The limit of the operator norm for random matrices with a variance profile[*]

Dimitris Cheliotis[†]   **Michail Louvaris**[‡]

April 22, 2024

Abstract

In this work we study symmetric random matrices with variance profile satisfying certain conditions. We establish the convergence of the operator norm of these matrices to the largest element of the support of the limiting empirical spectral distribution. We prove that it is sufficient for the entries of the matrix to have finite only the 4-th moment or the $4 + \epsilon$ moment in order for the convergence to hold in probability or almost surely respectively. Our approach determines the behaviour of the operator norm for random symmetric or non-symmetric matrices whose variance profile is given by a step or a continuous function, random band matrices whose bandwidth is proportional to their dimension, random Gram matrices, triangular matrices and more.

1 Introduction

The problem of understanding the operator norm of a large random matrix with independent entries is multidisciplinary, occupying mathematicians, statisticians, physicists. On the mathematical side, tools from classical probability, geometric analysis, combinatorics, free probability and more have been used. The problem dates back to 1981, where in [20] the convergence of the largest eigenvalue of renormalized 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 [3], the authors gave necessary and sufficient conditions for the entries of a Wigner matrix to converge. The crucial condition 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 [28]. Later, universality results were established for sparse random matrix models, for example in [23] for random graphs and in [26] for random banded matrices. Moreover, sharp non-asymptotic results for a general class of matrices were established in [5] and in [4].

All the models mentioned above can be considered as random matrices with general variance profile, i.e.,

[†]National and Kapodistrian University of Athens, Department of Mathematics, Panepistimiopolis, Athens 15784, Greece, dcheliotis@math.uoa.gr

[‡]Laboratoire dinformatique Gaspard Monge (LIGM / UMR 8049), Université Gustave Eiffell, Marne-la-Vallée michail.louvaris@univ-eiffel.fr

random matrices whose entries' variances can depend on the dimension of the matrix and the location of the element in the matrix. These models have also drawn a lot of attention lately, see for example [13], [14], where non-Hermitian models are considered. More specifically, assume that $A_N = (a_{i,j}^{(N)})$, $N \in \mathbb{N}^+$, is a sequence of symmetric random matrices, with $a_{i,j}^{(N)}$ real valued having mean zero and variance $s_{i,j}^{(N)}$ bounded by a fixed number, say 1. Classically, the first question is whether the empirical spectral distribution of an appropriate normalization of A_N (e.g., $A_N / \sqrt{N}$) converges to a nontrivial probability measure, as in Wigner's theorem. Nothing guarantees that, and one can construct examples where the sequence of the empirical spectral distributions does not converge. The work [30], using the notion of graphons, gave conditions on the variance profile $s_{i,j}^{(N)}$, $i, j \in [N]$, $N \in \mathbb{N}^+$ so that convergence takes place.

The next, natural, question concerns the convergence of the largest eigenvalue to the largest element of the support of the limiting distribution. Again, this in not automatic but requires additional assumptions. It was established in the recent works [1], [18], [24], [17] (whose focus however is not this question) for some class of random matrices with a general variance profile under the assumption that the entries of the matrices have finite all moments (the first two works assume that each $a_{i,j}^{(N)}$ is sharp sub-Gaussian, the last two assume that for each $k \in \mathbb{N}^+$ there is a constant bounding the $2k$ moment of each $a_{i,j}^{(N)}$). In this paper, we generalize these 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 empirical spectral distribution under general assumptions for the variance profile of the matrices. Regarding finiteness of moments, we assume only that $\sup_{N \in \mathbb{N}^+, i,j \in [N]} \mathbf{E}|a_{i,j}^{(N)}|^4 < \infty$.

Notation. For any $N \times N$ matrix $A = (a_{i,j})_{i,j \in [N]} \in \mathbb{R}^{N \times N}$ with eigenvalues $\{\lambda_i(A)\}_{i \in [N]}$, the measure

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

will be called the Empirical Spectral Distribution (E.S.D.) of A. When the eigenvalues are real, write $\lambda_{\max}(A)$ for the maximum among them. We will use the following two norms on square matrices. For $A \in \mathbb{R}^{N \times N}$,

$$|A|_{\text{op}} := \max_{x \in \mathbb{R}^N : \|x\|_2 = 1} \|Ax\|_2 = \sqrt{\lambda_{\max}(AA^T)} \tag{1.1}$$

$$\|A\|_{\max} := \max_{i,j \in [N]} |a_{i,j}|. \tag{1.2}$$

It is easy to see that $|A|_{\text{op}} \leq N\|A\|_{\max}$ and if the matrix A is symmetric, then

$$|A|_{\text{op}} := \max_{i \in [N]} |\lambda_i(A)|. \tag{1.3}$$

2 Statement of the results

Throughout this section, $(A_N)_{N \in \mathbb{N}^+}$ is a sequence of symmetric random matrices with independent entries (up to symmetry), $A_N = (a_{i,j}^{(N)})_{i,j \in [N]}$ is an $N \times N$ matrix, and all $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i, j \in [N]\}$ are defined on the same probability space and take real values.

A standard assumption for the sequence is the following (see relation (2.2.1) in [2]).

Assumption 2.1.

1. $\mathbf{E}a_{i,j}^{(N)} = 0$ for all $N \in \mathbb{N}^+$, $i, j \in [N]$, and $\sup_{N \in \mathbb{N}^+, i,j \in [N]} \mathbf{E}|a_{i,j}^{(N)}|^2 < \infty$.

2. For any $\varepsilon > 0$,

$$\lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j \in [N]} \mathbf{E}\left\{|a_{i,j}^{(N)}|^2 \mathbf{1}_{|a_{i,j}^{(N)}| \geq \varepsilon \sqrt{N}}\right\} = 0. \tag{2.1}$$

This is satisfied in the case that $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i, j \in [N], i \leq j\}$ are i.i.d. with mean 0 and finite variance. But it is not enough to guarantee that the ESD of the appropriately normalized A_N converges to a nontrivial limit. To state a sufficient condition for this, we introduce some notation that will be used throughout the work. We let

$$s_{i,j}^{(N)} := \mathbf{E}\{|a_{i,j}^{(N)}|^2\} \tag{2.2}$$

for all $N \in \mathbb{N}^+$, $i, j \in [N]$ and $V_0 := \sup_{N \in \mathbb{N}^+, i,j \in [N]} s_{i,j}^{(N)} \in [0, \infty)$.

Also, let $\mathbf{C}_k$ be the set of ordered rooted trees with k edges (where $k \in \mathbb{N}$) 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. The number of such trees is the k−th Catalan number, i.e.,

$$|\mathbf{C}_k| = \frac{1}{k + 1} \binom{2k}{k}, \tag{2.3}$$

and a trivial bound that we will use is $|\mathbf{C}_k| \leq 2^{2k}$. For each such tree, we consider its vertices ordered $v_0 < v_1 < \cdots < v_k$ so that v_0 is the root, each parent is smaller than its children, and the children keep the order they have as vertices of an ordered tree. A labeling of such a tree is an ordered $k+1$-tuple $(\ell_0, \ell_1, \cdots, \ell_k)$ of different objects, the object ℓ_i is the label of vertex v_i.

A quantity of fundamental importance for the sequel is the following sum

$$M_N(k) := \sum_{T \in \mathbf{C}_k} \sum_{\substack{\mathbf{i} \in [N]^{k+1} \\ \text{labeling of } T}} \prod_{\{i,j\} \in E(T)} s_{i,j}^{(N)}. \tag{2.4}$$

$E(T)$ denotes the set of edges of the tree T. Note that $M_N(0) = N$ since by convention the product over an empty index set equals 1.

Assumption 2.2. There is a probability measure μ on $\mathbb{R}$ such that for each $k \in \mathbb{N}$ it holds

$$\lim_{N \to \infty} \frac{M_N(k)}{N^{k+1}} = \int x^{2k} \, d\mu(x). \tag{2.5}$$

A tool for checking this assumption is explained in Remark 2.13 below.

If the sequence $(A_N)_{N \in \mathbb{N}^+}$ satisfies both Assumptions 2.1 and 2.2, then $\mu_{A_N / \sqrt{N}} \Rightarrow \mu$ with probability one (see the proof of Theorem 3.2 of [30]). The measure μ is symmetric with compact support contained in $[-2\sqrt{V_0}, 2\sqrt{V_0}]$. The compactness of the support follows from (2.5), $M_N(k) \leq |\mathbf{C}_k| N^{k+1} V_0^k$, and $|\mathbf{C}_k| \leq 2^{2k}$. Let

$$\mu_\infty := \sup \operatorname{suppt} \mu. \tag{2.6}$$

We seek conditions under which the maximum eigenvalue of $A_N / \sqrt{N}$ converges to μ_∞ in probability. An easy argument will give us the lower bound, and since $\lambda_{\max}(A) \leq |A|_{\mathrm{op}}$ for any symmetric matrix $A \in \mathbb{R}^{N \times N}$, it will be enough to prove the upper bound for the operator norm of $A_N / \sqrt{N}$.

For this purpose, we need stronger assumptions. The following is stronger than Assumption 2.1.

Assumption 2.3.

(a) $\mathbf{E} a_{i,j}^{(N)} = 0$ for all $N \in \mathbb{N}^+$, $i, j \in [N]$, $\sup_{N \in \mathbb{N}^+, i,j \in [N]} \mathbf{E} |a_{i,j}^{(N)}|^2 \leq 1$, and $\sup_{N \in \mathbb{N}^+, i,j \in [N]} \mathbf{E} |a_{i,j}^{(N)}|^4 < \infty$.

(b) For any $\varepsilon > 0$ it is true that

$$\lim_{N \to \infty} \sum_{i,j} \mathbf{P}(|a_{i,j}^{(N)}| \geq \varepsilon \sqrt{N}) = 0. \tag{2.7}$$

Note that condition (2.7) is satisfied if we assume that all $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i, j \in [N]\}$ have the same distribution with finite 4-th moment.

We gain control over $|A_N|_{\mathrm{op}}$ through the traces of high moments of A_N, and the main difficulty, which the next conditions (Assumption 2.4 and Assumption 2.6) try to address, is how to connect these traces with μ_∞, which emerges out of $\{\lambda_i(A_N) : i \in [N]\}$ only after we take $N \to \infty$.

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

$$s_{i,j}^{(N)} \leq \min\{s_{2i,2j}^{(2N)}, s_{2i-1,2j}^{(2N)}, s_{2i-1,2j-1}^{(2N)}\}. \tag{2.8}$$

For example, this assumption is satisfied if $s_{i,j}^{(N)} = h(i/N, j/N)$ for all $N \in \mathbb{N}^+$, $i, j \in [N]$, where $h : [0, 1]^2 \to [0, \infty)$ is a function decreasing separately in each variable.

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

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

(ii) For any bounded graphon W and any multigraph $G = (V, E)$, we call isomorphism density from G to W the quantity

$$t(G, W) := \int_{[0,1]^{|V|}} \prod_{\{i,j\} \in E} W(x_i, x_j) \prod_{i \in V} dx_i. \tag{2.9}$$

Now, let $(A_N)_{N \in \mathbb{N}^+}$ be a sequence of random matrices with elements having finite second moment. Each A_N defines a graphon, W_N, through the relation

$$W_N(x, y) := s^{(N)}_{\lceil Nx \rceil, \lceil Ny \rceil} \tag{2.10}$$

for each $(x, y) \in [0, 1] \times [0, 1]$. For this relation, $\lceil 0 \rceil$ denotes 1.

Assumption 2.6. There exists a graphon W such that the W_N of (2.10) satisfies

$$\lim_{N \to \infty} t(T, W_N) = t(T, W) \tag{2.11}$$

for any finite tree T. Moreover, for any $D > 0$ there exists some $C = C(D) \in (0, \infty)$ and $N_0 = N_0(D) \in \mathbb{N}^+$ such that for any $N \geq N_0$ it holds

$$\int_{[0,1]^2} |W_N(x, y) - W(x, y)| \, dx \, dy \leq CN^{-D}. \tag{2.12}$$

This assumption together with Assumption 2.3 implies Assumption 2.2 (This will be explained in Lemma 4.2). Again, we denote by μ_∞ the maximum of the support of μ.

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

Assumption 2.7. $(A_N)_{N \in \mathbb{N}^+}$ is a sequence of symmetric random matrices, the entries of each A_N are independent (up to symmetry), and there exists a random variable X with mean 0, variance 1, and finite $4 + \delta$ moment for some $\delta > 0$, which stochastically dominates the entries of A_N in the following sense

$$\mathbf{P}(|\{A_N\}_{i,j}| \geq t) \leq \mathbf{P}(|X| \geq t), \text{ for all } t \in [0, \infty), N \in \mathbb{N}^+, i, j \in [N]. \tag{2.13}$$

We are now ready to present our first main result.

Theorem 2.8. *Let $(A_N)_{N \in \mathbb{N}^+}$ be a sequence of matrices satisfying Assumption 2.3. Then if either Assumptions 2.2 and 2.4 hold or Assumption 2.6 holds, it is true that*

$$\lim_{N \to \infty} \frac{|A_N|_{\text{op}}}{\sqrt{N}} = \mu_\infty \text{ in probability} \tag{2.14}$$

where μ_∞ is defined in (2.6). *Moreover, if the sequence $(A_N)_{N \in \mathbb{N}^+}$ satisfies Assumption 2.7, the convergence in* (2.14) *holds in the almost sure sense.*

Note that Assumption 2.4 is restrictive and does not cover several of the well-known and studied models. Thus, in what follows, we try to extend the domain of validity of Theorem 2.8. We first give two definitions. For $N \in \mathbb{N}^+$ and $U \subset [N]^2$:

• We call a $(x, y) \in U$ *internal point* of U if $\{(x + d_1, y + d_2) : d_1, d_2 \in \{-1, 0, 1\}\} \subset U$. We denote by U^o the set of internal points of U.

• We say that U is *axially convex* if $(i, j) \in U, (i, j') \in U, r \in [N], (r - j)(r - j') < 0$ imply $(i, r) \in U$ and $(i, j) \in U, (i', j) \in U, r \in [N], (r - i)(r - i') < 0$ imply $(r, j) \in U$.

Definition 2.9 (Generalized step function variance profile). Let $(A_N)_{N \in \mathbb{N}^+}$ be a sequence of symmetric random matrices, A_N of dimension $N \times N$, with each element having zero mean and finite second moment. Moreover, suppose that there exists an $\mathbb{N}^+$−valued sequence $(d_N)_{N \in \mathbb{N}^+}$ with $\lim_{N \to \infty} d_N / N = 0$ and such that for each N there is a partition $\mathcal{P}_N := \{\mathcal{B}_i^{(N)} : i = 1, 2, \dots, d_N\}$ of the grid $[N]^2$ consisting of d_N axially convex sets with the following properties.

(a) If $A \in \mathcal{P}_N$ then $R(A) := \{(i,j) : (j,i) \in A\} \in \mathcal{P}_N$.

(b) For any $m \in [d_N]$ there exists $f \in [d_{2N}]$ such that

$$2\mathcal{B}_m^{(N)} \subset \mathcal{B}_f^{(2N)}. \tag{2.15}$$

(c) For any $N \in \mathbb{N}$, $m \in [d_N]$ and $i \in [N]$ the line segment $x = i$ intersects $\mathcal{B}_m^{(N)} \setminus (\mathcal{B}_m^{(N)})^\circ$ at most 2 times.

Then if for all $(i,j) \in [N]^2$ the variance of the (i,j)-entry of A_N is given by

$$s_{i,j}^{(N)} := \sum_{m \in [d_N]} s_m^{(N)} 1_{(i,j) \in \mathcal{B}_m^{(N)}} \tag{2.16}$$

for some set of numbers $\{s_i\}_{i \in [d_N]}$ so that $s_m^{(N)} = s_k^{(N)}$ if $R(\mathcal{B}_m^{(N)}) = \mathcal{B}_k^{(N)}$, we will call the sequence of matrices $(A_N)_{N \geq 1}$ random matrix model whose variance profile is given by a generalized step function.

The following Theorem is a corollary of Theorem 2.8 and gives results of the type (2.14) for the operator norm of the matrix

• A_N when A_N is a non-periodic band matrix with band size proportional to N or has a step or continuous profile.

• $A_N A_N^T$ (i.e., Gram matrix) when A_N is a rectangular matrix with step or continuous variance profile.

Details are given after the next theorem and in subsection 8.2.

Theorem 2.10. *Let $(A_N)_{N \in \mathbb{N}^+}$ be a random matrix model whose variance profile is given by a generalized step function. If it also satisfies Assumptions 2.2, 2.3, and for every $N \in \mathbb{N}$ and $(i,j) \in [N]^2$ it is true that*

$$s_{i,j}^{(N)} \leq s_{2i,2j}^{(2N)}, \tag{2.17}$$

then

$$\lim_{N \to \infty} \frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} = \mu_\infty \quad \text{in probability,} \tag{2.18}$$

where μ_∞ is defined in (2.6). Moreover, if the sequence $(A_N)_{N \in \mathbb{N}^+}$ satisfies Assumption 2.7 the convergence in (2.18) holds in the almost sure sense.

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

$$\{A \odot B\}_{i,j} = \{A\}_{i,j} \{B\}_{i,j} \text{ for all } i, j \in [N]. \tag{2.19}$$

Note that Assumption 2.7 is satisfied if A_N can be written as

$$A_N = \Sigma_N \odot A_N', \tag{2.20}$$

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 $4 + \delta$ moment for some $\delta > 0$ and for each N the entries of Σ_N are elements of $[0, 1]$.

Next, we study the operator norm of two widespread random matrix models.

Definition 2.11 (Step function variance profile). Consider

    a) $m \in \mathbb{N}^+$ and numbers $\{\sigma_{p,q}\}_{p,q \in [m]} \in [0, 1]^{m \times m}$ with $\sigma_{p,q} = \sigma_{q,p}$ for all $p, q \in [m]$.

    b) For each $N \in \mathbb{N}^+$, a partition of $[N]$ into m intervals $\{I_p^{(N)}\}_{p \in [m]}$. The numbering of the intervals is such that $x < y$ whenever $x \in I_p^{(N)}, y \in I_q^{(N)}$ and $p < q$. Let $L_p^{(N)}$ and $R_p^{(N)}$ be the left and right endpoint respectively of $I_p^{(N)}$.

    c) Numbers $0 = a_0 < a_1 < \cdots < a_{m-1} < a_m := 1$. We assume that $\lim_{N \to \infty} R_p^{(N)}/N = a_p$ for each $p \in [m]$.

    d) A random variable X_0 with $\mathbf{E}(X_0) = 0, \mathbf{E}(X_0^2) = 1$.

For each $N \in \mathbb{N}^+$, define the matrix $\Sigma_N \in \mathbb{R}^{N \times N}$ by $(\Sigma_N)_{i,j} = \sigma_{p,q}$ if $i \in I_p^{(N)}, j \in I_q^{(N)}$, and let $\{A_N\}_{N \in \mathbb{N}^+}$ be the sequence of symmetric random matrices defined by

$$A_N = \Sigma_N \odot A_N' \tag{2.21}$$

where A_N' is symmetric and its entries are independent (up to symmetry) random variables all with distribution the same as X_0. Then $(A_N)_{N \in \mathbb{N}^+}$ will be called *symmetric random matrix model whose variance profile is given by a step function.*

Let $\hat{I}_p := [a_{p-1}, a_p)$ for $p \in [m-1]$, and $\hat{I}_m := [a_{m-1}, 1]$. These intervals together with the numbers from a) determine a function $\sigma : [0,1]^2 \to [0,1]$ as follows

$$\sigma(x, y) := \sigma_{p,q} \text{ if } x \in \hat{I}_p, y \in \hat{I}_q. \tag{2.22}$$

We call the function σ^2 the variance profile of the model.

Definition 2.12 (Continuous function variance profile). For

    a) a continuous and symmetric function $\sigma : [0,1]^2 \to [0,1]$ (i.e. $\sigma(x,y) = \sigma(y,x)$ for all $x, y \in [0,1]$),

    b) a sequence $(\Sigma_N)_{N \in \mathbb{N}^+}$ of symmetric matrices, $\Sigma_N \in [0,1]^{N \times N}$, with the property

$$\lim_{N \to \infty} \sup_{1 \le i,j \le N} \left| (\Sigma_N)_{i,j} - \sigma(i/N, j/N) \right| = 0, \tag{2.23}$$

    c) a random variable X_0 with $\mathbf{E}(X_0) = 0, \mathbf{E}(X_0^2) = 1$,

consider the sequence $\{A_N\}_{N \in \mathbb{N}^+}$ of symmetric random matrices, A_N of dimension $N \times N$, defined by

$$A_N = \Sigma_N \odot A_N' \tag{2.24}$$

where the entries of A_N' are independent (up to symmetry) random variables all with distribution the same as X_0. Then we say that $(A_N)_{N \in \mathbb{N}^+}$ is a *random matrix model whose variance profile is given by a continuous function.* Again, we call the function σ^2 the variance profile.

Remark 2.13 (Checking Assumption 2.2). A sufficient condition for the validity of Assumption 2.2 is that $(A_N)_{N \in \mathbb{N}^+}$ satisfies Assumption 2.1 and there is a graphon W such that $W_N \to W$ almost everywhere in $[0,1] \times [0,1]$.

Indeed, the bounded convergence theorem gives that $t(T, W_N) \to t(T, W)$ for all trees. Then Theorem 3.2 (a) of [30] shows that the ESD of $A_N / \sqrt{N}$ converges almost surely weakly to a probability measure $\mu^{\sqrt{W}}$ whose $2k$ moment equals

$$\lim_{N \to \infty} \sum_{T \in \mathbf{C}_k} t(T, W_N) \tag{2.25}$$

while the moments of odd order are 0. Then, for each $T \in \mathbf{C}_k$,

$$0 \le t(T, W_N) - \sum_{\substack{\mathbf{i} \in [N]^{K+1} \\ \text{labeling of } T}} N^{-k-1} \prod_{\{i,j\} \in E(T)} s_{i,j}^{(N)} = O(1/N), \tag{2.26}$$

because $t(T, W_N)$ is simply the same as the sum in the previous relation with the only difference that $\mathbf{i}$ is not required to be a labeling, i.e., it can have repetitions. It follows that Assumption 2.2 holds. As we remarked after (2.5), $\mu^{\sqrt{W}}$ is symmetric and has bounded support. Denote by $\mu_\infty^{\sqrt{W}}$ the largest element of the support.

If, in the two models above, X_0 has finite $4 + \delta$ moment for some small $\delta > 0$, then it is easy to see that the sequence $(A_N)_{N \in \mathbb{N}^+}$ satisfies Assumptions 2.3. It also satisfies Assumption 2.2 because it satisfies Assumption 2.1 and, in both cases, $W_N(x, y)$ converges to $\sigma^2(x, y)$ for almost all $(x, y) \in [0,1] \times [0,1]$, thus the preceding remark applies.

Our result for the model (2.21) is the following.

Theorem 2.14. *Let $(A_N)_{N \in \mathbb{N}^+}$ be a random matrix model whose variance profile is given by a step function as above. Assume that X_0 has mean value 0, variance 1, and finite $4 + \delta$ moment, for some small $\delta > 0$. Then it is true that*

$$\lim_{N \to \infty} \frac{|A_N|_{op}}{\sqrt{N}} = \mu_\infty^\sigma \quad a.s. \tag{2.27}$$

The previous theorem together with an approximation result that we prove in Section 7 (Proposition 7.1) has the following consequence for the model (2.24).

Corollary 2.15. *Let $(A_N)_{N \in \mathbb{N}^+}$ be a sequence of matrices whose variance profile is given by a continuous function. If X_0 has mean zero, variance one, and finite $4 + \delta$ moment, then*

$$\lim_{N \to \infty} \frac{|A_N|_{op}}{\sqrt{N}} = \mu_\infty^\sigma \quad a.s. \tag{2.28}$$

Remark 2.16. 1) Theorem 2.14 covers the cases in the Wigner matrix model [i. e., $A_N := (a_{i,j})_{i,j \in [N]}$ with $\{a_{i,j} : 1 \le i \le j \le N, N \in \mathbb{N}^+\}$ i.i.d. with $\mathbf{E}(a_{1,1}) = 0, \mathbf{E}(a_{1,1}^2) = 1$] where $\mathbf{E}(|a_{1,1}|^{4+\delta}) < \infty$ for some $\delta > 0$. Recall that the necessary and sufficient condition for the validity of (2.27) in that model is $\mathbf{E}(|a_{1,1}|^4) < \infty$.
2) Corrolary 2.15 holds also in the case that the function σ of Definition 2.12 is piecewise continuous in a sense explained in the end of Section 7.

3 Analysis of high order moments

Assume at the moment that the entries of A_N have finite moments of all orders.

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_∞. In general, for $k \in \mathbb{N}$, it is true that

$$\mathbf{E} \operatorname{tr}(A_N^{2k}) = \sum_{i_1, i_2, \ldots, i_{2k} \in [N]} \mathbf{E} \left(\prod_{l=1}^{2k} a_{i_l, i_{l+1}}^{(N)} \right) \tag{3.1}$$

with the conventions that $i_{2k+1} = i_1$, when $k = 0$ the sum is only over $i_1 \in [N]$, and the product over an empty set equals 1.

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

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

and set of edges

$$\{\{i_r, i_{r+1}\} : r = 1, 2, \ldots, 2k\}. \tag{3.2}$$

As explained in [2] (in the proof of relation (3.1.6) there, pages 49, 50 or in Theorem 3.2 of [30]), the limit

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

remains the same if in the sum of (3.1) we keep only the summands such that

$$\text{the graph } G(\mathbf{i}) \text{ is a tree with } k + 1 \text{ vertices} \tag{3.3}$$

Then, necessarily, 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. Such a $G(\mathbf{i})$ becomes an ordered rooted tree if we mark i_1 as the root and order children of the same vertex according to the order they appear in the cycle.

Cycles $\mathbf{i}$ that don't satisfy (3.3) we call bad cycles. So, for $k \in \mathbb{N}$, the sum in (3.1) can be written as

$$\mathbf{E} \operatorname{tr}(A^{2k}) = M_N(k) + B_N(k), \tag{3.4}$$

where

$$M_N(k) := \sum_{T \in \mathbf{C}_k} \sum_{\mathbf{i} \in [N]^{(2k) \vee 1} : G(\mathbf{i}) \sim T} \prod_{\{i,j\} \in E(G(\mathbf{i}))} s_{i,j}^{(N)}, \tag{3.5}$$

$$B_N(k) := \sum_{\mathbf{i} \in [N]^{2k} : \text{bad cycle}} \mathbf{E} X(\mathbf{i}). \tag{3.6}$$

Recall that $\mathbf{C}_k$ are the ordered rooted trees with k edges and $G(\mathbf{i}) \sim T$ means that the graphs are isomorphic as ordered rooted trees. Note also that $M_N(k)$ has already been defined in (2.4) but the two definitions for it agree. Also, $M_N(0) = N, B_N(0) = 0$.

The plan is to control the expectation of the trace in (3.4) 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 [27].

Proposition 3.1. *Let A_N be an $N \times N$ symmetric random matrix with independent entries (up to symmetry) and with $\mathbf{E}(a_{i,j}^{(N)}) = 0, s_{i,j}^{(N)} \leq 1$ for all $N \in \mathbb{N}, i, j \in [N]$. 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 integers $1 \leq k < N$ it is true that*

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

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

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

Additionally, in case $t \geq 2$, 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. In case $t = 1$, $T(\mathbf{i})$ is the graph with one vertex, i_1, and one edge (loop) with end vertices i_1, i_1. Thus, $T(\mathbf{i})$ has t vertices and $1 \vee (t-1)$ edges.

To bound $|\mathbf{E}X(\mathbf{i})|$, notice that if any of $a_1, a_2, \ldots, a_s$ 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 $a \geq 2$ it holds $\mathbf{E}(|a_{i,j}^{(N)}|^a) \leq (C_1 N^{1/2-\epsilon})^{a-2}s_{i,j}^{(N)}$. Thus

$$\mathbf{E}|X(\mathbf{i})| = \prod_{q=1}^{s}\mathbf{E}|X_{e_q}|^{a_q} \leq (CN^{1/2-\epsilon})^{a_1+\cdots+a_s-2s}\prod_{\{i,j\}\in E(G(\mathbf{i}))}s_{i,j}^{(N)} \leq (CN^{1/2-\epsilon})^{2k-2s}\prod_{\{i,j\}\in E(T(\mathbf{i}))}s_{i,j}^{(N)}. \tag{3.8}$$

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

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

Using the bound on $N_{T,a_1,a_2,\ldots,a_s}$ provided by Lemma 9.1, we obtain

$$|B_N(k)| \leq \sum_{s=1}^{k}\sum_{t=1}^{k\wedge(s+1)}\sum_{a_1,a_2,\ldots,a_s}(CN^{1/2-\epsilon})^{2k-2s}\left\{\mathbf{1}_{t=1}\sum_{i\in[N]}s_{i,i}^{(N)} + \mathbf{1}_{t\geq 2}\sum_{T\in\mathbf{C}_{t-1}}N_{T,a_1,a_2,\ldots,a_s}\sum_{\mathbf{i}\in[N]^{2k}:T(\mathbf{i})\sim T}\prod_{\{i,j\}\in E(T(\mathbf{i}))}s_{i,j}^{(N)}\right\} \tag{3.10}$$

$$\leq \sum_{s=1}^{k}\sum_{t=1}^{k\wedge(s+1)}\sum_{a_1,a_2,\ldots,a_s}(CN^{1/2-\epsilon})^{2k-2s}\left\{\mathbf{1}_{t=1}\sum_{i\in[N]}s_{i,i}^{(N)} + \mathbf{1}_{t\geq 2}\sum_{T\in\mathbf{C}_{t-1}}N_{T,a_1,a_2,\ldots,a_s}\sum_{\mathbf{i}\in[N]^{2(t-1)}:T(\mathbf{i})\sim T}\prod_{\{i,j\}\in E(T(\mathbf{i}))}s_{i,j}^{(N)}\right\} \tag{3.11}$$

$$\leq \sum_{s=1}^{k}\sum_{t=1}^{k\wedge(s+1)}\sum_{a_1,a_2,\ldots,a_s}(CN^{1/2-\epsilon})^{2k-2s}(4k^4)^{4(s+1-t)+2(k-s)}M_N(t-1). \tag{3.12}$$

We used here the fact that $s_{i,i}^{(N)} \leq 1$, so that $\sum_{i \in [N]} s_{i,i}^{(N)} \leq N = M_N(0)$. The inside sum in (3.12) 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 $s^{2(k-s)} \leq k^{2(k-s)}$. Thus the above sum is bounded by

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

$\square$

Proposition 3.2. *Let $(A_N)_{N \in \mathbb{N}^+}$ be a sequence of symmetric matrices and $R > 0$ so that the sequence satisfies Assumption 2.3 and the following condition $\Sigma(R)$:*
For each $C_1 > 0$ there are $C_2 > 0$ and $N_0 \in \mathbb{N}^+$ such that

$$M_N(k) \leq C_2 N^{k+1} R^{2k} \tag{3.14}$$

for all $N, k \in \mathbb{N}^+$ with $N \geq N_0$ and $1 \leq k \leq C_1 \log N$.
Then for each $\epsilon > 0$, it holds

$$\lim_{N \to \infty} \mathbb{P}\left(\frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} \geq R(1 + \epsilon)\right) = 0. \tag{3.15}$$

Proof. Fix $\eta \in (0, 1/8)$ and define the $N \times N$ matrices $A_N^{\leq}, A_N^{>}$ by

$$(A_N^{\leq})_{i,j} := a_{i,j}^{(N)} \mathbb{1}_{|a_{i,j}| \leq N^{\frac{1}{2}-\eta}}, \tag{3.16}$$

$$(A_N^{>})_{i,j} := a_{i,j}^{(N)} \mathbb{1}_{|a_{i,j}| > N^{\frac{1}{2}-\eta}} \tag{3.17}$$

for all $i, j \in [N]$. 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. Note that

$$\frac{1}{\sqrt{N}} |A_N|_{\mathrm{op}} \leq \frac{1}{\sqrt{N}} \left(|A_N^{\leq} - \mathbf{E}A_N^{\leq}|_{\mathrm{op}} + |\mathbf{E}A_N^{\leq}|_{\mathrm{op}} + |A_N^{>}|_{\mathrm{op}}\right). \tag{3.18}$$

We will bound the three terms in the right hand side of the last inequality. For the first two, we use only Assumption 2.3 and the arguments in the proof of Theorem 2.3.23 in [27].

1) The term $N^{-\frac{1}{2}} |\mathbf{E}(A_N^{\leq})|_{\mathrm{op}}$ is a deterministic sequence that converges to 0 because, since $a_{i,j}^{(N)}$ is centered, we have

$$|(\mathbf{E}A_N^{\leq})_{i,j}| = |(\mathbf{E}A_N^{>})_{i,j}| \leq N^{-3(\frac{1}{2}-\eta)} \sup_N \max_{i,j} \mathbf{E}|a_{i,j}^{(N)}|^4. \tag{3.19}$$

And using the inequality $|C|_{\mathrm{op}} \leq N\|C\|_{\max}$, we get that

$$|\mathbf{E}(A_N^{\leq})|_{\mathrm{op}} \leq N^{3\eta-\frac{1}{2}} \sup_N \max_{i,j} \mathbf{E}|a_{i,j}^{(N)}|^4 \overset{N \to \infty}{\to} 0.$$

2) The term $N^{-\frac{1}{2}} |A_N^{>}|_{\mathrm{op}}$ converges to 0 in probability. Indeed, for any $\delta_1 > 0$,

$$\mathbf{P}(|A_N^{>}|_{\mathrm{op}} > \delta_1 \sqrt{N}) \leq \mathbf{P}(|a_{i,j}^{(N)}| > \delta_1 \sqrt{N} \text{ for some } i,j \in [N]) \tag{3.20}$$

$$+ \mathbf{P}(|A_N^{>}|_{\mathrm{op}} > \delta_1 \sqrt{N} \text{ and } |a_{i,j}^{(N)}| \leq \delta_1 \sqrt{N} \text{ for all } i,j \in [N]). \tag{3.21}$$

The first quantity goes to zero as $N \to \infty$ because of (2.7). For the second, it is an easy exercise to show that if each entry of a matrix M has absolute value at most a and each row and column of M has at most one non-zero element then $|M|_{\mathrm{op}} \leq a$ (use the expression $|M|_{\mathrm{op}} = \sup_{x:\|x\|_2=1} \|Mx\|_2$). Consequently, the probability in (3.21) is at most

$$\sum_{i=1}^{N} \sum_{1 \leq j_1 < j_2 \leq N} \mathbf{P}(|a_{i,j_1}| > N^{\frac{1}{2}-\eta}, |a_{i,j_2}| > N^{\frac{1}{2}-\eta}) + \sum_{j=1}^{N} \sum_{1 \leq i_1 < i_2 \leq N} \mathbf{P}(|a_{i_1,j}| > N^{\frac{1}{2}-\eta}, |a_{i_2,j}| > N^{\frac{1}{2}-\eta}) \tag{3.22}$$

$$\leq 2N \binom{N}{2} \frac{\sup_{N \in \mathbb{N}^+, i,j \in [N]} (\mathbf{E}\{|a_{i,j}^{(N)}|^4\})^2}{N^{4-8\eta}} \leq \frac{1}{N^{1-8\eta}} \sup_{N \in \mathbb{N}^+, i,j \in [N]} (\mathbf{E}\{|a_{i,j}^{(N)}|^4\})^2 \overset{N \to \infty}{\to} 0. \tag{3.23}$$

We used the independence of the entries in each row or column and Markov's inequality.

3) To deal with $|A_N^{\leq} - \mathbf{E}A_N^{\leq}|_{\mathrm{op}}$, we will use Proposition 3.1. Let

$$\tilde{A}_N := A_N^{\leq} - \mathbf{E}A_N^{\leq}, \tag{3.24}$$

$$s_{i,j}^{(N),\leq} := \mathbf{E}\{(\tilde{A}_N)_{i,j}^2\}. \tag{3.25}$$

Proposition 3.1 applies to $\tilde{A}_N$ because any element of the matrix, say $(\tilde{A}_N)_{i,j}$, has zero mean and variance $s_{i,j}^{(N),\leq} \leq \mathbf{E}\{(A_N^{\leq})^2\} \leq \mathbf{E}(A_N^2) = s_{i,j}^{(N)} \leq 1$. Thus, if we denote by $\tilde{M}_N(m)$ the terms (2.4) for $m \in [N]$ and for the matrix $\tilde{A}_N$, we will have $\tilde{M}_N(m) \leq M_N(m)$ for all $m \in \mathbb{N}$, and Proposition 3.1, gives that for any $1 \leq k < N$,

$$\mathbf{E}\operatorname{tr}(\tilde{A}_N^{2k}) \leq M_N(k) + \sum_{s=1}^k (4k^5)^{2k-2s}\left(2N^{\frac{1}{2}-\eta}\right)^{2k-2s} \sum_{t=1}^{(s+1)\wedge k} (4k^4)^{4(s+1-t)} M_N(t-1) \tag{3.26}$$

Now fix $C_1 > 0$, its value will be determined in (3.32) below. For $1 \leq k \leq C_1 \log N$,

$$\mathbf{E}\operatorname{tr}(\tilde{A}_N^{2k}) \leq C_2 N^{k+1} R^{2k} + C_2 \sum_{s=1}^k (4k^5)^{2k-2s}\left(2N^{\frac{1}{2}-\eta}\right)^{2k-2s} \sum_{t=1}^{(s+1)\wedge k} (4k^4)^{4(s+1-t)} N^t R^{2(t-1)}. \tag{3.27}$$

Next, we focus on the second summand in the right hand side of the previous inequality for N large enough. In the sum in t we factor out $(4k^4)^{4(s+1)} R^{-2}$, and in the resulting sum of geometric progression with ratio a larger than 1 we use the bound $a + a^2 + \cdots + a^{(s+1)\wedge k} \leq ka^{(s+1)\wedge k}$. Thus the sum in (3.27) is bounded by

$$C_2 \frac{k}{R^2} \sum_{s=1}^k (4k^5)^{2k-2s}\left(2N^{\frac{1}{2}-\eta}\right)^{2k-2s} (4k^4)^{4(s+1)} \left(\frac{NR^2}{(4k^4)^4}\right)^{(s+1)\wedge k} \tag{3.28}$$

$$= 2^8 C_2 k^{17} N^k (R^2)^{k-1} + C_2 \frac{k}{R^2}(NR^2)^{k+1} \sum_{s=1}^{k-1}\left(\frac{4(4k^5)^2}{N^{2\eta}R^2}\right)^{k-s} \tag{3.29}$$

$$\leq 2^8 C_2 k^{17} N^k (R^2)^{k-1} + 2^7 C_2 k^{11}(R^2)^{(k-1)} N^{k+1-2\eta} \leq 2^9 C_2 k^{17} N^{k+1-2\eta} R^{2k-2} \tag{3.30}$$

[in summing the geometric series in (3.29), we used the bound $c + c^2 + \cdots + c^r \leq 2c$ if $0 \leq c < 1/2$]. Thus, returning to (3.27),

$$\mathbf{E}\operatorname{tr}(\tilde{A}_N^{2k}) \leq N^{k+1} R^{2k}\{1 + o(1)\}^{2k}. \tag{3.31}$$

with $o(1)$ depending on R, C_2, η.

Fix $\epsilon > 0$, pick

$$C_1 > \frac{2 + \epsilon}{\log(1 + \epsilon)}, \tag{3.32}$$

and apply the above for $k := [C_1 \log N]$. Relation (3.31) implies

$$\mathbb{P}\left(\frac{|\tilde{A}_N|_{\mathrm{op}}}{\sqrt{N}} \geq R(1 + \epsilon)\right) \leq \mathbb{P}\left(\frac{|\tilde{A}_N|_{\mathrm{op}}^{2k}}{N^k} \geq R^{2k}(1 + \epsilon)^{2k}\right) \leq \frac{1}{R^{2k}(1+\epsilon)^{2k}} \frac{1}{N^k} \mathbf{E}|\tilde{A}_N|_{\mathrm{op}}^{2k}$$

$$\leq N\left(\frac{1 + o(1)}{1 + \epsilon}\right)^{2k} = O\left(\frac{1}{N^{1+\epsilon}}\right), \tag{3.33}$$

for any N large enough. The last equality is true because of the choice of k and C_1. $\qquad\square$

A tool for proving almost sure convergence of the sequence $|A_N|_{\mathrm{op}}/\sqrt{N}$ is the following lemma.

Lemma 3.3. *Let $(A_N)_{N\in\mathbb{N}^+}$ be a sequence of matrices, A_N is $N \times N$, and $R > 0$ so that the sequence satisfies Assumption 2.3(a), condition $\Sigma(R)$, and Assumption 2.7. Then*

$$\limsup_{N\to\infty} \frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} \leq R \ a.s. \tag{3.34}$$

Proof. Pick $\eta \in (0, 1/8)$, its exact value will be determined below, and define the matrices $A_N^{\leqslant}, \mathbf{E}A_N^{\leqslant}$ as in the proof of Proposition 3.2. The proof will be accomplished once we show that

$$\limsup_N \frac{|A_N^{\leqslant}|_{\mathrm{op}}}{\sqrt{N}} \leq R, \quad \text{a.s., and} \tag{3.35}$$

$$\mathbf{P}\left(A_N \neq A_N^{\leqslant} \text{ for infinitely many } N\right) = 0. \tag{3.36}$$

PROOF OF (3.35). Since $\Sigma(R)$ holds for the sequence $(A_N)_{N \geq 1}$, the proof of Proposition 3.2 (the part with heading 3) shows that

$$\limsup_N \frac{|A_N^{\leqslant} - \mathbf{E}A_N^{\leqslant}|_{\mathrm{op}}}{\sqrt{N}} \leq \mu_\infty \quad \text{a.s.} \tag{3.37}$$

because the upper bound in (3.33) is summable with respect to N. In the same proof it is shown that

$$\limsup_N \frac{|\mathbf{E}A_N^{\leqslant}|_{\mathrm{op}}}{\sqrt{N}} = 0 \text{ a.s.}$$

Using these two facts and the triangle inequality we get (3.35).

PROOF OF (3.36). Let X be the random variable that stochastically dominates the entries of A_N in the sense of (2.13). Let X_N be a sequence of symmetric random matrices after an appropriate coupling such that for all $N \in \mathbb{N}$ and $i, j \in [N]$ it is true that

$$|a_{i,j}^{(N)}| \leq |(X_N)_{i,j}| \tag{3.38}$$

and the entries of X_N are independent up to symmetry and all following the same law as X. It is an easy exercise to show that for any $a, c > 1$ and Y real valued random variable we have

$$\sum_{k=1}^\infty a^k \mathbf{P}(|Y| \geq c^k) \leq \frac{1}{a-1} \mathbf{E}\{|Y|^{\frac{\log a}{\log c}}\}. \tag{3.39}$$

Using this inequality and the fact that the random variable X has finite $4 + \delta$ moment, we get that all $\eta \leq \frac{\delta}{\delta+4}$ satisfy

$$\sum_{m=1}^\infty 2^{2m} \mathbf{P}(|X| \geq 2^{\frac{m}{2}(1-\eta)}) < \infty. \tag{3.40}$$

Thus, picking in the beginning of the proof an arbitrary η with $0 < \eta < (1/8) \wedge (\delta/(4+\delta))$, we have

$$\mathbf{P}\left(A_N \neq A_N^{\leqslant} \text{ for infinitely many } N\right) = \mathbf{P}\left(\text{ for infinitely many } N \text{ there are } i, j \in [N] : |a_{i,j}^{(N)}| > CN^{\frac{1}{2}-\eta}\right) \tag{3.41}$$

$$\leq \mathbf{P}\left(\text{ for infinitely many } N \text{ there are } i, j \in [N] : |X_{i,j}^{(N)}| > CN^{\frac{1}{2}-\eta}\right) = \mathbf{P}\left(X_N \neq X_N^{\leqslant} \text{ for infinitely many } N\right). \tag{3.42}$$

In the second line, the inequality is a consequence of (3.38), and the matrix $X_N^{\leqslant}$ is the matrix whose (i, j) element is $(X_N)_{i,j} \mathbf{1}_{|(X_N)_{i,j}| \leq CN^{\frac{1}{2}-\eta}}$. The convergence of the series in (3.40). implies that the probability in the right hand side of (3.42) is 0 (see [2], pages 94 and 95) and finishes the proof of (3.36). $\qquad \square$

4 Proof of Theorem 2.8

The convergence $\mu_{A_N/\sqrt{N}} \Rightarrow \mu$ in probability implies that

$$\liminf_N \frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} \geq \mu_\infty \quad \text{in probability,} \tag{4.1}$$

that is, for all $\epsilon > 0, \lim_{N \to \infty} \mathbf{P}(|A_N|_{\mathrm{op}}/\sqrt{N} < \mu_\infty - \epsilon) = 0$. So in order to prove Theorem 2.8 one needs to prove that

$$\limsup_N \frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} \leq \mu_\infty \tag{4.2}$$

in probability. By Proposition 3.2, it is enough to prove that condition $\Sigma(\mu_\infty)$ is satisfied.

We will prove condition $\Sigma(\mu_\infty)$ separately for each of the Assumptions 2.4 and 2.6 in the next two lemmas.

Lemma 4.1. *Let $(A_N)_{N \in \mathbb{N}^+}$ be a sequence of matrices that satisfies Assumptions 2.2, 2.3, and 2.4. Then for every $k, N \in \mathbb{N}^+$ such that $k < N$ it is true that*

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

In case $\mu_\infty > 0$, the inequality is true (as equality) for $k = 0$ also.

Proof. Fix $N, k \in \mathbb{N}^+$ with $k < N$ 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

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

with

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

for all $i_1, i_2, \ldots, i_{k+1} \in [N]$. Each $\varphi_{\mathbf{d}}$ is one to one and, for different vectors $\mathbf{d}, \mathbf{d}' \in \{-1, 0\}^{k+1}$, the image of $\varphi_{\mathbf{d}}$ is disjoint from that of $\varphi_{\mathbf{d}'}$. If G is a plane rooted tree whose vertices in order of appearance in a depth first search are $(i_1, i_2, \ldots, i_{k+1}) \in [N]^{k+1}$, and $\varphi_{\mathbf{d}}(i_1, i_2, \ldots, i_{k+1}) = (j_1, j_2, \ldots, j_{k+1})$, we denote by $\varphi_{\mathbf{d}}(G)$ the plane rooted tree with vertex set $\{j_1, j_2, \ldots, j_{k+1}\}$, root j_1, and edges $\{\{j_a, j_b\} : \{a, b\}$ is an edge of $G\}$. Note that if all coordinates of $\mathbf{i} \in [N]^{k+1}$ are different, the same is true for the coordinates of $\varphi_{\mathbf{d}}(\mathbf{i})$.

Lastly, by assumption 2.4, for any $T \in \mathbf{C}_k, \mathbf{i} \in [N]^{2k}$ such that $G(\mathbf{i}) \sim T$ and $\mathbf{d} \in \{-1, 0\}^{k+1}$, it is true that

$$\prod_{\{i,j\} \in E(G(\mathbf{i}))} s_{i,j}^{(N)} \leq \prod_{\{i,j\} \in E(\varphi_{\mathbf{d}}(G(\mathbf{i})))} s_{i,j}^{(2N)}. \tag{4.3}$$

So if one sums over all possible trees in $\mathbf{C}_k$ and $d \in \{-1, 1\}^{k+1}$, (4.3) implies that

$$2^{k+1} M_N(k) = \sum_{\mathbf{d} \in \{-1,0\}^{k+1}} \sum_{T \in \mathbf{C}_k} \sum_{\mathbf{i} \in [N]^{2k}: G(\mathbf{i}) \sim T} \prod_{\{i,j\} \in E(G(\mathbf{i}))} s_{i,j}^{(N)} \leq \sum_{\mathbf{d} \in \{-1,0\}^{k+1}} \sum_{T \in \mathbf{C}_k} \sum_{\mathbf{i} \in [N]^{2k}: G(\mathbf{i}) \sim T} \prod_{\{i,j\} \in E(\varphi_{\mathbf{d}}(G(\mathbf{i})))} s_{i,j}^{(2N)} \leq M_{2N}(k). \tag{4.4}$$

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

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

is increasing in the variable m. So by (2.5) it is true that

$$\sup_m q_m = \lim_{m \to \infty} q_m = \int x^{2k} d\mu(x) \leq \mu_\infty^{2k}.$$

In particular, $q_0 \leq \mu_\infty^{2k}$, completing the proof. $\square$

Lemma 4.2. *Suppose $(A_N)_{N \in \mathbb{N}^+}$ is a sequence of matrices such that Assumptions 2.3, 2.6 hold. Then for each $C_1 > 0$ there is $C_2 > 0$ such that*

$$M_N(k) \leq C_2 N^{k+1} \mu_\infty^{2k} \tag{4.5}$$

for all $N \in \mathbb{N}^+$ and $1 \leq k \leq C_1 \log N$.

Proof. Note that for $1 \leq k < N$,

$$\frac{1}{N^{k+1}} M_N(k) \leq \sum_{T \in \mathbf{C}_k} \int_{[0,1]^{k+1}} \left(\prod_{\{i,j\} \in E(T)} W_N(x_i, x_j) \right) dx_1 dx_2 \cdots dx_{k+1} =: \Xi_N(k). \tag{4.6}$$

The inequality holds because the left hand side results if on the right hand side we restrict the domain of integration to the union of the sets $\prod_{r=1}^{k+1}((i_r - 1)/N, i_r/N]$ where all $i_1, i_2, \ldots, i_{k+1} \in [N]$ are different. Thus, it is enough to show (4.5) with the left hand side replaced by $N^{k+1} \Xi_N(k)$.

Fix $T \in \mathbf{C}_k$ and enumerate the edges of T in the order of first appearance during a depth first search algorithm. For $\{i, j\} \in E(T)$, let $\{i, j\}_{\text{ord}}$ be its enumeration. Then for any integer $l \in [0, k]$ define the following quantities.

$$\mu_N^{(l)}(k, T) = \int_{[0,1]^{k+1}} \prod_{\{i,j\} \in E(T):\{i,j\}_{\text{ord}} \leq l} W_N(x_i, x_j) \prod_{\{i,j\} \in E(T):\{i,j\}_{\text{ord}} \geq l+1} W(x_i, x_j) dx_1 dx_2 \cdots dx_{k+1}. \tag{4.7}$$

Note that

$$\sum_{T \in \mathbf{C}_k} \mu_N^{(0)}(k, T) = \mu_{2k}, \quad \sum_{T \in \mathbf{C}_k} \mu_N^{(k)}(k, T) = \Xi_N(k) \tag{4.8}$$

Fix $D > 0$. Since all the variances are uniformly bounded by 1, Assumption (2.12) implies that there exists some $N_0(D)$ and $C > 0$ such that for $N \geq N_0(D)$ and any $1 \leq l \leq k < N$ it is true that

$$|\mu_N^{(l)}(k, T) - \mu_N^{(l-1)}(k, T)| \leq \int_{[0,1]^2} |W_N(x, y) - W(x, y)| \, dx \, dy \leq C \frac{1}{N^D}. \tag{4.9}$$

Consequently, since $|\mathbf{C}_k| \leq 2^{2k}$, for $k < N$ we have

$$|\Xi_N(k) - \mu_{2k}| \leq \sum_{T \in \mathbf{C}} \left| \sum_{l=1}^k \{\mu_N^{(l)}(k, T) - \mu_N^{(l-1)}(k, T)\} \right| \leq \sum_{T \in \mathbf{C}_k} \sum_{l=1}^k |\mu_N^{(l)}(k, T) - \mu_N^{(l-1)}(k, T)| \leq \frac{Ck2^{2k}}{N^D}. \tag{4.10}$$

Pick any $D > -2C_1 \log(\mu_\infty/2)$. Then there is $N_0' \in \mathbb{N}^+, N_0' > N_0(D)$ such that $Ck2^{2k}/N^D \leq \mu_\infty^{2k}$ for all $N > N_0'$ and $1 \leq k \leq C_1 \log N$. And since $\mu_{2k} \leq \mu_\infty^{2k}$, we will have $\Xi_N(k) \leq 2\mu_\infty^{2k}$ for the same N and k. If we choose a constant $C_2 \geq 2$ so that (4.5) is satisfied for $N \in [N_0']$ and $1 \leq k \leq C_1 \log N$, then we will have (4.5) for all N, k claimed. $\qquad \square$

4.1 Proof of almost sure convergence under the additional Assumption 2.7

The convergence in probability that we have proven so far gives

$$\liminf_{N \to \infty} \frac{|A_N|_{op}}{\sqrt{N}} \geq \mu_\infty \text{ a.s.} \tag{4.11}$$

The opposite inequality follows from Lemma 3.3 whose assumptions are satisfied, with $R = \mu_\infty$, because, under both scenarios of the Theorem, assumption $\Sigma(\mu_\infty)$ holds.

5 Proof of Theorem 2.10

The plan is to write the matrix A_N as $A_N^{(1)} + A_N^{(2)}$ so that for the sequence $\{A_N^{(1)}\}_{N \geq 1}$ we can apply Theorem 2.8 while for $\{A_N^{(2)}/\sqrt{N}\}_{N \geq 1}$ the operator norm will tend to zero.

Let

$$\mathcal{D}_N := \{(i, j) \in [N]^2 : \text{ there exists } m \in [d_N] : (i, j) \in (\mathcal{B}_m^{(N)})^\circ\}. \tag{5.1}$$

Then define the matrices

$$\{A_N^{(1)}\}_{i,j} := \mathbf{1}_{(i,j) \in \mathcal{D}_N} \{A_N\}_{i,j}, \tag{5.2}$$

$$\{A_N^{(2)}\}_{i,j} = \mathbf{1}_{(i,j) \notin \mathcal{D}_N} \{A_N\}_{i,j}. \tag{5.3}$$

The proof follows from the triangle inequality for the operator norm and the following two statements, which we are going to prove next.

$$\lim_{N \to \infty} \frac{|A_N^{(2)}|_{op}}{\sqrt{N}} = 0 \quad \text{in probability.} \tag{5.4}$$

$$\lim_{N \to \infty} \frac{|A_N^{(1)}|_{op}}{\sqrt{N}} = \mu_\infty. \tag{5.5}$$

PROOF OF (5.4). For any $k \in \mathbb{N}$ denote by $M_N^{(2)}(k)$ the quantity (2.4) but with the role of A_N played by $A_N^{(2)}$, i.e., $s_{i,j}^{(N)}$ is replaced by $s_{i,j}^{(N)} \mathbf{1}_{(i,j) \notin \mathcal{D}_N}$ By Proposition 3.2 it is sufficient to prove that for any constant $C_1 > 0$ it is true that for any $k \leq C_1 \log N$,

$$M_N^{(2)}(k) \leq N(8d_N)^k. \tag{5.6}$$

This is true because each product in (2.4) is at most 1, then the inner sum has at most $N(2d_N)^k$ non zero terms [there are N choices for i_1, and then, for each choice of i_1 there are at most $2d_N$ choices for i_2 that

have $s_{i_1,i_2}^{(N)} \neq 0$ due to condition (c) of Definition 2.9, and the same restriction holds for $i_3, \ldots, i_{k+1}]$ and the outer sum has $|\mathbf{C}_k| \leq 4^k$ terms.

PROOF OF (5.5). We will show that Theorem 2.8 can be applied to the sequence $\{A_N^{(1)}\}_{N \geq 1}$. First we prove that

$$\mu_{A_N^{(1)}/\sqrt{N}} \Rightarrow \mu \text{ in probability as } N \to \infty. \tag{5.7}$$

As remarked after relation (2.5), $\mu_{A_N/\sqrt{N}} \Rightarrow \mu$ in probability as $N \to \infty$. Then, from a well known inequality (Corollary A.41 in [2]), the Levy distance between $\mu_{A_N/\sqrt{N}}$ and $\mu_{A_N^{(1)}/\sqrt{N}}$ is bounded as follows.

$$L^3(\mu_{A_N/\sqrt{N}}, \mu_{A_N^{(1)}/\sqrt{N}}) \leq \frac{1}{N}\mathrm{tr}\left\{\left(\frac{1}{\sqrt{N}}A_N - \frac{1}{\sqrt{N}}A_N^{(1)}\right)^2\right\} = \frac{1}{N^2}\sum_{i,j \in [N]}\{(A_N^{(2)})_{i,j}\}^2. \tag{5.8}$$

The expectation of the rightmost quantity is at most $N^{-2}N2d_N$ (since each row of $A_N^{(2)}$ has at most $2d_N$ elements that are not identically zero random variables and these random variables have second moment at most 1), which tends to 0 as $N \to 0$ because of the assumption on d_N.

Then the sequence $\{A_N^{(1)}\}_{N \geq 1}$ satisfies:

• Assumption 2.2 with the same measure as $\{A_N\}_{N \geq 1}$. This follows from Lemma 9.2. Assumption (c) of that lemma is satisfied because of (5.7).

• Assumption 2.3, this is clear,

• Assumption 2.4. Indeed, fix $(i,j) \in [N]^2$. If $(i,j) \in \mathcal{D}_N$, there exists some $m \in [d_N]$ such that

$$\{(i + d_1, j + d_2) : d_1, d_2 \in \{-1, 0, 1\}\} \subseteq \mathcal{B}_m^{(N)}.$$

Then Assumption (2.15) implies that there exists some $f \in [d_{2N}]$ such that

$$\{(2i + d_1, 2j + d_2) : d_1, d_2 \in \{-2, 0, 2\}\} \subseteq \mathcal{B}_f^{(2N)}.$$

But since $\mathcal{B}_f^{(2N)}$ is axially convex (see before Definition 2.9), one can conclude that

$$\{(2i + d_1, 2j + d_2) : d_1, d_2 \in \{-2, -1, 0, 1, 2\}\} \subseteq \mathcal{B}_f^{(2N)}.$$

Now since $(k, \ell) \mapsto s_{k,\ell}^{(2N)}$ is constant in $\mathcal{B}_f^{(2N)}$ [see (2.16)] and we assumed (2.17), our claim follows.

Thus, all the Assumptions of Theorem 2.8 hold for $A_N^{(1)}$, and hence (5.5) holds.

Almost sure convergence under the additional Assumption 2.7. Using Lemma 3.3, we will prove that

$$\limsup_{N \to \infty} \frac{|A_N^{(1)}|_{\mathrm{op}}}{\sqrt{N}} \leq \mu_\infty \quad \text{a.s.} \tag{5.9}$$

$$\limsup_{N \to \infty} \frac{|A_N^{(2)}|_{\mathrm{op}}}{\sqrt{N}} \leq \epsilon \quad \text{a.s. for any } \epsilon > 0. \tag{5.10}$$

And these are enough to prove our claim.

Notice that the validity of Assumptions 2.3(a) and 2.7 for the sequence $(A_N)_{N \in \mathbb{N}^+}$ implies the validity of the same assumptions for the sequences $(A_N^{(1)})_{N \in \mathbb{N}^+}$ and $(A_N^{(2)})_{N \in \mathbb{N}^+}$. As was mentioned above, the sequence $\{A_N^{(1)}\}_{N \geq 1}$ satisfies Assumption 2.2 with the same measure as $\{A_N\}_{N \geq 1}$. And then Lemma 4.1 implies that the sequence $(A_N^{(1)})_{N \in \mathbb{N}^+}$ satisfies condition $\Sigma(\mu_\infty)$, while (5.6) and $\lim_{N \to \infty} d_N/n = 0$ imply that, for any $\epsilon > 0$, the sequence $(A_N^{(2)})_{N \in \mathbb{N}^+}$ satisfies condition $\Sigma(\epsilon)$. Thus, Lemma 3.3 applies and gives the desired inequalities.

6 Step function profile. Proof of Theorem 2.14

Proof of Theorem 2.14. The inequality $\liminf_{N \to \infty} |A_N|_{\mathrm{op}}/\sqrt{N} \geq \mu_\infty^\sigma$ almost surely is justified with the same argument as (4.1) with the only difference that here we have $\mu_{A_N/\sqrt{N}} \Rightarrow \mu^\sigma$ a.s., and so the inequality will be true in the a.s. sense.

For the reverse inequality, we will apply Lemma 3.3. To check Assumptions 2.3(a) and 2.7, required by that lemma, note that the (i,j) element of A_N is of the form $\sigma_{p,q}X_0'$ for a constant $\sigma_{p,q} \in [0, 1]$ and $X_0' \stackrel{d}{=} X_0$, and

clearly X_0 can play the role of X in relation (2.13). We will prove that $(A_N)_{N\in\mathbb{N}^+}$ satisfies condition $\Sigma(c)$ for all $c > \mu_\infty$, and this will finish the proof.

Define the matrix $\hat{\Sigma}_N \in \mathbb{R}^{N\times N}$ by

$$(\hat{\Sigma}_N)_{i,j} = \sigma_{p,q} \qquad \text{if} \quad \begin{aligned} &a_{p-1} + (1/N) \le i/N < a_p \text{ and} \\ &a_{q-1} + (1/N) \le j/N < a_q, \end{aligned} \tag{6.1}$$

and $\hat{A}_N := \hat{\Sigma}_N \odot A'_N$.

Also, let

$$\varepsilon_N := \max\left\{ \left| \frac{R_p^{(N)}}{N} - a_p \right| : p \in [m-1] \right\}. \tag{6.2}$$

By Definition 2.11, it holds $\lim_{N\to\infty} \varepsilon_N = 0$.

Claim 1: a) With probability one, $\hat{A}_N / \sqrt{N}$ has the same limiting ESD as $A_N / \sqrt{N}$.

b) For $\hat{A}_N$, Lemma 4.1 applies.

Consequently,

$$\hat{M}_N(k) \le N^{k+1} (\mu_\infty^\sigma)^{2k} \tag{6.3}$$

for all $1 \le k < N$.

Proof of Claim 1:

a) This is true because by Theorem A.43 in the Appendix A of [2], the Kolmogorov distance between $\mu_{A_N/\sqrt{N}}$ and $\mu_{\hat{A}_N/\sqrt{N}}$ is at most

$$\frac{1}{N} \operatorname{rank}(A_N - \hat{A}_N) \le \frac{m}{N} \max_{p\in[m]} (\max\{R_p, Na_p\} - \min\{R_p, Na_p\}) = m\varepsilon_N \overset{N\to\infty}{\to} 0. \tag{6.4}$$

b) Assumption 2.3 is satisfied because $\mathbf{E}(|X_0|^{4+\delta})$ and $\sigma_{p,q} \le 1$ for all $p, q \in [m]$. To show that Assumption 2.2 is satisfied, we repeat the argument just before the statement of the Theorem. For the sequence $(\hat{A}_N)_{N\in\mathbb{N}^+}$, the corresponding $W_N(x,y)$, as $N \to \infty$, converges to $\sigma^2(x,y)$ for almost all $(x,y) \in [0,1]^2$. Assumption 2.4 is satisfied because if for some i,j we have $\operatorname{Var}[(\hat{A}_N)_{i,j}] > 0$, then this equals $\sigma_{p,q}^2$ for the unique p, q as in (6.1). Then

$$\frac{2i-1}{2N} \in [a_{p-1} + \frac{1}{2N}, a_p - \frac{1}{2N}), \frac{2i}{2N} \in [a_{p-1} + \frac{1}{N}, a_p), \tag{6.5}$$

$$\frac{2j-1}{2N} \in [a_{q-1} + \frac{1}{2N}, a_q - \frac{1}{2N}), \frac{2j}{2N} \in [a_{q-1} + \frac{1}{N}, a_q). \tag{6.6}$$

Thus, (2.8) holds as equality.

Claim 2: There is $\partial \in (0, \infty)$ so that $M_N(k) \le e^{\partial(k+1)\varepsilon_N} \hat{M}_N(k)$ for all $k < N$.

Proof of Claim 2: Define the following sets of indices.

$$\Delta_p^{(N)} := I_p^{(N)} \cap [a_{p-1}N + 1, a_pN), \tag{6.7}$$

$$\Delta^{(N)} := \cup_{p=1}^m \Delta^{(N)}. \tag{6.8}$$

When $p = m$, the interval in the intersection becomes closed on the right also. Then

$$M_N(k) \le \hat{M}_N(k) + \sum_{\emptyset \neq J \subset [k+1]} \sum_{T\in\mathbf{C}_k} \sum_{i_1\cdots i_{k+1}} \mathbf{1}\left(i_l \notin \Delta^{(N)} \text{ if } l \in J \text{ and } i_l \in \Delta^{(N)} \text{ if } l \notin J \right) \prod_{\{i,j\}\in E(T)} s_{i,j}^{(N)} \tag{6.9}$$

$$= \hat{M}_N(k) + \sum_{\emptyset \neq J \subset [k+1]} \sum_{T\in\mathbf{C}_k} \sum_{m_1,\ldots,m_{k+1}\in[m]^{k+1}} a(T, J, m_1, m_1, \ldots, m_{k+1}) \tag{6.10}$$

where

$$a(T, J, m_1 \cdots m_{k+1}) := \sum_{i_1\in I_{m_1}^{(N)}, \cdots i_{k+1}\in I_{m_{k+1}}^{(N)}} \mathbf{1}(i_l \notin \Delta^{(N)} \text{ if } l \in J, i_l \in \Delta^{(N)} \text{ if } l \notin J, (i_\ell)_{\ell\in[k+1]} \text{distinct}) \prod_{\{i,j\}\in E(T)} s_{i,j}^{(N)}. \tag{6.11}$$

Note that

$$\sum_{T\in\mathbf{C}_k} \sum_{m_1,m_2\cdots m_{k+1}\in[m]^{k+1}} a(T, \emptyset, m_1 \cdots m_{k+1}) \le \hat{M}_N(k). \tag{6.12}$$

We will show that for some constant $\partial = \partial(I_1, I_2, \ldots, I_m) \in (0, \infty)$ we have

$$a(T, J, m_1 \cdots m_{k+1}) \le (\partial \varepsilon_N)^{|J|} a(T, \emptyset, m_1, \cdots m_{k+1}). \tag{6.13}$$

In the definition of $a(T, J, m_1, m_2, \ldots, m_{k+1})$, the product is common to all summands [recall the rectangles of constancy of the map $(i,j) \mapsto s_{i,j}^{(N)}$]. We write $a(T, J, m_1 \cdots m_{k+1})$ and $a(T, \emptyset, m_1, \ldots m_{k+1})$ as

$$\sum_{i_\ell \in I_{m_\ell}^{(N)} \text{ for } \ell \notin J} \mathbf{1}(i_l \in \Delta^{(N)} \text{ for } l \notin J, (i_\ell)_{\ell \in [k+1]\setminus J} \text{ distinct}) \sum_{i_\ell \in I_{m_\ell}^{(N)} \text{ for all } \ell \in J} \mathbf{1}(i_l \notin \Delta^{(N)} \text{ for all } l \in J, (i_\ell)_{\ell \in [k+1]} \text{distinct}) \prod_{\{i,j\} \in E(T)} s_{i,j}^{(N)}, \tag{6.14}$$

$$\sum_{i_\ell \in I_{m_\ell}^{(N)} \text{ for } \ell \notin J} \mathbf{1}(i_l \in \Delta^{(N)} \text{ for } l \notin J, (i_\ell)_{\ell \in [k+1]\setminus J} \text{ distinct}) \sum_{i_\ell \in I_{m_\ell}^{(N)} \text{ for all } \ell \in J} \mathbf{1}(i_l \in \Delta^{(N)} \text{ for all } l \in J, (i_\ell)_{\ell \in [k+1]} \text{distinct}) \prod_{\{i,j\} \in E(T)} s_{i,j}^{(N)} \tag{6.15}$$

We will compare the inner sums in the two expressions. Notice that there are $C_1, C_2 > 0$ that depend on $a_1, a_2, \ldots, a_m$ only so that

$$|\Delta_p^{(N)}| \ge C_1 N, \tag{6.16}$$

$$|I_p^{(N)} \setminus \Delta^{(N)}| \le C_2 \varepsilon_N N \tag{6.17}$$

for all $p \in [m]$. For each fixed collection $(i_\ell)_{\ell \notin J}$, the inner sum in (6.14) is at most $(C_2 \varepsilon_N N)^{|J|}$ while the inner sum in (6.15) is at least $(C_1 N/2)^{|J|}$. The ratio of the first over the second bound is $(2C_2 \varepsilon_N / C_1)^{|J|}$. Thus, we get (6.13) with $\partial := 2C_2/C_1$.

Taking into account (6.13) and (6.12), we get that the second summand in (6.10) is bounded above by

$$\sum_{t=1}^{k+1} \sum_{J \subset [k+1]: |J|=t} \hat{M}_N(k)(\partial \varepsilon_N)^t = M_N^{(1)}(k) \sum_{t=1}^{k+1} \binom{k+1}{t} (\partial \varepsilon_N)^t = \hat{M}_N(k) \left\{ (1 + \partial \varepsilon_N)^{k+1} - 1 \right\} \tag{6.18}$$

Consequently, $M_N(k) \le (1 + \partial \varepsilon_N)^{k+1} \hat{M}_N(k) \le e^{\partial(k+1)\varepsilon_N} \hat{M}_N(k)$, and this proves Claim 2.

Now, combining this with (6.3), we get that condition $\Sigma((1 + \varepsilon)\mu_\infty)$ is satisfied for each $\varepsilon > 0$. □

7 An approximation result and proof of Corollary 2.15

Proposition 7.1. *Let $(A_N)_{N \in \mathbb{N}^+}$ be a sequence of symmetric random matrices, A_N of dimension $N \times N$, of the form*

$$A_N = \Sigma_N \odot A_N', \tag{7.1}$$

where $\Sigma_N \in [0, \infty)^{N \times N}$ and A_N' is a random $N \times N$ symmetric matrix with independent entries (up to symmetry) all with zero mean and unit variance.

For every $n \in \mathbb{N}^+$ consider a sequence $(\Sigma_N^{(n)})_{N \in \mathbb{N}^+}$ of matrices, with $\Sigma_N^{(n)} \in [0, \infty)^{N \times N}$, and define

$$A_N^{(n)} := \Sigma_N^{(n)} \odot A_N' \quad \text{for each } N \in \mathbb{N}^+. \tag{7.2}$$

(a) Assume that

(i) the sequence $(A_N')_{N \in \mathbb{N}^+}$ satisfies Assumption 2.3,

(ii) for each $n \in \mathbb{N}^+$ it holds

$$\lim_{N \to \infty} \frac{|A_N^{(n)}|_{\text{op}}}{\sqrt{N}} = \mu_\infty^{(n)} \text{ in probability,} \tag{7.3}$$

where $\mu_\infty^{(n)}$ is a finite constant,

(iii)

$$\lim_{n \to \infty} \mu_\infty^{(n)} =: \mu_\infty \in \mathbb{R}, \tag{7.4}$$

(iv)

$$\lim_{n\to\infty}\limsup_{N\to\infty}|\Sigma_N - \Sigma_N^{(n)}|_{\max} = 0. \tag{7.5}$$

Then

$$\lim_{N\to\infty}\frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} = \mu_\infty \qquad \text{in probability.} \tag{7.6}$$

(b) Assume that, in addition to the assumptions of (a), the convergence in (7.3) holds in the a.s. sense and Assumption 2.7 holds for the sequence $(A'_N)_{N\in\mathbb{N}^+}$. Then the limit in (7.6) holds in the a.s. sense.

Proof. (a) Fix $\epsilon \in (0, 1/2)$ and n_0 large enough such that for every $n \geq n_0$ it is true that

$$|\mu_\infty - \mu_\infty^{(n)}| \leq \epsilon \text{ and } \limsup_{N\to\infty}|\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon.$$

Fix an $n \geq n_0$. There is an $N_0 = N_0(n) \in \mathbb{N}^+$ so that $|\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$ for all $N \geq N_0$. Then for $N \geq N_0$ we have

$$\mathbf{P}\left(\left|\frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} - \mu_\infty\right| \geq 5\epsilon\right) \leq \mathbf{P}\left(\left|\frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} - \mu_\infty^{(n)}\right| \geq 4\epsilon\right) \leq \mathbf{P}\left(\frac{|A_N - A_N^{(n)}|_{\mathrm{op}}}{\sqrt{N}} \geq 3\epsilon\right) + \mathbf{P}\left(\left|\frac{|A_N^{(n)}|_{\mathrm{op}}}{\sqrt{N}} - \mu_\infty^{(n)}\right| \geq \epsilon\right). \tag{7.7}$$

The last term in (7.7) converges to zero as $N \to \infty$ due to (7.3). For the previous term we will apply Proposition 3.2. Notice that the sequence $(A_N - A_N^{(n)})_{N\in\mathbb{N}^+}$ satisfies

- Assumption 2.3 because $(A_N - A_N^{(n)})_{i,j} = ((\Sigma_N)_{i,j} - (\Sigma_N^{(n)})_{i,j})(A'_N)_{i,j}$ and $|(A_N - A_N^{(n)})_{i,j}| \leq |(A'_N)_{i,j}|$ (for all $N \in \mathbb{N}^+, i,j \in [N]$) and we assumed that $(A'_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 2.3

- condition $\Sigma(2\varepsilon)$ because if, for $t \in \mathbb{N}^+$ with $t < N$, we call $M'_N(t)$ the quantity defined in (2.4) for the matrix $A_N - A_N^{(n)}$, and note that the (i,j) element of $A_N - A_N^{(n)}$ has mean zero and variance $\{(\Sigma_N)_{i,j} - (\Sigma_N^{(n)})_{i,j}\}^2$, we obtain that

$$M'_N(t) \leq N^{t+1}2^{2t}(|\Sigma_N - \Sigma_N^{(n)}|_{\max})^{2t} < N^{t+1}(2\epsilon)^{2t}. \tag{7.8}$$

Since $3\epsilon > 2\epsilon(1 + \epsilon)$, Proposition 3.2 implies that the penultimate term in (7.7) goes to zero as $N \to \infty$.

(b) It is enough to prove that with probability 1 it holds $\overline{\lim}_{N\to\infty}\frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} \leq \mu_\infty$. Because of (7.3) (holding a.s.) and (7.4), it is enough to prove that for all $\epsilon > 0$ and all n large enough, with probability 1, it holds

$$\overline{\lim}_{N\to\infty}\frac{|A_N - A_N^{(n)}|_{\mathrm{op}}}{\sqrt{N}} \leq 2\epsilon. \tag{7.9}$$

To prove this, we will apply Lemma 3.3. Take n_0 so that for all $n \geq n_0$ it holds $\limsup_{N\to\infty}|\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$. Now fix $n \geq n_0$. There is $N_0 = N_0(n) \in \mathbb{N}^+$ so that $|\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$ for all $N \geq N_0$. Then the sequence $(A_N - A_N^{(n)})_{N\geq N_0}$ satisfies Assumption 2.3(a) as we saw in part a) of the proposition, Assumption 2.7 (because A'_N does so and $|\Sigma_N - \Sigma_N^{(n)}|_{\max} < 1$), and assumption $\Sigma(2\epsilon)$ because of (7.8). Then Lemma 3.3 gives the desired inequality. □

Proof of Corollary 2.15. We will apply Proposition 7.1(b) for the sequence $(A_N)_{N\in\mathbb{N}^+}$. The sequence $(A'_N)_{N\in\mathbb{N}^+}$ mentioned in that Proposition is exactly the sequence $(A'_N)_{N\in\mathbb{N}^+}$ of relation (2.24) and it satisfies Assumption 2.3 because for it the discussion following Assumption 2.3 applies (X_0 has finite fourth moment).

For each $n \in \mathbb{N}^+$, we define the following obvious approximation to σ.

$$\sigma^{(n)}(x, y) := n^2 \int_{I_k}\int_{I_\ell}\sigma(a, b)\,da\,db \text{ if } (x, y) \in I_k \times I_\ell \text{ for } k, \ell \in [n], \tag{7.10}$$

where $I_k := [\frac{k-1}{n}, \frac{k}{n})$ for $k \in [n-1]$ and $I_n := [(n-1)/n, 1]$. Then, we define the matrices $\Sigma_N^{(n)}$ through the relation $(\Sigma_N^{(n)})_{i,j} := \sigma^{(n)}(i/N, j/N)$.

For each $n \in \mathbb{N}^+$, the sequence of matrices $(\Sigma_N^{(n)} \odot A'_N)_{N\geq 1}$ satisfies the assumptions of Theorem 2.14. Consequently, as $N \to \infty$, the sequence $(\mu_{A_N^{(n)}/\sqrt{N}})_{N\in\mathbb{N}^+}$ converges almost surely weakly to a symmetric measure,

say $\mu^{(n)}$, with support contained in $[-\mu_\infty^{(n)}, \mu_\infty^{(n)}]$ and (7.3) holds in the a.s. sense. In a claim below we prove that condition (7.4) is satisfied. Finally, to check (7.5), note that

$$\left|(\Sigma_N)_{i,j} - (\Sigma_N^{(n)})_{i,j}\right| \leq \left|(\Sigma_N)_{i,j} - \sigma(i/N, j/N)\right| + \left|\sigma(i/N, j/N) - \sigma^{(n)}(i/N, j/N)\right|. \tag{7.11}$$

In the right hand side of the last inequality, the first term converges to zero as $N \to \infty$ due to (8.2), and the second term is at most the supremum norm of $\sigma - \sigma^{(n)}$, which goes to zero as $n \to \infty$ because σ is uniformly continuous in $[0, 1]^2$. Thus, Proposition 7.1(b) applies and completes the proof.

CLAIM: Condition (7.4) is satisfied.

We modify the proof of Lemma 6.4 of [24]. Call μ the weak limit as $N \to \infty$ of $\mu_{A_N / \sqrt{N}}$, then $F_N, F_N^{(n)}$ the distribution function of $\mu_{A_N / \sqrt{N}}$ and $\mu_{A_N^{(n)} / \sqrt{N}}$ respectively, and $F, F^{(n)}$ the distribution function of $\mu, \mu^{(n)}$ respectively. Let

$$\lambda_{N,1} \leq \lambda_{N,2} \leq \cdots \leq \lambda_{N,N},$$
$$\lambda_{N,1}^{(n)} \leq \lambda_{N,2}^{(n)} \leq \cdots \leq \lambda_{N,N}^{(n)}$$

the eigenvalues of $A_N / \sqrt{N}, A_N^{(n)} / \sqrt{N}$ respectively.

Let $\epsilon \in (0, 1/2)$. There is $n_0 = n_0(\epsilon)$ so that for all $n \geq n_0$ it holds $\limsup_{N \to \infty} |\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$. Take now $n \geq n_0$ fixed. There is $N_0 = N_0(n) \in \mathbb{N}^+$ so that $|\Sigma_N - \Sigma_N^{(n)}|_{\max} < \epsilon^2$ for all $N \geq N_0$. As explained in the proof of Proposition 7.1, $\lim_{N \to \infty} \mathbf{P}\left(|A_N - A_N^{(n)}|_{op} \geq 3\epsilon \sqrt{N}\right) = 0$. There is sequence $(N_k)_{k \geq 1}$ so that in a set Ω_ϵ of probability 1, eventually for all k we have $|A_{N_k} - A_{N_k}^{(n)}|_{op} < 3\epsilon \sqrt{N_k}$. Since

$$\max_{i \in [N]} |\lambda_{N_k,i}^{(n)} - \lambda_{N_k,i}| \leq |A_{N_k} - A_{N_k}^{(n)}|_{op} / \sqrt{N_k},$$

in Ω_ϵ (the inequality is true by Theorem A46 in [2]), we will have eventually for all $k \in \mathbb{N}^+$ that

$$F_{N_k}^{(n)}(a - 3\epsilon) \leq F_{N_k}(a) \leq F_{N_k}^{(n)}(a + 3\epsilon) \tag{7.12}$$

for all $a \in \mathbb{R}$. From here, using the convergence as $N \to \infty$ of F_N to F and of $F_N^{(n)}$ to $F^{(n)}$, we have that for all $a \in \mathbb{R}$ it holds

$$F^{(n)}(a - 3\epsilon) \leq F(a) \leq F^{(n)}(a + 3\epsilon). \tag{7.13}$$

[First we get this for all a outside a countable subset of $\mathbb{R}$ and then using the right continuity of $F, F^{(n)}$ we get it for all $a \in \mathbb{R}$.] This implies that $|\mu_\infty^{(n)} - \mu_\infty| \leq 3\epsilon$ and finishes the proof of the claim. $\square$

Remark 7.2. The above proof easily generalizes to the case that the function σ is piecewise continuous in the following sense. There are $m \in \mathbb{N}^+$, $0 = a_0 < a_1 < \cdots < a_{m-1} < a_m = 1$ so that letting $I_p := [a_{p-1}, a_p)$ for $p = 1, 2, \ldots, m - 1$, and $I_m = [a_{m-1}, 1]$ the function $\sigma|I_p \times I_q$ is uniformly continuous for all $p, q \in [m]$ (i.e., when σ extends continuously in the closure of each rectangle $I_p \times I_q$. Recall that to handle the last term in (7.11) all we needed was the uniform continuity of σ.

8 Examples

8.1 Random Gram matrices

Let $(X_N)_{N \in \mathbb{N}^+}$ be a sequence of matrices so that X_N is an $M(N) \times N$ matrix with independent, centered entries with unit variance, and $M : \mathbb{N}^+ \to \mathbb{N}^+$ a function with $\lim_{N \to \infty} \frac{M(N)}{N} = c \in (0, \infty)$. It is known that the empirical spectral distribution of XX^T, after rescaling, converges to the Marchenko-Pastur law μ_{MP} [25]. Moreover, the convergence of the rescaled largest eigenvalue to the largest element of the support of μ_{MP} has been established in [29] under the assumption of finite fourth moment for the entries. However, some applications in wireless communication require understanding the spectrum of XX^T, where X has a variance profile, see for example [21] or [15]. 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 empirical spectral distribution for specific variance profiles. Firstly we give some definitions.

Definition 8.1 (Step function variance profile). Consider

    a) $m, n \in \mathbb{N}^+$ and numbers $\{\sigma_{p,q}\}_{p\in[m],q\in[n]} \in [0,\infty)^{mn}$.

    b) For each $K \in \mathbb{N}^+$, two partitions $\{I_p^{(K)}\}_{p\in[m]}, \{J_p^{(K)}\}_{p\in[n]}$ of $[K]$ in m and n intervals respectively. The numbering of the intervals is such that $x < y$ whenever $x \in I_p^{(K)}, y \in I_q^{(K)}$ or $x \in J_p^{(K)}, y \in J_q^{(K)}$ with $p < q$. Let $LI_p^{(K)}$ and $RI_p^{(K)}$ be the left and right endpoint respectively of $I_p^{(K)}$ and similarly $LJ_p^{(K)}$ and $RJ_p^{(K)}$ for $J_p^{(K)}$.

    c) Numbers $0 = a_0 < a_1 < \cdots < a_{m-1} < a_m := 1$. We assume that $\lim_{M\to\infty} RI_p^{(M)}/M = a_p$ for each $p \in [m]$.

    d) Numbers $0 = \beta_0 < \beta_1 < \cdots < \beta_{n-1} < \beta_n := 1$. We assume that $\lim_{N\to\infty} RJ_q^{(N)}/N = \beta_q$ for each $q \in [n]$.

    e) $M : \mathbb{N}^+ \to \mathbb{N}^+$ a function,

    f) A random variable X_0 with $\mathbf{E}(X_0) = 0, \mathbf{E}(X_0^2) = 1$.

For each $M, N \in \mathbb{N}^+$, define the matrix $\Sigma_{M,N} \in \mathbb{R}^{M\times N}$ by $(\Sigma_{M,N})_{i,j} = \sigma_{p,q}$ if $i \in I_p^{(M)}, j \in I_q^{(N)}$, and let $\{A_N\}_{N\in\mathbb{N}^+}$ be the sequence of random matrices defined by

$$A_N = \Sigma_{M(N),N} \odot A'_{M(N),N} \tag{8.1}$$

where $A'_{M(N),N}$ is an $M(N) \times N$ matrix whose elements are independent random variables all with distribution the same as X_0. We say that A_N in (8.1) is a *random matrix model whose variance profile is given by a step function*.

Definition 8.2 (Continuous function variance profile). For

    a) a continuous function $\sigma : [0,1]^2 \to [0,1]$,

    b) $M : \mathbb{N}^+ \to \mathbb{N}^+$ a function

    c) a sequence $(\Sigma_{M(N),N})_{N\in\mathbb{N}^+}$ of matrices, $\Sigma_{M(N),N} \in [0,1]^{M(N)\times N}$, with the property

$$\lim_{N\to\infty} \sup_{i\in[M(N)],j\in[N]} \left|(\Sigma_{M(N),N})_{i,j} - \sigma(i/M(N),j/N)\right| = 0, \tag{8.2}$$

    d) a random variable X_0 with $\mathbf{E}(X_0) = 0, \mathbf{E}(X_0^2) = 1$,

consider the sequence $\{A_N\}_{N\in\mathbb{N}^+}$ of random matrices, $A_N \in \mathbb{R}^{M(N)\times N}$, defined by

$$A_N = \Sigma_{M(N),N} \odot A'_N \tag{8.3}$$

where the entries of A'_N are independent random variables all with distribution the same as X_0. Then we say that $(A_N)_{N\in\mathbb{N}^+}$ is a *random matrix model whose variance profile is given by a continuous function*. Again, we call σ the variance profile.

SYMMETRIZATION To study the eigenvalues of $A_N A_N^T$, where A_N falls in one of the cases of the two last definitions, we use the trick of symmetrization. If A is an $M \times N$ matrix, where $M, N \in \mathbb{N}^+$, we call symmetrization of A the $(M+N) \times (M+N)$ symmetric matrix $\tilde{A}$ defined by

$$\tilde{A} := \begin{bmatrix} \mathbf{O}_{M,M} & A \\ A^T & \mathbf{O}_{N,N} \end{bmatrix} \tag{8.4}$$

where, for any $k, l \in \mathbb{N}^+$, $\mathbf{O}_{k,l}$ denotes the $k \times l$ matrix with all of its entries equal to 0. The characteristic polynomials of $AA^T, \tilde{A}$ are connected through the relation

$$\jmath^M \det(\jmath \mathbb{I}_{M+N} - \tilde{A}) = \jmath^N \det(\jmath^2 \mathbb{I}_M - AA^T) \tag{8.5}$$

for all $\jmath \in \mathbb{C}$. Thus, in the case $M \leq N$, if we call $(t_1, t_2, \ldots, t_{M+N})$ the eigenvalues of the symmetric matrix $\tilde{A}$, then the vector $(t_1^2, t_2^2, \ldots, t_{M+N}^2)$ contains twice each eigenvalue of AA^T and $N - M$ times the eigenvalue 0 (multiple eigenvalues appear in the previous vectors according to their multiplicities). Thus, the empirical spectral distributions of $AA^T, \tilde{A}$ are related through

$$\mu_{\tilde{A}} \circ T^{-1} = \frac{2M}{M+N}\mu_{AA^T} + \frac{N-M}{M+N}\delta_0 \tag{8.6}$$

with $T : \mathbb{R} \to [0, \infty)$ being the map $x \mapsto x^2$.

STEP FUNCTION PROFILE: If $(A_N)_{N \in \mathbb{N}^+}$ is as in Definition 8.1 with $M(N) := \lceil cN \rceil$ for some $c \in (0, 1]$, then the sequence $(\tilde{A}_N)_{N \in \mathbb{N}^+}$ is of the form given in Definition 2.11 with the following modification. We require that there is some $\gamma : \mathbb{N}^+ \to \mathbb{N}^+$ with $\lim_{N \to \infty} \gamma(N) = \infty$ so that the N-th matrix is of dimension $\gamma(N) \times \gamma(N)$ and, for each $N \in \mathbb{N}^+$, the family $(I_p^{(N)})_{p \in [m]}$ is a partition of $[\gamma(N)]$. The numbers a_p satisfy $\lim_{N \to \infty} R_p^{(N)} / \gamma(N) = a_p$. With this modification, Theorem 2.14 holds if the denominator in (2.27) is replaced by $\sqrt{\gamma(N)}$.

The sequence $(\tilde{A}_N)_{N \in \mathbb{N}^+}$ fits into this framework. We have $\gamma(N) = \lceil cN \rceil + N$, the role of m (of Definition 2.11) is played by $m + n$ (m, n from Definition 8.1), the $(m + n)^2$ constants are

$$\tilde{\sigma}_{p,q} := \begin{cases} 0 & \text{if } p \in [m], q \in [m], \\ \sigma_{p,q-m} & \text{if } p \in [m], q \in [m+n] \backslash [m], \\ \sigma_{q,p-m} & \text{if } p \in [m+n] \backslash [m], q \in [m], \\ 0 & \text{if } p \in [m+n] \backslash [m], q \in [m+n] \backslash [m] \end{cases} \tag{8.7}$$

for each N, and the partition of $[\gamma(N)]$ into $m + n$ intervals consists of the intervals (we write M instead of $\lceil cN \rceil$)

$$\left\{ [Ma_{p-1}, Ma_p) \cap \mathbb{N}^+ : p \in [m] \right\}, \tag{8.8}$$

$$\left\{ [M + N\beta_{q-1}, M + N\beta_q) \cap \mathbb{N}^+ : q \in [n] \right\}. \tag{8.9}$$

Dividing the right endpoints of the intervals by $\gamma(N)$ and taking $N \to \infty$, we get the $m + n$ numbers

$$\frac{c}{1+c} a_1 < \frac{c}{1+c} a_2 < \cdots < \frac{c}{1+c} a_m < \frac{c}{1+c} + \frac{1}{1+c} \beta_1 < \cdots < \frac{c}{1+c} + \frac{1}{1+c} \beta_n. \tag{8.10}$$

If we feed these data to the recipe of Definition 2.11, relation (2.21) will give as A_N the matrix $\tilde{A}_N$ where A_N is given by (8.1). The discussion preceding Theorem 2.14 applied to the sequence $(\tilde{A}_N)_{N \geq 1}$ gives that the ESD of $\tilde{A}_N / \sqrt{\gamma(N)}$ converges almost surely weakly to a symmetric probability measure $\tilde{\mu}^\sigma$ with compact support. Call $\tilde{\mu}_\infty^\sigma$ the largest element of the support. Relation (8.6) implies that the ESD of $A_N A_N^T / N$ converges to a measure with compact support contained in $[0, \infty)$ and the largest element of this support is $\mu_\infty = (1+c)(\tilde{\mu}_\infty^\sigma)^2$. Then Theorem 2.14 has the following corollary.

Corollary 8.3. *Assume that $(A_N)_{N \geq 1}$ is as in Definition 8.1 with $M := \lceil cN \rceil$ for some $c \in (0, 1]$ and $\mathbf{E}(|X_0|^{4+\delta}) < \infty$ for some $\delta > 0$. Then it is true that*

$$\lim_{N \to \infty} \frac{|A_N A_N^T|_{\mathrm{op}}}{N} = \mu_\infty \quad a.s. \tag{8.11}$$

CONTINUOUS FUNCTION PROFILE: If $(A_N)_{N \in \mathbb{N}^+}$ is as in Definition 8.2 with $M(N) := \lceil cN \rceil$ for some $c \in (0, 1]$, then we apply the discussion preceding Theorem 2.14 to the sequence $(\tilde{A}_N)_{N \in \mathbb{N}^+}$. The graphon, W_N, corresponding to $\tilde{A}_N$ converges pointwise in $[0, 1]^2$ to the graphon $\tilde{\sigma}$ with

$$\tilde{\sigma}(x, y) := \begin{cases} 0 & \text{if } (x, y) \in [0, c/(1+c)]^2 \cup (c/(1+c), 1]^2, \\ \sigma(x(1+c)/c, (1+c)y - c) & \text{if } (x, y) \in [0, c/(1+c)] \times (c/(1+c), 1], \\ \sigma(y(1+c)/c, (1+c)x - c) & \text{if } (x, y) \in (c/(1+c), 1] \times [0, x/(1+c)]. \end{cases} \tag{8.12}$$

We used (8.2) and the continuity of σ. Since $(\tilde{A}_N)_{N \in \mathbb{N}^+}$ also satisfies Assumption 2.1, we get that the ESD of $\tilde{A}_N / \sqrt{\gamma(N)}$ converges almost surely weakly to a symmetric probability measure $\tilde{\mu}^\sigma$ with compact support. Call $\tilde{\mu}_\infty^\sigma$ the largest element of the support. As above, the ESD of $A_N A_N^T / N$ converges to a measure with compact support contained in $[0, \infty)$, and the largest element of this support is $\mu_\infty = (1+c)(\tilde{\mu}_\infty^\sigma)^2$.

Corollary 8.4. *Assume that $(A_N)_{N \geq 1}$ is as in Definition 8.2 with $M := \lceil cN \rceil$ for some $c \in (0, 1]$ and $\mathbf{E}(|X_0|^{4+\delta}) < \infty$ for some $\delta > 0$. Then it is true that*

$$\lim_{N \to \infty} \frac{|A_N A_N^T|_{\mathrm{op}}}{N} = \mu_\infty \quad a.s. \tag{8.13}$$

Proof. The proof does not follow directly from Corollary 2.15 because the sequence $(\tilde{A}_N)_{N\in\mathbb{N}^+}$ does not necessarily have a continuous variance profile in the sense of Definition 2.12. Instead, we mimic the proof of that corollary. We define $\sigma^{(n)}$ as in (7.10), and the $M \times N$ matrix $\Sigma_N^{(n)}$ as $(\Sigma_N^{(n)})_{i,j} := \sigma^{(n)}(i/M, j/N)$ for all $i \in [M], j \in [N]$. Then we apply an obvious modification of Proposition 7.1 (the N-th matrix is of dimension $\gamma(N) \times \gamma(N)$, with $\gamma(N) = \lceil cN \rceil + N$) with the role of Σ_N and $\Sigma_N^{(n)}$ played by $\tilde{\Sigma}_{M,N}, \tilde{\Sigma}_N^{(n)}$ (the symmetrizations of $\Sigma_{M,N}$ and $\Sigma_N^{(n)}$, defined in (8.4). The proof continues by adopting the proof of Corollary 2.15 to this setting. Note that $|\tilde{\Sigma}_{M,N} - \tilde{\Sigma}_N^{(n)}|_{\max} = |\Sigma_{M,N} - \Sigma_N^{(n)}|_{\max}$, which has $\lim_{n\to\infty} \limsup_{N\to\infty} |\Sigma_{M,N} - \Sigma_{M,N}^{(n)}|_{\max} = 0$. $\quad\square$

Remark 8.5. In [21] 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 8.3 we prove the convergence of the largest eigenvalue of these models as well. The authors in [21] 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.

8.2 Further applications of Theorem 2.10

In the Random Matrix Theory literature what are commonly described as Random matrices with variance-profile given by a step function are more or less what we describe in Theorem 2.14. In this subsection we give some examples which are covered by the generalized version of this variance-profile matrices (Definition 2.9) but not from the "standard" step functions.

Let $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i, j \in [N]\}$ identically distributed random variables, $a_{i,j}^{(N)} = a_{j,i}^{(N)}$ for all $N \in \mathbb{N}^+, i, j \in [N]$, $\{a_{i,j}^{(N)} : 1 \le j \le i \le N\}$ independent for each N, and $a_{1,1}^{(1)}$ has mean 0 and variance 1. Fix $p \in (0, 1]$ and let A_N be the matrix with entries

$$\{A_N\}_{i,j} = a_{i,j}^{(N)} \mathbf{1}_{|i-j| \le pN}, \quad i, j \in [N], \tag{8.14}$$

The sequence $(A_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 2.1 (easy to check) and also Assumption 2.2. To see the last point, we follow Remark 2.13. The graphon corresponding to A_N is $W_N(x, y) = \mathbf{1}_{|\lceil Nx \rceil - \lceil Ny \rceil| \le pN}$ which converges to the graphon $W(x, y) = \mathbf{1}_{|x-y| \le p}$ at least on the set $\{(x, y) \in [0, 1]^2 : |x - y| \ne pn\}$, which has measure 1. Thus, with probability one, the ESD of $A_N / \sqrt{N}$ converges weakly to a symmetric measure μ with compact support. Call μ_∞ the supremum of the support of μ.

Corollary 8.6 (Non-Periodic Band Matrices with Bandwidth proportional to the dimension)**.** *Assume that for the matrix defined in* (8.14) *we have that* $a_{1,1}^{(N)}$ *has* 0 *mean, unit variance and finite* $4 + \delta$ *moment for some* $\delta > 0$. *Then*

$$\lim_{N\to\infty} \frac{|A_N|_{\mathrm{op}}}{\sqrt{N}} = \mu_\infty \quad a.s.$$

Proof. The sequence $(A_N)_{N\in\mathbb{N}^+}$ satisfies Assumption 2.2, as we saw above, and also Assumption 2.3 because $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, i, j \in [N]\}$ are identically distributed and $a_{1,1}$ has mean zero, variance one, and finite fourth moment. The corollary then is a straightforward application of Theorem 2.10, where $d_N = 3$, the partition of $[N]^2$ required by Definition 2.9 consists of the sets

$$\mathcal{B}_1^{(N)} := \{(i, j) \in [N]^2 : |(i/N) - (j/N)| \le p\}, \tag{8.15}$$
$$\mathcal{B}_2^{(N)} := \{(i, j) \in [N]^2 : (i/N) > (j/N) + p\}, \tag{8.16}$$
$$\mathcal{B}_3^{(N)} := \{(i, j) \in [N]^2 : (j/N) > (i/N) + p\}, \tag{8.17}$$

and $s_1^{(N)} = 1, s_2^{(N)} = s_3^{(N)} = 0$. Condition (b) of that definition is satisfied by $f := m$ for each $m \in [3]$. $\quad\square$

Remark 8.7. The random band matrix models have been extensively studied after the novel work in [8] 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 has been extensively studied, see for example [26] or the survey [11]. Moreover when the bandwidth of such matrices

is non-periodic but the bandwidth (the maximum number of non identically zero entries per row) is $o(N)$ but also tends to infinity has also been examined in [6]. To the best of our knowledge, the convergence of the largest eigenvalue of non-periodic Band Matrices with bandwidth proportional to the dimension has not been established.

Our next result concerns the singular values of triangular random matrices. It is well known under various assumptions for the entries, but we record it here as another application of our main theorem.

Let $\{a_{i,j}^{(N)} : N \in \mathbb{N}^+, 1 \leq i \leq j \in [N]\}$ identically distributed random variables, $\{a_{i,j}^{(N)} : 1 \leq j \leq i \leq N\}$ independent for each N, and $a_{1,1}^{(1)}$ has mean 0 and variance 1. Let A_N be the matrix with entries

$$\{A_N\}_{i,j} = a_{i,j}^{(N)} 1_{i \leq j}, \quad i,j \in [N], \tag{8.18}$$

Corollary 8.8 (Triangular matrices)**.** *Assume that for the matrix defined in* (8.18) *we have that* $a_{1,1}^{(N)}$ *has* 0 *mean, unit variance and finite* $4 + \delta$ *moment for some* $\delta > 0$*. Then*

$$\lim_{N \to \infty} \frac{|A_N A_N^T|_{\mathrm{op}}}{N} = e \quad a.s.$$

Proof. As in the case of Gram matrices, we denote by $\tilde{A}_N$ the symmetrization of A_N, defined in (8.4). We have $|A_N A_N^T|_{\mathrm{op}} = |\tilde{A}_N|_{\mathrm{op}}^2$. We will apply Theorem 2.10 to the sequence $(\tilde{A}_N)_{N \in \mathbb{N}^+}$. The partition of $[2N]^2$ required by Definition 2.9 consists of the following three sets (i.e., $d_N = 3$)

$$\mathcal{B}_1^{(N)} := \{(i,j) \in [2N]^2 : |i - j| \leq N - 1\}, \tag{8.19}$$

$$\mathcal{B}_2^{(N)} := \{(i,j) \in [2N]^2 : i \geq N + j\}, \tag{8.20}$$

$$\mathcal{B}_3^{(N)} := \{(i,j) \in [2N]^2 : j \geq N + i\}, \tag{8.21}$$

and the corresponding values of the variance are $s_1^{(N)} = 0, s_2^{(N)} = s_3^{(N)} = 1$. Assumption 2.2 follows as an application of Remark 2.13, in the same way as in the previous corollary. The measure μ of that assumption satisfies $\mu \circ T^{-1} = \nu$, where ν is the limit of the E.S.D of $N^{-1} A_N A^T$ [recall (8.6)]. It was shown in [12] that ν has support $[0, e]$. It follows that μ has support $[-\sqrt{e}, \sqrt{e}]$ [See Remark 2.2 of [10] for a more detailed discussion of this phenomenon].

Assumption 2.3 is satisfied because the elements of $\tilde{A}_N$ with indices in $\mathcal{B}_2^{(N)} \cup \mathcal{B}_3^{(N)}$ are identically distributed with zero mean, unit variance and finite fourth moment (the remaining elements of the matrix are identically zero random variables). Finally, condition (2.17) is satisfied as equality.

Thus, the corollary follows from Theorem 2.10. $\qquad\square$

9 Two technical lemmas

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 [2].

Lemma 9.1. $N_{T,a_1,a_2,\ldots,a_s} \leq (4k^4)^{4(s+1-t)+2(k-s)}$ *if* $t \geq 2$ *and* $N_{T,a_1,a_2,\ldots,a_s} = 1$ *if* $t = 1$.

Proof. When $t = 1$, since the cycle is bad, we have $s = 1$ and $a_1 = 2k$, and there is only one cycle with these s, t and vertex set $\{1\}$.

For the case $t \geq 2$, take a cycle $\mathbf{i} := (i_1, i_2, \ldots, i_{2k})$ as in (3.9) 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 $\{(r, (i_r, i_{r+1})) : r = 1, 2, \ldots, 2k\}$. Edges of the cycle we call the edges of $G(\mathbf{i})$, and the multiplicity of each edge is computed from $\mathbf{i}$. The graph $G(\mathbf{i})$ does not have multiple edges.

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 $G(\mathbf{i})$ 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 $G(\mathbf{i})$ 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

$$r \leq 2m, \tag{9.1}$$

$$t = s + 1 - m \leq k, \tag{9.2}$$

$$|T_4| = 2m + 2(k - s). \tag{9.3}$$

The first relation is Lemma 5.6 in [2]. 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 (9.2) too, we have $|T_4| = 2k - 2(t - 1) = 2k - 2(s - m)$.

Now back to the task of bounding $N_{T, a_1, \ldots, a_s}$. We fix a cycle as in the beginning of the proof and 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 $Q_1 \subset \{1, 2, \ldots, 2k\} \times (\{1, 2, \ldots t\}^2 \cup \{1, 2, \ldots t\}^3)$ 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 $Q_2 \subset \{1, 2, \ldots, 2k\} \times \{1, 2, \ldots, t\}^2$ 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 and knowing that $T(\mathbf{i}) = T$, 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_4 (if $i_2 = i_1$) or T_1. The set Q_1 will tell us if we are in the first case and will give us all we want. If we are in the second case, the initial vertex is 1 and the final 2. 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 neighbors in $G(\mathbf{i})$ that we haven't encountered up to the leg ℓ_{i-1}, then ℓ_i is T_3 irregular, and by the defining property of T_3 irregular legs, we can determine the index of its final vertex. If the vertex v_i does have such neighbors, call z the one that appears earlier in the cycle.

- If $z \in U$, then in case it was included in U because of ℓ_{i-1} (this can be read off from Q_1. Note that z could not have been included because of an earlier leg because z has not appeared earlier than v_i), we have that
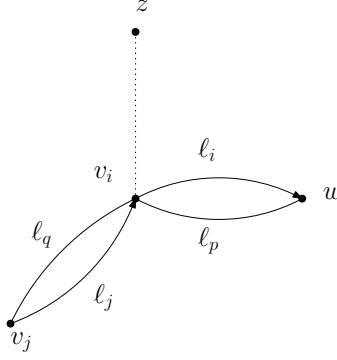
Figure 1: The case $z \notin U$. The legs $\ell_i, \ell_j (i < j)$ are T_3, while ℓ_p, ℓ_q are T_1.

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 neighbor 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 (recall that $p < i < j$), 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 (i.e., 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 neighbor 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 T, Q_1, Q_2 allows to determine **i**.

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

$$(2kt^2(t + 1))^{|T_4|}(2kt^2)^r \leq (4k^4)^{r+|T_4|}. \tag{9.4}$$

Then (9.1) and (9.3) give $r + |T_4| \leq 4m + 2(k - s)$, and finally using (9.2), we get the desired bound. □

The next lemma is used in the proof of Theorem 2.10.

Lemma 9.2. *Let $(A_N)_{N \in \mathbb{N}^+}$ be a sequence of matrices, A_N of dimension $N \times N$, that satisfies Assumption 2.1 and Assumption 2.2 with measure μ. Suppose that there are two sequences of matrices $(A_N^{(1)})_{N \in \mathbb{N}^+}$ and $(A_N^{(2)})_{N \in \mathbb{N}^+}$ such that*

(a) $A_N = A_N^{(1)} + A_N^{(2)}$,

(b) For all $N \in \mathbb{N}^+$ and $i, j \in [N]$, at least one of $(A_N^{(1)})_{i,j}, (A_N^{(2)})_{i,j}$ is identically zero random variable.

(c) $\mu_{A_N^{(1)}/\sqrt{N}} \Rightarrow \mu$ in probability as $N \to \infty$.

Then $(A_N^{(1)})_{N \in \mathbb{N}^+}$ also satisfies Assumptions 2.1, 2.2 with the measure μ.

Proof. We only need to check the validity of Assumption 2.2 as the validity of Assumption 2.1 is immediate. Because $A_N^{(1)}$ satisfies Assumption 2.1, there is a decreasing sequence $(\eta_N)_{N \in \mathbb{N}^+}$ of positive reals converging to 0 so that

$$\lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j \in [N]} \mathbf{E}\left[(\{A_N^{(1)}\}_{i,j})^2 \mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\right] = 0. \tag{9.5}$$

Set $A_N^{(1),\leq}$ to be the matrix whose (i,j) entry is

$$\{A_N^{(1)}\}_{i,j}\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| \leq \eta_N N^{\frac{1}{2}}\right) - \mathbf{E}\left[\{A_N^{(1)}\}_{i,j}\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| \leq \eta_N N^{\frac{1}{2}}\right)\right] \tag{9.6}$$

and $\mu_{N,i,j} := \mathbf{E}\left[\{A_N^{(1)}\}_{i,j}\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| \leq \eta_N N^{\frac{1}{2}}\right)\right]$.

CLAIM:

$$\mu_{N^{-1/2}A_N^{(1),\leq}} \Rightarrow \mu \quad \text{in probability as } N \to \infty \tag{9.7}$$

The Levy distance between $\mu_{A_N^{(1)}/\sqrt{N}}$ and $\mu_{A_N^{(1),\leq}/\sqrt{N}}$ is bounded as follows.

$$L^3(\mu_{A_N^{(1)}/\sqrt{N}}, \mu_{A_N^{(1),\leq}/\sqrt{N}}) \leq \frac{1}{N}\text{tr}\left\{\left(\frac{1}{\sqrt{N}}A_N^{(1)} - \frac{1}{\sqrt{N}}A_N^{(1),\leq}\right)^2\right\} \tag{9.8}$$

$$= \frac{1}{N^2}\sum_{i,j\in[N]}\{\mu_{N,i,j}^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| \leq \eta_N N^{\frac{1}{2}}\right) + (\{A_N^{(1)}\}_{i,j} + \mu_{N,i,j})^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\} \tag{9.9}$$

$$\leq \frac{3}{N^2}\sum_{i,j\in[N]}\mu_{N,i,j}^2 + \frac{2}{N^2}\sum_{i,j\in[N]}(\{A_N^{(1)}\}_{i,j})^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right). \tag{9.10}$$

Since the entries of $A_N^{(1)}$ have mean 0, we have

$$\mu_{N,i,j}^2 = \left(\mathbf{E}\left[\{A_N^{(1)}\}_{i,j}\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\right]\right)^2 \leq \mathbf{E}\left[(\{A_N^{(1)}\}_{i,j})^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\right].$$

Thus, the expectation of the expression in (9.10) is at most

$$\frac{5}{N^2}\sum_{i,j\in[N]}\mathbf{E}\left\{(\{A_N^{(1)}\}_{i,j})^2\mathbf{1}\left(|\{A_N^{(1)}\}_{i,j}| > \eta_N N^{\frac{1}{2}}\right)\right\},$$

which tends to zero as $N \to \infty$ due to (9.5). This, combined with assumption (c), proves the claim.

Fix $k \in \mathbb{N}^+$ and set $M_N(k), M_N^{(1),\leq}(k)$ the asymptotic contributing terms (see (2.4)) of A_N and $A_N^{(1),\leq}$ respectively. Notice that

$$M_N^{(1),\leq}(k) \leq M_N^{(1)}(k) \leq M_N(k). \tag{9.11}$$

The rightmost inequality is true because the variance of $\{A_N^{(1)}\}_{i,j}$ is either zero or $s_{i,j}^{(N)}$ due to assumption (b) of the lemma. The leftmost inequality is true because if W is a real valued random variable with mean 0 and finite variance and $\tilde{W}$ is a variable with $|\tilde{W}| \leq |W|$, then $\text{Var}(\tilde{W}) \leq \text{Var}(W)$.

Lemma 3.6 of [30] implies that

$$\frac{M_N^{(1),\leq}(k)}{N^{k+1}} = \frac{1}{N^{k+1}}\mathbf{E}\,\text{tr}\{(A_N^{(1),\leq})^{2k}\} + o(1) \tag{9.12}$$

as $N \to \infty$. We will prove that the right hand side converges to $\int x^{2k}d\mu$ as $N \to \infty$. It will be convenient to let $B_N := A_N^{(1),\leq}/\sqrt{N}$ and $\{\hat{\lambda}_i(B_N) : i \in [N]\}$ its eigenvalues.

Pick some $C > \mu_\infty$ and consider the function $g_C(x) = (|x| \wedge C)^{2k}$, which is bounded and continuous. Then,

$$\frac{1}{N^{k+1}}\mathbf{E}\,\text{tr}\{(A_N^{(1),\leq})^{2k}\} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\{(\hat{\lambda}_i\{B_N\})^{2k}\} \tag{9.13}$$

and the right hand side can be estimated as follows.

$$\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\{(\hat{\lambda}_i\{B_N\})^{2k}\} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\,g_C(\hat{\lambda}_i\{B_N\})\right| \leq \frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\{\hat{\lambda}_i^{2k}(B_N)\mathbf{1}_{|\hat{\lambda}_i(B_N)|\geq C}\}$$

$$\leq \frac{1}{N}\sum_{i=1}^{N}\sqrt{\mathbf{E}\hat{\lambda}_i^{4k}(B_N)}\sqrt{\mathbb{P}(|\hat{\lambda}_i(B_N)| \geq C)} \leq \sqrt{\sum_{i=1}^{N}\frac{\mathbf{E}\hat{\lambda}_i^{4k}(B_N)}{N}}\sqrt{\frac{\mathbf{E}\sum_{i=1}^{N}\mathbf{1}_{|\hat{\lambda}_i(B_N)|\geq C}}{N}} \overset{N\to\infty}{\to} 0. \tag{9.14}$$

To justify the convergence to zero, note that the quantity in the second square root converges to zero by our choice of $C > \mu_\infty$ and the in probability weak convergence of the E.S.D. of B_N to μ. The quantity in the first

square root is bounded in N because, due to (9.12), its difference from $M_N^{(1),\leq}(2k)/N^{2k+1}$ is bounded and the latter is less than $M_N(2k)/N^{2k+1}$ which is bounded in N since it converges to $\int x^{4k}\,d\mu$.

The in probability weak convergence (9.7) implies that

$$\frac{1}{N}\sum_{i=1}^{N} g_C(\lambda_i\{B_N\}) \to \int x^{2k}d\mu \text{ in probability,} \tag{9.15}$$

and the boundedness of g_C allows to conclude that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}g_C(\lambda_i\{B_N\}) = \int x^{2k}d\mu. \tag{9.16}$$

Thus, relations (9.12), (9.13), (9.14),(9.16) show that

$$\lim_{N\to\infty}\frac{M_N^{(1),\leq}(k)}{N^{k+1}} = \int x^{2k}d\mu. \tag{9.17}$$

And this combined with (9.11) concludes the proof. □

References

[1] J. Alt, L. Erdős, T. Krüger, and Y. Nemish. Location of the spectrum of kronecker random matrices. 2019.

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

[3] 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.

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

[5] 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.

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

[7] P. Billingsley. *Convergence of probability measures*. John Wiley and Sons, second edition, 1999.

[8] 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.

[9] 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.

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

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

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

[13] 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.

[14] 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.

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

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

[17] R. Ducatez, J. Husson, and A. Guionnet. Large deviation principle for the largest eigenvalue of random matrices with a variance profile. *arXiv preprint arXiv:2403.05413*, 2024.

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

[19] 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.

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

[21] 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.

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

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

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

[25] 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.

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

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

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

[29] Y.-Q. Yin, Z.-D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability theory and related fields*, 78:509–521, 1988.

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