Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2009, Vol. 45, No. 3, 589–610
DOI: 10.1214/08-AIHP188
© Association des Publications de l'Institut Henri Poincaré, 2009

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Poisson convergence for the largest eigenvalues of heavy tailed random matrices

Antonio Auffinger[a,1], Gérard Ben Arous[a,1] and Sandrine Péché[b]

[a]*Courant Institute of the Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA.*
E-mail: auffing@cims.nyu.edu; benarous@cims.nyu.edu
[b]*Institut Fourier, Université Joseph Fourier – Grenoble, 100 rue des Maths, BP 74, 38402 St Martin d'Heres, France.*
E-mail: sandrine.peche@ujf-grenoble.fr

Received 6 November 2007; revised 10 May 2008; accepted 28 March 2008

Abstract. We study the statistics of the largest eigenvalues of real symmetric and sample covariance matrices when the entries are heavy tailed. Extending the result obtained by Soshnikov in (*Electron. Commun. Probab.* **9** (2004) 82–91), we prove that, in the absence of the fourth moment, the asymptotic behavior of the top eigenvalues is determined by the behavior of the largest entries of the matrix.

Résumé. On étudie la loi des plus grandes valeurs propres de matrices aléatoires symétriques réelles et de covariance empirique quand les coefficients des matrices sont à queue lourde. On étend le résultat obtenu par Soshnikov dans (*Electron. Commun. Probab.* **9** (2004) 82–91) et on montre que le comportement asymptotique des plus grandes valeurs propres est déterminé par les plus grandes entrées de la matrice.

MSC: 15A52; 62G32; 60G55

Keywords: Largest eigenvalues statistics; Extreme values; Random matrices; Heavy tails

1. Introduction and notation

We study the statistics of the largest eigenvalues of symmetric and sample covariance matrices when the entries are heavy tailed. Extending the result obtained by Soshnikov in [16], we prove that in the absence of a finite fourth moment, the asymptotic behavior of the top eigenvalues is determined by the behavior of the largest entries of the matrix, i.e. that the point process of the largest eigenvalues (properly normalized) converges to a Poisson Point Process, as in the usual extreme value theory for i.i.d. random variables. This result was predicted in the physics literature by Biroli, Bouchaud and Potters [7].

We first consider the case of random real symmetric matrices with independent and heavy tailed entries. Let (a_{ij}), $1 \le i \le n$, $1 \le j \le n$ be i.i.d. random variables such that:

$$1 - F(x) = \bar{F}(x) = \mathbb{P}\big(|a_{ij}| > x\big) = L(x)x^{-\alpha}, \tag{1}$$

where $\alpha > 0$ and L is a slowly varying function, i.e., for all $t > 0$

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1.$$

[1]Supported in part by NSF Grant OISE-0730136.

Consider the $n \times n$ real symmetric random matrix A_n whose entries above the diagonal are the (a_{ij}), $1 \le i \le j \le n$. Hypothesis (1) would be natural in the theory of extreme values for i.i.d. random variables. It simply asserts that the distribution of the entries is in the max-domain of attraction of the Fréchet distribution with exponent α (see [12], p. 54). Thus, for any $\alpha > 0$ the point process of extreme values of the entries of A_n (properly normalized) is asymptotically Poissonian. More precisely, let

$$b_n = \inf\left\{x\colon\ 1 - F(x) \le \frac{2}{n(n+1)}\right\}, \tag{2}$$

then the Point Process

$$\hat{\mathcal{P}}_n = \sum_{1 \le i \le j \le n} \delta_{b_n^{-1}|a_{ij}|}$$

converges to a Poisson Point Process with intensity

$$\rho(x) = \frac{\alpha}{x^{1+\alpha}}.$$

It is also classical that there exists another slowly varying function L_o such that

$$b_n \sim L_o(n) n^{2/\alpha}. \tag{3}$$

When $L \equiv 1$, then $b_n = (\frac{n(n+1)}{2})^{1/\alpha}$.

We denote by $\lambda_1 \ge \cdots \ge \lambda_n$ the n (real) eigenvalues of A_n and we consider the point process on $(0, \infty)$ of (normalized) positive eigenvalues of A_n:

$$\mathcal{P}_n = \sum \delta_{b_n^{-1}\lambda_i} \mathbf{1}_{\lambda_i > 0}.$$

Theorem 1. *We assume* (1) *with* $0 < \alpha < 4$. *For* $2 \le \alpha < 4$ *we also assume that the entries are centered, i.e.* $\mathbb{E}(a_{ij}) = 0$. *The random point process* $\mathcal{P}_n$ *converges in distribution to the Poisson Point Process* $\mathcal{P}$ *defined on* $(0, \infty)$ *with intensity* $\rho(x) = \frac{\alpha}{x^{1+\alpha}}$.

This result thus shows that the largest eigenvalues of A_n behave as the largest entries of the matrix A_n when $0 < \alpha < 4$. It was proved in the range $0 < \alpha < 2$ by Soshnikov [16]. It implies for instance that the maximum eigenvalue has a Fréchet limit distribution:

Corollary 1.

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{1}{b_n}\lambda_1 \le x\right) = \exp\left(-x^{-\alpha}\right). \tag{4}$$

One word of comment is in order here. When the entries have light tails, it is well known that the random field of largest eigenvalues is not Poissonian but determinantal, and that the fluctuations of the top eigenvalue are asymptotically distributed as in the GOE, i.e. have a Tracy–Widom distribution [14]. We actually believe that the universal Tracy–Widom picture holds as soon as $\alpha > 4$, see [7] for a discussion and simulation. Some steps in this direction have been achieved by Ruzmaikina [13], who proves that the Tracy–Widom limit holds for α large enough. (She claimed that $\alpha > 18$ is enough, we believe that the arguments of [13] only work if $\alpha > 36$, see Remark 3.)

Let us first consider the case $\alpha > 2$. It is well known that, if

$$\mu_n = \frac{1}{n} \sum \delta_{\lambda_i/\sqrt{n}} \tag{5}$$

denotes the spectral measure of $\frac{A_n}{\sqrt{n}}$, μ_n converges weakly almost surely to a non-random limit, the semi-circle law,

$$\nu(x) = \begin{cases} \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2} & \text{if } |x| \le 2\sigma, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

which depends only on the variance σ^2 of the entries. In the case where $\alpha > 4$, Bai and Yin [1] have proved that the top eigenvalue sticks to the bulk, i.e. that for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\lambda_1 - 2\sigma\right| \le \varepsilon\right) \to 1, \quad n \to \infty.$$

This shows that for $\alpha > 4$, $b_n^{-1}\lambda_1 \to \infty$ so that our result in Corollary 1 ceases to be true. Our result shows that $\frac{1}{\sqrt{n}}\lambda_1$ is, roughly speaking, of order $n^{2/\alpha-1/2}$ and thus diverges. This is in agreement with and sharpens Bai and Yin's result, who have shown that a finite fourth moment is necessary to have the convergence to the edge of the bulk.

The case $\alpha = 4$ with infinite fourth moment seems very interesting and still open. It might exhibit an interesting transition between the Tracy–Widom regime and the Poissonian one.

Coming back now to the case $\alpha < 2$, the situation is different. The bulk itself is not a semi-circle. It was recently proved by Ben Arous and Guionnet [4] that the spectral measure

$$\hat{\mu}_n = n^{-1}\sum \delta_{\lambda_i/c_n},$$

where

$$c_n = \inf\left\{x\colon\ 1 - F(x) \le \frac{1}{n}\right\}$$

converges to a limiting distribution μ_α. This limit probability distribution μ_α is not compactly supported and has a polynomial tail of type $\frac{C_\alpha}{x^{1+\alpha}}\,\mathrm{d}x$ for some constant C_α. In this case ($\alpha < 2$), this is perfectly compatible with the present result: the extreme values of i.i.d. random variables with that distribution μ_α would have exactly the behavior we have given.

We also study in this paper the behavior of the top of the spectrum for another very important family of random matrices, i.e. the ensemble of large random sample covariance matrices.

In this setting, we consider A_n a $n \times p$ random matrix with i.i.d. centered entries (a_{ij}), $1 \le i \le n$, $1 \le j \le p$ and we define as usual the sample covariance matrix $X_n = \frac{1}{p}A_nA_n^t$. The asymptotic behavior of the bulk of X_n is also well known by the classical result of Marchenko and Pastur [10], in the case where we assume a finite second moment. The case where $0 < \alpha < 2$ is treated in [3].

Similarly to (2), let

$$b_{np} = \inf\left\{x\colon\ 1 - F(x) \le \frac{1}{np}\right\}.$$

If $\lambda_1 \ge \cdots \ge \lambda_n$ are the ordered eigenvalues of $A_nA_n^t$ and $\lim_{n\to\infty}\frac{p}{n} = \gamma$ for some positive constant $\gamma \ge 1$ defining

$$\mathcal{P}_n = \sum_i \delta_{b_{np}^{-2}\lambda_i}$$

we have the following:

Theorem 2. *We assume* (1) *with* $0 < \alpha < 4$. *For* $2 \le \alpha < 4$, *we also assume that the entries are centered, i.e.,* $\mathbb{E}(a_{ij}) = 0$. *The random point process* $\mathcal{P}_n$ *converges in distribution, as* p *goes to infinity, to the Poisson Point Process* $\mathcal{P}$ *defined on* $(0,\infty)$ *with intensity* $\rho(x) = \frac{\alpha}{2x^{1+\alpha/2}}$.

Again, as a simple corollary, we obtain the behavior of the maximal eigenvalue:

Corollary 2.

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{1}{b_{np}^2}\lambda_1 \le x\right) = \exp\left(-x^{-\alpha/2}\right). \tag{7}$$

The rest of the paper is organized as follows. First, in Section 2 we recall briefly the main results contained in [16], i.e., the proof of Theorem 1 in the case $0 < \alpha < 2$. We then prove Theorem 2 in the case $0 < \alpha < 2$ in Section 3. We then study in Sections 4 and 5 the case where $2 \le \alpha < 4$ for the Wigner and the sample covariance matrix cases respectively. This is a bit different in nature since we now have to perform a more subtle separation of scales. This is done through an estimate of traces of high powers of our random matrices properly truncated. This combinatorial part of the proof draws on the work of Soshnikov [14] and on the recent work by Péché and Soshnikov [11].

2. Wigner matrices when $0 < \alpha < 2$

In this section, we will recall the results in the paper of Soshnikov [16] (see also [15] for precise statement of Lemma 3).

Let A_n be $n \times n$ random (real) symmetric matrix with i.i.d. entries satisfying (1) with $0 < \alpha < 2$. Also let $\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \cdots \ge \lambda_n$ be its eigenvalues and $a_{i_l j_l}$ its l-th largest entry in absolute value.

In order to prove Theorem 1 in this case, Soshnikov [16] proceeds as follows. The basic idea is to show that for each finite k, and for each given $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{\lambda_k}{|a_{i_k j_k}|} - 1\right| > \varepsilon\right) \to 0 \quad \text{as } n \to \infty. \tag{8}$$

We first consider the case where $k = 1$, which implies Corollary 1. The following crucial lemma is purely probabilistic. It describes how the largest entries are placed in the matrix. This lemma will be adapted in all other sections.

Lemma 3. (a) *With probability going to one, there are no diagonal entries greater in absolute value than* $b_n^{11/20}$.

(b) *With probability going to one, there is no pair* (i, j) *such that* $|a_{ij}| > b_n^{99/100}$ *and* $|a_{ii}| + |a_{jj}| > b_n^{1/10}$.

(c) *For any positive* $\delta > 0$ *with probability going to one there is no row that has at least two entries greater in absolute value than* $b_n^{3/4} + \delta$.

(d) *With probability going to one, there is no row such that its maximum and the sum of the absolute value of the remaining elements in the row are both greater than* $b_n^{3/4+\alpha/8}$.

Once one has proved the previous lemma, the next step is to relate the entries of the matrix with its maximum eigenvalue λ_1. This can be done in two steps.

First, one can bound from below the top eigenvalue using the Rayleigh–Ritz representation of λ_1:

$$\lambda_1 = \sup_{v:|v|=1} \langle A_n v, v\rangle. \tag{9}$$

Considering a well-chosen vector v in terms of the position of the largest entry of A_n, (9) will provide the inequality

$$\lambda_1 \ge a_{i_1 j_1}\big(1 + \mathrm{o}(1)\big).$$

Secondly, one studies the following norm of the matrix A_n, that is the norm of A_n as a linear operator from l_∞ to l_∞:

$$\|A_n\|_\infty \equiv \max_i \sum_{j=1}^n |a_{ij}|. \tag{10}$$

Since A_n is symmetric, we can show that $\|A_n\|_\infty$ is an upper bound for λ_1. Lemma 3 then relates $\|A_n\|_\infty$ to the maximum entry of the matrix A_n. In fact, Soshnikov showed that, given $\varepsilon > 0$, there exists $\theta > 0$ such that for n sufficiently large, one has:

$$\mathbb{P}\left(\left|\frac{\|A_n\|_\infty}{\max_{ij}|a_{ij}|} - 1\right| > \varepsilon\right) \le n\exp\big(-n^\theta\big).$$

This proves (8) when $k = 1$.

Now, for any finite k, using Lemma 3, it is possible to find some well-chosen unit vectors v_k, such that,

$$A_n v_k = |a_{i_k j_k}| v_k + r_k \quad \text{and} \quad \|r_k\| = \mathrm{o}(1).$$

This fact together with a standard result in perturbation theory of symmetric matrices (see for instance [5], p. 77) imply that A_n has eigenvalues $|a_{i_l j_l}|(1 + \mathrm{o}(1))$, $1 \le l \le k$, for any finite k.

To finally get Theorem 1 we use induction on k in (8), supported by Corollary 1 and the following very classical result about symmetric matrices, that can be found for instance in [5], p. 59.

Proposition 4 (Cauchy interlacing theorem). *Let A_n be an $n \times n$ Symmetric matrix and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ its ordered (real) eigenvalues. If one considers the restriction B of A_n to any subspace of co-dimension 1 and denotes by $\mu_1 \ge \mu_2 \ge \cdots \ge \mu_{n-1}$ the eigenvalues of B then*

$$\lambda_1 \ge \mu_1 \ge \lambda_2 \ge \mu_2 \ge \cdots \ge \mu_{n-1} \ge \lambda_n.$$

We briefly explain how to use the last statement in this setting. For simplicity, instead of describing the general step of the induction, we describe only the step $k = 2$. First, one considers the submatrix obtained by removing the i_1-th row and the j_1-th column from A_n. Clearly, this submatrix has $a_{i_2 j_2}$ as largest entry in absolute value, and by the interlacing property its largest eigenvalue will be greater than λ_2. Thus, one can apply Corollary 1 to finally get (8) for $k = 2$.

To get Theorem 1 as stated, i.e., to prove tightness and the convergence of $\mathcal{P}_n$ to $\mathcal{P}$, it suffices to prove, see [9] Theorem 16.16, that for all intervals (a, b), where $0 < a < b$ one has that the random variable $\mathcal{P}_n(a, b)$ converges in distribution to $\mathcal{P}(a, b)$. This can be verified as follows.

Since

$$\mathbb{P}\big(\mathcal{P}_n(a, \infty) > k\big) = \mathbb{P}\left(\frac{\lambda_{k+1}}{b_n} > a\right),$$

it is easy to see that (8) proves the convergence in distribution of $\mathcal{P}_n$ when restricted to an interval (a, ∞), $a > 0$, i.e. the random variable $\mathcal{P}_n(a, \infty)$ converges in distribution to $\mathcal{P}(a, \infty)$ which is a Poisson Process with parameter $a^{-\alpha}$.

To derive the result for a general interval, we first note that since

$$\mathbb{P}\big(\mathcal{P}\big(\{b\}\big) > 0\big) = 0,$$

we can consider intervals $(a, b]$ which can be written as the difference of $I_a = (a, \infty)$ and $I_b = (b, \infty)$. Now,

$$\mathbb{P}\big(\mathcal{P}_n(a, b] = l\big) = \sum_{k=l}^{\infty} \mathbb{P}\big(\mathcal{P}_n(a, \infty) = k, \mathcal{P}_n(b, \infty) = k - l\big). \tag{11}$$

Each term inside the sum converges to $\mathbb{P}(\mathcal{P}(a, \infty) = k, \mathcal{P}(b, \infty) = k - l)$ and is also bounded by $\mathbb{P}(\mathcal{P}_n(a, \infty) = k)$. Since the sum

$$\sum_{k=l}^{\infty} \mathbb{P}\big(\mathcal{P}_n(a, \infty) = k\big)$$

is finite, an application of Fatou's lemma shows that (11) converges to

$$\mathbb{P}\big(\mathcal{P}(a, b] = l\big) = \sum_{k=l}^{\infty} \mathbb{P}\big(\mathcal{P}(a, \infty) = k, \mathcal{P}(b, \infty) = k - l\big).$$

The details omitted here will be considered in the next sections, specially the next one, where we follow the same strategy in the case of Sample Covariance matrices with $0 < \alpha < 2$.

3. Sample covariance matrices when $0 < \alpha < 2$

The proof of Theorem 2 in this case is based on the following lemma, which is almost identical to Lemma 3 given in last section.

Lemma 5. *Let A_n be a $n \times p$ random matrix with i.i.d. entries a_{ij}, $1 \le i \le n$, $1 \le j \le p$, satisfying (1). Also assume that $\lim_{n\to\infty} \frac{p}{n} = \gamma$ for some constant $\gamma \ge 1$. Then*:

(a) *If B_{np}^{δ} is the event 'There is a row with 2 entries greater than b_{np}^{δ} in absolute value' then*

$$\forall \delta > 3/4 \qquad \lim_{p\to\infty} \mathbb{P}\big(B_{np}^{\delta}\big) = 0.$$

(b) *Also*,

$$\lim_{n\to\infty} \mathbb{P}\left(\exists i, 1 \le i \le n,\ \max_{1\le j\le p} |a_{ij}| > b_{np}^{3/4+\alpha/8} \text{ and } \sum_{j=1}^{p} |a_{ij}| - \max_{1\le j\le p} |a_{ij}| > b_{np}^{3/4+\alpha/8}\right) = 0.$$

(c) *Similarly*,

$$\lim_{n\to\infty} \mathbb{P}\left(\exists j, 1 \le j \le p,\ \max_{1\le i\le n} |a_{ij}| > b_{np}^{3/4+\alpha/8} \text{ and } \sum_{i=1}^{n} |a_{ij}| - \max_{1\le i\le n} |a_{ij}| > b_{np}^{3/4+\alpha/8}\right) = 0.$$

3.1. *Proof of Lemma 5*

Proof of (a). First, a basic fact from slowly varying functions that will be used repeatedly in this paper (see for instance [6], Chapter I). Given any $\delta > 0$, one has that

$$x^{-\delta} \ll L(x) \ll x^{\delta} \quad \text{as } x \to \infty. \tag{12}$$

We recall that $f(x) \ll g(x)$ means that the ratio $\frac{f(x)}{g(x)}$ tends to 0 as x tends to infinity. Hence, using (12) and choosing $\delta = \frac{3}{4} + \varepsilon$,

$$\begin{aligned}
\mathbb{P}\big(B_{np}^{\delta}\big) &= \mathbb{P}\big(\exists i \le n, \exists j, k, j \ne k, \text{ s.t. } |a_{ij}| \ge b_{np}^{3/4+\varepsilon} \text{ and } |a_{ik}| \ge b_{np}^{3/4+\varepsilon}\big) \\
&\le p^2 n\big(1 - F\big(b_{np}^{3/4+\varepsilon}\big)\big)^2 \\
&\le p^2 n L\big(b_{np}^{3/4+\varepsilon}\big)^2 \frac{1}{b_{np}^{3\alpha/2+2\varepsilon\alpha}} \\
&= \mathrm{o}\big(n^{-4\varepsilon+\theta}\big)
\end{aligned}$$

for a small enough θ, since for any $\varepsilon > 0$, $b_{np}^{3\alpha/2+2\varepsilon\alpha} \gg n^3$ by the definition of b_{np}. □

Proof of (b). We split the proof in two cases. The idea in both cases is the same and the computation almost identical.

We start by assuming that $1 < \alpha < 2$.

Let $T \in \mathbb{N}$ be such that $1/(2T+1) < 1/4 - \alpha/8$.

Proposition 6. *Assume that $1 < \alpha < 2$. There exists $\theta > 0$ such that, for n sufficiently large*,

$$\mathbb{P}\Bigg(\sum_{j:|a_{ij}|<b_{np}^{(T+1)/(2T+1)}} |a_{ij}| \le \frac{1}{2} b_{np}^{3/4+\alpha/8}\Bigg) \ge 1 - n\exp\big(-n^{\theta}\big).$$

Proof. In order to prove the last proposition we introduce $Y_i^0 = p$, and for $k > 1$:

$$Y_i^k \equiv \#\big(1 \le j \le p\colon\ |a_{ij}| \ge b_{np}^{k/(2T+1)}\big),$$

so that

$$\sum \mathbf{1}_{\{j:|a_{ij}|<b_{np}^{(T+1)/(2T+1)}\}}|a_{ij}| \le \sum_{k=0}^{T} Y_i^k b_{np}^{(k+1)/(2T+1)}. \tag{13}$$

Lemma 7. *Let $k \le T$. There exists $\theta > 0$ such that $\mathbb{P}(Y_i^k \ge 2\mathbb{E}Y_i^k) \le \exp(-p^\theta)$.*

Proof. By definition of Y_i^k, and setting by convention $b_{np}^{0/(2T+1)} = 0$, for all $1 \le i \le n$ and $0 \le k \le T$, $k \in \mathbb{N}$ we have

$$\mathbb{E}Y_i^k = p\bar{F}\big(b_{np}^{k/(2T+1)}\big).$$

Also, using Chernoff's inequality, we have that

$$\mathbb{P}\big(Y_i^k \ge 2\mathbb{E}Y_i^k\big) \le \exp\bigg(-\frac{1}{4}\mathbb{E}Y_i^k\bigg). \tag{14}$$

Thus, we can argue as follows:

First, replace the value $\mathbb{E}Y_i^k$ in (14) and use the expression for $\bar{F}(x)$ given in (1). Hence, by (12) there exists ε sufficiently small such that

$$\begin{aligned}
\mathbb{P}\big(Y_i^k \ge 2\mathbb{E}Y_i^k\big) &\le \exp\bigg(-\frac{1}{4}p\bar{F}\big(b_{np}^{k/(2T+1)}\big)\bigg) \le \exp\bigg(-\frac{1}{4}p\bar{F}\big(b_{np}^{T/(2T+1)}\big)\bigg)\\
&\le \exp\big(-\frac{1}{4}pL\big(b_{np}^{T/(2T+1)}\big)/b_{np}^{T\alpha/(2T+1)}\big) \le \exp\big(-p/4b_{np}^{T\alpha/(2T+1)}b_{np}^{T\varepsilon/(2T+1)}\big)\\
&\le \exp\big(-p/4(np)^{T/(2T+1)(\alpha+\varepsilon)/\alpha}\big) \le \exp\big(-p^\theta\big),
\end{aligned}$$

where the last inequality is justified by the hypothesis $\lim_{n\to\infty}\frac{p}{n} = \gamma$. □

Lemma 8. *Assume that $1 < \alpha < 2$. Then,*

$$\sum_{k=0}^{T}\mathbb{E}Y_i^k b_{np}^{(k+1)/(2T+1)} \le \frac{1}{4}b_{np}^{3/4+\alpha/8}. \tag{15}$$

Proof. One has that

$$\begin{aligned}
\mathbb{E}\bigg(\sum_{j:|a_{ij}|<b_{np}^{(T+1)/(2T+1)}}|a_{ij}|\bigg) &\le \sum_{k=0}^{T}\mathbb{E}Y_i^k b_{np}^{(k+1)/(2T+1)} \le \sum_{k=0}^{T}2p\bar{F}\big(b_{np}^{k/(2T+1)}\big)b_{np}^{(k+1)/(2T+1)}\\
&\le 2p\sum_{k=0}^{T}L\big(b_{np}^{k/(2T+1)}\big)b_{np}^{(k+1)/(2T+1)}b_{np}^{-k\alpha/(2T+1)}\\
&\le 2pb_{np}^{1/(2T+1)}\sum_{k=0}^{T}b_{np}^{k(1+\delta-\alpha)/(2T+1)},
\end{aligned}$$

where $\delta > 0$ can be chosen such that $1 + \delta - \alpha < 0$. Therefore,

$$\mathbb{E}\Bigg(\sum_{j:|a_{ij}|<b_{np}^{(T+1)/(2T+1)}} |a_{ij}|\Bigg) \le 2pb_{np}^{1/(2T+1)} \sum_{k=0}^{\infty} b_{np}^{k(1+\delta-\alpha)/(2T+1)} \le pb_{np}^{1/4-\alpha/8}$$
$$\le b_{np}^{1/4-\alpha/8+\alpha/2} \le \frac{1}{4} b_{np}^{3/4+\alpha/8}. \qquad \square \tag{16}$$

Combining the last two lemmas with (13), we get Proposition 6. $\square$

Now we turn back to the sum of all terms which have absolute value between $b_{np}^{(T+1)/(2T+1)}$ and $b_{np}^{3/4+\alpha/16}$. This sum is easier to handle since we have fewer entries. To simplify a little bit the notation below, put $\mu = \frac{T+1}{2T+1}$.

Proposition 9. *There exists $\kappa > 0$ such that*

$$\mathbb{P}\Bigg(\exists i, \sum \mathbf{1}_{\{j:b_{np}^{\mu}<|a_{ij}|<b_{np}^{3/4+\alpha/16}\}} |a_{ij}| \ge \frac{1}{2} b_{np}^{3/4+\alpha/8}\Bigg) \le \exp\bigl(-n^{\kappa}\bigr).$$

Proof.

$$\begin{aligned}
&\mathbb{P}\Bigg(\exists i, \sum \mathbf{1}_{\{j:b_{np}^{\mu}<|a_{ij}|<b_{np}^{3/4+\alpha/16}\}} |a_{ij}| \ge \frac{1}{2} b_{np}^{3/4+\alpha/8}\Bigg) \\
&\quad \le n\mathbb{P}\Bigg(\#\bigl\{j : b_{np}^{\mu} < |a_{ij}|\bigr\} \ge \frac{1}{2} b_{np}^{\alpha/16}\Bigg) \\
&\quad \le n n^{1/2b_{np}^{\alpha/16}} \bar{F}\bigl(b_{np}^{\mu}\bigr)^{1/2b_{np}^{\alpha/16}} \\
&\quad \le n\Bigg(\frac{nL(b_{np}^{\mu})}{b_{np}^{\mu\alpha}}\Bigg)^{1/2b_{np}^{\alpha/16}} \\
&\quad \le \exp\bigl(-n^{\kappa}\bigr),
\end{aligned} \tag{17}$$

for some sufficiently small $\kappa > 0$ since $\mu > 1/2$. $\square$

Let us finish the proof of statement (b) when $1 < \alpha < 2$. By part (a) we know that, with probability going to 1 as p goes to infinity, there is at most one term in each line that exceeds $b_{np}^{3/4+\alpha/16}$ in absolute value. So it is enough to consider the sum of all entries less than or equal to $b_{np}^{3/4+\alpha/16}$ and prove that in fact the probability that this sum is less than $b_{np}^{3/4+\alpha/8}$ goes to one as p tends to infinity. We proved this statement in two parts, analyzed in Proposition 6 and Proposition 9 respectively.

Case $0 < \alpha \le 1$:

We repeat the same argument and computation used in the other case. We begin by proving the counterpart of Proposition 6 in this case.

Proposition 10. *There exists $\theta > 0$ such that for n large enough*

$$\mathbb{P}\Bigg(\sum_j \mathbf{1}_{\{j:|a_{ij}|<b_{np}^{(T+1)/(2T+1)}\}} |a_{ij}| \le \frac{1}{2} b_{np}^{3/4+\alpha/8}, \forall i\Bigg) \ge 1 - n\exp\bigl(-n^{\theta}\bigr).$$

Proof. Lemma 7 is valid when $0 < \alpha < 2$ so that it is enough to prove

Lemma 11. *Let* $0 < \alpha \le 1$ *and* T *such that* $\frac{1}{2T+1} < \frac{\alpha}{8}$. *Then,*

$$\sum_{k=0}^{T} \mathbb{E} Y_i^k b_{np}^{(k+1)/(2T+1)} \le \frac{1}{2} b_{np}^{3/4+\alpha/8}. \tag{18}$$

Proof. One has that

$$\sum_{k=0}^{T} \mathbb{E} Y_i^k b_{np}^{(k+1)/(2T+1)} \le \sum_{k=0}^{T} 2p\bar{F}\big(b_{np}^{k/(2T+1)}\big) b_{np}^{(k+1)/(2T+1)} \le 2p \sum_{k=0}^{T} L\big(b_{np}^{k/(2T+1)}\big) b_{np}^{(k+1)/(2T+1)} b_{np}^{-k\alpha/(2T+1)}$$

$$\le 2p b_{np}^{1/(2T+1)} \sum_{k=0}^{T} b_{np}^{k(1+\delta-\alpha)/(2T+1)} \le b_{np}^{1/2+\alpha/4} \le \frac{1}{2} b_{np}^{3/4+\alpha/8}.$$

□

Using Proposition 10 and Proposition 9 as before it is easy to prove statement (b) of Lemma 5. □

□

Proof of (c). As one can easily see, the proof of item (c) of the lemma is identical to the proof of part (b) up to a permutation of p's and n's. □

Remark 1. *Recalling definition* (10), *statement* (*b*) *of Lemma* 5, *Proposition* 6 *and Proposition* 10 *show that for every* $\varepsilon > 0$ *there exists* $n_o(\varepsilon)$ *and* $\theta > 0$ *such that for all* $n > n_o$ *one has*

$$\mathbb{P}\left(\left| \frac{\|A_n\|_\infty}{\max_{ij} |a_{ij}|} - 1 \right| > \varepsilon \right) \le n \exp\big(-n^\theta\big).$$

By part (c) *of Lemma* 5, *the same is valid if one replaces* $\|A_n\|_\infty$ *above by* $\|A_n\|_1 \equiv \sup_j \sum_{i=1}^p |a_{ij}|$.

Remark 2. *From now on, if* X_n *and* Y_n *are two sequences of random variables defined on the same probability space, we will use the notation* $X_n = Y_n(1 + \mathrm{o}(1))$ *to indicate that for all* $\varepsilon > 0$, *the probability* $\mathbb{P}(|\frac{X_n}{Y_n} - 1| > \varepsilon)$ *goes to* 0 *as* n *goes to infinity, i.e. the ratio* $\frac{X_n}{Y_n}$ *converges in probability to* 1.

3.2. *Proof of Theorem* 2

Proof of Corollary 2. We begin by the proof of Corollary 2. The main thing to show is that

$$\frac{\lambda_1}{a_{i_1 j_1}^2} \longrightarrow 1, \tag{19}$$

in probability as n tends to infinity (we recall that $|a_{i_1 j_1}| = \max |a_{ij}|$). In fact, if we assume (19), extreme value theory for i.i.d. random variables tell us that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{a_{i_1 j_1}^2}{b_{np}^2} \le x \right) = \exp\big(-x^{-\alpha/2}\big), \tag{20}$$

so (19) and (20) will imply Corollary 2. Thus, our task is to prove (19) and the idea is as follows: Given $\varepsilon > 0$, we want to show that for n sufficiently large we have

$$\lambda_1 \ge a_{i_1 j_1}^2 \big(1 + \mathrm{o}(1)\big), \tag{21}$$

$$\lambda_1 \le a_{i_1 j_1}^2 \big(1 + \mathrm{o}(1)\big), \tag{22}$$

with probability greater than $1-\varepsilon$. The main tool used to prove both equations will be Lemma 5, and we will start with the easiest inequality (21).

Since for all unit vectors v we have the bound $\langle X_n v, v\rangle \le \lambda_1$, our task is the following: we must find a suitable vector that gives us (21). Therefore, let (i_1, j_1) be the position of $a_{i_1 j_1}$ in A_n as the notation suggests. If one takes $v = (0, \dots, 0, 1, 0, \dots, 0)$ where the sole non-zero entry of v is in the position i_1, the vector v will do the job. In fact,

$$\langle X_n v, v\rangle = \sum_{j=1}^{p} a_{i_1 j}^2 = a_{i_1 j_1}^2 + \sum_{j=1, j\neq j_1}^{p} a_{i_1 j}^2 = a_{i_1 j_1}^2\big(1+\mathrm{o}(1)\big),$$

by part (b) of Lemma 5 and the fact that $a_{i_1 j_1}$ is the maximum of np i.i.d. random variables. This proves (21).

To obtain (22) we first recall the definition of

$$\|X_n\|_\infty \equiv \sup_i \sum_{j=1}^{n} |X_{ij}|.$$

The eigenvector equation for λ_1

$$\sum_{j=1}^{n} X_{ij} v_j = \lambda_1 v_i$$

implies that

$$\lambda_1 |v_i| \le \sum_{j=1}^{n} |X_{ij}||v_j| \le \sup_l |v_l| \sum_{j=1}^{n} |X_{ij}|$$

so,

$$\lambda_1 \sup_i |v_i| \le \Big(\sup_l |v_l|\Big) \sup_i \sum_{j=1}^{n} |X_{ij}|.$$

Therefore, $\|X_n\|_\infty$ is an upper bound for λ_1.

Hence, with probability going to one,

$$\begin{aligned}
\lambda_1 &\le \|X_n\|_\infty \le \sup_i \left\{\sum_{j=1}^{n}\sum_{k=1}^{p} |a_{ik}||a_{jk}|\right\} \le \sup_i \left\{\sum_{k=1}^{p} |a_{ik}| \sum_{j=1}^{n} |a_{jk}|\right\} \\
&\le \sup_i \left\{\sum_{k=1}^{p} |a_{ik}|\right\} \sup_l \left\{\sum_{j=1}^{n} |a_{jl}|\right\} \le \|A_n\|_\infty \|A_n\|_{1\to 1} \\
&\le a_{i_1 j_1}^2\big(1+\mathrm{o}(1)\big),
\end{aligned}$$

where the last inequality comes from Remark 1. □

Proof of Theorem 2. It is enough to show that for any finite k we have for all $1 \le l \le k$

$$\lim_{p\to\infty} \mathbb{P}\left(\frac{1}{b_{np}^2}\lambda_l \le x\right) = \lim_{p\to\infty} \mathbb{P}\left(\frac{1}{b_{np}^2} a_{i_l j_l}^2 \le x\right), \tag{23}$$

where $a_{i_l j_l}$ is the l-th term of the sequence $|a_{ij}|$ in the decreasing order.

Let $e_1, \ldots, e_p$ be the standard orthonormal basis of $\mathbb{R}^p$. If we compute $X_n.e_{i_l}$ we get:

$$X_n e_{i_l} = \sum_{i=1}^{n} X_{ii_l} e_i = X_{i_l i_l} e_{i_l} + r_l, \tag{24}$$

for some vector r_l in $\mathbb{R}^n$.

Also, since X_n is symmetric, one can find an orthogonal matrix U and a diagonal matrix D such that $X_n = UDU^{-1}$. Now, suppose that $X_{i_l i_l}$ is not an eigenvalue of D. Then $D - X_{i_l i_l} I$ is invertible and one can use Eq. (24) to get:

$$1 = \|e_{i_l}\| = \left\| U(D - X_{i_l i_l} I)^{-1} U^{-1} r_l \right\| \leq \|U\| \left\| U^{-1} \right\| \left\| (D - X_{i_l i_l} I)^{-1} \right\| \|r_l\|, \tag{25}$$

which implies

$$\min_i |\lambda_i - X_{i_l i_l}| \leq \|r_l\|, \tag{26}$$

so there exists an eigenvalue λ of X_n such that

$$|\lambda - X_{i_l i_l}| \leq \|r_l\|. \tag{27}$$

If $X_{i_l i_l}$ is an eigenvalue of $D - X_{i_l i_l} I$, (27) is clearly satisfied.

We now know, by Lemma 5, that $X_{i_l i_l} = a^2_{i_l j_l}(1 + \mathrm{o}(1))$. Therefore, if we manage to prove that r_l has a norm that is negligible with respect to $a^2_{i_l j_l}$, we will be able to say that X_n has eigenvalues $a^2_{i_l j_l}(1 + \mathrm{o}(1))$, $1 \leq l \leq k$ for any finite k.

Bounding the norm of r_l, one gets

$$\begin{aligned}
\|r\| &= \left(\sum_{i=1, i \neq i_l}^{n} X^2_{ii_l} \right)^{1/2} \leq \sum_{i=1, i \neq i_l}^{n} |X_{ii_l}| \\
&\leq \sum_{i=1, i \neq i_l}^{n} \sum_{k=1}^{p} |a_{ik}||a_{i_l k}| \\
&= \sum_{k=1}^{p} |a_{i_l k}| \sum_{i=1, i \neq i_l}^{n} |a_{ik}| \equiv S_1.
\end{aligned}$$

We cannot estimate S_1 directly as we did in part (a) but if we define

$$S_2 = \sum_{k=1, k \neq j_l}^{p} |a_{i_l k}| \sum_{i=1, i \neq i_l}^{n} |a_{ik}|, \tag{28}$$

then

$$S_2 \leq \left(\sup_k \sum_{i=1, i \neq i_l}^{n} |a_{ik}| \right) \sum_{k=1, k \neq j_l}^{p} |a_{i_l k}|, \tag{29}$$

which tells us that S_2 is negligible with respect to $a^2_{i_l j_l}$, again by Lemma 5. Now,

$$S_1 - S_2 = |a_{i_l j_l}| \sum_{k=1, k \neq i_l}^{n} |a_{i_l k}|, \tag{30}$$

which is also negligible with respect to $a_{i_l j_l}^2$ since $\sum_{k=1,k\neq i_l}^n |a_{i_l k}|$ is negligible by a direct application of part (b) of Lemma 5. Hence, S_1 is also negligible with respect to $a_{i_l j_l}^2$.

We now know that $a_{i_l j_l}^2(1+\mathrm{o}(1))$, $1 \le l \le k$, are eigenvalues of X_n. However, this does not imply that they are exactly the k top eigenvalues. At the moment this is true only for the maximum, by Corollary 2, and we need to check it for $1 < l \le k$. In other words, what we get from the last statement is that for all $1 < l \le k$, for p large enough:

$$\lambda_l \ge a_{i_l j_l}^2\big(1+\mathrm{o}(1)\big), \tag{31}$$

and we need to prove the reverse inequality. To achieve our goal, we consider the compression of the matrix X_n step by step, i.e., we cut from A_n the row i_1 and from A_n^t the column i_1 and then we compute their product. By part (a) of Lemma 5, the entry $a_{i_2 j_2}$ is still in the matrix and the product $X_n^{(2)}$ is just the matrix X_n without the row and column i_1. Now we know by the Cauchy Interlacing Theorem, see [5], Corollary 3.1.5, that if $\kappa_1 \ge \cdots \ge \kappa_{p-1}$ are the eigenvalues of $X_n^{(2)}$, we have:

$$\lambda_1 \ge \kappa_1 \ge \lambda_2 \ge \kappa_2 \ge \cdots \ge \kappa_{p-1} \ge \lambda_p. \tag{32}$$

Combining (32) with Corollary 2 applied for the matrix $X_n^{(2)}$, we get the desired inequality for λ_2. Repeating the same argument for all $1 \le l \le k$, the proof is complete. □

4. Wigner matrices when $2 \le \alpha < 4$

Now we consider the symmetric random matrix $(a_{ij})_{i,j=1}^n$ where the entries of A_n are centered i.i.d. and satisfy (1) with $2 \le \alpha < 4$. We treat separately the case where $\alpha = 2$.

4.1. *Truncation*

The main difference between the proof of this section to the previous one is that here we should care about the contribution given by the bulk of the spectra, that is we should control in some way the smaller entries of the matrix A_n and then proceed as before. Thus, to investigate the behavior of the largest eigenvalue λ_1, we split the above random matrix as follows. Let β be such that

$$\frac{1}{\alpha} < \beta < \frac{2(8-\alpha)}{\alpha(10-\alpha)}, \tag{33}$$

and we define

$$A_1 = (A_{ij}\mathbf{1}_{|a_{ij}|\le n^\beta})_{i,j=1}^n, \qquad A_2 = A_n - A_1. \tag{34}$$

Since $\beta < \frac{2}{\alpha}$, it is also clear that with probability going to 1, the largest entry of A_n is the largest entry of A_2. The condition that $\beta > 1/\alpha$ is assumed to guarantee that we can study the asymptotic behavior of the eigenvalues of A_2 in a similar way of the previous section. On the other hand, the condition that $\beta < \frac{2(8-\alpha)}{\alpha(10-\alpha)}$ is assumed to guarantee that the spectrum of A_1, properly normalized, remains bounded.

4.2. *Bounding the spectrum of A_1*

We first investigate the behavior of the largest eigenvalue of A_1 referring the reader to the results of [2] and [1]. These papers deal with the case of random Wigner matrices with the presence of a finite fourth moment and they prove boundedness of the spectra. Fix some $\varepsilon > 0$ such that

$$\varepsilon < \min\left\{\frac{1}{\alpha} - \frac{1}{4}, \frac{1}{\alpha} - \frac{\beta}{2}, \frac{1}{16}\left(\frac{8}{\alpha} - 1 - \beta\left(5 - \frac{\alpha}{2}\right)\right)\right\}.$$

Here, we will prove that the largest eigenvalue and the smallest eigenvalue of $\frac{1}{n^{2/\alpha-\varepsilon}}A_1$ are bounded on a set of probability arbitrarily close to 1. In this direction, our main result in this subsection will be:

Proposition 12. *Let s_n be some sequence going to infinity in such a way that $\log n \ll s_n \ll n^\gamma$ where $0 < \gamma \leq \min\{\frac{1}{8}(\frac{8}{\alpha} - 1 - \beta(5 - \frac{\alpha}{2})), \frac{1}{2\alpha} - \frac{\beta}{4}\}$. Then there exists a constant $C > 0$ such that $\mathbb{E}(\mathrm{Tr}(\frac{A_1}{n^{2/\alpha-\varepsilon}})^{2s_n}) < C(2\sigma)^{2s_n}\frac{n}{s_n^{3/2}}$.*

Before giving the proof, we indicate how to use Proposition 12 to deduce the desired result, that is the boundedness of the largest eigenvalue of $\frac{1}{n^{2/\alpha-\varepsilon}}A_1$. We have

$$\mathbb{P}\left(\lambda_1\left(\frac{1}{n^{2/\alpha-\varepsilon}}A_1\right) \geq 4\sigma\right) \leq \frac{E(\lambda_1(1/(n^{2/\alpha-\varepsilon})A_1)^{2s_n})}{(4\sigma)^{2s_n}} \leq \frac{\mathbb{E}(\mathrm{Tr}(A_1/(n^{2/\alpha-\varepsilon}))^{2s_n})}{(4\sigma)^{2s_n}}$$
$$\leq \exp(-\eta s_n)$$

for some constant $\eta > 0$, proving that $\lambda_1(\frac{1}{n^{2/\alpha-\varepsilon}}A_1)$ is bounded in probability. By symmetry, one gets the same result for the smallest eigenvalue.

Proof of Proposition 12. To estimate $\mathbb{E}(\mathrm{Tr}(\frac{A_1}{n^{2/\alpha-\varepsilon}})^{2s_n})$, we use the moment method. Developing the expectation, we have that

$$\mathbb{E}\big(\mathrm{Tr}(A_1)^{2s_n}\big) = \sum_{\mathcal{P}} \mathbb{E}\hat{a}_{i_0i_1}\hat{a}_{i_1i_2}\hat{a}_{i_2i_3}\hat{a}_{i_3i_4}\hat{a}_{i_4i_5}\cdots\hat{a}_{i_{2s_n-2}i_{2s_n-1}}\hat{a}_{i_{2s_n-1}i_0}, \tag{35}$$

where $\hat{a}_{ij} = a_{ij}\mathbf{1}_{\{|a_{ij}|\leq n^\beta\}}$ and $\mathcal{P}$ denotes the set of all closed paths $P = \{i_0, i_1, \ldots, i_{2s_n-1}, i_0\}$ with a distinguished origin, in the set $\{1, 2, \ldots, n\}$.

The following two lemmas are a direct consequence of [8], Chapter VIII.9, Theorem 2.23.

Lemma 13. *Let $C_n = \mathbb{E}(\hat{a}_{ij}) = \mathbb{E}(a_{ij}\mathbf{1}_{|a_{ij}|\leq n^\beta})$. Then for n large enough, $|C_n| \leq L(n^\beta)n^{\beta(1-\alpha)}$ where L is defined in (1).*

Lemma 14. *For any $k \geq 2$, let $D_n^{2k} = \mathbb{E}(\hat{a}_{ij}^{2k}) = \mathbb{E}(a_{ij}^{2k}\mathbf{1}_{|a_{ij}|\leq n^\beta})$. Then for n large enough, there exists a slowly varying function l_0 such that $D_n^{2k} \leq l_0(n^\beta)n^{\beta(2k-\alpha)}$.*

Thus it follows that

$$\big|\lambda_1(A_1 - \mathbb{E}A_1) - \lambda_1(A_1)\big| \leq |C_n|n \leq L\big(n^\beta\big)n^{\beta(1-\alpha)+1}, \tag{36}$$

so one can write

$$\lambda_1\left(\frac{1}{b_n}A_n\right) \leq \lambda_1\left(\frac{1}{b_n}(A_1 - \mathbb{E}A_1)\right) + \lambda_1\left(\frac{1}{b_n}(A_2)\right) + \frac{1}{b_n}\big|\lambda_1(A_1 - \mathbb{E}A_1) - \lambda_1(A_1)\big|, \tag{37}$$

where the last term tends to zero as n tends to infinity since $\beta(1-\alpha) + 1 - \frac{2}{\alpha} < 0$. Thus, we may assume that A_1 is also centered, i.e., the truncated variables $\hat{a}_{ij}$ are centered.

Now we move back to Eq. (35), to compute the expected value of the trace of $\frac{A_1^{2s_n}}{n^{(4/\alpha-2\varepsilon)s_n}}$. The first step after the centering is to consider the contribution of even paths, i.e. paths such that each edge occurs an even number of times.

We refer to the paper of Soshnikov [17] for most of the details and further notation. To each path $P = i_o \to i_1 \to i_2 \to \cdots \to i_{2s_{n-1}} \to i_0$, we first associate a set of s_n "marked instants" as follows. We read the edges of P successively. The instant at which an edge $i \to j$ is read is then said to be marked if up to that moment (inclusive) the edge (i, j) was read an odd number of times. Other instants are said to be unmarked. Now, the number of possible arrangements of marked/unmarked instants in a path of length $2s_n$ is equal to the number of Dick paths, i.e., the number of simple random walks of length $2s_n$, starting and ending at 0, and conditioned to remain in the positive quadrant. The number of Dyck paths is known to be the Catalan number: $\frac{(2s_n)!}{s_n!(s_n+1)!}$.

We say that a vertex is marked if it occurs at a marked instant. For any $0 \leq k \leq s_n$, we then define N_k to be the subset of vertices in $\{1, \ldots, n\}$ occurring k times as a marked vertex. Any vertex belonging to N_k is said to be a vertex

of self-intersection of type k. For any $0 \le k \le s_n$, we denote by n_k the cardinality of N_k and call $(n_o, n_1, \dots, n_{s_n})$ the type of P. Note that all the vertices that appear in P belong to $N_k, k \ge 1$, except possibly the origin i_o.

Then, it is easy to see that $\{1, \dots, n\}$ splits as the disjoint union of the sets $N_o, N_1, \dots, N_{s_n}$. From these definitions, one can see that

$$\sum_{k=0}^{s_n} n_k = n \quad \text{and} \quad \sum_{k=0}^{s_n} k n_k = s_n. \tag{38}$$

To estimate the number of possible paths and their contribution to the expectation, we proceed as follows. We first determine the set of marked instants and the type of the path. Then, we assign labels chosen in $\{1, \dots, n\}$ to each marked instant and to the origin of the path. Finally, we assign labels to each unmarked instant and consider the expectation of the corresponding path.

Given the set of marked instants and the type of the path $(n_o, n_1, \dots, n_{s_n})$, one has exactly $\frac{s_n!}{\prod_{k=2}^{s_n}(k!)^{n_k}}$ ways to distribute the marked instants into the possible classes of self-intersection. The number of ways to distribute the vertices of $\{1, \dots, n\}$ into the set of possible classes $N_o, N_1, \dots, N_{s_n}$ and determine the origin of the path is at most $\frac{n!}{n_0! n_1! \cdots n_{s_n}!} n$. This is because the origin is in general a non-marked vertex. There now remains to give an upper bound on the number of ways to determine vertices at unmarked instants, that is fill in the blanks of the path. It was proved in [17] that the number of ways to assign labels at unmarked instants is not greater than $\prod_{k=2}^{s_n}(2k)^{kn_k}$. Indeed, the number of possible ways to determine the right endpoint of an edge starting from a vertex of type k at an unmarked instant is at most $2k$.

To consider the expectation of a path P of type $(n_o, n_1, \dots, n_{s_n})$, we will need the following lemma.

Lemma 15. *Consider an even path of type* $(n_0, n_1, \dots, n_{s_n})$. *One has*

$$\mathbb{E}(\hat{a}_{i_0 i_1}\hat{a}_{i_1 i_2}\hat{a}_{i_2 i_3}\hat{a}_{i_3 i_4}\hat{a}_{i_4 i_5}\cdots\hat{a}_{i_{2s_n-2} i_{2s_n-1}}\hat{a}_{i_{2s_n-1} i_0}) \le \sigma^{2s_n}\prod_{i\ge 2}\left(l_o\left(n^\beta\right)^i n^{\beta(2i-(\alpha/2-1))}\right)^{n_i}. \tag{39}$$

Proof. Assume that a non-oriented edge (ij) is seen $2l(ij)$ times. We denote by $l(i;ij)$ (resp. $l(j;ij)$) the number of times i (resp. j) is a marked vertex in (ij). We also set $L(ij) = \max\{l(i;ij), l(j;ij)\}$ and $L'(ij) = \min\{l(i;ij), l(j;ij)\}$.

First, using Lemma 14, we deduce that

$$\begin{aligned}\prod_{(ij):l(ij)>1}\mathbb{E}\hat{a}_{ij}^{2l(ij)} &\le \prod_{(ij):l(ij)>1} l_o\left(n^\beta\right)n^{\beta(2l(ij)-\alpha)}\\ &\le \prod_{(ij):l(ij)>1} l_o\left(n^\beta\right)n^{\beta(2L(ij)+(2-\alpha)/2+2L'(ij)-2+(2-\alpha)/2)}.\end{aligned} \tag{40}$$

Second, we change the product in (40) over all edges to a product over all vertices. In fact, one can associate to each edge occurring 4 times at least one marked occurrence of a vertex of self-intersection. To deal with vertices where $L'(ij) = L(ij) = 1$, we say that $L'(ij)$ is associated to the vertex which has the smallest multiplicity in the path. Using the fact that the number of marked occurrences of any vertex in edges seen at least 4 times cannot exceed the type of the vertex and (38), we get:

$$\begin{aligned}\prod_{(ij):l(ij)>1}\mathbb{E}\hat{a}_{ij}^{2l(ij)} &\le \prod_{(ij):l(ij)>1} l_o\left(n^\beta\right)n^{\beta(2l(ij)-\alpha)}\\ &\le n^{\beta\sum_{(ij):l(ij)>1}(2L(ij)+(2-\alpha)/2)+\beta\sum_{(ij):l(ij)>2}(2L'(ij)-2+(2-\alpha)/2)}\prod_{k\ge 2} l_o\left(n^\beta\right)^{kn_k}\\ &\le \prod_{i\ge 2}\left(l_o\left(n^\beta\right)^i n^{\beta(2i-(\alpha-2)/2)}\right)^{n_i}.\end{aligned} \tag{41}$$

□

Remark 3. *Lemma 15 plays the role of formula 4.7 in [13] to bound the contribution of a single path of type $(n_0, n_1, \dots, n_{s_n})$. Here is where we cannot understand the arguments of [13]. Indeed, with the notation of [13], we believe that formula 4.7,*

$$\left(\mathbb{E}\prod_{u=0}^{2s_n-1}\xi_{i_u i_{u+1}}\Big|\Omega_{1-\varepsilon_n}\right)W_n \le \frac{1}{4^{s_n}}4^r\prod_{k=3}^{p/2}(2kC)^{kn_k}\prod_{k=p/2}^{s_n}\left(2k\Lambda_n^2\right)^{kn_k},$$

should be written as

$$\left(\mathbb{E}\prod_{u=0}^{2s_n-1}\xi_{i_u i_{u+1}}\Big|\Omega_{1-\varepsilon_n}\right)W_n \le \frac{1}{4^{s_n}}4^r\prod_{k=3}^{p/4}(2kC)^{kn_k}\prod_{k=p/4}^{s_n}\left(2k\Lambda_n^4\right)^{kn_k},$$

since, as argued in Lemma 15, an edge seen p times implies the occurrence of a marked vertex of type at least $p/4$ but not necessarily of type $p/2$. Mutatis mutandis, the other arguments of Ruzmaikina carry out to show that the universal Tracy–Widom limit holds if $\alpha > 36$.

We now call Z_e the contribution of all even paths of type $(n_o, n_1, \dots, n_{s_n})$ to the expectation $\mathbb{E}\operatorname{Tr}(\frac{A_1}{n^{2/\alpha-\varepsilon}})^{2s_n}$. Writing

$$\mathbb{E}\operatorname{Tr}\left(\frac{A_1}{n^{2/\alpha-\varepsilon}}\right)^{2s_n} = \mathbb{E}\operatorname{Tr}\left(\frac{A_1}{\sqrt{n}}\right)^{2s_n}\left(\frac{1}{n^{2/\alpha-1/2-\varepsilon}}\right)^{2s_n},$$

we deduce that Z_e is bounded by:

$$\begin{aligned}
Z_e &\le \frac{1}{n^{s_n(4/\alpha-2\varepsilon)}}\frac{n!}{n_0!n_1!\cdots n_{s_n}!}n\frac{(2s_n)!}{s_n!(s_n+1)!}\frac{s_n!}{\prod_{k=2}^{s_n}(k!)^{n_k}}\prod_{k=2}^{s_n}(2k)^{kn_k}\\
&\quad\times\sigma^{2s_n}\prod_{k=2}^{s_n}\left(l_o^k n^{\beta(2k-(\alpha/2-1))}\right)^{n_k}\\
&\le \frac{1}{n^{(4/\alpha-1-2\varepsilon)\sum_{i\ge1}in_i}}\frac{n\cdots(n_0+1)}{n^{s_n}}n\frac{(2s_n)!}{s_n!(s_n+1)!}\frac{1}{\prod_{k=2}^{s_n}n_k!}\frac{s_n^{s_n-n_1}}{\prod_{k=2}^{s_n}(k!)^{n_k}}\prod_{k=2}^{s_n}(2k)^{kn_k}\\
&\quad\times\sigma^{2s_n}\prod_{k=2}^{s_n}\left(l_o^k n^{\beta(2k-(\alpha/2-1))}\right)^{n_k}\\
&\le \frac{1}{n^{(4/\alpha-2\varepsilon-1)\sum_{i\ge1}in_i}}\frac{(2s_n)!}{s_n!(s_n+1)!}n\frac{1}{\prod_{k=2}^{s_n}n_k!}\frac{s_n^{s_n-n_1}}{n^{s_n+n_0-n}}\frac{1}{\prod_{k=2}^{s_n}(ke^{-1})^{kn_k}}\prod_{k=2}^{s_n}(2k)^{kn_k}\\
&\quad\times\sigma^{2s_n}\prod_{k=2}^{s_n}\left(l_o^k n^{\beta(2k-(\alpha/2-1))}\right)^{n_k}\\
&\le \frac{(2s_n)!}{s_n!(s_n+1)!}n\sigma^{2s_n}\prod_{k=2}^{s_n}\frac{1}{n_k!}\left[\frac{l_o^k s_n^k n^{\beta(2k-(\alpha/2-1))}}{n^{4k/\alpha-2k\varepsilon-1}}\right]^{n_k}\frac{1}{n^{n_1(4/\alpha-2\varepsilon-1)}}.
\end{aligned}\tag{42}$$

First, we consider the contribution of simple paths, that is paths of type $(n_o, n_1, 0, \dots, 0)$. We denote $Z_{e,s}$ this contribution. Then

$$Z_{e,s} \le \frac{(2s_n)!}{s_n!(s_n+1)!}n\sigma^{2s_n}\frac{1}{n^{s_n(4/\alpha-1-2\varepsilon)}}.$$

This follows from the fact that in simple even paths, any edge is seen twice and the choice of the origin and marked vertices determines the path.

We next turn to the contribution of paths with self-intersections, which we denote by $Z_{e,i}$. Now, if we take the sum over all non-negative integers $n_2, n_3, \dots, n_{s_n}$ such that

$$\sum_{k=2}^{s_n} n_k > 0, \tag{43}$$

we have that:

$$Z_{e,i} \le \frac{(2s_n)!}{s_n!(s_n+1)!} n\sigma^{2s_n} \sum_{n_2,\dots,n_{s_n}} \prod_{k=2}^{s_n} \frac{1}{n_k!} \left[\frac{l_o^k s_n^k n^{\beta(2k-(\alpha/2-1))}}{n^{4k/\alpha-2k\varepsilon-1}} \right]^{n_k}. \tag{44}$$

Now as $\varepsilon < 2/\alpha - \beta$ and $s_n n^{2\beta} \ll n^{4/\alpha}$, we deduce that

$$\begin{aligned}
(44) &\le \frac{(2s_n)!}{s_n!(s_n+1)!} n\sigma^{2s_n} \sum_{M>0} \frac{1}{M!} \left[\frac{C s_n^2 n^{\beta(4-(\alpha/2-1))}}{n^{8/\alpha-4\varepsilon-1}} \right]^M \\
&\le C \frac{(2s_n)!}{s_n!(s_n+1)!} n\sigma^{2s_n} \frac{l_o^2 s_n^2 n^{\beta(5-\alpha/2))}}{n^{8/\alpha-4\varepsilon-1}} \\
&\le C \frac{(2s_n)!}{s_n!(s_n+1)!} n\sigma^{2s_n} l_o^2 s_n^2 n^{\beta(5-\alpha/2)+1-8/\alpha+4\varepsilon} \\
&= \mathrm{o}(1) \frac{(2s_n)!}{s_n!(s_n+1)!} n\sigma^{2s_n}.
\end{aligned} \tag{45}$$

In the last line, we have used the fact that $s_n \ll n^{(8/\alpha-1-\beta(5-\alpha/2))/8}$.

Now, we will see that this is also true for a path such that an edge occurs an odd number of times, proving Proposition 12. The necessary tools are the gluing and the insertion procedures developed in [11]. We refer to this article for details and notation.

In [11], one can prove that given a path P of length $2s_n$ with $2l$ non-returned edges, it is possible to construct a sequence $(P_0, P_1, P_2, \dots, P_J)$, $1 \le J \le 2l$, of subpaths of P such that the concatenated path $W = \bigcup_{i=0}^{J} P_i$, i.e. the path defined as if we read P_i in order, has length $2s_n - 2l$ and belongs to one of the following classes:

A W is a closed even path.
B W is a sequence of $I \le 2$ closed even paths where each origin is a marked vertex of P.
C W is a sequence of $I \le 2$ paths where each origin is a marked vertex of P and the union of these paths has only even edges.

The surgery in P consists only to remove the last occurrence of odd edges and reorder the remaining subpaths with the possibility of choosing the direction in which each subpath is read. To estimate the contribution of odd paths, one can reverse the above procedure, defining an onto map from paths of classes A, B, C to the set of odd paths. This was done in [11] and we concisely describe the method.

In case A, the simplest one, given a closed path W of length $2s_n - 2l$ one needs only to choose J vertices to split W, choose the order and direction of each subpath and how to assign and insert the $2l$ unreturned edges. Also, since $\hat{a}_{ij}$ are bounded by n^β, adding these $2l$ repetitions we multiply the contribution of the original path at most by $n^{2l\beta}$. Briefly, the contribution of paths with $2l$ odd edges such that $W(P) \in A$ can be bounded by:

$$C_1 \frac{(2s_n-2l)!}{(s_n-l)!(s_n-l+1)!} n\sigma^{2(s_n-l)} \sum_{J=1}^{2l} \binom{2s_n-2l}{J} J!2^J \binom{2l}{J} \frac{(2s_n-2l)!}{(2s_n-4l+J)!} \left(\frac{n^\beta}{n^{2/\alpha-\varepsilon}} \right)^{2l}, \tag{46}$$

which can be bounded by:

$$C_2 \frac{(2s_n - 2l)!}{(s_n - l)!(s_n - l + 1)!} n\sigma^{2(s_n - l)} (C_3 s_n)^{2l} \left(\frac{n^\beta}{n^{2/\alpha - \varepsilon}} \right)^{2l}. \tag{47}$$

This is less than or equal to:

$$C_2 \frac{(2s_n - 2l)!}{(s_n - l)!(s_n - l + 1)!} n\sigma^{2s_n} \left(\frac{C_3 s_n n^\beta}{n^{2/\alpha - \varepsilon}} \right)^{2l}. \tag{48}$$

The summation over l of the above gives a contribution which is negligible with respect to $(2\sigma)^{2s_n} n s_n^{-3/2}$ since $s_n^2 \ll n^{\frac{2}{\alpha} - \varepsilon - \beta}$.

Also, [11] gives us the following estimate for the contribution of odd paths coming from the class B:

$$\sum_{l=1}^{s_n - 1} \sum_{J=1}^{2l} C^{2l} 4^J J! \binom{2l}{J} \binom{2s_n - 2l}{J} \frac{(2s_n - 2l)!}{(s_n - l)!(s_n - l + 1)!} n\sigma^{2(s_n - l)} \left(\frac{n^\beta}{n^{2/\alpha - \varepsilon}} \right)^{2l}. \tag{49}$$

Again, proceeding as before there is a constant $K > 0$ such that (49) divided by the contribution of even paths can be bounded by

$$\sum_{l=1}^{s_n - 1} \left(K s_n n^{\beta - 2/\alpha + \varepsilon} \right)^{2l}, \tag{50}$$

which tends to 0 as n goes to infinity. As shown in [11], with a little bit of effort, the counting in case C can be reduced to the one in case A or B, which finishes the proof of the proposition.

Remark 4 (Case $\alpha = 2$). *When $\alpha = 2$, we do not use Proposition 12 as it is written to bound the largest eigenvalue of the truncated matrix. In fact, since one knows that, see [8],*

$$f(x) = \mathbb{E}\left(a_{ij}^2 \mathbf{1}_{\{|a_{ij}| < x\}} \right)$$

is a slowly varying function, we can show that for any $0 < \delta < \varepsilon$

$$\mathbb{P}\left(\lambda_1 \left(\frac{A_1}{n^{2/\alpha - \varepsilon + \delta}} \right) \geq 4 \right) \to 0 \quad (n \to \infty).$$

□

4.3. *The largest eigenvalue of A_2*

In this subsection we will prove that the largest eigenvalue of A_2 is actually asymptotically given by its largest entry in absolute value. For short, we denote by $\hat{A}_{ij}$, $i, j = 1, \ldots, n$, the entries of the matrix A_2 and by $\hat{A}_{i_1 j_1}$ the largest one in absolute value. The aim of this subsection is to prove:

Proposition 16. *One has that for any $\varepsilon > 0$*

$$\mathbb{P}\left(\left| \frac{\lambda_1(A_2)}{|\hat{A}_{i_1 j_1}|} - 1 \right| > \varepsilon \right) \to 0 \quad \textit{as } n \to \infty.$$

Proposition 16 will be enough to end the proof of the Corollary 1 as we will explain now. First, we point that since the matrices that we are dealing are symmetric, $\|A\| = \max\{|\lambda_1(A)|, |\lambda_n(A)|\}$ is exactly the operator norm of the matrix. Then triangular inequality implies:

$$\lambda_{\max}\left(\frac{1}{b_n} A_2 \right) + \lambda_{\min}\left(\frac{1}{b_n} A_1 \right) \leq \lambda_1\left(\frac{1}{b_n} A_n \right) \leq \lambda_{\max}\left(\frac{1}{b_n} A_1 \right) + \lambda_{\max}\left(\frac{1}{b_n} A_2 \right). \tag{51}$$

Now, since $\frac{n^{2/\alpha-\varepsilon}}{b_n}$ goes to 0 as n goes to infinity, the boundedness in probability of the largest eigenvalue of $\frac{1}{n^{2/\alpha-\varepsilon}}A_1$ as proved in the last subsection implies that $\lambda_{\max}(\frac{1}{b_n}A_1)$ and $\lambda_{\min}(\frac{1}{b_n}A_1)$ go to 0 in probability. Then, by Proposition 16, we have that the largest eigenvalue of A_n behave just as the largest eigenvalue of A_2, i.e.,

$$\lim_{n\to\infty}\mathbb{P}\left(\lambda_{\max}\left(\frac{1}{b_n}A_n\right)\le x\right)=\lim_{n\to\infty}\mathbb{P}\left(\lambda_{\max}\left(\frac{1}{b_n}A_2\right)\le x\right)=\lim_{n\to\infty}\mathbb{P}\left(\frac{a_{i_1j_1}}{b_n}\le x\right)=\exp\bigl(-x^{-\alpha}\bigr). \tag{52}$$

Proof of Proposition 16. The proof of Proposition 16 relies on the two following lemmas, repeating the arguments of Lemma 5.

Lemma 17. *Let $\varepsilon>0$ be fixed. With probability going to one, one has that*

(a) *There are no diagonal entry $\hat{A}_{ii}$ greater in absolute value than $b_n^{1/2+\varepsilon}$.*
(b) *There is no pair (i,j) such that $|\hat{A}_{ij}|\ge b_n^{99/100}$ and $|\hat{A}_{ii}|+|\hat{A}_{jj}|\ge b_n^{1/10}$.*
(c) *For any $\delta>0$, there is no row that has two entries greater than $b_n^{3/4+\delta}$.*

Proof. (a) follows from basics results about extremes of n independent variables. For (b), one recalls (12) to compute

$$\begin{aligned}\mathbb{P}\bigl(\exists(i,j)\|\hat{A}_{ij}|\ge b_n^{99/100}\text{ and }|\hat{A}_{ii}|+|\hat{A}_{jj}|\ge b_n^{1/10}\bigr)&\le 2\binom{n}{2}\bar{F}\bigl(b_n^{99/100}\bigr)\bar{F}\bigl(b_n^{1/10}\bigr)\\&\le n^2L\bigl(b_n^{99/100}\bigr)L\bigl(b_n^{1/10}\bigr)\bigl(b_n^{-109\alpha/100}\bigr)\\&=\mathrm{o}\bigl(n^{-18/100+\theta}\bigr)\end{aligned}$$

for some $\theta>0$ small enough.

Similarly, another application of (12) yields

$$\begin{aligned}&\mathbb{P}\bigl(\exists i\le n,\exists j,k\le n,\ j\ne k,\ \text{s.t. }|\hat{A}_{ij}|\ge b_n^{3/4+\delta}\text{ and }|\hat{A}_{ik}|\ge b_n^{3/4+\delta}\bigr)\\&\quad\le n^3\bar{F}\bigl(b_n^{3/4+\delta}\bigr)\bar{F}\bigl(b_n^{3/4+\delta}\bigr)\\&\quad\le n^3L\bigl(b_n^{3/4+\delta}\bigr)^2\frac{1}{b_n^{3\alpha/2+2\delta\alpha}}\\&\quad=\mathrm{o}\bigl(n^{-4\delta+\theta}\bigr)\end{aligned}$$

for some $\theta>0$ small enough. In the last line we have used that for any $\delta>0$, $b_n^{3\alpha/2+2\delta\alpha}\gg n^3$. □

We now show that the largest entry $\hat{A}_{i_1j_1}$ determines the largest eigenvalue of A_2. The idea is similar to the first part with minor changes. Introduce the vector

$$v_1=\frac{1}{\sqrt{2}}(e_{i_1}\pm e_{j_1}),$$

where the sign $\pm$ is determined by the following rule: $\pm e_{j_1}=+e_{j_1}$ if $\hat{A}_{i_1j_1}\ge 0$ and $-e_{j_1}$ otherwise. Then, with probability going to one,

$$\langle A_2v_1,v_1\rangle=\frac{1}{2}\hat{A}_{i_1i_1}+\frac{1}{2}\hat{A}_{j_1j_1}+|\hat{A}_{i_1j_1}|=|\hat{A}_{i_1j_1}|\bigl(1+\mathrm{o}(1)\bigr),$$

in view of item (b) of the preceding lemma and the fact that $\hat{A}_{j_1j_1}$ is the largest entry of the matrix A_2. Thus, again with probability going to one,

$$\lambda_{\max}(A_2)\ge|\hat{A}_{i_1j_1}|\bigl(1+\mathrm{o}(1)\bigr).$$

We now turn to the upper bound which follows from the following lemma.

Lemma 18. *One has that*

$$\lim_{n\to\infty} \mathbb{P}\left(\exists i, 1 \le i \le n, \max_{1\le j\le n} |\hat{A}_{ij}| > b_n^{3/4+\alpha/16} \text{ and } \sum_{j=1}^{n} |\hat{A}_{ij}| - \max_{1\le j\le n} |\hat{A}_{ij}| > b_n^{3/4+\alpha/16}\right) = 0.$$

Proof. Recall that by (33) we have $\alpha\beta > 1$.

Now, a trivial union bound gives us that

$$\begin{aligned}\mathbb{P}\left(\exists i, \sum_{j:n^\beta<|\hat{A}_{ij}|<b_n^{3/4+\alpha/32}} |\hat{A}_{ij}| \ge \frac{1}{2} b_n^{3/4+\alpha/16}\right) &\le n\mathbb{P}\left(\#\left\{j\colon n^\beta < |\hat{A}_{ij}| < b_n^{3/4+\alpha/32}\right\} \ge \frac{b_n^{\alpha/32}}{2}\right) \\ &\le n^{1+b_n^{\alpha/32}} \bar{F}\left(n^\beta\right)^{b_n^{\alpha/32}} \\ &\le n^{1+b_n^{\alpha/32}(1-\alpha\beta)} \le \exp\left(-n^\kappa\right) \end{aligned} \tag{53}$$

for some sufficiently small $\kappa > 0$. Thus, (53) together with part (c) of Lemma 17 yields Lemma 18. □

Now, combining Lemma 18 with (c) of Lemma 17 and repeating the same argument done in last section to prove part (b) in Lemma 5, Proposition 16 holds. □

Proof of Theorem 1 when $2 < \alpha < 4$. By Proposition 12, with probability going to 1, we know that the spectrum of $\frac{1}{\sqrt{n^{2/\alpha-\varepsilon}}} A_1$ is bounded. Also, let $A_{i_l j_l}$ be the l-th largest entry in absolute value of A_2. If one sets

$$v_l = e_{i_l} \pm e_{j_l},$$

where $e_{i_l} \pm e_{j_l} = e_{i_l} + e_{j_l}$ if $A_{i_l j_l} \ge 0$ and $e_{i_l} - e_{j_l}$ otherwise, then

$$A_2 v_l = \sum_{i=1}^{n} (A_{ii_l} \pm A_{ij_l}) e_i = |A_{i_l j_l}| v_l + r,$$

where

$$r = A_{i_l i_l} e_{i_l} \pm A_{j_l j_l} e_{j_l} + \sum_{i=1, i\ne i_l, j_l}^{n} (A_{ii_l} \pm A_{ij_l}) e_i.$$

By the same arguments that we used in Theorem 2, one can show that $\frac{\|r\|}{b_n}$ tends to 0 as n tends to infinity. This implies an existence of an eigenvalue λ of A_2 such that $\lambda = |A_{i_l j_l}|(1 + o(1))$. An application of Cauchy Interlacing theorem just as before shows that in fact $\lambda = \lambda_l(A_2)$, where $\lambda_l(A_2)$ represents the l-th eigenvalue of A_2 in descending order. Since by Weyl's inequalities, one has for all l , $1 \le l \le n$,

$$\lambda_l(A_2) + \lambda_n(A_1) \le \lambda_l(A_1 + A_2) \le \lambda_l(A_2) + \lambda_1(A_1). \tag{54}$$

Dividing Eq. (54) by b_n^2 and taking the limit one gets

$$\lim_{n\to\infty} \mathbb{P}\left(\lambda_l(A_1 + A_2) \le b_n x\right) = \lim_{n\to\infty} \mathbb{P}\left(\lambda_l(A_2) \le b_n x\right) = \lim_{n\to\infty} \mathbb{P}(a_{i_l j_l} \le b_n x),$$

which ends the proof of (b). □

5. Sample covariance matrices when $2 \le \alpha < 4$

This section heavily uses the previous ones. We repeat the arguments of Section 3 using the results of Section 4. As before, the first step is to truncate our matrix.

Define

$$A_1 = (\hat{A}_{ij}\mathbf{1}_{|\hat{A}_{ij}|\le n^{\beta}})_{i,j}, \qquad A_2 = A_n - A_1, \tag{55}$$

and also

$$X_1 = A_1 A_1^t, \qquad X_2 = X_n - X_1. \tag{56}$$

It is clear that $X_2 = A_1 A_2^t + A_2 A_1^t + A_2 A_2^t$ and that the largest entries of A_n belong to A_2. We are going to study the eigenvalues of X_2. The following lemma follows directly from the proof of Lemma 5 and Lemma 18.

Lemma 19. *Let A_2 be defined as above. Also assume that $p = \lfloor \gamma n \rfloor$ for some constant $\gamma \ge 1$. Then*:

(a) *If B_n^{δ} is the event 'There is a row with 2 entries greater than b_{np}^{δ} in absolute value' then*

$$\forall \delta > 3/4 \quad \lim_{p\to\infty} \mathbb{P}\big(B_n^{\delta}\big) = 0.$$

(b) *One has that*

$$\lim_{p\to\infty} \mathbb{P}\left(\exists i, 1 \le i \le n,\ \max_{1\le j\le p} |a_{ij}| > b_{np}^{3/4+\alpha/16} \text{ and } \sum_{j=1}^{p} |a_{ij}| - \max_{1\le j\le p} |a_{ij}| > b_{np}^{3/4+\alpha/16}\right) = 0.$$

(c) *One has that*

$$\lim_{p\to\infty} \mathbb{P}\left(\exists j, 1 \le j \le p,\ \max_{1\le i\le n} |a_{ij}| > b_{np}^{3/4+\alpha/16} \text{ and } \sum_{j=1}^{p} |a_{ij}| - \max_{1\le j\le p} |a_{ij}| > b_{np}^{3/4+\alpha/16}\right) = 0.$$

If one takes $v = (0, \dots, 0, 1, 0, \dots, 0)$ where the only non-zero entry of v is in the position i_1, $\langle X_2 v, v\rangle$ will give us the following bound:

$$\lambda_{\max}(X_2) \ge a_{i_1 j_1}^2 \big(1 + \mathrm{o}(1)\big). \tag{57}$$

This inequality is justified by the preceding lemma, just as in Theorem 2, and by the fact that the diagonal of $A_1 A_2^t + A_2 A_1^t$ has only zeros. More than that, as in Section 3, we can also infer from Lemma 19 that

$$\lambda_{\max}\big(A_2 A_2^t\big) = a_{i_1 j_1}^2 \big(1 + \mathrm{o}(1)\big). \tag{58}$$

Now, using Rayleigh–Ritz representation and linearity of the scalar product, we have that

$$\begin{aligned}\lambda_{\max}(X_n) &= \max_{v:|v|=1} \langle X_n v, v\rangle = \max_{v:|v|=1} \big\langle (A_1 + A_2)(A_1 + A_2)^t v, v\big\rangle \\ &= \max_{v:|v|=1} \big(\langle X_1 v, v\rangle + \big\langle A_2 A_2^t v, v\big\rangle + \big\langle A_1 A_2^t v, v\big\rangle + \big\langle A_2 A_1^t v, v\big\rangle\big)\end{aligned}$$

which, by Cauchy–Schwarz, yields

$$\begin{aligned}\lambda_{\max}(X_n) &\le \max_{v:|v|=1} \langle X_1 v, v\rangle + \max_{v:|v|=1} \big\langle A_2 A_2^t v, v\big\rangle + 2 \max_{v:|v|=1} \big\| A_1^t v\big\| \max_{v:|v|=1} \big\| A_2^t v\big\| \\ &\le \lambda_{\max}(X_1) + \lambda_{\max}\big(A_2 A_2^t\big) + 2\big(\lambda_{\max}(X_1)\lambda_{\max}\big(A_2 A_2^t\big)\big)^{1/2}.\end{aligned} \tag{59}$$

In view of (57)–(59) and Weyl's inequality for $X_n = X_1 + X_2$, namely,

$$\lambda_{\max}(X_2) + \lambda_{\min}(X_1) \le \lambda_{\max}(X_n),$$

it remains to show that the largest eigenvalue of X_1 is negligible with respect to $a_{i_1 j_1}^2$ to conclude that

$$\lambda_{\max}(X_n) = a_{i_1 j_1}^2 \big(1 + \mathrm{o}(1)\big)$$

and, therefore, finish the proof of Corollary 2.

Thus, we turn back our attention to the truncated matrix $X_1 = A_1 A_1^t$. As we did before, we will show that its largest eigenvalue properly normalized remains bounded on a set of probability arbitrarily close to 1. Again, we study the asymptotics of some expectations of X_1.

By the results of the last section, we just need to control the expected value of some traces of the matrix X_1.

Let s_n be as in Section 4.2. One can write

$$\mathbb{E}\big(\operatorname{Tr} X_1^{s_n}\big) = \sum_{\mathcal{P}} \mathbb{E} a_{i_1 i_0} a_{i_1 i_2} a_{i_3 i_2} a_{i_3 i_4} \cdots a_{i_{2s_n-1} i_{2s_n-2}} a_{i_{2s_n-1} i_0}, \tag{60}$$

where $\mathcal{P}$ denotes the set of all closed paths $P = \{i_0, i_1, \dots, i_{2s_n-1}, i_0\}$ with a distinguished origin, in the set $\{1, 2, \dots, p\}$ with the restriction $i_t \in \{1, 2, \dots, n\}$ for odd t. Now, since $a_{ij} \ne a_{ji}$ orientation of an edge plays a role.

We say that a path is odd if the number of passages in the direction $i \to j$ plus the number of passages in the direction $j \to i$ is odd for some i and j. P is even if it is not odd. Since we can center the random variables $a_{ij} \mathbf{1}_{|a_{ij}| \le n^\beta}$ just as in Section 4, odd paths with a non-zero contribution have at least 3 passages in one direction of an odd edge. We will prove that the contribution of odd paths to the sum (60) is negligible and that contribution of even paths can be easily bounded by the results on the Wigner case. To do so, one can proceed as follows.

Construct a $p \times p$ random symmetric matrix $M = (y_{ij})_{1 \le i, j \le p}$ such that y_{ij} are independent identically distributed random variables with the same distribution as a_{11}. Hence, if we denote $\mathbb{E}(P)$ as the contribution of the path,

$$\sum_{P \in \mathcal{P},\, P_{\text{even}}} \mathbb{E}(P) \le \mathbb{E}\big(\operatorname{Tr} M^{2s_n}\big). \tag{61}$$

In fact, one has a 1–1 relation between paths in the LHS and paths that give a non-zero contribution to the sum in the RHS. Furthermore, if an edge (i, j) is read r times from left to right and s times in the opposite direction, then $p + s = 2q$ for some integer q and we can use the inequality

$$\mathbb{E} a_{ij}^r \mathbb{E} a_{ji}^s \le \mathbb{E}|a_{ij}|^r \mathbb{E}|a_{ji}|^s \le \mathbb{E} y_{ij}^{2q}, \tag{62}$$

leading to (61).

Therefore, proceeding as in Section 4, Eq. (61) implies:

Proposition 20. *There exists a constant $C > 0$ such that $\mathbb{E}(\operatorname{Tr}(\frac{X_1}{n^{4/\alpha - 2\varepsilon}})^{2s_n})) < (2\sigma)^{4s_n} \frac{Cn}{s_n^{3/2}}$. Moreover,*

$$\mathbb{P}\big(\lambda_{\max}(X_1) \ge 8\sigma^2 n^{4/\alpha - 2\varepsilon}\big) \le \frac{\mathbb{E}(\operatorname{Tr}(X_1/n^{4/(\alpha-2\varepsilon)})^{2s_n})}{(8\sigma^2)^{2s_n}} \le \exp(-\eta s_n),$$

for some constant $\eta > 0$.

To prove that the contribution of odd paths is negligible with respect to $(2\sigma)^{4s_n} \frac{Cn}{s_n^{3/2}}$, we can still bound the contribution of each path using the inequality (62). Since to an odd path $\mathcal{P}$, there corresponds an odd path in the expansion of $\mathbb{E}(\operatorname{Tr} M^{2s_n})$, we analyze their contribution as in Section 4.

$$\sum_{P \in \mathcal{P},\, P_{\text{odd}}} \mathbb{E}(P) \ll (2\sigma)^{4s_n} \frac{Cn}{s_n^{3/2}}. \tag{63}$$

This ends the proof Corollary 2 in the case of $2 \leq \alpha < 4$. The proof of Theorem 2 in this case is identical to the proof of the case $0 < \alpha < 2$.

Acknowledgments

The authors would like to thank two anonymous referees for the careful and meticulous reading of the paper. Their comments and corrections helped in making this paper more coherent and error free.

References

[1] Z. D. Bai, P. R. Krishnaiah and Y. Q. Yin. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** (1988) 509–521. MR0950344

[2] Z. D. Bai and Y. Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.* **16** (1988) 1729–1741. MR0958213

[3] S. Belinschi, A. Dembo and A. Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Commun. Math. Phys.* (2009). To appear.

[4] G. Ben Arous and A. Guionnet. The spectrum of heavy tailed random matrices. *Comm. Math. Phys.* **278** (2008) 715–751. MR2373441

[5] R. Bhatia. *Matrix Analysis.* Springer, New York, 1996. MR1477662

[6] N. H. Bingham, C. M. Goldie and J. L. Teugels. *Regular Variation*. Cambridge Univ. Press, Cambridge, 1987. MR0898871

[7] G. Biroli, J. P. Bouchaud and M. Potters. On the top eigenvalue of heavy-tailed random matrices. *Europhys. Lett.* **78** (2007) 10001. MR2371333

[8] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. II.* Wiley, New York, 1966. MR0210154

[9] O. Kallenberg. *Foundations of Modern Probability.* Springer, New York, 2001. MR1876169

[10] V. Marchenko and L. Pastur. The distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* **72** (1967) 507–536. MR0208649

[11] S. Péché and A. Soshnikov. Wigner random matrices with non-symmetrically distributed entries. *J. Stat. Phys.* **129** (2007) 857–884. MR2363385

[12] R. Resnick. *Extreme Values, Regular Variation and Point Processes* **4**. Springer, New York, 1987. MR0900810

[13] A. Ruzmaikina. Universality of the edge distribution of eigenvalues of Wigner random matrices with polynomially decaying distributions of entries. *Comm. Math. Phys.* **261** (2006) 277–296. MR2191882

[14] A. Soshnikov. A note on universality of the distribution of largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** (2002) 1033–1056. MR1933444

[15] A. Soshnikov. Poisson statistics for the largest eigenvalues in random matrix ensembles. In *Mathematical Physics of Quantum Mechanics* 351–364. *Lecture Notes in Phys.* **690**. Springer, Berlin, 2006. MR2234922

[16] A. Soshnikov. Poisson statistics for the largest eigenvalue of Wigner random matrices with heavy tails. *Electron. Comm. Probab.* **9** (2004) 82–91. MR2081462

[17] A. Soshnikov. Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** (1999) 697–733. MR1727234