On the spectral radius of a random matrix: an upper bound without fourth moment

Charles Bordenave, Pietro Caputo, Djalil Chafaï, Konstantin Tikhomirov

ON THE SPECTRAL RADIUS OF A RANDOM MATRIX

CH. BORDENAVE, P. CAPUTO, D. CHAFAÏ, AND K. TIKHOMIROV

Abstract. Consider a square matrix with independent and identically distributed entries of zero mean and unit variance. It is well known that if the entries have a finite fourth moment, then, in high dimension, with high probability, the spectral radius is close to the square root of the dimension. We conjecture that this holds true under the sole assumption of zero mean and unit variance, in other words that there are no outliers in the circular law. In this work we establish the conjecture in the case of symmetrically distributed entries with a finite moment of order larger than two. The proof uses the method of moments combined with a novel truncation technique for cycle weights that might be of independent interest.

1. Introduction

Let X_N denote the random $N \times N$ matrix $(X_{i,j})_{i,j=1,\dots,N}$, where $X_{i,j}$ are independent copies of a given complex valued random variable $\mathbf{x}$ with mean zero and unit variance:

$$\mathbb{E}[\mathbf{x}] = 0 \quad \text{and} \quad \mathbb{E}[|\mathbf{x}|^2] = 1. \tag{1.1}$$

Let $\rho(X_N)$ denote the spectral radius of X_N:

$$\rho(X_N) := \max\left\{|\lambda| : \lambda \text{ eigenvalue of } X_N\right\}. \tag{1.2}$$

The well known circular law states that, in probability, the empirical distribution of the eigenvalues of $N^{-1/2}X_N$ weakly converges to the uniform law on the unit disc of the complex plane [TV, BC2]. In particular, it follows that with high probability

$$\rho(X_N) \geqslant (1-\delta)\sqrt{N}\,, \tag{1.3}$$

for any $\delta > 0$ and large enough N. Here and below we say that a sequence of events holds with high probability if their probabilities converge to one. The corresponding upper bound on $\rho(X_N)$ has been established by Bai and Yin [BY] under a finite fourth moment assumption: if $\mathbb{E}[|\mathbf{x}|^4] < \infty$, then with high probability $\rho(X_N) \leqslant (1+\delta)\sqrt{N}$, for any $\delta > 0$ and large enough N; see also Geman and Hwang [GH] and Geman [G] for an independent proof under stronger assumptions. Together with (1.3), this says that if $\mathbb{E}[|\mathbf{x}|^4] < \infty$ then, in probability, $\rho(X_N)/\sqrt{N} \to 1$, as $N \to \infty$. We refer to [G, BY] and references therein for related estimates and more background and applications concerning the spectral radius of a random matrix. Surprisingly, there seems to be little or no discussion at all in the literature – even in the recent works [T] and [BC1] – about the necessity of the fourth moment assumption for the behavior $\rho(X_N) \sim \sqrt{N}$. We propose the following conjecture, which is illustrated by Figure 1.

Conjecture 1.1. *The convergence in probability*

$$\lim_{N\to\infty} \frac{\rho(X_N)}{\sqrt{N}} = 1, \tag{1.4}$$

Date: Summer 2016, compiled July 18, 2016.

1991 *Mathematics Subject Classification.* 05C20; 15B52; 47A10; 05C80.

Key words and phrases. Combinatorics; Digraph; Spectral Radius; Random matrix; Heavy Tail.

Partial support: A*MIDEX project ANR-11-IDEX-0001-02 funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR).

holds under the sole assumptions (1.1).

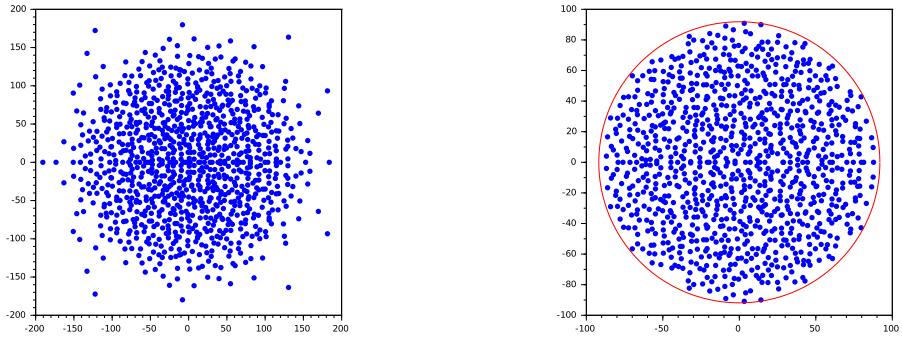

FIGURE 1. The dots are the eigenvalues of a single realization of X_N where $N = 1000$ and $\mathbf{x}$ is real with distribution given by

$$\mathbb{P}(\mathbf{x} > t) = \mathbb{P}(\mathbf{x} < -t) = \frac{1}{2t^\alpha}, \quad t \geqslant 1,$$

with $\alpha = 1.8$ (left) and $\alpha = 2.2$ (right). The circle has radius $\sqrt{(\mathbb{E}|\mathbf{x}|^2)N}$.

Another way to put this is to say that there are no outliers in the circular law. This phenomenon reveals a striking contrast between eigenvalues and singular values of X_N, the latter exhibiting Poisson distributed outliers in absence of a fourth moment, see for instance [S, ABP]. A tentative heuristic explanation of this phenomenon may proceed as follows. Suppose $\mathbf{x}$ has a heavy tail of index α, that is $\mathbb{P}(|\mathbf{x}| > t) \sim t^{-\alpha}$, as $t \to \infty$. If $\alpha \in (2, 4)$, then with high probability in the matrix $X = X_N$ there are elements $X_{i,j}$ with $|X_{i,j}| > N^\beta$, for any $1/2 < \beta < 2/\alpha$. Any such element is sufficient to produce a singular value diverging as fast as N^β. On the other hand, to create a large eigenvalue, a single large entry is not sufficient. Roughly speaking one rather needs at least one sequence of indices $i_1, i_2, \ldots, i_{k+1}$ with $i_1 = i_{k+1}$ with a large product $\prod_j |X_{i_j, i_{j+1}}|$, i.e. one cycle with a large weight if we view the matrix as an adjacency matrix of an oriented and weighted graph. It is not difficult to see that the sparse matrix consisting of all entries $X_{i,j}$ with $|X_{i,j}| > N^\beta$ is acyclic with high probability, as long as $\alpha\beta > 1$.

Somewhat similar phenomena should be expected for heavy tails with index $\alpha \in (0, 2)$. As shown in [BCC], in that case the circular law must be replaced by a new limiting law μ_α in the complex plane. More precisely, the empirical distribution of the eigenvalues of $X/N^{1/\alpha}$ tends weakly as $N \to \infty$ to a rotationally invariant light tailed law μ_α, while the empirical distribution of the singular values of $X/N^{1/\alpha}$ tends weakly as $N \to \infty$ to a heavy tailed law ν_α. By the above reasoning, no significant outliers should appear in the spectrum. The precise analogue of (1.4) in this case is however less obvious since the support of μ_α is unbounded. From the tail of μ_α, one might expect that the spectral radius is of order $N^{1/\alpha}(\log N)^{1/\alpha + o(1)}$ while typical eigenvalues are of order $N^{1/\alpha}$.

In this paper we prove that the conjectured behavior (1.4) holds if $\mathbf{x}$ is symmetric and has a finite moment of order $2 + \varepsilon$ for an arbitrary $\varepsilon > 0$. We say that $\mathbf{x}$ is symmetric if the law of $\mathbf{x}$ coincides with the law of $-\mathbf{x}$.

Theorem 1.2. *Suppose that* $\mathbf{x}$ *is symmetric and that* $\mathbb{E}\left[|\mathbf{x}|^2\right] = 1$. *Suppose further that* $\mathbb{E}\left[|\mathbf{x}|^{2+\varepsilon}\right] < \infty$ *for some* $\varepsilon > 0$. *Then, in probability,*

$$\lim_{N \to \infty} \frac{\rho(X_N)}{\sqrt{N}} = 1. \tag{1.5}$$

In view of (1.3), to prove the theorem one only needs to establish the upper bound $\rho(X_N) \leqslant (1+\delta)\sqrt{N}$ with high probability, for every $\delta > 0$. We shall prove the following stronger non-asymptotic estimate, covering variables $\mathbf{x}$ whose law may depend on N.

Theorem 1.3. *For any $\varepsilon, \delta > 0$ and $B > 0$, there exists a constant $C = C(\varepsilon, \delta, B) > 0$ such that for any $N \in \mathbb{N}$, for any symmetric complex random variable $\mathbf{x}$ with $\mathbb{E}\left[|\mathbf{x}|^2\right] \leqslant 1$ and $\mathbb{E}\left[|\mathbf{x}|^{2+\varepsilon}\right] \leqslant B$, we have*

$$\mathbb{P}(\rho(X_N) \geqslant (1+\delta)\sqrt{N}) \leqslant \frac{C}{(\log N)^2}. \tag{1.6}$$

The rest of this note is concerned with the proof of Theorem 1.3. We finish this introduction with a brief overview of the main arguments involved.

1.1. **Overview of the proof.** The proof of Theorem 1.3 combines the classical method of moments with a novel cycle weight truncation technique. For lightness of notation, we write X instead of X_N. The starting point is a standard general bound on $\rho(X)$ in terms of the trace of a product of powers of X and X^*. Let $\|X\|$ denote the operator norm of X, that is the maximal eigenvalue of $\sqrt{X^*X}$, which is also the largest singular value of X. Recall the Weyl inequality $\rho(X) \leqslant \|X\|$. For any integer $m \geqslant 1$ one has

$$\rho(X) = \rho(X^m)^{1/m} \leqslant \|X^m\|^{1/m} \quad \text{and} \quad \|X^m\|^2 \leqslant \mathrm{Tr}((X^*)^m X^m).$$

It follows that for any integer $k \geqslant 2$, setting $m = k-1$,

$$\rho(X_N)^{2k-2} \leqslant \mathrm{Tr}((X^*)^{k-1}X^{k-1}) = \sum_{i,j} [X^{k-1}]_{i,j}[(X^*)^{k-1}]_{j,i}. \tag{1.7}$$

Expanding the summands in (1.7) one obtains

$$\rho(X_N)^{2k-2} \leqslant \sum_{i,j} \sum_{P_1, P_2 : i \mapsto j} w(P_1)\bar{w}(P_2), \tag{1.8}$$

where the internal sum ranges over all paths P_1 and P_2 of length $k-1$ from i to j, the weight $w(P)$ of a path $(i_1, \ldots, i_k)$ is defined by

$$w(P) := \prod_{\ell=1}^{k-1} X_{i_\ell, i_{\ell+1}}\,, \tag{1.9}$$

and $\bar{w}(P)$ denotes the complex conjugate of $w(P)$. So far we have not used any specific form of the matrix entries.

As a warm up, it may be instructive to analyze the following simple special case. Assume that $X_{i,j}$ has the distribution

$$\mathbf{x} = \begin{cases} \pm q^{-\frac{1-\varepsilon}{2}} & \text{with probability } \frac{q}{2}, \\ 0 & \text{with probability } 1-q, \end{cases} \tag{1.10}$$

where $q = q_N \in (0,1]$ is a parameter that may depend on N, while $\varepsilon \in (0,1)$ is a fixed small constant. If $q_N \equiv 1$, then we have a uniformly random ± 1 matrix, while if $q_N \to 0$, $N \to \infty$ one has a matrix that may serve as a toy model for the sparse matrices from the heuristic discussion given above. Notice that the assumptions of Theorem 1.3 are satisfied with the same parameter ε and with $B = 1$, since

$$\mathbb{E}[|\mathbf{x}|^2] = q^{\varepsilon} \quad \text{and} \quad \mathbb{E}[|\mathbf{x}|^{2+\varepsilon}] \leqslant q^{\varepsilon/2}.$$

We can now take expectation in (1.8). Using the symmetry of $\mathbf{x}$ we may restrict the sum over paths P_1, P_2 satisfying the constraint that in the union $P_1 \cup P_2$ each directed edge $(i_\ell, i_{\ell+1})$ appears an even number of times. We say that $P_1 \cup P_2$ is even. In this case $\mathbb{E}[w(P_1)\bar{w}(P_2)] = q^{-(1-\varepsilon)(k-1)}q^n$, where n is the number of edges in $P_1 \cup P_2$ without

counting multiplicities. Let P denote the closed path obtained as follows: start at i, follow P_1, then add the edge (j,i), then follow P_2, then end with the edge (j,i) again. Thus, P is an even closed path of length $2k$. Notice that

$$\mathbb{E}[w(P_1)\bar{w}(P_2)] \leqslant q^{-\varepsilon}\mathbb{E}[w(P)].$$

Since the map $(P_1,P_2)\mapsto P$ is injective we have obtained

$$\mathbb{E}[\rho(X_N)^{2k-2}] \leqslant q^{-\varepsilon}\sum_P \mathbb{E}[w(P)], \tag{1.11}$$

where the sum ranges over all even closed paths of length $2k$. Observe that

$$\mathbb{E}[w(P)] \leqslant q^{-(1-\varepsilon)k}q^{\ell},$$

where ℓ is the number of distinct vertices in P. Therefore, letting $\mathcal{N}(k,\ell)$ denote the number of even closed paths of length $2k$ with ℓ vertices, (1.11) is bounded above by

$$\sum_{\ell=1}^{k}\mathcal{N}(k,\ell)q^{-\varepsilon}q^{-(1-\varepsilon)k}q^{\ell}. \tag{1.12}$$

Combinatorial estimates to be derived below, see Lemma 2.2 and Lemma 2.3, imply that $\mathcal{N}(k,\ell)\leqslant k^2(4k)^{6(k-\ell)}N^{\ell}$. Putting all together we have found

$$\mathbb{E}[\rho(X_N)^{2k-2}] \leqslant k^2N^k\sum_{\ell=1}^{k}a(k,N,q)^{k-\ell} \tag{1.13}$$

where $a(k,N,q)=(4k)^6(Nq^{(1-\varepsilon)})^{-1}$. We choose $k\sim(\log N)^2$. Suppose that $q\geqslant N^{-1-\varepsilon}$. Then $Nq^{(1-\varepsilon)}\geqslant N^{\varepsilon^2}$ and therefore $a(k,N,q)\leqslant 1$ if N is large enough. It follows that $\mathbb{E}[\rho(X_N)^{2k-2}]\leqslant k^3N^k$, and by Markov's inequality, for all fixed $\delta>0$:

$$\begin{aligned}\mathbb{P}\left(\rho(X_N)\geqslant(1+\delta)\sqrt{N}\right) &\leqslant (1+\delta)^{-2k+2}N^{-k+1}\mathbb{E}[\rho(X_N)^{2k-2}]\\ &\leqslant (1+\delta)^{-2k+2}k^3N.\end{aligned} \tag{1.14}$$

Since $k\sim(\log N)^2$ this vanishes faster than $N^{-\gamma}$ for any $\gamma>0$. On the other hand, if $q\leqslant N^{-1-\varepsilon}$, then a different, simpler argument can be used. Indeed, since an acyclic matrix is nilpotent, it follows that if $\rho(X_N)>0$ then there must exist a cycle with nonzero entries from the matrix X. The probability of a given such cycle is q^{ℓ} where ℓ is the number of vertices of the cycle. Estimating by N^{ℓ} the number of cycles with ℓ vertices one has

$$\mathbb{P}[\rho(X_N)>0]\leqslant\sum_{\ell=1}^{\infty}(qN)^{\ell}. \tag{1.15}$$

Thus, if $q\leqslant N^{-1-\varepsilon}$, then $\mathbb{P}[\rho(X_N)>0]\leqslant 2qN\leqslant 2N^{-\varepsilon}$. This concludes the proof of (1.6) in the special case of the model (1.10).

The given argument displays, albeit in a strongly simplified form, some of the main features of the proof of Theorem 1.3: the role of symmetry, the role of combinatorics, and the fact that cycles with too high weights have to be ruled out with a separate probabilistic estimate. The latter point requires a much more careful handling in the general case. Since it represents the main technical novelty of this work, let us briefly illustrate the main idea here. Consider the collection $\mathcal{C}_m$ of all possible oriented cycles with m edges of the form $C=(i_1,\dots,i_{m+1})$ with $i_j\in\{1,\dots,N\}$, and with no repeated vertex except for $i_1=i_{m+1}$. Let ν_m denote the uniform distribution over the set $\mathcal{C}_m$. Given the matrix X_N, we look at the weight $|w(C)|^{2t}$ corresponding to the cycle C repeated $2t$ times, where $w(C)$ is defined in (1.9). Since one can restrict to even closed

paths, and each such path can be decomposed into cycles that are repeated an even number of times, it is crucial to estimate the empirical averages

$$\nu_m[|w(C)|^{2t}] = \frac{1}{|\mathcal{C}_m|} \sum_{C \sim \mathcal{C}_m} |w(C)|^{2t},$$

where the sum runs over all cycles with m edges and $|\mathcal{C}_m|$ denotes the total number of them. Broadly speaking, we will define an event $\mathcal{E}_k$ by requiring that

$$\nu_m[|w(C)|^2] \leqslant k^2\,, \quad \text{and} \quad \nu_m[|w(C)|^{2+\varepsilon}] \leqslant k^2 B^m, \tag{1.16}$$

for all $m \leqslant k$, where as before $k \sim (\log N)^2$. The assumptions of Theorem 1.3 ensure that $\mathcal{E}_k$ has large probability by a first moment argument. Thus, in computing the expected values of $w(P)$ we may now condition on the event $\mathcal{E}_k$. Actually, on the event $\mathcal{E}_k$ we will be able to estimate deterministically the quantities $\nu_m\left[|w(C)|^{2t}\right]$. To see this, observe that if

$$w_{\max}(m) := \max_{C \sim \mathcal{C}_m} |w(C)|$$

denotes the maximum weight for a cycle with m edges, then

$$w_{\max}^2 \leqslant \Big(\sum_{C \sim \mathcal{C}_m} |w(C)|^{2+\varepsilon} \Big)^{\frac{1}{1+\varepsilon/2}}.$$

If ε is small enough, on the event $\mathcal{E}_k$, from (1.16) one has $w_{\max}^2 \leqslant (|\mathcal{C}_m| k^2 B^m)^{1-\varepsilon/4}$. Since $|\mathcal{C}_m| \leqslant N^m$, a simple iteration proves that for any $t \geqslant 1$:

$$\nu_m[|w(C)|^{2t}] \leqslant (k^2 N^m B^m)^{t(1-\varepsilon/4)} \leqslant N^{mt(1-\varepsilon/8)}, \tag{1.17}$$

for all N large enough. The bound (1.17) turns out to be sufficient to handle all paths P of the form of a cycle $C \sim \mathcal{C}_m$ repeated $2t$ times, for all $m \leqslant k$. To control more general even closed paths P one needs a more careful analysis involving the estimate of larger empirical averages corresponding to various distinct cycles at the same time. We refer to Section 3.3 below for the details. The combinatorial estimates are worked out in Section 2. Finally, in Section 4 we complete the proof of Theorem 1.3.

2. Counting paths and digraphs

We first introduce the basic graph theoretic terminology and then prove some combinatorial estimates.

2.1. **Multi digraphs and even digraphs.** For each natural N, $[N]$ denotes the set $\{1, 2, \dots, N\}$. A directed graph, or simply digraph, on $[N]$, is a pair $G = (V, E)$, where $V \subset [N]$ is the set of vertices and $E \subset [N] \times [N]$ is the set of directed edges. We also consider multisets E, where a directed edge $e \in E$ appears with its own multiplicity $n_e \in \mathbb{N}$. In this case we say that $G = (V, E)$ is a *multi digraph*. Given a vertex v of a multi digraph, the out-degree $\deg_+(v)$ is the number of edges of the form $(v, j) \in E$, counting multiplicities. Similarly, the in-degree $\deg_-(v)$ is the number of edges of the form $(j, v) \in E$, counting multiplicities. Notice that each loop of the form (v, v) is counted once both in $\deg_+(v)$ and $\deg_-(v)$.

Given natural m, a *path* of length m is a sequence $(i_1, \dots, i_{m+1}) \in [N]^{m+1}$. The path P is *closed* if the first and the last vertex coincide. Each path $P = (i_1, \dots, i_{m+1})$ naturally generates a multi digraph $G_P = (V, E)$, where $V = \{i_1, \dots, i_{m+1}\}$ and E contains the edge (i, j) with multiplicity n if and only if the path P contains exactly n times the adjacent pair (i, j). Notice that in general there is more than one path generating the same multi digraph. If the path P is closed, then G_P is strongly connected, that is for any $u, v \in V$ one can travel from u to v by following edges from E. A closed path without repeated vertices except for the first and last vertices is called a *cycle*. A loop

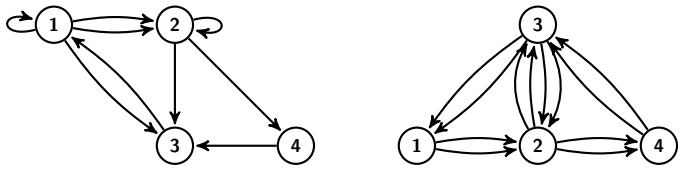

FIGURE 2. Two examples of multi digraphs. In the first case $\deg_+(1) = 4$ and $\deg_-(1) = 2$. The second example is an even digraph: it is generated by the even path $(1, 2, 3, 2, 4, 3, 1, 2, 3, 2, 4, 3, 1)$, and it can be decomposed into two double cycles, e.g. $(1, 2, 3, 1, 2, 3, 1)$ and $(2, 4, 3, 2, 4, 3, 2)$.

(i, i) is considered a cycle of length 1. A multi digraph will be called a *double cycle* if it is obtained by repeating two times a given cycle. In particular, a double cycle is not allowed to have loops unless its vertex set consists of just one vertex. We say that P is an *even path* if it is closed and every adjacent pair (i, j) is repeated in P an even number of times. A multi digraph is called an *even digraph* if it is generated by an even path; see Figure 2.1 for an example. Thus, an even digraph is always strongly connected. The following lemma can be proved by adapting the classical theorems of Euler and Veblen.

Lemma 2.1. *For a strongly connected multi digraph G, the following are equivalent:*

1) G is an even digraph;
2) $\deg_+(v) = \deg_-(v)$ is even for every vertex v;
3) G can be partitioned into a collection of double cycles.

2.2. **Equivalence classes and rooted digraphs.** Two multi digraphs $G = (V, E)$ and $G' = (V', E')$ are called *isomorphic* if there is a bijection $f : V \to V'$ such that $(i, j) \in E$ if and only if $(f(i), f(j)) \in E'$ and the multiplicities of the corresponding edges coincide. The associated equivalence classes are regarded as unlabeled multi digraphs. Given an unlabeled multi digraph $\mathcal{U}$, we will write $G \sim \mathcal{U}$ for any multi digraph G belonging to the class $\mathcal{U}$. An edge-rooted multi digraph $G = (V, E, \rho)$, or simply a *rooted digraph*, is defined as a multi digraph with a distinguished directed edge $\rho \in E$. The definition of equivalence classes is extended to rooted digraphs as follows. Two rooted digraphs $G = (V, E, \rho)$ and $G' = (V', E', \rho')$ are called isomorphic if there is a bijection $f : V \to V'$ such that $(i, j) \in E$ if and only if $(f(i), f(j)) \in E'$, multiplicities of corresponding edges coincide, and $f(\rho) = \rho'$. With minor abuse of notation we will use the same terminology as above, and write $G \sim \mathcal{U}$ for rooted digraphs G belonging to the equivalence class $\mathcal{U}$.

2.3. **Counting.** We turn to an estimate on the number of paths generating a given even digraph. Let $G = (V, E)$ be an even digraph with $|E| = 2k$ edges. Unless otherwise specified, multiplicities are always included in the edge count $|E|$. By Lemma 2.1 every vertex v has even in- and out-degrees satisfying

$$\deg_+(v) = \deg_-(v). \tag{2.1}$$

Thus G has at most k vertices. Moreover, since the number of edges in G is $2k$, we have

$$\sum_{v \in V} \deg_+(v) = \sum_{v \in V} \deg_-(v) = 2k. \tag{2.2}$$

Lemma 2.2 (Counting paths on digraphs)**.** *Let $G = (V, E)$ be an even digraph with $|E| = 2k$ and $|V| = \ell$. The number of paths generating G does not exceed*

$$\ell(4k - 4\ell)!$$

Proof. There are ℓ possibilities for the starting points of the path. The path is then characterized by the order in which neighboring vertices are visited. At each vertex v, there are $\deg_+(v)$ visits, and at most $\deg_+(v)/2$ out-neighbors. If $\deg_+(v) = 2$, there is only one possible choice for the next neighbor. If $\deg_+(v) \geqslant 4$, then there are at most $\deg_+(v)!$ possible choices considering all visits to the vertex v. Hence, the number of paths generating G is bounded by

$$\ell \prod_{v:\,\deg_+(v)\geqslant 4} (\deg_+(v)!) \leqslant \ell \left(\sum_{v:\,\deg_+(v)\geqslant 4} \deg_+(v)\right)!$$

where we have used that the product of factorials does not exceed the factorial of the sum. Now, let q be the number of vertices v such that $\deg_+(v) \geqslant 4$. From (2.2), we have

$$\sum_{v:\,\deg_+(v)\geqslant 4} \deg_+(v) + 2(\ell - q) = 2k. \tag{2.3}$$

Estimating the sum in (2.3) from below by $4q$ one has $4q + 2(\ell - q) \leqslant 2k$. Hence,

$$q \leqslant k - \ell. \tag{2.4}$$

Using (2.4) in (2.3) one finds

$$\sum_{v:\,\deg_+(v)\geqslant 4} \deg_+(v) \leqslant 4k - 4\ell.$$

□

For integers $1 \leqslant \ell \leqslant \min\{k, N\}$, let $\mathcal{G}_N(k, \ell)$ be the set of rooted even digraphs $G = (V, E)$ with $V \subset [N]$ such that $|V| = \ell$ and $|E| = 2k$.

Lemma 2.3 (Graphs counting)**.** *For any $k, N \in \mathbb{N}$, $1 \leqslant \ell \leqslant \min\{k, N\}$, the cardinality of $\mathcal{G}_N(k, \ell)$ satisfies*

$$|\mathcal{G}_N(k, \ell)| \leqslant N^\ell k^{2(k-\ell)+1}. \tag{2.5}$$

Proof. We first choose ℓ vertices among N. There are

$$\binom{N}{\ell} \leqslant \frac{N^\ell}{\ell!}$$

choices. Without loss of generality we assume that the set of vertices is given by $\{1, \dots, \ell\}$. Next, we assign an admissible degree to each vertex of $\{1, \dots, \ell\}$. Let $m(j) \in \mathbb{N}$ be defined as $m(j) = \deg_\pm(j)/2$. In view of (2.1) and (2.2), one has $m(j) \geqslant 1$ and $\sum_{j=1}^{\ell} m(j) = k$. Thus there are

$$\binom{k-1}{\ell-1} \leqslant k^{k-\ell}$$

choices for the vector $(m(1), \dots, m(\ell))$. Next, we need to count the number of multi digraphs with the given degree sequence. To this end, we may use the configuration model. Namely, we think of every vertex j as having $m(j)$ heads and $m(j)$ tails. Altogether, there will be k heads and k tails. Each head is thought of as a pair of loose out-edges (without an assigned out-neighbor) while each tail is thought of as a pair of loose in-edges (without an assigned in-neighbor). The number of multi digraphs with the given degree sequence is bounded by the number of bipartite matchings of heads and tails, which gives $k!$ possible choices. Thus, using $k!/\ell! \leqslant k^{k-\ell}$, we see that the total number of even multi digraphs with ℓ vertices and $2k$ edges is bounded above by

$$N^\ell k^{2(k-\ell)}.$$

It remains to choose the root edge. Since there are at most k choices, the proof is complete. □

3. Statistics of even digraphs

Every edge $(i,j) \in [N] \times [N]$ is given the random weight $X_{i,j}$, where $X_{i,j}$ are independent copies of a random variable $\mathbf{x}$ satisfying the assumptions of Theorem 1.3. The weight of an even digraph $G = (V, E)$, is defined as

$$p(G) := \prod_{(i,j)\in E} |X_{i,j}|^{n_{i,j}}, \tag{3.1}$$

where each edge $(i,j) \in E$ has multiplicity $n_{i,j} \geqslant 2$. Note that in this formula we interpret "$(i,j) \in E$" without taking into account the multiplicity in the multiset E. Given an unlabeled even graph $\mathcal{U}$, consider the equivalence class of even digraphs $\{G : G \sim \mathcal{U}\}$. We are interested in estimating

$$\mathcal{S}_h(\mathcal{U}) := \frac{2^h |\{G \sim \mathcal{U} :\, p(G) \geqslant 2^h\}|}{|\{G :\, G \sim \mathcal{U}\}|}, \tag{3.2}$$

for $h = 0, 1, 2, \ldots$ Moreover, we define

$$\mathcal{S}(\mathcal{U}) := \max(1, \max_{h\in\{0,1,2,\ldots\}} \mathcal{S}_h(\mathcal{U}))\,. \tag{3.3}$$

We refer to $S(\mathcal{U})$ as the *statistics* of the unlabeled even digraph $\mathcal{U}$.

We extend the above definitions to rooted even digraphs as follows. The weight of a rooted even digraph $G = (V, E, \rho)$ is defined by

$$p_r(G) = \prod_{(i,j)\in E} |X_{i,j}|^{n_{i,j} - 2\mathbf{1}_{(i,j)=\rho}}. \tag{3.4}$$

Note that

$$p_r(V, E, \rho) = |X_\rho|^{-2} p(V, E),$$

is well defined even if $X_\rho = 0$ since the root edge ρ satisfies $\rho \in E$ and thus $n_\rho \geqslant 2$. If $\mathcal{U}$ is an unlabeled rooted even digraph, that is an equivalence class of rooted even digraphs, then $\mathcal{S}_h(\mathcal{U})$ and $\mathcal{S}(\mathcal{U})$ are defined as in (3.2) and (3.3), provided $p(G)$ is replaced by $p_r(G)$ in that expression.

Estimates for the statistics $\mathcal{S}(\mathcal{U})$ will be derived from a basic estimate for double cycles. Let $\mathcal{C}_m$ be the unlabeled double cycle with $2m$ edges. Similarly, $\mathcal{C}_m^\star$ will denote the unlabeled rooted double cycle with $2m$ edges. From the assumptions of Theorem 1.3, for any double cycle $C \sim \mathcal{C}_m$ we have

$$\mathbb{E}[p(C)] \leqslant 1\,, \quad \mathbb{E}[p(C)^{1+\varepsilon/2}] \leqslant B^m. \tag{3.5}$$

Note that the same bounds apply for any rooted double cycle $C \sim \mathcal{C}_m^\star$, with the weights $p(C)$ replaced by $p_r(C)$.

Lemma 3.1 (Cycles statistics)**.** *For any $k \geqslant 1$, and $m \leqslant k$, define the event*

$$\mathcal{A}_k := \mathcal{A}_k^1 \cap \mathcal{A}_k^2 \cap \mathcal{A}_k^3$$

where

$$\mathcal{A}_k^1 := \bigcap_{m=1}^k \Big\{ \sum_{h=0}^\infty \mathcal{S}_h(\mathcal{C}_m) \leqslant k^2 \Big\},$$

$$\mathcal{A}_k^2 := \bigcap_{m=1}^k \Big\{ \sum_{h=0}^\infty \mathcal{S}_h(\mathcal{C}_m^\star) \leqslant k^2 \Big\},$$

$$\mathcal{A}_k^3 := \bigcap_{m=1}^k \Big\{ \sum_{h=0}^\infty 2^{h\varepsilon/2} \mathcal{S}_h(\mathcal{C}_m) \leqslant k^2 B^m \Big\}.$$

Then

$$\mathbb{P}(\mathcal{A}_k) \geqslant 1 - \frac{6}{k}.$$

Proof. For any $a \geqslant 0$ one has

$$\frac{1}{2}\sum_{h=0}^{\infty} 2^h \mathbf{1}_{a \geqslant 2^h} \leqslant a \leqslant 1 + 2\sum_{h=0}^{\infty} 2^h \mathbf{1}_{a \geqslant 2^h}. \tag{3.6}$$

Take any $C \sim \mathcal{C}_m$. The first inequality in (3.6) yields

$$\frac{1}{2}\sum_{h=0}^{\infty} 2^h \mathbf{1}_{p(C) \geqslant 2^h} \leqslant p(C). \tag{3.7}$$

Taking the expectation, (3.5) implies

$$\sum_{h=0}^{\infty} 2^h \mathbb{P}(p(C) \geqslant 2^h) \leqslant 2.$$

On the other hand, by symmetry any $C \sim \mathcal{C}_m$ satisfies

$$2^h \mathbb{P}(p(C) \geqslant 2^h) = \mathbb{E}[\mathcal{S}_h(\mathcal{C}_m)]. \tag{3.8}$$

Hence, from Markov's inequality and a union bound over $1 \leqslant m \leqslant k$, one has

$$\mathbb{P}(\mathcal{A}_k^1) \geqslant 1 - \frac{2}{k}. \tag{3.9}$$

for all $m \leqslant k$. Next, as in (3.7) one shows that

$$p(C)^{1+\varepsilon/2} \geqslant \frac{1}{2}\sum_{h=0}^{\infty} 2^{h(1+\varepsilon/2)} \mathbf{1}_{p(C) \geqslant 2^h}.$$

Then (3.5) and (3.8) imply

$$\sum_{h=0}^{\infty} 2^{h\varepsilon/2} \mathbb{E}[\mathcal{S}_h(\mathcal{C}_m)] = \sum_{h=0}^{\infty} 2^{h+h\varepsilon/2} \mathbb{P}(p(C) \geqslant 2^h) \leqslant 2\,\mathbb{E}\left[p(C)^{1+\varepsilon/2}\right] \leqslant 2B^m.$$

Therefore, from Markov's inequality and a union bound over $1 \leqslant m \leqslant k$,

$$\mathbb{P}(\mathcal{A}_k^3) \geqslant 1 - \frac{2}{k}. \tag{3.10}$$

Finally, we observe that the same argument leading to (3.9) can be repeated for rooted cycles, with no modifications. It follows that

$$\mathbb{P}(\mathcal{A}_k^2) \geqslant 1 - \frac{2}{k}. \tag{3.11}$$

From (3.9)-(3.11) and the union bound over $i = 1, 2, 3$, it follows that

$$\mathbb{P}(\mathcal{A}_k) \geqslant 1 - \frac{6}{k}.$$

□

In the remainder of this section, on the event $\mathcal{A}_k$, we will deterministically upper bound the statistics of any unlabeled rooted even digraph; see Proposition 3.3 below. The proof will use the following induction statement.

Lemma 3.2 (Induction)**.** *Fix integers $1 \leqslant r \leqslant m \leqslant k \ll \sqrt{N}$. Let $\mathcal{U}'$ be an unlabeled rooted even digraph with at most k vertices and assume that $\mathcal{U}'$ can be decomposed as $\mathcal{U}' = \mathcal{U} \cup \mathcal{C}_m$ for some unlabeled rooted even digraph $\mathcal{U}$ and a double cycle $\mathcal{C}_m$ of length $2m$ having r common vertices with $\mathcal{U}$. Suppose that $\mathcal{A}_k$ holds. Then*

1) $\mathcal{S}(\mathcal{U}') \leqslant 3ek^2 N^r \mathcal{S}(\mathcal{U})$;

2) If $m \log B \leqslant \frac{\varepsilon}{4} r \log N$, then $\mathcal{S}(\mathcal{U}') \leqslant 5ek^2 N^{r(1-\varepsilon/8)} \mathcal{S}(\mathcal{U})$.

Proof. Fix an even rooted digraph $G' \sim \mathcal{U}'$ and denote by $C \sim \mathcal{C}_m$ and $G \sim \mathcal{U}$, respectively, the double cycle with $2m$ edges and the even rooted digraph isomorphic to $\mathcal{U}$ so that $G' = G \cup C$. Further, let π be a uniform random permutation of $[N]$, which we assume to be defined on a different probability space. Any permutation induces a mapping on rooted digraphs via vertex relabeling, so that the rooted digraph $\pi[G']$ is uniformly distributed on the set $\{H : H \sim \mathcal{U}'\}$. Hence we may write

$$\mathcal{S}_h(\mathcal{U}') = 2^h \mathbb{P}_\pi(p_r(\pi[G']) \geqslant 2^h), \quad h = 0, 1, \dots \tag{3.12}$$

where $\mathbb{P}_\pi$ denotes the probability w.r.t. the random permutation π. For any $a, b \geqslant 0$,

$$\begin{aligned}\mathbf{1}_{ab \geqslant 2^h} &= \mathbf{1}_{ab \geqslant 2^h} \Big(\sum_{\ell=1}^{h} \mathbf{1}_{2^{\ell-1} \leqslant a < 2^\ell} + \mathbf{1}_{a<1} + \mathbf{1}_{a \geqslant 2^h} \Big) \\ &\leqslant \sum_{\ell=1}^{h} \mathbf{1}_{b \geqslant 2^{h-\ell};\, a \geqslant 2^{\ell-1}} + \mathbf{1}_{b \geqslant 2^h} + \mathbf{1}_{a \geqslant 2^h}.\end{aligned}$$

Using this and $p_r(\pi[G']) = p_r(\pi[G])\, p(\pi[C])$, one may estimate

$$\begin{aligned}\mathbb{P}_\pi \left(p_r(\pi[G']) \geqslant 2^h \right) \leqslant \sum_{\ell=1}^{h} \mathbb{P}_\pi \left(p_r(\pi[G]) \geqslant 2^{h-\ell};\, p(\pi[C]) \geqslant 2^{\ell-1} \right) \\ + \mathbb{P}_\pi \left(p(\pi[C]) \geqslant 2^h \right) + \mathbb{P}_\pi \left(p_r(\pi[G]) \geqslant 2^h \right).\end{aligned} \tag{3.13}$$

Let us condition on a fixed realization R of π restricted to the vertices V of G. Thus, $\mathbb{P}_\pi(\cdot \,|\, R)$ represents a uniform average over all permutations that agree with the given R on V. We write $C' \sim (C; R)$ for any digraph C' that has the form $C' = \pi[C]$ for some π that agrees with R on V. Since C has $m - r$ free vertices (those which do not fall into V), and we can pick them among $N - |V|$ available vertices, the cardinality of $\{C' \sim (C; R)\}$ is at least

$$(N - |V|)(N - |V| - 1) \cdots (N - |V| - (m - r - 1)) \geqslant (N - k)^{(m-r)},$$

where we use that the total number of vertices satisfies $|V| + (m - r) \leqslant k$. Since the number of double cycles of length $2m$ is $\binom{N}{m}(m-1)! \leqslant N^m$, we can write for any $\tau > 0$:

$$\begin{aligned}\mathbb{P}_\pi(p(\pi[C]) \geqslant \tau \,|\, R) &= \frac{|\{C' \sim (C; R) :\, p(C') \geqslant \tau\}|}{|\{C' \sim (C; R)\}|} \\ &\leqslant (N - k)^{r-m} |\{C' \sim \mathcal{C}_m :\, p(C') \geqslant \tau\}| \\ &\leqslant (N - k)^{r-m} N^m \mathbb{P}_\pi(p(\pi[C]) \geqslant \tau) \leqslant e N^r \mathbb{P}_\pi(p(\pi[C]) \geqslant \tau),\end{aligned}$$

where we use $r \leqslant m \leqslant k \ll \sqrt{N}$ to bound $(1 - \frac{k}{N})^{r-m} \leqslant e$. Since the above estimate is uniform over the realization R, for any $\ell = 1, 2, \dots, h$ we have

$$\begin{aligned}&\mathbb{P}_\pi \left(p_r(\pi[G]) \geqslant 2^{h-\ell};\, p(\pi[C]) \geqslant 2^{\ell-1} \right) \\ &\quad \leqslant \mathbb{P}_\pi \left(p_r(\pi[G]) \geqslant 2^{h-\ell} \right) \sup_R \mathbb{P}_\pi \left(\pi[C] \geqslant 2^{\ell-1} \,|\, R \right) \\ &\quad \leqslant e N^r \mathbb{P}_\pi \left(p_r(\pi[G]) \geqslant 2^{h-\ell} \right) \mathbb{P}_\pi \left(p(\pi[C]) \geqslant 2^{\ell-1} \right).\end{aligned}$$

Using the definition of $\mathcal{S}(\mathcal{U})$ and the identity (3.12) applied to G and C we obtain, for all $\ell = 1, \dots, h$:

$$\mathbb{P}_\pi \left(p_r(\pi[G]) \geqslant 2^{h-\ell};\, p(\pi[C]) \geqslant 2^{\ell-1} \right) \leqslant e N^r 2^{1-h} \mathcal{S}(\mathcal{U}) \mathcal{S}_{\ell-1}(\mathcal{C}_m). \tag{3.14}$$

From (3.13) one has

$$\mathbb{P}_\pi(p(\pi[G']) \geqslant 2^h) \leqslant eN^r 2^{1-h}\mathcal{S}(\mathcal{U}) \sum_{\ell=0}^{h-1} \mathcal{S}_\ell(\mathcal{C}_m) + 2^{-h}\mathcal{S}_h(\mathcal{C}_m) + 2^{-h}\mathcal{S}(\mathcal{U}).$$

Since $\mathcal{S}(\mathcal{U}) \geqslant 1$, on the event $\mathcal{A}_k$ of Lemma 3.1 one can estimate

$$2^h\mathbb{P}_\pi(p(\pi[G']) \geqslant 2^h) \leqslant 2eN^r\mathcal{S}(\mathcal{U}) \sum_{\ell=0}^{\infty} \mathcal{S}_\ell(\mathcal{C}_m) + \mathcal{S}(\mathcal{U}) \leqslant 3ek^2N^r\mathcal{S}(\mathcal{U}).$$

Taking the supremum over h, the above relation proves the first assertion of the lemma.

Let us prove the second assertion. On the event $\mathcal{A}_k$ of Lemma 3.1, for any $T \in \mathbb{N}$,

$$\sum_{\ell=T}^{\infty} \mathcal{S}_\ell(\mathcal{C}_m) \leqslant 2^{-\varepsilon T/2}k^2B^m.$$

Fix $T = \lceil \log_2(N^{r(1-\varepsilon/8)}) \rceil$. If $m \log B \leqslant \frac{\varepsilon}{4} r \log N$, then

$$\sum_{\ell=T}^{\infty} \mathcal{S}_\ell(\mathcal{C}_m) \leqslant k^2N^{-\varepsilon r/8}.$$

Estimating as in (3.14) for all $\ell \geqslant T+1$, we obtain

$$\sum_{\ell=T+1}^{h} \mathbb{P}_\pi\left(p_r(\pi[G]) \geqslant 2^{h-\ell};\ p(\pi[C]) \geqslant 2^{\ell-1}\right) \leqslant 2^{-h+1}ek^2\mathcal{S}(\mathcal{U})N^{r(1-\varepsilon/8)}.$$

On the other hand, using $\mathbb{P}_\pi\left(p_r(\pi[G]) \geqslant 2^{h-\ell}\right) \leqslant 2^{-h+\ell}\mathcal{S}(\mathcal{U})$, we find

$$\sum_{\ell=1}^{T} \mathbb{P}_\pi(p_r(\pi[G]) \geqslant 2^{h-\ell};\ p(\pi[C]) \geqslant 2^{\ell-1}) \leqslant 2^{-h}\mathcal{S}(\mathcal{U})2^{T+1} \leqslant 2^{-h+2}\mathcal{S}(\mathcal{U})N^{r(1-\varepsilon/8)}.$$

From (3.13) it follows that

$$\mathbb{P}_\pi(p(\pi[G']) \geqslant 2^h) \leqslant 2^{-h+2}ek^2\mathcal{S}(\mathcal{U})N^{r(1-\varepsilon/8)} + 2^{-h}\mathcal{S}_h(\mathcal{C}_m) + 2^{-h}\mathcal{S}(\mathcal{U}).$$

On the event $\mathcal{A}_k$ one has $\mathcal{S}_h(\mathcal{C}_m) \leqslant k^2 \leqslant k^2\mathcal{S}(\mathcal{U})$, and therefore

$$2^h\mathbb{P}_\pi(p(\pi[G']) \geqslant 2^h) \leqslant 5ek^2N^{r(1-\varepsilon/8)}\mathcal{S}(\mathcal{U}).$$

Taking the supremum over h, we obtain the second assertion of the lemma. □

We turn to the main statement of this section

Proposition 3.3 (Main estimate)**.** *Suppose $N^{\varepsilon/16} \geqslant 5ek^2$, and let $\mathcal{U}$ be an unlabeled rooted even graph with $2k$ edges and x vertices. Define*

$$y_x := \max\left(0, k - x - \frac{4k\log B}{\varepsilon \log N}\right).$$

Then, on the event $\mathcal{A}_k$ we have

$$\mathcal{S}(\mathcal{U}) \leqslant N^{k-x}N^{-\varepsilon y_x/16}k^2\left(3ek^2\right)^{\frac{4k\log B}{\varepsilon\log N}}.$$

Proof. By Lemma 2.1 we may represent $\mathcal{U}$ as the union of double cycles $C_1, \dots, C_q$, such that:

1) C_1 is rooted;
2) for all $i \in [q]$, C_i has $2m_i$ edges;
3) for $i \geqslant 2$, C_i has $r_i \geqslant 1$ common vertices with $\cup_{j=1}^{i-1} C_j$.

Define the rooted even digraphs $U_i = \bigcup_{j=1}^{i} C_j$, $i = 1, 2, \dots, q$. Let $\mathcal{U}_i$ denote the associated equivalence classes. Let J be the set of indices $i \geqslant 2$ such that

$$m_i \log B \leqslant \frac{\varepsilon}{4} r_i \log N.$$

Since $m_i > \frac{\varepsilon \log N}{4 \log B}$ for any $i \geqslant 2$, $i \notin J$, using $\sum_{i \geqslant 2} m_i \leqslant k$ we see that

$$|\{2, \dots, q\} \setminus J| \leqslant \frac{4k \log B}{\varepsilon \log N}.$$

Since $\mathcal{U}_1$ is a rooted double cycle with at most $2k$ edges, and we are assuming the validity of the event $\mathcal{A}_k$, by Lemma 3.1 we have $\mathcal{S}(\mathcal{U}_1) \leqslant k^2$. Moreover, by Lemma 3.2, one has

$$\mathcal{S}(\mathcal{U}_i) \leqslant 3ek^2 \mathcal{S}(\mathcal{U}_{i-1}) N^{r_i}, \qquad i \in \{2, \dots, q\} \setminus J$$
$$\mathcal{S}(\mathcal{U}_i) \leqslant 5ek^2 \mathcal{S}(\mathcal{U}_{i-1}) N^{r_i - r_i \varepsilon/8} \leqslant \mathcal{S}(\mathcal{U}_{i-1}) N^{r_i - r_i \varepsilon/16}, \qquad i \in J,$$

where we used the assumption $5ek^2 \leqslant N^{\varepsilon/16}$. Next, observe that

$$\sum_{i=2}^{q} r_i = k - x.$$

Thus, combining the above estimates one has

$$\mathcal{S}(\mathcal{U}) \leqslant N^{k-x} N^{-\varepsilon y'/16} k^2 (3ek^2)^{\frac{4k \log B}{\varepsilon \log N}},$$

where $y' = \sum_{i \in J} r_i$. Note that

$$\sum_{i \notin J} r_i \leqslant \sum_{i \notin J} \frac{4 m_i \log B}{\varepsilon \log N} \leqslant \frac{4k \log B}{\varepsilon \log N},$$

implying that $y' \geqslant k - x - \frac{4k \log B}{\varepsilon \log N}$. The proof is complete. □

4. Proof of Theorem 1.3

Let $\mathcal{B}$ denote the event that $|X_{ij}| \leqslant N^2$ for all $(i, j) \in [N] \times [N]$. An application of Markov's inequality and the assumption $\mathbb{E}[|X_{ij}|^2] \leqslant 1$ shows that $\mathbb{P}(\mathcal{B}) \geqslant 1 - 1/N^2$. Thus, if we define $\mathcal{E}_k := \mathcal{A}_k \cap \mathcal{B}$, where $\mathcal{A}_k$ is the event from Lemma 3.1, then

$$\mathbb{P}(\mathcal{E}_k) \geqslant 1 - N^{-2} - 6k^{-1}. \tag{4.1}$$

We are going to choose eventually $k \sim (\log N)^2$. Therefore, thanks to (4.1), to prove the theorem it will be sufficient to prove the conditional statement

$$\mathbb{P}\left(\rho(X_N) \geqslant (1 + \delta)\sqrt{N} \mid \mathcal{E}_k\right) \leqslant C(\log N)^{-2}. \tag{4.2}$$

To prove this, we estimate the conditional moments $\mathbb{E}[\rho(X_N)^{2k-2} \mid \mathcal{E}_k]$. From the expansion in (1.8) one has

$$\mathbb{E}[\rho(X_N)^{2k-2} \mid \mathcal{E}_k] \leqslant \sum_{i,j} \sum_{P_1, P_2 : i \mapsto j} \mathbb{E}[w(P_1) \bar{w}(P_2) \mid \mathcal{E}_k], \tag{4.3}$$

where the internal sum ranges over all paths P_1 and P_2 of length $k-1$ from i to j, the weight $w(P)$ of a path is defined by (1.9), and $\bar{w}(P)$ denotes the complex conjugate of $w(P)$.

Notice that since $|X_{i,j}| \leqslant N^2$ on the event $\mathcal{E}_k$, all expected values appearing above are well defined. By the symmetry assumption we can replace the variables $X_{i,j}$ by

$$X'_{i,j} = \theta_{i,j} X_{i,j}$$

where $\theta_{i,j} \in \{-1, +1\}$ are symmetric i.i.d. random variables, independent from the $\{X_{i,j}\}$. Conditioning on $\mathcal{E}_k$ the entries X'_{ij} are no longer independent. However, since

$\mathcal{E}_k$ is measurable with respect to the absolute values $\{|X_{ij}|\}$, the signs $\theta_{i,j}$ are still symmetric and i.i.d. after conditioning on $\mathcal{E}_k$. It follows that

$$\mathbb{E}\left[w(P_1)\bar{w}(P_2) \mid \mathcal{E}_k\right] = 0,$$

whenever there is an edge with odd multiplicity in $P_1 \cup P_2$. Thus, in (4.3) we may restrict to P_1, P_2 such that each edge in $P_1 \cup P_2$ has even multiplicity. Let P denote the closed path obtained as follows: start at i, follow P_1, then add the edge (j,i), then follow P_2, then end with the edge (j,i) again. Thus, P is an even closed path of length $2k$. Note that according to our definition (3.4), if G_P is the rooted even digraph generated by the path P, with root at the edge (j,i), then

$$|w(P_1)\bar{w}(P_2)| = p_r(G_P).$$

Since the map $(P_1, P_2) \mapsto P$ is injective, (4.3) allows us to estimate

$$\mathbb{E}\left[\rho(X_N)^{2k-2} \mid \mathcal{E}_k\right] \leqslant \sum_{P} \mathbb{E}\left[p_r(G_P) \mid \mathcal{E}_k\right], \tag{4.4}$$

where the sum ranges over all even closed paths $P = (i_1, \ldots, i_{2k+1})$ of length $2k$ and G_P is defined as the rooted even digraph generated by the path P, with root at the edge (i_k, i_{k+1}). By Lemma 2.2, the sum in (4.4) can be further estimated by

$$k \sum_{x=1}^{k} (4k)^{4(k-x)} \sum_{G \in \mathcal{G}_N(k,x)} \mathbb{E}\left[p_r(G) \mid \mathcal{E}_k\right], \tag{4.5}$$

where we used $x(4k-4x)! \leqslant k(4k)^{4(k-x)}$, and $\mathcal{G}_N(k,x)$ denotes the set of all rooted even digraphs with $2k$ edges and x vertices. Below we estimate $\sum_{G \in \mathcal{G}_N(k,x)} p_r(G)$ deterministically on the set $\mathcal{E}_k$. Using the second inequality in (3.6) one has, for any $G \in \mathcal{G}_N(k,x)$:

$$p_r(G) \leqslant 1 + 2\sum_{h=0}^{\infty} 2^h \mathbf{1}_{p_r(G) \geqslant 2^h}.$$

Since on the event $\mathcal{E}_k$ all entries satisfy $|X_{i,j}| \leqslant N^2$, it follows that $p_r(G) \leqslant N^{4k-4}$. Therefore the above sum can be truncated at

$$H := \lfloor 4k \log_2 N \rfloor.$$

Let $\mathcal{U}$ be a given equivalence class of rooted even digraphs with x vertices and $2k$ edges. Summing over all $G \sim \mathcal{U}$, and recalling (3.3),

$$\sum_{G \sim \mathcal{U}} p_r(G) \leqslant \Big(1 + 2\sum_{h=0}^{H} \mathcal{S}_h(\mathcal{U})\Big) \left|\{G \sim \mathcal{U}\}\right| \leqslant 3H\mathcal{S}(\mathcal{U}) \left|\{G \sim \mathcal{U}\}\right|.$$

From Proposition 3.3, on the event $\mathcal{E}_k$ we can then estimate

$$\sum_{G \sim \mathcal{U}} p_r(G) \leqslant 3HN^{k-x}N^{-\varepsilon y_x/16}k^2\big(3ek^2\big)^{\frac{4k\log B}{\varepsilon \log N}} \left|\{G \sim \mathcal{U}\}\right|,$$

where $y_x = \max(0, k - x - \frac{4k\log B}{\varepsilon \log N})$. Summing over all equivalence classes $\mathcal{U}$ of rooted even digraphs with x vertices with $2k$ edges, on the event $\mathcal{E}_k$ one obtains

$$\sum_{G \in \mathcal{G}_N(k,x)} p_r(G) \leqslant 3HN^{k-x}N^{-\varepsilon y_x/16}k^2\big(3ek^2\big)^{\frac{4k\log B}{\varepsilon \log N}} \left|\mathcal{G}_N(k,x)\right|. \tag{4.6}$$

Going back to (4.5), using (4.6), and Lemma 2.3 to estimate $|\mathcal{G}_N(k,x)|$, one finds

$$\mathbb{E}\left[\rho(X_N)^{2k-2} \mid \mathcal{E}_k\right] \leqslant 3Hk^4N^k\big(3ek^2\big)^{\frac{4k\log B}{\varepsilon \log N}} \sum_{x=1}^{k} (4k)^{6(k-x)}N^{-\varepsilon y_x/16} \tag{4.7}$$

Fix $k \sim (\log N)^2$. If $x \leqslant k - \frac{8k \log B}{\varepsilon \log N}$, then $y_x \geqslant (k-x)/2$ and therefore

$$(4k)^{6(k-x)} N^{-\varepsilon y_x/16} \leqslant (4k)^{6(k-x)} N^{-\varepsilon(k-x)/32} \leqslant 1,$$

provided that N is sufficiently large. It follows that

$$\sum_{x=1}^{k} (4k)^{6(k-x)} N^{-\varepsilon y_x/16} \leqslant k + \tfrac{8k \log B}{\varepsilon \log N} (4k)^{\frac{48k \log B}{\varepsilon \log N}}.$$

From (4.7), for large enough N and $k \sim (\log N)^2$, one has

$$\mathbb{E}\left[\rho(X_N)^{2k-2} \mid \mathcal{E}_k\right] \leqslant N^k (\log N)^{C \log N}, \tag{4.8}$$

where $C = C(\varepsilon, B) > 0$ is a constant depending only on ε, B. The proof of (4.2) is concluded by using Markov's inequality: for any $\delta > 0$,

$$\begin{aligned}\mathbb{P}(\rho(X_N) \geqslant (1+\delta)\sqrt{N} \mid \mathcal{E}_k) &\leqslant (1+\delta)^{-2k+2} N^{-k+1} \mathbb{E}[\rho(X_N)^{2k-2} \mid \mathcal{E}_k] \\ &\leqslant (1+\delta)^{-2k+2} N (\log N)^{C \log N}.\end{aligned}$$

Since $k \sim (\log N)^2$, for fixed $\delta > 0$, the expression above is $\mathcal{O}(N^{-\gamma})$ for any $\gamma > 0$. This ends the proof of Theorem 1.3.

References

[ABP] A. Auffinger, G. Ben Arous, and S. Péché. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* 45(3):589–610, 2009.

[BY] Z.D. Bai and Y.Q. Yin. Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang. *Probab. Theory Related Fields* 73(4):555–569, 1986.

[BC1] Ch. Bordenave and M. Capitaine. Outlier eigenvalues for deformed i.i.d random matrices. To appear in *Comm. Pure Appl. Math.* (2016) preprint available at arXiv:1403.6001

[BC2] Ch. Bordenave and D. Chafaï. Around the circular law. *Probab. Surv.* 9:1–89 (2012).

[BCC] Ch. Bordenave, P. Caputo, and D. Chafaï. Spectrum of non-Hermitian heavy tailed random matrices. *Comm. Math. Phys.* 307(2):513–560, 2011.

[G] S. Geman. The spectral radius of large random matrices. *Ann. Probab.* 14(4):1318–1328, 1986.

[GH] S. Geman and C.-R. Hwang. A chaos hypothesis for some large systems of random equations. *Z. Wahrsch. Verw. Gebiete* 60(3):291–314 (1982).

[S] A. Soshnikov. Poisson statistics for the largest eigenvalues of wigner random matrices with heavy tails. *Electron. Comm. Probab.* 9:82–91, 2004.

[T] T. Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields* 155(1-2):231–263 (2013).

[TV] T. Tao and V. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.* 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

(Charles Bordenave) CNRS & Université de Toulouse, France

(Pietro Caputo) Università Roma Tre, Italy

(Djalil Chafaï) Université Paris-Dauphine, France

(Konstantin Tikhomirov) University of Alberta, Canada