the $(k,l)-$entry and 0 in every other entry and $A_N^{(k,l)}$ is an $\lfloor a_k^{(N)}\rfloor \times \lfloor a_l^{(N)}\rfloor$ matrix under the extra convention that $\left(A_N^{(k,l)}\right)^T = A_N^{(l,k)}$. Each of the matrices $A_N^{(k,l)}$ has i.i.d. entries whose distribution does not dependent on N and has 0 mean, $s_{k,l}$ variance and bounded $4-th$ moment. Moreover the sequences $a_k^{(d_N)}$ are such that

$$\lim_{N\to\infty} \frac{a_k^{(d_N)}}{N} = a_k, \text{ for } k \in \mathbb{N} \text{ and some } a_k \in [0,1]$$

such that

$$\sum_{i=1}^{\infty} a_i = 1.$$

Then the sequence of matrices A_N will be called random block matrix model with increasing block number. {thmrandombl

Theorem 7.17. *Let A_N be a random block matrix model with increasing block number. Assume that the matrices A_N satisfy the Assumptions 2.1, 2.3. Moreover the variances of the blocks of are fixed for all N, i.e. the variance of the entries of the k,l block are all equal $s_{k,l}^{(N)} = s_{k,l}$. the matrices each $\{M,L\}\in\mathbb{N}$ assign a number $s_{M,L}$ assume that for each N, there are some numbers $\{s_{i,j}^{(d_N)}\}_{(i,j)\in[d_N]^2}$ such that for the variance of the entries of the $(k,l)-th$ block of A_N is equal to $s_{k,l}^{(N)}$, so that $s^{(2N)_{2k,2l}} = s_{k,l}^{(N)}$.*

Furthermore assume that,

- *The number of blocks does not grow too fast, i.e.,*

$$\lim_{N\to\infty} \frac{d_N}{N} = 0$$

- *The growth of the sequences $a_k^{(d_N)}$ is almost linear meaning that there exists a N_0 such that for all $N \geq N_0$ it is true that for all $k \in [d_N]$*

$$\{2m : m \in \left(\sum_{i<k}\lfloor a_i^{(d_N)}\rfloor, \sum_{i\leq k}\lfloor a_i^{(d_N)}\rfloor\right] \cap \mathbb{N}\} = \{m \in \left(\sum_{i<k}\lfloor a_i^{(d_{2N})}\rfloor, \sum_{i\leq k}\lfloor a_i^{(d_{2N})}\rfloor\right] \cap 2\mathbb{N}\} \quad (7.4)$$ {growrateofb

Set μ_∞ to be the largest element in the support of μ. Then

$$\lim_{N\to\infty} \lambda_{\max}(\frac{A_N}{\sqrt{N}}) = \mu_\infty \ a.s.$$

{remnotfastg

Remark 7.18. The equality (7.4) allows the number of blocks to grow to infinity but the "growth" of each new interval will be very slow. For example say that the matrix A_{N_0} has d–blocks. For the first interval

$$(0, \lfloor a_1^{(d_{N_0})} \rfloor)$$

the value of $a^{(2N_0)}$ can be a number for which it is true that $\lfloor 2a^{(d_{N_0})} \rfloor = \lfloor x \rfloor$. So assuming that $a_1^{(d_{N_0})}$ is not a natural number, there exists some small $\epsilon > 0$ for which we can set

$$a_1^{(d_{2N_0})} := 2a_1^{(d_{2N_0})} - \epsilon$$

This allows us to add any number of disjoint new intervals whose union will be the interval $(a_1^{(d_{2N_0})}, 2a_1^{(d_{N_0})}]$ and their length will add up to ϵ, and still (7.4) will hold. For simplicity say we add 1 new interval. We can continue with that procedure and create new intervals of small length between all the previous d intervals. Again for simplicity assume that their length is all equal to ϵ. But these new intervals will not contain any natural numbers for any $N < O(\frac{1}{\epsilon})N_0$ and so they will not contribute to the blocks of the matrices A_N for all that N.

Proof of Theorem 7.17. The proof is a direct application of Theorem 2.15 and therefore it is ommited.

□

8 A lemma

{CountLemma}

In the next lemma, we prove the crucial estimate we invoked in the proof of Proposition 3.1. We adopt and present the terminology of Section 5.1.1 of [1].

{CycleCountL

Lemma 8.1. $N_{T,a_1,a_2,\ldots,a_s} \leq (4k^4)^{4(s+1-t)+2(k-s)}$

Note to us: We don't use the values of $a_1, a_2, \ldots, a_s$.

Proof. Take a cycle $\mathbf{i} := (i_1, i_2, \ldots, i_{2k})$ and assume that it has edge multiplicities $a_1, a_2, \ldots, a_s \geq 2$. Each step in the cycle we call a *leg*. More formally, legs are the elements of the set $\{(r, (i_r, i_{r+1})) : r = 1, 2, \ldots, 2k\}$, which become exactly the edges of $G(\mathbf{i})$ if we replace (i_a, i_{a+1}) with $\{i_a, i_{a+1}\}$.

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

T_1: if $i_{a+1} \notin \{i_1, \ldots, i_a\}$. I. e., the leg leads to a new vertex.

T_3: if there is a T_1 leg $(b, (i_b, i_{b+1}))$ with $b < a$ so that $a = \min\{c > b : \{i_c, i_{c+1}\} = \{i_b, i_{b+1}\}\}$. I. e., at the time of its appearance, it increases the multiplicity of a T_1 edge of $G(\mathbf{i})$ 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 $\{i_a, i_{a+1}\} = \{i_b, i_{b+1}\}$.
I.e., at the time of its appearance, it creates a new edge but leads to a vertex that has appeared already.

Moreover, a T_3 leg $(a, (i_a, i_{a+1}))$ is called *irregular* if there is exactly one T_1 leg $(b, (i_b, i_{b+1}))$ which has $b < a$, $v_a \in \{i_b, i_{b+1}\}$, 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 $G_1(\mathbf{i})$ is s and the number of its vertices is t (since $T(\mathbf{i}) \sim T \in \mathbf{C}_{t-1}$). Call

ℓ: the number of edges of $G_1(\mathbf{i})$ 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

$$r \leq 2m, \tag{8.1}$$ {rBound}

$$t = s + 1 - m \leq k, \tag{8.2}$$ {tBound}

$$|T_4| = 2m + 2(k - s). \tag{8.3}$$ {T4Bound}

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 $s \leq k$ (all edges of $G(\mathbf{i})$ have multiplicity at least 2) and if $s = k$ then $m \geq 1$ since the cycle is bad. For the last relation, note that $|T_3| = |T_1| = t - 1$ and thus, using (8.2) too, we have $|T_4| = 2k - 2(t - 1) = 2k - 2(s - m)$.

Now back to the task of bounding $N_{T,a_1,\dots,a_s}$. 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.

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.

We claim that, having $Q_1, Q_2, T(\mathbf{i})$ we can reconstruct the cycle $\mathbf{i}$.

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).

- 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

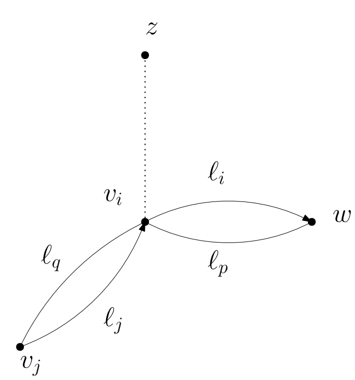

Figure 1: The case $z \notin U$.The legs $\ell_i, \ell_j (i < j)$ are T_3, while ℓ_p, ℓ_q are T_1. {graphCase}

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 1. 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 (**??**) is the elements i_r, j_r in the legs. This is determined in the next two steps.

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

$$(2kt^2(t+1))^{|T_4|}(2kt^2)^r \leq (4k^4)^{r+|T_4|}. \tag{8.4}$$ {MapCount}

Then (8.1) and (8.3) give $r + |T_4| \leq 4m + 2(k-s)$, and finally using (8.2), and we get the desired bound. □

References

[1] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.

[2] Z.-D. Bai and Y.-Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix. *The Annals of Probability*, pages 1729–1741, 1988.

[3] A. S. Bandeira, M. T. Boedihardjo, and R. van Handel. Matrix concentration inequalities and free probability. *arXiv preprint arXiv:2108.06312*, 2021.

[4] A. S. Bandeira and R. Van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479–2506, 2016.

[5] F. Benaych-Georges and S. Péché. Largest eigenvalues and eigenvectors of band or sparse random matrices. *Electronic Communications in Probability*, 19:1–9, 2014.

[6] P. Billingsley. *Convergence of probability measures*. Second edition, 1999.

[7] L. V. Bogachev, S. A. Molchanov, and L. A. Pastur. On the level density of random band matrices. *Mathematical notes of the Academy of Sciences of the USSR*, 50(6):1232–1242, 1991.

[8] A. Bose, K. Saha, A. Sen, and P. Sen. Random matrices with independent entries: beyond non-crossing partitions. *Random Matrices: Theory and Applications*, 11(02):2250021, 2022.

[9] A. Bose and P. Sen. XX^t matrices with independent entries. *arXiv preprint arXiv:2205.10811*, 2022.

[10] P. Bourgade. Random band matrices. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 2759–2783. World Scientific, 2018.

[11] D. Cheliotis. Singular values of triangular random matrices through the moment method. 2022.

[12] N. Cook, W. Hachem, J. Najim, and D. Renfrew. Non-hermitian random matrices with a variance profile (i): deterministic equivalents and limiting esds. *Electronic Journal of Probability*, 23:1–61, 2018.

[13] N. Cook, W. Hachem, J. Najim, and D. Renfrew. Non-hermitian random matrices with a variance profile (ii): Properties and examples. *Journal of Theoretical Probability*, 35(4):2343–2382, 2022.

[14] R. Couillet and M. Debbah. *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

[15] X. Ding. Spectral analysis of large block random matrices with rectangular blocks. *Lithuanian Mathematical Journal*, 54(2):115–126, 2014.

[16] L. Erdős, T. Krüger, and D. Schröder. Random matrices with slow correlation decay. In *Forum of Mathematics, Sigma*, volume 7. Cambridge University Press, 2019.

[17] R. R. Far, T. Oraby, W. Bryc, and R. Speicher. On slow-fading mimo systems with nonseparable correlation. *IEEE Transactions on Information Theory*, 54(2):544–553, 2008.

[18] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[19] W. Hachem, P. Loubaton, and J. Najim. A clt for information-theoretic statistics of gram random matrices with a given variance profile. *The Annals of Applied Probability*, 18(6):2071–2130, 2008.

[20] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

[21] J. Huang and H.-T. Yau. Edge universality of sparse random matrices. *arXiv preprint arXiv:2206.06580*, 2022.

[22] J. Husson. Large deviations for the largest eigenvalue of matrices with variance profiles. *Electronic Journal of Probability*, 27:1–44, 2022.

[23] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

[24] S. Sodin. The spectral edge of some random band matrices. *Annals of mathematics*, pages 2223–2251, 2010.

[25] T. Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[26] C. A. Tracy and H. Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.

[27] Y. Zhu. A graphon approach to limiting spectral distributions of wigner-type matrices. *Random Structures & Algorithms*, 56(1):251–279, 2020.