Localization and delocalization for heavy tailed band matrices

Florent Benaych-Georges, Sandrine Péché

▶ **To cite this version:**

Florent Benaych-Georges, Sandrine Péché. Localization and delocalization for heavy tailed band matrices. 2012. hal-00746679v1

HAL Id: hal-00746679

https://hal.science/hal-00746679v1

Preprint submitted on 29 Oct 2012 (v1), last revised 24 Jun 2015 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LOCALIZATION AND DELOCALIZATION FOR HEAVY TAILED BAND MATRICES

FLORENT BENAYCH-GEORGES AND SANDRINE PÉCHÉ

Abstract. We consider some random band matrices with band-width N^μ whose entries are independent random variables with distribution tail in $x^{-\alpha}$. We consider the largest eigenvalues and the associated eigenvectors and prove the following phase transition. On the one hand, when $\alpha < 2(1+\mu^{-1})$, the largest eigenvalues have order $N^{\frac{1+\mu}{\alpha}}$, are asymptotically distributed as a Poisson process and their associated eigenvectors are essentially carried by two coordinates (this phenomenon has already been remarked by Soshnikov in [26, 27] for full matrices with heavy tailed entries, *i.e.* when $\alpha < 2$, and by Auffinger *et al* in [1] when $\alpha < 4$). On the other hand, when $\alpha > 2(1+\mu^{-1})$, the largest eigenvalues have order $N^{\frac{\mu}{2}}$ and most eigenvectors of the matrix are delocalized, *i.e.* approximately uniformly distributed on their N coordinates.

Introduction

Recently some growing interest has been laid on the understanding of the asymptotic behavior of both eigenvalues and eigenvectors of random matrices in the large size limit. For Wigner random matrices, that is $N \times N$ Hermitian or real symmetric random matrices with i.i.d. entries (modulo the symmetry assumption), the large-N-asymptotic behavior is now well understood, provided the distribution of the entries has sub-exponential decay (or at least a large enough number of moments). It is indeed known from the works of Erdös-Schlein-Yau, Tao-Vu and Knowles-Yin ([12, 13, 14, 29, 30, 17] and see also references therein) that :

– eigenvalues are very close to their theoretical prediction given by well-chosen quantiles of the semi-circle distribution (the proof is based on a *strong semi-circle law*). This also yields universality of local statistics in the bulk and at the edge under some appropriate moment assumptions (see Erdös [8] e.g. for a review of the recent results).

– eigenvectors are fully delocalized in the following sense. The *localization length*, L, of an eigenvector $\mathbf{v}$ is the typical number of coordinates bearing most of its ℓ^2 norm. Then it is proved that with very "high probability" there does not exist an eigenvector with

Date: October 29, 2012.

2000 *Mathematics Subject Classification.* 15A52;60F05.

Key words and phrases. Random matrices, band matrices, heavy tailed random variables.

This work was partly accomplished during the first named author's stay at New York University Abu Dhabi, Abu Dhabi (U.A.E.).

localization length $L \ll N$. Or roughly speaking all coordinates are in the order of $N^{-1/2}$.

In this article, we want to fill in the gap of understanding the role of moments in the delocalization properties of eigenvectors. We will be interested in a model of random matrices that we believe to be quite rich, namely random band matrices with heavy-tailed entries.

More precisely, the matrices under consideration in this paper are Hermitian random matrices with at most N^μ non zero entries per row. In other words, we force some of the entries of a Wigner matrix to be zero. This model is believed to be more complicated than Wigner ensembles due to the fact that there is no reference ensemble: there does not exist a "simple" band random matrix ensemble for which eigenvalue/eigenvector statistics can be explicitly computed as for the GUE/GOE in Wigner matrices. Thus usual comparison methods (four moments theorem, Green function comparison method) cannot be used directly in this setting.
Such a model is also believed to exhibit a phase transition, depending on μ. On a physical level of rigor, Fyodorov and Mirlin [16] e.g. have explained that for Gaussian entries, the localization length of a typical eigenvector in the bulk of the spectrum shall be of order $L = O(N^{2\mu})$ so that eigenvectors should be localized (resp. delocalized or extended) if $\mu < 1/2$ (resp. $> 1/2$). The only rigorous result in the direction of localization is by Schenker [22]. Therein it is proved that $L \ll N^{8\mu}$ for all eigenvectors of random band matrices with i.i.d. Gaussian entries on the band. On the other hand, delocalization in the bulk is proved by Erdös, Knowles, Yau and Yin [11] when $\mu > 4/5$. In both regimes, it is known from Erdös and Knowles [9, 10] that typically $L \geq N^{7\mu/6}$ for a certain class of random band matrices (with sub-exponential tails and symmetric distribution). We refer the reader to Spencer [28] and Erdös, Schlein and Yau [13] for a more detailed discussion on the localized/delocalized regime. Regarding the edges of the spectrum, much less is known about the typical localization length of the associated eigenvectors. The authors are not aware of a proof that eigenvectors at the edge are fully delocalized. However, Sodin's statement [24] combined with Erdos-Knowles-Yau-Yin'results [11] suggest that this should be true when $\mu > 5/6$.

We will also allow the i.i.d. non zero entries to admit only a finite number of moments (which can actually be zero). Allowing heavy-tailed entries allows some more localization, especially at the edge of the spectrum, as we can infer from Wigner matrices. This is discussed in particular in the seminal paper by Cizeau and Bouchaud [7]. It is known that the limiting spectral measure of such Wigner matrices is the semi-circle distribution provided that the variance of the entries is finite (otherwise another limiting distribution has been identified by Guionnet and Ben Arous [5]). Regarding eigenvectors, it was shown by Soshnikov [26, 27] and Auffinger, Ben Arous and Péché [1] that eigenvectors associated to the largest eigenvalues have a localization length of order 1 if the entries do not admit a finite fourth moment. The localization length is not so clear in the bulk but some progress has been obtained by Bordenave and Guionnet [6]. However it is commonly believed that

the fourth moment shall be a threshold for the localization of eigenvectors at the edge of the spectrum of full Wigner matrices. For band matrices when the bandwidth is negligible w.r.t. the size N of the matrix, no such threshold has been intuited. This is also a gap we intend to fill in here.

Specifically, we prove the following phase transition. On the one hand, when $\alpha < 2(1+\mu^{-1})$, the largest eigenvalues and associated eigenvectors are determined by the largest entries. Largest eigenvalues are in the order of $N^{\frac{1+\mu}{\alpha}}$ and asymptotically distributed as a Poisson process while the associated eigenvectors are essentially carried by two coordinates. This phenomenon has already been noted by Soshnikov in [26, 27] for full matrices ($\mu = 1$) when $\alpha < 2$, and by Auffinger *et al* in [1] when $\alpha < 4$. On the other hand, when $\alpha > 2(1 + \mu^{-1})$, so that $N^{\frac{1+\mu}{\alpha}} \ll N^{\mu/2}$, largest entries no longer play a role. Then the largest eigenvalues have order $N^{\frac{\mu}{2}}$ and most eigenvectors of the matrix are delocalized, *i.e.* approximately uniformly distributed on their N coordinates.

The paper is organized as follows. In Section 1, we state our two main theorems : Theorem 1.1 is the localization result mentioned above about the extreme eigenvalues and eigenvectors in the case $\alpha < 2(1 + \mu^{-1})$ and Theorem 1.5 is the delocalization result mentioned above about the extreme eigenvalues of the matrix and most of its eigenvectors in the case $\alpha > 2(1 + \mu^{-1})$. Sections 2, 3 and 4 are devoted to the proofs of these results and the appendix is devoted to the proof of several technical results, including Theorem 5.3, a general result whose idea goes back to papers of Soshnikov about the surprising phenomenon that certain Hermitian matrices have approximately equal largest eigenvalues and largest entries.

Notation. For any functions (or sequences) f, g, we write $f(x) \sim g(x)$ (resp. $f(x) \overset{sl.}{\sim} g(x)$) when $f(x)/g(x) \longrightarrow 1$ (resp. $f(x)/g(x)$ is *slowly varying*) as $x \to +\infty$. We denote by $\|\mathbf{v}\|$ the ℓ^2-norm of $\mathbf{v} \in \mathbb{C}^N$ and by $\|A\|$ the $\ell^2 \to \ell^2$ operator norm of a matrix A. When A is normal, then $\|A\| = \rho(A)$ where $\rho(A)$ is the spectral radius of A and we use equivalently both notations.

An event A_N depending on a parameter N is said to hold *with exponentially high probability* (abbreviated w.e.h.p. in the sequel) if $\mathbb{P}(A_N) \geq 1 - e^{-CN^\theta}$ for some $C, \theta > 0$.

1. The results

Let us fix two exponents $\alpha > 0$ and $\mu \in (0, 1]$ and let, for each N, $A_N = [a_{ij}]_{i,j=1}^N$ be a real symmetric (or complex Hermitian) random matrix such that for all i's in $\{1, \dots, N\}$ except possibly $o(N)$ of them,

$$\{j \,;\, a_{ij} \text{ is not almost surely zero}\} \sim N^\mu, \tag{1}$$

whose non almost surely null entries are i.i.d. modulo the symmetry assumption and such that for a certain $\alpha > 0$,

$$G(x) := \mathbb{P}(|a_{ij}| > x) \overset{sl.}{\sim} x^{-\alpha} \qquad \text{as } x \to \infty. \tag{2}$$

If $\alpha \geq 1+\mu^{-1}$, we also suppose that the a_{ij}'s are symmetrically distributed. This symmetry assumption simplifies the exposition of arguments but can be relaxed (we briefly indicate this possible extension in Remark 2.5 below). Note that for all fixed i, j, a_{ij} might depend on N (think for example of the case where A_N is a band matrix), hence should be denoted by $a_{ij}(N)$. However, we suppose that the estimate (2) is uniform in N and that if $\alpha > 2(1+\mu^{-1})$, the second moment of the non identically zero entries of A_N is equal to one.

The standard example of matrices satisfying (1) is given by *band matrices*, *i.e.* matrices with entries a_{ij} such that $a_{ij} = 0$ when $|i-j| > N^{\mu}/2$. Another very close example is the one of *cyclic band matrices*, *i.e.* matrices with entries a_{ij} such that $a_{ij} = 0$ when $|i-j| > N^{\mu}/2$ and $|i-j| > N - N^{\mu}/2$.

We denote by $\lambda_1 \geq \lambda_2 \geq \cdots$ the eigenvalues of A_N (they depend implicitly on N) and we choose some unit associated eigenvectors

$$\mathbf{v}_1, \mathbf{v}_2, \ldots$$

Let us also introduce a set of pairs of indices $(i_1 \leq j_1), (i_2 \leq j_2), \ldots$ such that for all k, $|a_{i_k j_k}|$ is the kth largest entry, in absolute value, of A_N. Let $\theta_k \in \mathbb{R}$ such that $a_{i_k j_k} = |a_{i_k j_k}| e^{2i\theta_k}$. The eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots$ are chosen such that for each k,

$$e^{-i\theta_k} \langle \mathbf{v}_k, \mathbf{e}_{i_k} \rangle \geq 0,$$

with $\mathbf{e}_1, \ldots, \mathbf{e}_N$ the vectors of the canonical basis.

As we shall see in the two following theorems, the asymptotic behavior of both the largest eigenvalues of A_N and their associated eigenvectors exhibit a phase transition with threshold $\alpha = 2(1+\mu^{-1})$.

Theorem 1.1 (Subcritical case)**.** *Let us suppose that $\alpha < 2(1+\mu^{-1})$. Then for each fixed $k \geq 1$, we have the convergences in probability, as $N \to \infty$,*

$$\frac{\lambda_k}{|a_{i_k j_k}|} \longrightarrow 1 \tag{3}$$

and

$$\mathbf{v}_k - \frac{1}{\sqrt{2}} \left(\mathbf{e}^{i\theta_k} e_{i_k} + \mathbf{e}^{-i\theta_k} e_{j_k} \right) \longrightarrow 0 \qquad \textit{(for the } \ell^2\textit{-norm).} \tag{4}$$

As a consequence of (3)*, for b_N the sequence defined by* (5) *below, the random point process*

$$\sum_{k\,;\,|a_{i_k j_k}|>0} \delta_{\lambda_k / b_N}$$

converges in law to the law of a Poisson point process on $(0, +\infty)$ with intensity measure $\frac{\alpha}{x^{\alpha+1}} \mathrm{d}x$.

The sequence b_N of the theorem is defined by

$$b_N := \inf\left\{ x \geq 0\,;\, G(x) \leq \frac{1}{\#\{\text{non identically zero independent entries of } A_N\}} \right\}, \tag{5}$$

where $G(x)$ is defined by (2). It can easily be deduced from (1) and (2) that

$$b_N \overset{sl.}{\sim} N^{\frac{1+\mu}{\alpha}}. \tag{6}$$

Roughly speaking, this theorem says that when $\alpha < 2(1+\mu^{-1})$, the largest eigenvalues of A_N have order $N^{\frac{1+\mu}{\alpha}}$, but no fixed limit when divided by $N^{\frac{1+\mu}{\alpha}}$, because the limiting object is a Poisson process. Moreover, the corresponding eigenvectors are essentially supported by two components. As we shall see in the following theorem, the case $\alpha > 2(1+\mu^{-1})$ is deeply different: in this case, the largest eigenvalues of A_N have order $N^{\frac{\mu}{2}}$ and tend to 2 when divided by $N^{\frac{\mu}{2}}$, whereas the eigenvectors are much more delocalized, *i.e.* supported by a large number of components.

To be more precise, we use the following Definition 7.1 from Erdös, Schlein and Yau [13].

Definition 1.2. *Let L be a positive integer and $\eta > 0$ be given. A unit vector $\mathbf{v} = (v_1, \ldots, v_N) \in \mathbb{C}^N$ is said to be* (L,η)-localized *if there exists a set $S \subset \{1, \ldots, N\}$ such that $|S| = L$ and $\sum_{j \in S^c} |v_j|^2 \leq \eta$.*

We shall also use the following slightly modified version of the above definition.

Definition 1.3. *Let L be a positive integer and $\eta > 0$ be given. A unit vector $\mathbf{v} = (v_1, \ldots, v_N) \in \mathbb{C}^N$ is said to be* (L,η)-successively localized *if there exists a set S which is an interval of the set $\{1, \ldots, N\}$ endowed with the cyclic order such that $|S| = L$ and $\sum_{j \in S^c} |v_j|^2 \leq \eta$.*

Remark 1.4. The larger L and η, the stronger a statement of the type *"There is non (L,η)-localized eigenvector"* is.

Theorem 1.5 (Supercritical case)**.** *Let us suppose that $\alpha > 2(1+\mu^{-1})$. Then for each fixed $k \geq 1$, as $N \to \infty$, we have the convergence in probability*

$$\frac{\lambda_k}{N^{\frac{\mu}{2}}} \longrightarrow 2. \tag{7}$$

Moreover, for $L := \lfloor N^c \rfloor$, with c such that

$$c < \frac{2}{5}\mu\frac{\alpha-2}{\alpha-1} \qquad (\text{resp. } c < \mu), \tag{8}$$

for any $\eta_0 < 1/2$, we have, as $N \to \infty$,

$$\mathbb{P}\left(\cup_{\eta < \eta_0} \{ \exists k, |\lambda_k| > \sqrt{2\eta}\rho(A) \text{ and } \mathbf{v}_k \text{ is } (L,\eta)\text{-localized} \} \right) \longrightarrow 0. \tag{9}$$

Remark 1.6. Note that this theorem does not only apply to the edges of the spectrum, as η runs from 0 to η_0 in (9).

Remark 1.7. Note that focusing on successively localized vectors, we would need to improve the bound $c < \mu$ in order to get some flavor of the usual threshold of the so-called *Anderson transition*. The localization length L of typical eigenvectors in the bulk is indeed supposed numerically to be in the order of $L \approx N^{2\mu}$ when $\mu < 1/2$ for entries with many moments. At the edge of the spectrum, the authors are not aware of any intuited (even at a physical level of rigor) localization length in the localized regime.

To prove both above theorems, we shall also use the following result, which had not appeared at this level of generality yet.

Theorem 1.8. *We suppose that the hypotheses* (1) *and* (2) *hold with $\alpha > 2$ and that the first and second moments of the non identically zero entries of A_N are respectively equal to* 0 *and* 1*. Then the empirical spectral measure of $A_N/N^{\frac{\mu}{2}}$ converges almost surely to the semi-circle law with support $[-2, 2]$.*

Proof. The proof relies on a classical cutt-off and moments method, copying the proof of the convergence to the semi-circle distribution for standard Wigner matrices (see for example [2, Th. 2.5]). □

2. A preliminary result: general upper-bound on the moments

Theorem 2.1. *Consider some positive exponents $\gamma, \gamma', \gamma''$ such that*

$$\frac{\mu}{2} \leq \gamma' \qquad \text{and} \qquad \frac{\mu}{4} + \gamma + \gamma'' < \gamma' \tag{10}$$

and define the truncated matrix $\hat{A}_N = [a_{ij} 1_{|a_{ij}| \leq N^{\gamma}}]_{i,j=1}^{N}$. Then for $s_N \leq N^{\gamma''}$, there exists a slowly varying function L such that

$$\mathbb{E}[\mathrm{Tr}(\hat{A}_N^{2s_N})] \leq L(N) N^{1+2\gamma} s_N^{-3/2} (2N^{\gamma'})^{2s_N}.$$

The following corollary follows directly from the theorem and from the Chebichev inequality.

Corollary 2.2. *For any $\kappa < 1$ (possibly depending on N), we have, up to a polynomial factor in the RHT,*

$$\mathbb{P}\left(\|\hat{A}_N\| \geq \kappa \times 2N^{\gamma'} \right) \leq \kappa^{-2s_N}. \tag{11}$$

Remark 2.3. Roughly speaking, this theorem says that for any $\epsilon > 0$,

$$\|\hat{A}_N\| \leq (2+\epsilon) N^{\max\{\frac{\mu}{2}, \frac{\mu}{4}+\gamma+\epsilon\}} \qquad \text{for } N \gg 1.$$

Remark 2.4. Note that for the theorem and its corollary to be true, one does not really need the size of the matrix to be N, but just to be not more than a fixed power of N. This remark will allow us to apply the estimate (11) to submatrices of A_N.

Proof of the theorem. Our strategy will be to use the ideas of Soshnikov, well explained in [23] (see also [25] or [1]). We shall also need an estimate on the moments of the truncated variables $\hat{a}_{ij} := a_{ij}1_{|a_{ij}|\leq N^\gamma}$. By [15], Chap. VIII.9, Th. 2.23, we have that for any $k \geq 0$, for any (non identically null) a_{ij},

$$\mathbb{E}[\hat{a}_{ij}^{2k}] \overset{sl.}{\sim} \begin{cases} 1 & \text{if } k \leq \alpha, \\ N^{\gamma(k-\alpha)} & \text{if } k > \alpha. \end{cases} \tag{12}$$

We have, suppressing the dependence on N to simplify the notation,

$$\operatorname{Tr} \hat{A}^{2s} = \sum_{\substack{1\leq i_0,\dots,i_{2s}\leq N \\ i_0=i_{2s}}} \hat{a}_{i_0 i_1} \cdots \hat{a}_{i_{2s-1} i_{2s}}.$$

To any $\mathbf{i} = (i_0, \dots, i_{2s})$ such that $i_0 = i_{2s}$, we associate the non oriented graph $G_{\mathbf{i}} := (V_{\mathbf{i}}, E_{\mathbf{i}})$ with vertex set $\{i_0, \dots, i_{2s}\}$ and edges $\{i_{\ell-1}, i_\ell\}$, $1 \leq \ell \leq 2s$ and the closed path $P_{\mathbf{i}} = i_0 \to i_1 \to \cdots \to i_{2s}$ on this graph.

Since the a_{ij}'s are symmetrically distributed, each edge of $G_{\mathbf{i}}$ has to be visited an even number of times by $P_{\mathbf{i}}$ for the contribution of $\mathbf{i}$ to $\mathbb{E} \operatorname{Tr} \hat{A}^{2s}$ to be nonzero.

To such a $\mathbf{i}$, we associate a set $M_{\mathbf{i}}$ of s *marked instants* as follows. We read the edges of $P_{\mathbf{i}}$ successively. The instant at which an edge $\{i, j\}$ 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 (note that the instant are counted from 1 to $2s$, hence the instant where, for example, the edge $i_0 \to i_1$ is read is the instant 1). Other instants are said to be *unmarked*. Since each edge of $G_{\mathbf{i}}$ is visited an even number of times by $P_{\mathbf{i}}$, it is clear that

$$\#M_{\mathbf{i}} = s.$$

Now, for each $0 \leq k \leq s$, we define $N_{\mathbf{i}}(k)$ to be the set of i's in $\{1, \dots, N\}$ occurring exactly k times as the current vertex of a marked instant and n_k be its cardinality. Let the family $(n_0, \dots, n_s)$ be called the *type* of $\mathbf{i}$. Note that we have

$$\sum_{k=0}^{s} n_k = N \qquad \text{and} \qquad \sum_{k=0}^{s} k n_k = s. \tag{13}$$

Let us now count the number of $\mathbf{i}$'s with fixed type $(n_0, \dots, n_s)$ (where the n_i's satisfy (13)). To define such a $\mathbf{i}$, one first has to choose the set $M_{\mathbf{i}}$ of marked instants : there are as many possibilities as Dick paths, *i.e.* the Catalan number

$$C_s := \frac{1}{s+1}\binom{2s}{s}$$

Then one has to choose an unlabelled partition of $M_{\mathbf{i}}$ defined by the fact that two marked instants are in the same class if and only if the path $P_{\mathbf{i}}$ is at the same vertex of $G_{\mathbf{i}}$ at both

of these instants. Such a partition is only required to have n_k blocks of cardinality k for each $k = 1, \ldots, s$. Hence there are

$$\frac{s!}{\prod_{k=1}^{s}(k!)^{n_k}} \times \frac{1}{\prod_{k=1}^{s} n_k!}$$

possibilities (the first factor counting the labelled partitions and the second one "delabelling"). At this point, one has to choose the vertices of $G_{\mathbf{i}}$. For i_0, there are N possibilities. For each other vertex, there are at most W possibilities. There are at most $n_1 + \cdots + n_s$ other vertices. Indeed, except possibly i_0, each vertex is occurring a certain number of times as the current vertex of a marked instant (for example at the first time the vertex is visited by the path $P_{\mathbf{i}}$). Hence there are at most $Nd_N^{n_1+\cdots+n_s}$ possibilities for the choices of the vertices of $G_{\mathbf{i}}$. There now remains to give an upper-bound on the number of ways to determine vertices at unmarked instants (such vertices will not be new, but still have to be chosen among the before chosen vertices). Soshnikov proved in [25] that this number is not larger than $\prod_{k=2}^{s}(2k)^{kn_k}$ (the idea is that the number of ways to determine the endpoint of an edge starting from a vertex of type k at an unmarked instant is at most $2k$).

To sum up, the number of $\mathbf{i}$'s with fixed type $(n_0, \ldots, n_s)$ is at most

$$C_s \frac{s!}{\prod_{k=1}^{s}(k!)^{n_k}} \times \frac{1}{\prod_{k=1}^{s} n_k!} \times Nd_N^{n_1+\cdots+n_s} \times \prod_{k=2}^{s}(2k)^{kn_k}.$$

Let us now give an upper bound on the expectation $\mathbb{E}[\hat{a}_{i_0 i_1} \cdots \hat{a}_{i_{2s-1} i_{2s}}]$ depending on the type of $\mathbf{i}$. For $i, j \in V_{\mathbf{i}}$, let $\overline{ij}$ denote the edge $\{i, j\}$ of $G_{\mathbf{i}}$ (this edge is unoriented, so $\overline{ij} = \overline{ji}$) and let $k(\overline{ij})$ denote the half of number of times that this edge is visited by $P_{\mathbf{i}}$, i.e. the number of marked instants along edge $\overline{ij}$. We also introduce $k_{i;\overline{ij}}$ to be the number of times that the vertex i is marked along the edge $\overline{ij}$. Clearly,

$$k(\overline{ij}) = k_{i;\overline{ij}} + k_{j;\overline{ij}} \quad \text{and} \quad \text{type}(i) = \sum_j k_{i;\overline{ij}}.$$

We know, by (12), that for a certain slowly varying sequence $L(n)$ (that can change at every line)

$$\mathbb{E}[\hat{a}_{i_0 i_1} \cdots \hat{a}_{i_{2s-1} i_{2s}}] = L(n) \prod_{\substack{e \in E_{\mathbf{i}}, \\ k(e) \geq \alpha}} N^{\gamma(2k_e - \alpha)} \leq L(n) \prod_{\substack{e \in E_{\mathbf{i}}, \\ k(e) \geq 2}} N^{\gamma(2k_e - 2)} = L(n) N^{-2\gamma E} \prod_{\substack{e \in E_{\mathbf{i}}, \\ k(e) \geq 2}} N^{2\gamma k_e},$$

where E denotes the number of edges e such that $k(e) \geq 2$. Let us now enumerate the edges via their extremities. Then

$$\begin{aligned}\sum_{\substack{e\in E_{\mathbf{i}},\\ k(e)\geq 2}} k(e) &= \#\{\text{marked instants along edges } e \text{ such that } k(e)\geq 2\}\\ &= \sum_{\substack{(v,w)\in V_{\mathbf{i}}^2\\ k(\overline{vw})\geq 2}} k_{v;\overline{vw}} = \sum_{\substack{(v,w)\in V_{\mathbf{i}}^2\\ k(\overline{vw})\geq 2,\mathrm{type}(v)=1}} k_{v;\overline{vw}} + \sum_{\substack{(v,w)\in V_{\mathbf{i}}^2\\ k(\overline{vw})\geq 2,\mathrm{type}(v)\geq 2}} k_{v;\overline{vw}}.\end{aligned}$$

Let us now use the fact, well known from [25, 20, 1] that if an edge $\overline{vw}$ is visited at least 4 times by the path $P_{\mathbf{i}}$, then at least one of v and w have type ≥ 2, except for the first visited vertex i_0. It follows that the first sum above is $\leq E+1$. Hence

$$\begin{aligned}\sum_{\substack{e\in E_{\mathbf{i}},\\ k(e)\geq 2}} k(e) &\leq E+1 + \sum_{v\in V_{\mathbf{i}},\mathrm{type}(v)\geq 2}\sum_{w} k_{v;\overline{vw}}\\ &= E+1 + \sum_{k=2}^{s} kn_k\end{aligned}$$

Hence $\mathbb{E}[\hat{a}_{i_0i_1}\cdots\hat{a}_{i_{2s-1}i_{2s}}] \leq L(N)N^{2\gamma(1+\sum_{k=2}^s kn_k)}$.

As a consequence,

$$\mathbb{E}[\mathrm{Tr}(\hat{A}^{2s})]/N^{2s\gamma'} \leq$$

$$L(N)N^{1+2\gamma-2s\gamma'}C_s s! \sum_{\substack{n_1,\dots,n_s\\ \text{s. t. (13) holds}}} \frac{1}{\prod_{k=1}^s (k!)^{n_k}} \times \frac{1}{\prod_{k=1}^s n_k!} \times d_N^{n_1+\cdots+n_s} \times \prod_{k=2}^{s}(2k)^{kn_k} \times N^{2\gamma\sum_{k=2}^s kn_k}.$$

Let us now use the fact $s! \leq n_1! s^{(s-n_1)} = n_1! s^{\sum_{k=2}^s kn_k} \leq n_1! N^{\gamma''\sum_{k=2}^s kn_k}$, $d_N \leq N^{\mu}$ and $N^{-2s\gamma'} = N^{-2\gamma'\sum_{k=1}^s kn_k}$. We get

$$\mathbb{E}[\mathrm{Tr}(\hat{A}^{2s})]/N^{2s\gamma'} \leq$$

$$L(N)N^{1+2\gamma}C_s \sum_{\substack{n_1,\dots,n_s\\ \text{s. t. (13) holds}}} N^{(\mu-2\gamma')n_1} \prod_{k=2}^{s} \frac{1}{n_k!}\left(\frac{N^{\mu}(2kN^{2(\gamma-\gamma')+\gamma''})^k}{k!}\right)^{n_k}.$$

But by the hypothesis (10), $\mu - 2\gamma' \leq 0$, hence the first factor is ≤ 1, so, using the fact that by (13),n_1 is determined by the other n_j's, we get

$$\begin{aligned}\mathbb{E}[\mathrm{Tr}(\hat{A}^{2s})]/N^{2s\gamma'} &\leq L(N)N^{1+2\gamma}C_s \sum_{n_2,\dots,n_s\geq 0} \prod_{k=2}^{s} \frac{1}{n_k!}\left(\frac{N^{\mu}(2kN^{2(\gamma-\gamma')+\gamma''})^k}{k!}\right)^{n_k}\\ &\leq L(N)N^{1+2\gamma}C_s \sum_{n_2,\dots,n_s\geq 0} \prod_{k=2}^{s} \frac{1}{n_k!}\left(\frac{N^{\mu}(2N^{2(\gamma-\gamma'+\gamma'')})^k}{k!}\right)^{n_k}\\ &\leq L(N)N^{1+2\gamma}C_s \exp\left(\sum_{k=2}^{s}\frac{N^{\mu}(2N^{2(\gamma-\gamma'+\gamma'')})^k}{k!}\right)\end{aligned}$$

To conclude,

$$\mathbb{E}[\mathrm{Tr}(\hat{A}^{2s})]/N^{2s\gamma'} \leq L(N)N^{1+2\gamma}C_s \exp\left(\sum_{k=2}^{s} \frac{N^{\mu}(2N^{2(\gamma-\gamma'+\gamma'')})^k}{k!}\right).$$

By the hypothesis made at Equation (10), we have $\frac{\mu}{4}+\gamma+\gamma'' < \gamma'$, so that the exponential term stays bounded as $N \to \infty$. Using that $C_s \sim 4^s(\pi s^3)^{-1/2}$, we get Theorem 2.1. □

Remark 2.5. In the case where the entries $a_{ij}, 1 \leq i,j \leq N$ are not symmetrically distributed, one can prove a similar statement as in Theorem 2.1. The proof is based on arguments already given in Section 4 of [1] and [20]. One can indeed assume that the truncated entries are centered. Then, the main modification in evaluating $\mathbb{E}[\mathrm{Tr}(\hat{A}^{2s})]/N^{2s\gamma'}$ is that one has to take into account the contribution of paths with edges seen an odd number of times. However any such edge is seen at least 3 times, because the entries are centered. It can then be shown that the contribution of such paths is negligible (provided s is small enough as in Theorem 2.1), as to each such edge corresponds a vertex of type > 1.

3. Proof of Theorem 1.1

We are going to prove Theorem 1.1 as an application of Theorem 5.3 of the appendix, for $c_n := b_N$, the sequence defined by (5). More precisely, we shall use its "random versions": the case $\alpha < 1+\mu^{-1}$ will be a consequence of Corollary 5.4, whereas in the case $1+\mu^{-1} \leq \alpha < 2(1+\mu^{-1})$, we need to truncate the entries, and the conclusion will follow from Corollary 5.5.

Hypothesis (2) implies that the distribution of the non-zero entries is in the max-domain of attraction of the Fréchet distribution with exponent α (see [21], p. 54). By e.g. [19, Th. 2.3.1], it implies that as $N \to \infty$, the point process

$$\sum_{k\,;\,|a_{i_kj_k}|>0} \delta_{|a_{i_kj_k}|/b_N}$$

converges in distribution to a Poisson point process on $(0,+\infty)$ with intensity measure $\frac{\alpha}{x^{\alpha+1}}\mathrm{d}x$. It explains why the second part of Theorem 1.1 is a consequence of its first part and why Hypothesis (25) of the corollaries 5.4 and 5.5 is satisfied by the $|a_{i_kj_k}|$'s.

Note that by (2), for any $\theta > 0$, for any non indenticaly null a_{ij},

$$\mathbb{P}(|a_{ij}| > b_N^{\theta}) \overset{sl.}{\sim} b_N^{-\alpha\theta} \overset{sl.}{\sim} N^{-\theta(1+\mu)}. \tag{14}$$

The following claim (valid without any assumption on α) is a direct consequence of (14) and of the union bound.

Claim 3.1. *For any $\eta > 0$, with probability going to one as $N \to \infty$, we have:*

a) *no row of A_N has two entries larger, in absolute value, than $b_N^{\frac{1+2\mu}{2(1+\mu)}+\eta}$ (the exponent $\frac{1+2\mu}{2(1+\mu)}$ increases from $\frac{1}{2}$ to $\frac{3}{4}$ as μ increases from 0 to 1),*

b) *the matrix A_N has no diagonal entry larger, in absolute value, than $b_N^{\frac{1}{1+\mu}+\eta}$.*

So for any positive δ, δ', parts (b.i) and (b.ii) of the random version of Hypothesis 5.2 are satisfied with

$$\kappa := \frac{1+2\mu}{2(1+\mu)} + \delta, \qquad \tau := \frac{1}{1+\mu} + \delta'. \tag{15}$$

Let us now verify Part (b.iii).

3.1. **Case where $\alpha < 1 + \mu^{-1}$.** Set

$$S := \sum_{j\,;\,|a_{ij}|<b_N^\kappa} |a_{ij}| \tag{16}$$

(we suppress the dependence in N to simplify notation). We shall prove that there is $\nu < 1$ such that $\mathbb{P}(S_N > b_N^\nu)$ is exponentially small with some bounds that are uniform on i. The sum S can be rewritten $S = S_1 + S_2 + S_3$ as follows :

$$S = \sum_{|a_{ij}|\le N^{\frac{\mu}{\alpha}-\eta}} |a_{ij}| + \sum_{N^{\frac{\mu}{\alpha}-\eta}<|a_{ij}|\le N^{\frac{\mu}{\alpha}+\eta}} |a_{1i}| + \sum_{N^{\frac{\mu}{\alpha}+\eta}<|a_{ij}|\le b_N^\kappa} |a_{ij}|$$

The sums S_1, S_2, S_3 can be treated with respectively parts a), c) and d) of Proposition 5.6 of the appendix. The treatment of S_1 uses the facts $b_N \overset{sl.}{\sim} N^{\frac{1+\mu}{\alpha}}$ and that $\mu + \frac{\mu}{\alpha}(1-\alpha)^+ < \frac{1+\mu}{\alpha}$, which is always true when $\alpha \le 1$ and which is a consequence of $\alpha < 1 + \mu^{-1}$ when $\alpha > 1$.

3.2. **Case where $1 + \mu^{-1} \le \alpha < 2(1 + \mu^{-1})$.** We have seen at (6) that $b_N \overset{sl.}{\sim} N^{\frac{1+\mu}{\alpha}}$. So to apply Corollary 5.5 for $c_n = b_N$, we have to find a cut-off exponent γ satisfying both following constraints:

1) for κ defined by (15), there is $\varepsilon > 0$ such that with exponentially high probability,

$$S := \sum_{j\,;\,N^\gamma<|a_{ij}|\le b_N^\kappa} |a_{ij}| \le N^{\frac{1+\mu}{\alpha}-\varepsilon}, \tag{17}$$

2) there is $\varepsilon' > 0$ such that with probability tending to one, we have:

$$\|\hat{A}_N\| \le N^{\frac{1+\mu}{\alpha}-\varepsilon'} \qquad (\text{with } \hat{A}_N := [a_{ij}1_{|a_{ij}|\le N^\gamma}]_{1\le i,j\le N}).$$

By Corollary 2.2, the second condition is satisfied when $\max\{\frac{\mu}{2}, \frac{\mu}{4} + \gamma\} < \frac{1+\mu}{\alpha}$. As $\alpha < 2(1 + \mu^{-1})$, we have $\frac{\mu}{2} < \frac{1+\mu}{\alpha}$, so for condition 2) to be verified, one only needs

$$\gamma < \frac{1+\mu}{\alpha} - \frac{\mu}{4}.$$

To treat the sum S of (17), one proceeds as we did to treat the sum S defined at (16) in the case $\alpha < 1 + \mu^{-1}$, except that now,

$$S_1 = \sum_{N^\gamma < |a_{1j}| \leq N^{\frac{\mu}{\alpha} - \eta}} |a_{1j}|$$

and S_1 is treated thanks to Part b) of Proposition 5.6. Indeed, Part b) of Proposition 5.6 implies that w.e.h.p., $S_1 \leq N^{\mu - \gamma(\alpha-1) + \eta}$ with $\eta > 0$ as small as we need. Hence to fulfill Condition 1), one needs γ to satisfy

$$\mu - \gamma(\alpha - 1) < \frac{1+\mu}{\alpha},$$

i.e. $\gamma > \frac{\mu}{\alpha} - \frac{1}{\alpha(\alpha-1)}$. To sum up, we need to find a cut-off exponent γ such that

$$\frac{\mu}{\alpha} - \frac{1}{\alpha(\alpha-1)} < \gamma < \frac{1+\mu}{\alpha} - \frac{\mu}{4}.$$

Hence to conclude, it suffices to remark that $\alpha < 2(1 + \mu^{-1})$ implies

$$\frac{\mu}{\alpha} - \frac{1}{\alpha(\alpha-1)} < \frac{1+\mu}{\alpha} - \frac{\mu}{4}. \tag{18}$$

4. Proof of Theorem 1.5

4.1. **Eigenvalues.** Let us first prove the part about the eigenvalues, *i.e.* Equation (7). First, by Theorem 1.8, for any fixed $k \geq 1$, we have

$$\liminf \frac{\lambda_k}{N^{\frac{\mu}{2}}} \geq 2.$$

Let us now prove that

$$\limsup \frac{\lambda_k}{N^{\frac{\mu}{2}}} \leq 2. \tag{19}$$

To do that, we will prove that one can find a cut-off exponent γ such that for $\hat{A}_N := [a_{ij} 1_{|a_{ij}| \leq N^\gamma}]_{1 \leq i,j \leq N}$, we have

$$\limsup \frac{\|\hat{A}_N\|}{N^{\frac{\mu}{2}}} \leq 2 \qquad \text{and} \qquad \|A_N - \hat{A}_N\| = o(N^{\frac{\mu}{2}}). \tag{20}$$

To treat the first part of (20), we apply Corollary 2.2 with $\gamma' = \frac{\mu}{2}$ and $\gamma'' > 0$ such that

$$\frac{\mu}{4} + \gamma + \gamma'' < \gamma'.$$

For such a γ'' to exist, the constraint on γ is that $\gamma < \frac{\mu}{4}$. To treat the second part of (20), we use the following claim and the fact that

$$\|A_N - \hat{A}_N\| = \sup_{\lambda \text{ eig. of } A_N - \hat{A}_N} |\lambda| \leq \|A_N - \hat{A}_N\|_{\ell^\infty \to \ell^\infty} \leq \max_i \sum_{j\,;\,|a_{ij}| > N^\gamma} |a_{ij}|. \tag{21}$$

Claim 4.1. *Under the hypothesis that $\alpha > 2(1+\mu^{-1})$, for any $\gamma > \frac{\mu}{2(\alpha-1)}$, there is $\eta > 0$ such that with probability tending to one, we have:*

$$\max_i \sum_{j\,;\,|a_{ij}|>N^\gamma} |a_{ij}| \leq N^{\frac{\mu}{2}-\eta}.$$

Let us conclude the proof of the eigenvalues part of Theorem 1.5 before proving the claim. All we need is to find a cut-off exponent γ such that

$$\frac{\mu}{2(\alpha-1)} < \gamma < \frac{\mu}{4}.$$

The existence of such a γ is equivalent to the fact that $\alpha - 1 > 2$, which is true because $\alpha > 2(1+\mu^{-1}) \geq 4$.

Proof of the claim. Let $S(i)$ be the sum in the statement. By (2), it is easy to see that for any $\theta > \frac{1+\mu}{\alpha}$, with probability tending to one, we have

$$\max_{ij} |a_{ij}| \leq N^\theta.$$

Hence by Part a) of Claim 3.1 (using the fact that $b_N \overset{sl.}{\sim} N^{\frac{1+\mu}{\alpha}}$), for such a θ, with probability tending to one, for all i,

$$S(i) \leq \sum_{j\,;\,N^\gamma<|a_{ij}|\leq N^\theta} |a_{ij}|.$$

Using parts b), c), d) of Proposition 5.6 and cutting the sum in three pieces, it is easy to see that for any $\phi > \max\{\mu - \gamma(\alpha-1), \theta\}$, we have, w.e.h.p., uniformly on i,

$$\sum_{j\,;\,N^\gamma<|a_{ij}|\leq N^\theta} |a_{ij}| \leq N^\phi.$$

Now, to conclude the proof of the claim, it suffices to notice that the hypotheses $\gamma > \frac{\mu}{2(\alpha-1)}$ and $\alpha > 2(1+\mu^{-1})$ are respectively equivalent to $\mu - \gamma(\alpha-1) < \frac{\mu}{2}$ and $\frac{1+\mu}{\alpha} < \frac{\mu}{2}$, so that one can find some exponents θ, ϕ satisfying

$$\frac{1+\mu}{\alpha} < \theta \qquad \text{and} \qquad \max\{\mu - \gamma(\alpha-1), \theta\} < \phi < \frac{\mu}{2}.$$

□

4.2. **Eigenvectors.** We shall first prove the following lemma. Let us recall that a *principal submatrix* of a matrix $H = [x_{ij}]_{1\leq i,j\leq N}$ is a matrix of the type $H = [x_{j_k j_\ell}]_{1\leq k,\ell\leq L}$, where $1 \leq L \leq N$ and $1 \leq j_1 < \cdots < j_L \leq N$. The submatrix will be said to be *successively extracted* if the indices $j_1, \ldots, j_L$ form an interval of the set $\{1, \ldots, N\}$ endowed with the cyclic order.

Lemma 4.2. *Let H be a Hermitian matrix and $\rho_L(H)$ (resp. $\rho_L^{succ}(H)$) be the maximum spectral radius of its $L \times L$ principal (reps. principal successively extracted) submatrices. Let λ be an eigenvalue of H and $\mathbf{v}$ an associated unit eigenvector.*

If $\mathbf{v}$ is (L,η)-localized, then $|\lambda| \leq \frac{\rho_L(\mathbf{X})+\sqrt{\eta}\rho(\mathbf{X})}{\sqrt{1-\eta}}$.

If $\mathbf{v}$ is (L,η)-successively localized, then $|\lambda| \leq \frac{\rho_L^{succ}(H)+\sqrt{\eta}\rho(H)}{\sqrt{1-\eta}}$.

Proof. Let $j_1 < \cdots < j_L$ be indices such that $\sum_{\ell=1}^{L} |v_{j_\ell}|^2 \geq 1-\eta$ and let P be the orthogonal projection onto the subspace generated by the vectors $\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_L}$ (the $\mathbf{e}_j$'s are the vectors of the canonical basis). We have

$$\lambda P\mathbf{v} = PH\mathbf{v} = PHP\mathbf{v} + PH(1-P)\mathbf{v}.$$

Then the conclusion follows directly from the following

$$|\lambda| \times \sqrt{1-\eta} \leq |\lambda| \times \|P\mathbf{v}\| \leq \rho(PHP) + \rho(H)\|(1-P)\mathbf{v}\| \leq \rho(PHP) + \sqrt{\eta}\rho(H).$$

□

Claim 4.3. *Let us suppose that $\alpha > 2(1+\mu^{-1})$.*

a) *Let us fix c such that*

$$c < \frac{2}{5}\mu\frac{\alpha-2}{\alpha-1}. \tag{22}$$

Then there is $\varepsilon > 0$ such that w.e.h.p., the following holds:

For any $\lfloor N^c \rfloor \times \lfloor N^c \rfloor$ principal submatrix B of A_N, $\|B\| \leq N^{\frac{\mu}{2}-\varepsilon}$.

b) *Let us fix c such that*

$$c < \mu. \tag{23}$$

Then there is $\varepsilon > 0$ such that w.e.h.p., the following holds:

For any $\lfloor N^c \rfloor \times \lfloor N^c \rfloor$ successively extracted principal submatrix B of A_N, we have $\|B\| \leq N^{\frac{\mu}{2}-\varepsilon}$.

Before proving the claim, let us conclude to the proof of the eigenvectors part of Theorem 1.5. We know that $\rho(A) \sim 2N^{\frac{\mu}{2}}$ and that there is $\varepsilon > 0$ such that with probability tending to one, $\rho_L(A)$ (resp. $\rho_L^{\mathrm{succ}}(A)$) is bounded from above by $N^{\frac{\mu}{2}-\varepsilon}$. Since $\frac{1}{\sqrt{1-\eta}} < \frac{1}{\sqrt{1-\eta_0}} < \sqrt{2}$, Lemma 4.2 allows to conclude.

Proof of the claim. We shall treat a) and b) in the same time. Let us first note that Equation (22) (resp. Equation (23)) is equivalent to (resp. implies that)

$$\frac{c}{4} + \frac{\mu}{2(\alpha-1)} + c < \frac{\mu}{2} \qquad (\text{resp. } \frac{c}{4} + \frac{\mu}{2(\alpha-1)} < \frac{\mu}{2}).$$

Hence one can choose some positive exponents $\varepsilon, \gamma, \gamma', \gamma''$ such that $\gamma' \geq c/2$ and

$$\gamma > \frac{\mu}{2(\alpha-1)}, \quad \gamma'' > c \quad (\text{resp. } \gamma'' > 0) \quad \text{and} \quad \frac{c}{4} + \gamma + \gamma'' < \gamma' < \frac{\mu}{2} - \varepsilon. \tag{24}$$

Any submatrix $B = [a_{j_k j_\ell}]_{1 \leq k,\ell \leq \lfloor N^c \rfloor}$ can be written $B = \hat{B} + (B - \hat{B})$, with $\hat{B} := [a_{j_k j_\ell} 1_{|a_{j_k j_\ell}| \leq N^\gamma}]_{k,\ell}$. We know (see e.g. (21)) that independently of the choice of the j_k's

$$\|B - \hat{B}\| \leq \max_{1 \leq i \leq N} \sum_{j \text{ s.t. } |a_{ij}| > N^\gamma} |a_{ij}|.$$

Hence by Claim 4.1, the condition $\gamma > \frac{\mu}{2(\alpha-1)}$ of Equation (24) ensures us that for a certain $\eta > 0$, with probability tending to one, independently of the choice of the j_k's $\|B - \hat{B}\| \leq N^{\frac{\mu}{2} - \eta}$. Hence one can focus on $\hat{B}$.

Let us now apply Corollary 2.2 and Remark 2.4. We get that for any choice of $j_1, \ldots, j_{\lfloor N^c \rfloor}$, up to a polynomial factor in the RHT,

$$\mathbb{P}(\|\hat{B}\| \geq N^{\frac{\mu}{2} - \varepsilon}) \leq N^{-(\frac{\mu}{2} - \varepsilon - \gamma')2\lfloor N^{\gamma''} \rfloor}.$$

But there are at most N^{N^c} (resp. N) ways to choose the indices $j_1, \ldots, j_{\lfloor N^c \rfloor}$ of the rows of the submatrix B (resp. of the successively extracted submatrix B). Hence the probability that $\|\hat{B}\| \geq N^{\frac{\mu}{2} - \varepsilon}$ for at least one of these choices is $\leq N^{-(\frac{\mu}{2} - \varepsilon - \gamma')2\lfloor N^{\gamma''} \rfloor + N^c}$ (resp. $\leq N^{-(\frac{\mu}{2} - \varepsilon - \gamma')2\lfloor N^{\gamma''} \rfloor + 1}$). Since by (24), $\gamma'' > c$ (resp. $\gamma'' > 0$) and $\frac{\mu}{2} - \varepsilon - \gamma' > 0$, the conclusion follows. □

5. Appendix

5.1. **Eigenvalues and eigenvectors under perturbation.** In this section, we state a result about eigenvectors and eigenvalues of perturbed Hermitian matrices. The part about eigenvalues can be found in the literature (see the books by Bhatia [3, 4]), but we did not find the part about the eigenvectors in the literature.

Proposition 5.1. *Let H be a Hermitian matrix and $\mathbf{v}$ be a unit vector such that for a certain $\lambda \in \mathbb{R}$,*

$$H\mathbf{v} = \lambda \mathbf{v} + \varepsilon \mathbf{w},$$

with $\mathbf{w}$ a unit vector such that $\mathbf{w} \perp \mathbf{v}$ and $\varepsilon > 0$.

a) *Then H has an eigenvalue λ_ε in the ball $\bar{B}(\lambda, \varepsilon)$.*

b) *Suppose moreover that H has only one eigenvalue (counted with multiplicity) in $\bar{B}(\lambda, \varepsilon)$ and that all other eigenvalues are at distance at least $d > \varepsilon$ of λ. Then for $\mathbf{v}_\varepsilon$ a unit eigenvector associated to λ_ε, we have*

$$\|\mathbf{v}_\varepsilon - P_{\mathbf{v}}(\mathbf{v}_\varepsilon)\| \leq \frac{2\varepsilon}{d - \varepsilon},$$

where $P_{\mathbf{v}}$ denotes the orthogonal projection onto $\mathrm{Span}(\mathbf{v})$.

Proof. Part b) is a simple consequence of perturbation theory (see e.g. Lemma A.2 in [18]). Let $\mathbf{v}_\varepsilon$ be a normalizes eigenvector associated to λ_ε. We decompose $\mathbf{v}_\varepsilon = \langle \mathbf{v}_\varepsilon, \mathbf{v}\rangle \mathbf{v} + \mathbf{r}$ with $\mathbf{r} \perp \mathbf{v}$. Then $H\mathbf{v}_\varepsilon = \langle \mathbf{v}_\varepsilon, \mathbf{v}\rangle(\lambda \mathbf{v} + \varepsilon \mathbf{w}) + H\mathbf{r}$. From this we deduce that

$$(\lambda - H)r = \langle \mathbf{v}_\varepsilon, \mathbf{v}\rangle \varepsilon \mathbf{w} + (\lambda - \lambda_\varepsilon)\mathbf{v}_\varepsilon.$$

This yields the resut, by considering the norm of $(\lambda - H)$ restricted to the subspace $v^\perp$. □

5.2. **Largest eigenvalues vs largest entries of matrices.** In this section, we present a synthetic version of some ideas first appeared in Soshnikov's paper [26]. We also extend these ideas to the eigenvectors level. Theorem 5.3 below gives a sufficient condition for a large deterministic Hermitian matrix to have its kth largest eigenvalue approximately equal to its kth largest entry in absolute value for all fixed k. Note that this is not what happens usually: in some way the large entries need to overwhelm the other entries.
We also give sufficient condition so that the corresponding eigenvector is approximately equal to the eigenvector of the symmetric matrix formed by forcing all but this kth largest entry to be 0. The sufficient condition is, roughly speaking, that the largest entries and their spacings have an order $c_n \gg 1$, are sufficiently well spread out in the matrix and that, up to the removing of these largest entries, the sum of the terms of each row of the matrix have order $\ll c_n$. In Corollary 5.4, we give the random matrix version of this theorem and in Corollary 5.5, we explain how one is allowed to first remove a part of the matrix which does not affect the largest entries.

For each n, let H_n be an $n \times n$ deterministic Hermitian matrix with entries h_{ij}. Let us denote by $\lambda_1 \geq \lambda_2 \geq \cdots$ the eigenvalues of H_n (they depend implicitly on n) and let us choose some unit associated eigenvectors

$$\mathbf{v}_1, \mathbf{v}_2, \ldots$$

Let us also introduce a set of pairs of indices $(i_1 \leq j_1), (i_2 \leq j_2), \ldots$ such that for all k, $|h_{i_k j_k}|$ is the kth largest entry, in absolute value, of H_n. Let $\theta_k \in \mathbb{R}$ such that $h_{i_k j_k} = |h_{i_k j_k}| e^{2i\theta_k}$ The eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots$ are chosen such that for each k, $e^{-i\theta_k}\langle \mathbf{v}_k, \mathbf{e}_{i_k}\rangle \geq 0$. We make the following hypotheses.

Hypothesis 5.2. (a) *There is a sequence $c_n \longrightarrow +\infty$ such that for any fixed k,*

(a.i) $c_{n+k} \sim c_n$,

(a.ii) $0 < \liminf \frac{|h_{i_k j_k}|}{c_n} \leq \limsup \frac{|h_{i_k j_k}|}{c_n} < \infty$ *and* $\liminf \frac{|h_{i_k j_k}| - |h_{i_{k+1} j_{k+1}}|}{c_n} > 0$

(b) *There exists three exponents $\kappa, \tau, \nu \in (0,1)$ such that for n large enough,*

(b.i) *no row of H_n has two entries larger, in absolute value, than c_n^κ,*

(b.ii) *no diagonal entry of H_n is larger, in absolute value, than c_n^τ,*

(b.iii) *for each $i \in \{1, \dots, n\}$,*

$$\sum_{j\,;\,|h_{ij}|<c_n^\kappa} |h_{ij}| \leq c_n^\nu.$$

Theorem 5.3. *Under Hypothesis 5.2, as $n \to \infty$, for any $k \geq 1$ fixed,*

$$\frac{\lambda_k}{|h_{i_k j_k}|} \longrightarrow 1 \qquad \text{and} \qquad \mathbf{v}_k - \frac{e^{i\theta_k}\mathbf{e}_{i_k} + e^{-i\theta_k}\mathbf{e}_{j_k}}{\sqrt{2}} \longrightarrow 0 \qquad \text{(for the } \ell^2\text{-norm).}$$

Before proving the theorem, let us state its two "random versions".

Corollary 5.4. *Suppose now that the matrix H_n is random, that the sequence b_n is deterministic and satisfies Hypothesis (a.i), replace Hypothesis (a.ii) by*

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \mathbb{P}\Big(\frac{|h_{i_k j_k}|}{c_n} < \varepsilon\Big) + \mathbb{P}\Big(\frac{|h_{i_k j_k}|}{c_n} > \frac{1}{\varepsilon}\Big) + \mathbb{P}\Big(\frac{|h_{i_k j_k}| - |h_{i_{k+1} j_{k+1}}|}{c_n} < \varepsilon\Big) = 0 \tag{25}$$

and suppose that Hypothesis (b) holds with probability tending to one. Then the conclusions of the theorem remain true for the convergence in probability.

Proof. Recall that a sequence of real random variables converges in probability to a deterministic limit if and only if from each of its subsequences, one can extract a subsequence converging almost surely. Hence it suffices to notice that the deterministic theorem also holds (obviously) if one replaces the sequence H_n of $n \times n$ matrices, by a sequence $H_{\varphi(n)}$ of $\varphi(n) \times \varphi(n)$ matrices with $\varphi(n) \to +\infty$. □

By Proposition 5.1, one directly deduces the following corollary.

Corollary 5.5. *Suppose that one can write $H_n = \hat{H}_n + (H_n - \hat{H}_n)$, where $H_n - \hat{H}_n$ satisfies the hypotheses of Corollary 5.4 and that for a certain $\rho < 1$,*

$$\frac{\|\hat{H}_n\|}{c_n^\rho} \text{ converges in probability to zero.} \tag{26}$$

Then the conclusions of Theorem 5.3 for H_n remain true for the convergence in probability.

Proof of Theorem 5.3. We suppress the dependence on n to simplify notation.

Fact 1: We have $\|H\|_{\ell^\infty \to \ell^\infty} = |h_{i_1 j_1}|(1+o(1))$ (and as a consequence, $\lambda_1 \leq |h_{i_1 j_1}|(1+o(1))$).

Indeed, $\|H\|_{\ell^\infty \to \ell^\infty} = \max_i \sum_j |h_{ij}| \geq |h_{i_1 j_1}|$, thus to prove Fact 1, it suffices to notice that for n large enough, for all i,

$$\sum_j |h_{ij}| \leq \sum_{j\,;\,|h_{ij}|<c_n^\kappa} |h_{ij}| + \max_j |h_{ij}| \leq c_n^\nu + \max_j |h_{ij}| \leq \underbrace{c_n^\nu}_{\ll |h_{i_1 j_1}| \text{ by (a.ii).}} + |h_{i_1 j_1}|.$$

Fact 2: For any fixed $k \geq 1$, $\lambda_k \leq |h_{i_k j_k}|(1 + o(1))$.

Indeed, the hypotheses ensure that for n large enough, the numbers $i_1, \dots, i_k$ are pairwise distinct, so that the largest entry, in absolute value, of the $(n-k+1)\times(n-k+1)$ matrix $H^{(k)}$, deduced from H by removing rows and columns with indices $i_1, \dots, i_{k-1}$, is $|h_{i_k j_k}|$. This matrix (more specifically : this *sequence of matrices*, because n is an implicit parameter here) also satisfies the previous hypotheses (for the sequence $c_n^{(k)} := c_{n+k-1}$). Hence by the previous fact, $\lambda_1(H^{(k)}) \leq |h_{i_k j_k}|(1+o(1))$. But by Weyl's interlacing inequalities, we have $\lambda_k(H) \leq \lambda_1(H^{(k)})$. It allows to conclude.

Fact 3: For any fixed $k \geq 1$, for $\mathbf{v} = \frac{e^{i\theta_k}\mathbf{e}_{i_k} + e^{-i\theta_k}\mathbf{e}_{j_k}}{\sqrt{2}}$, we have

$$H\mathbf{v} = |h_{i_k j_k}|\mathbf{v} + \mathbf{r}, \qquad \text{with } \|\mathbf{r}\| = o(c_n).$$

Indeed, for $\mathbf{r} := H\mathbf{v} - |h_{i_k j_k}|\mathbf{v}$, it is easy to see that

$$\|\mathbf{r}\| \leq \frac{1}{\sqrt{2}} \left(|h_{i_k i_k}| + |h_{j_k j_k}| + \sum_{i \notin \{i_k, j_k\}} (|h_{i_k i}| + |h_{i j_k}|) \right).$$

We have $\|\mathbf{r}\| = o(c_n)$ because by Hypothesis (b), $|h_{i_k i_k}| + |h_{j_k j_k}| \leq 2c_n^{\kappa}$ and

$$\sum_{i \notin \{i_k, j_k\}} (|h_{i_k i}| + |h_{i j_k}|) \leq 2c_n^{\nu}.$$

Let us now conclude the proof of the theorem. Since $|h_{i_k j_k}|$ has order c_n, Fact 3 and Part a) of Proposition 5.1 imply that for any fixed $k \geq 1$, H has an eigenvalue equal to $|h_{i_k j_k}|(1+o(1))$. Hence by Fact 2 and Hypothesis (a.ii), $\lambda_k = |h_{i_k j_k}|(1+o(1))$. By Hypothesis (a.ii) again, it follows that $|\lambda_k - \lambda_{k+1}|$ has order c_n and so one can apply Part b) of Proposition 5.1 to deduce from fact 3 that

$$\left\| \mathbf{v}_k - \frac{e^{i\theta_k}\mathbf{e}_{i_k} + e^{-i\theta_k}\mathbf{e}_{j_k}}{\sqrt{2}} \right\| \longrightarrow 0.$$

□

5.3. **Sums of truncated heavy-tailed random variables.** In this section, we give exponential estimates on the concentration of sums of truncated heavy-tailed variables. In the paper, these estimates are needed for example to give upper bounds on the spectral radius of matrices via the maximum of the sums of the entries along the rows.

Let us consider some i.i.d. variables $Y_i \geq 0$ such that for a certain $\alpha > 0$,

$$\mathbb{P}(Y_1 > y) \overset{sl.}{\sim} y^{-\alpha} \qquad \text{as } y \to \infty. \tag{27}$$

Let us also fix a sequence $d_n \overset{sl.}{\sim} n^{\mu}$ for a fixed $\mu > 0$.

Proposition 5.6. a) *For any sequence $\beta_n \overset{sl.}{\sim} n^b$ with $0 \leq b \leq \mu/\alpha$ and any $\varepsilon > 0$, we have w.e.h.p.*

$$\sum_{j=1}^{d_n} Y_j 1_{Y_j \leq \beta_n} \ \leq \ n^{\mu + b(1-\alpha)^+ + \varepsilon}.$$

b) *If $\alpha \geq 1$, for any sequences $\alpha_n \overset{sl.}{\sim} n^a$, $\beta_n \overset{sl.}{\sim} n^b$ with $0 \leq a < b \leq \mu/\alpha$ and any $\varepsilon > 0$, we have w.e.h.p.*

$$\sum_{j=1}^{d_n} Y_j 1_{\alpha_n < Y_j \leq \beta_n} \ \leq \ n^{\mu - a(\alpha-1) + \varepsilon}.$$

c) *For any sequences $\alpha_n \overset{sl.}{\sim} n^{\frac{\mu}{\alpha} - \eta}$, $\beta_n \overset{sl.}{\sim} n^{\frac{\mu}{\alpha} + \eta'}$, with $\eta, \eta' > 0$ and any $\varepsilon > \alpha\eta + \eta'$, we have w.e.h.p.*

$$\sum_{j=1}^{d_n} Y_j 1_{\alpha_n < Y_j \leq \beta_n} \ \leq \ n^{\frac{\mu}{\alpha} + \varepsilon}.$$

d) *For any sequences $\alpha_n \overset{sl.}{\sim} n^{\frac{\mu}{\alpha} + \eta}$ with $\eta > 0$, $\beta_n \overset{sl.}{\sim} n^{\beta}$, with $\beta > 0$ and any $\gamma > \beta$, we have w.e.h.p.*

$$\sum_{j=1}^{d_n} Y_j 1_{\alpha_n < Y_j \leq \beta_n} \ \leq \ n^{\gamma}.$$

Before proving the proposition, we shall first establish the following concentration result for sums of Bernoulli variables.

Lemma 5.7. *For each $n \geq 1$, let $X_1, \ldots, X_m$ be some independent Bernoulli variables with paramater p (m, the X_i's and p depending on the parameter n). Suppose that $mp \geq Cn^\theta$ for some constants $C, \theta > 0$. Then for any fixed $\eta > 0$, we have w.e.h.p.*

$$\left| \frac{1}{m} \sum_{i=1}^{m} X_i - p \right| \leq \eta p. \tag{28}$$

Proof. Set $S := \sum_{i=1}^m X_i$. Let us for example give an upper-bound on $\mathbb{P}(S \geq mp(1+\eta))$ (the other side can treated in the same way with $\lambda \leq 0$). Using the fact that for any $\lambda \geq 0$,

$$\log \mathbb{E}[e^{\lambda X_i}] = \log(1 + p(e^\lambda - 1)) \leq p(e^\lambda - 1),$$

we have, for any $\lambda \geq 0$,

$$\mathbb{P}(S \geq mp(1+\eta)) \leq \mathbb{E}[e^{\lambda S}] e^{-\lambda mp(1+\eta)} \leq e^{-mp\{\lambda(1+\eta) - (e^\lambda - 1)\}}.$$

We conclude remarking that $\max_{\lambda \geq 0} \{\lambda(1+\eta) - (e^\lambda - 1)\} = (1+\eta)\log(1+\eta) - \eta > 0$. □

Proof of Proposition 5.6. a) First, one gets rid of the j's such that $Y_j \leq 1$ because their sum is $\leq d_n$. Then, set

$$S_n := \sum_{j=1}^{d_n} Y_j 1_{1<Y_j\leq\beta_n}$$

and $k_\varepsilon := \lfloor b/\varepsilon \rfloor$. We have $\theta := \frac{\mu-\alpha k_\varepsilon \varepsilon}{2} > 0$ and for n large enough, $\beta_n \leq n^{(k_\varepsilon+1)\varepsilon}$, so

$$S_n \leq \sum_{k=0}^{k_\varepsilon}\sum_{j=1}^{d_n} Y_j 1_{n^{k\varepsilon}<Y_j\leq n^{(k+1)\varepsilon}} \leq \sum_{k=0}^{k_\varepsilon} \underbrace{\sum_{j=1}^{d_n} 1_{n^{k\varepsilon}<Y_j\leq n^{(k+1)\varepsilon}}}_{:=Z_k^{(n)}} n^{(k+1)\varepsilon}.$$

For each k, $Z_k^{(n)}$ is a sum of $d_n \overset{sl.}{\sim} n^\mu$ independent Bernoulli variables with parameter $p_k(n) \overset{sl.}{\sim} n^{-\alpha k\varepsilon}$. We have $d_n p_k(n) \overset{sl.}{\sim} n^{\mu-\alpha k\varepsilon}$ hence for n large enough, $d_n p_k(n) \geq n^\theta$ (where θ is defined above). As a consequence, by Lemma 5.7, w.e.h.p., for each k,

$$Z_k^{(n)} \leq 2 d_n p_k(n).$$

This implies that

$$S_n \leq \sum_{k=0}^{k_\varepsilon} 2 d_n p_k(n) n^{(k+1)\varepsilon} \leq n^{\mu+b(1-\alpha)^+ +\varepsilon}.$$

b) Let S_n be the considered sum. The proof works in the same way as the one of a), introducing $k_\varepsilon := \lfloor (b-a)/\varepsilon \rfloor$, $\theta := \frac{\mu-\alpha(a+k_\varepsilon\varepsilon)}{2} > 0$ and writing

$$S_n \leq \sum_{k=0}^{k_\varepsilon}\sum_{j=1}^{d_n} 1_{\alpha_n n^{k\varepsilon}<Y_j\leq\alpha_n n^{(k+1)\varepsilon}} \alpha_n n^{(k+1)\varepsilon}.$$

c) This is a direct application of Lemma 5.7, since the considered sum if $\leq \beta_n \times$ the sum of d_N Bernoulli variables with parameter $\overset{sl.}{\sim} n^{-\mu+\alpha\eta}$.

d) Note that if the considered sum is $> n^\gamma$, then there are at least n^γ/β_n non zero terms in the sum. By the union bound, this happens with probability at most

$$d_n^{\lceil n^\gamma/\beta_n\rceil} \times \mathbb{P}(Y_1 > \alpha_n)^{\lceil n^\gamma/\beta_n\rceil}$$

(indeed, there are at most $d_n^{\lceil n^\gamma/\beta_n\rceil}$ subsets of cardinality $\lceil n^\gamma/\beta_n\rceil$ in $\{1,\ldots,d_n\}$). Using (27), one can easily check that this probability is exponentially small. □

References

[1] A. Auffinger, G. Ben Arous, S. Péché *Poisson convergence for the largest eigenvalues of heavy tailed random matrices*. Ann. Inst. Henri Poincaré Probab. Stat. 45 (2009), no. 3, 589–610.

[2] Z.D. Bai, J.W. Silverstein *Spectral analysis of large dimensional random matrices*. Second Edition, Springer, New York, 2009.

[3] R. Bhatia *Matrix analysis*. Graduate Texts in Mathematics, 169. Springer-Verlag, New York, 1997.
[4] R. Bhatia *Perturbation bounds for matrix eigenvalues*. Reprint of the 1987 original. Classics in Applied Mathematics, 53. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007.
[5] G. Ben Arous, A. Guionnet, *The spectrum of heavy tailed random matrices.* Comm. Math. Phys. 278 (2007), no 3, 715–751.
[6] C. Bordenave, A. Guionnet *Localization and delocalization of eigenvectors for heavy-tailed random matrices*. Preprint arXiv:1201.1862.
[7] P. Cizeau, J-P. Bouchaud, *Theory of Lévy matrices* Phys. Rev. E 50 (1994).
[8] L. Erdös *Universality of Wigner random matrices: a Survey of Recent Results* Preprint arXiv:1004.0861.
[9] L. Erdös, A. Knowles *Quantum diffusion and eigenfunction delocalization in a random band matrix model*. Comm. Math. Phys. 303 (2011), no. 2, 509–554.
[10] L. Erdös, A. Knowles, *Quantum diffusion and delocalization for band matrices with general distribution.* Ann. Henri Poincaré 12 (2011), no. 7, 1227–1319.
[11] L. Erdös, A. Knowles, H-T. Yau, J. Yin, *Delocalization and Diffusion Profile for Random Band Matrices*. Preprint arXiv:1205.5669
[12] L. Erdös, B. Schlein, H.T. Yau *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. 287 (2009).
[13] L. Erdös, B. Schlein, H.T. Yau, *Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices*, Ann. Prob. 37 (2009).
[14] L. Erdös, B. Schlein, H.T. Yau *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not. 2010 (2010).
[15] W. Feller *An introduction to probability theory and its applications*, volume II, second edition, New York London Sydney : J. Wiley, 1966.
[16] Y.V. Fyodorov and A.D. Mirlin, *Scaling properties of localization in random band matrices: a σ-model approach*, Phys. Rev. Lett. 67 (1991).
[17] A. Knowles, J. Yin, *Eigenvector Distribution of Wigner Matrices* Preprint arXiv:1102.0057
[18] A. Knowles, J. Yin, *The Outliers of a Deformed Wigner Matrix*, arXiv.
[19] M.R. Leadbetter, G.Lindgren and H. Rootzén *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York, 1983.
[20] S. Péché, A. Soshnikov *Wigner random matrices with non-symmetrically distributed entries*. J. Stat. Phys. 129 (2007), no. 5-6, 857–884.
[21] R. Resnick *Extreme Values, Regular Variation and Point Processes*. Springer, New York, 1987.
[22] J. Schenker *Eigenvector localization for random band matrices with power law band width.* Comm. Math. Phys. 290 (2009), no. 3, 10651097.
[23] Y. Sinai, A. Soshnikov *A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices*. (Russian) Funktsional. Anal. i Prilozhen. 32 (1998), no. 2, 56–79, 96; translation in Funct. Anal. Appl. 32 (1998), no. 2, 114–131.
[24] S. Sodin *The spectral edge of some random band matrices*, Ann. Math. 172 (2010), 2223–2251.
[25] A. Soshnikov *Universality at the edge of the spectrum in Wigner random matrices*. Comm. Math. Phys. 207 (1999), no. 3, 697–733.
[26] A. Soshnikov *Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails*. Electron. Comm. Probab. 9 (2004), 82–91.
[27] A. Soshnikov *Poisson statistics for the largest eigenvalues in random matrix ensembles*. Mathematical physics of quantum mechanics, 351–364, Lecture Notes in Phys., 690, Springer, Berlin, 2006.
[28] T. Spencer *Random banded and sparse matrices* The Oxford handbook of random matrix theory. Oxford University Press (2011) 471–488.
[29] T. Tao, V. Vu *Random matrices: Universal properties of eigenvectors* Preprint arXiv:1103.2801
[30] T. Tao, V. Vu *Random matrices: universality of local eigenvalue statistics up to the edge*. Comm. Math. Phys. 298 (2010), no. 2, 549–572.

MAP 5, UMR CNRS 8145 - Université Paris Descartes, 45 rue des Saints-Pères 75270 Paris cedex 6, France and CMAP, École Polytechnique, route de Saclay, 91128 Palaiseau Cedex, France.

E-mail address: florent.benaych-georges@parisdescartes.fr

LPMA, Université Paris Diderot, Site Chevaleret 75205 Paris Cedex 13, France.

E-mail address: sandrine.peche@math.univ-paris-diderot.fr