# Spectrum of non-Hermitian heavy tailed random matrices

Charles Bordenave, Pietro Caputo, Djalil Chafai

# SPECTRUM OF NON-HERMITIAN HEAVY TAILED RANDOM MATRICES

CHARLES BORDENAVE, PIETRO CAPUTO, AND DJALIL CHAFAÏ

Abstract. Let $(X_{jk})_{j,k\geqslant 1}$ be i.i.d. complex random variables such that $|X_{jk}|$ is in the domain of attraction of an $\alpha$-stable law, with $0 < \alpha < 2$. Our main result is a heavy tailed counterpart of Girko's circular law. Namely, under some additional smoothness assumptions on the law of $X_{jk}$, we prove that there exist a deterministic sequence $a_n \sim n^{1/\alpha}$ and a probability measure $\mu_\alpha$ on $\mathbb{C}$ depending only on $\alpha$ such that with probability one, the empirical distribution of the eigenvalues of the rescaled matrix $(a_n^{-1}X_{jk})_{1\leqslant j,k\leqslant n}$ converges weakly to $\mu_\alpha$ as $n \to \infty$. Our approach combines Aldous & Steele's objective method with Girko's Hermitization using logarithmic potentials. The underlying limiting object is defined on a bipartized version of Aldous' Poisson Weighted Infinite Tree. Recursive relations on the tree provide some properties of $\mu_\alpha$. In contrast with the Hermitian case, we find that $\mu_\alpha$ is not heavy tailed.

Contents



## 1. Introduction

The *eigenvalues* of an $n\times n$ complex matrix $M$ are the roots in $\mathbb{C}$ of its characteristic polynomial. We label them $\lambda_1(M),\ldots,\lambda_n(M)$ so that $|\lambda_1(M)| \geqslant \cdots \geqslant |\lambda_n(M)| \geqslant 0$. We also denote by $s_1(M) \geqslant \cdots \geqslant s_n(M)$ the *singular values* of $M$, defined for every $1 \leqslant k \leqslant n$ by $s_k(M) :=$

*Date*: Expanded version of a paper published in Communications in Mathematical Physics, 307, 513-560 (2011).
2000 *Mathematics Subject Classification.* 47A10; 15A52; 05C80.
*Key words and phrases.* Spectral theory; Objective method; Operator convergence; Logarithmic potential; Random matrices; Random Graphs; Heavy tailed distributions; $\alpha$-stable laws.

$\lambda_k(\sqrt{MM^*})$ where $M^* = \overline{M}^\top$ is the conjugate transpose of $M$. We define the empirical spectral measure and the empirical singular values measure as

$$\mu_M = \frac{1}{n}\sum_{k=1}^n \delta_{\lambda_k(M)} \quad \text{and} \quad \nu_M = \frac{1}{n}\sum_{k=1}^n \delta_{s_k(M)}.$$

Let $(X_{ij})_{i,j\geqslant 1}$ be i.i.d. complex random variables with cumulative distribution function $F$. Consider the matrix $X = (X_{ij})_{1\leqslant i,j\leqslant n}$. Following Dozier and Silverstein [20, 19], if $F$ has finite positive variance $\sigma^2$, then for every $z \in \mathbb{C}$, there exists a probability measure $\mathcal{Q}_{\sigma,z}$ on $[0,\infty)$ depending only on $\sigma$ and $z$, with explicit Cauchy-Stieltjes transform, such that a.s. (almost surely)

$$\nu_{\frac{1}{\sqrt{n}}X - zI} \underset{n\to\infty}{\rightsquigarrow} \mathcal{Q}_{\sigma,z} \tag{1.1}$$

where $\rightsquigarrow$ denotes the weak convergence of probability measures. The proof of (1.1) is based on a classical approach for Hermitian random matrices with bounded second moment: truncation, centralization, recursion on the resolvent, and cubic equation for the limiting Cauchy-Stieltjes transform. In the special case $z = 0$, the statement (1.1) reduces to the quarter-circular law theorem (square version of the Marchenko-Pastur theorem, see [37, 52, 54]) and the probability measure $Q_{\sigma,0}$ is the quarter-circular law with Lebesgue density

$$x \mapsto \frac{1}{\pi\sigma^2}\sqrt{4\sigma^2 - x^2}\mathbb{1}_{[0,2\sigma]}(x). \tag{1.2}$$

Girko's famous circular law theorem [25] states under the same assumptions that a.s.

$$\mu_{\frac{1}{\sqrt{n}}X} \underset{n\to\infty}{\rightsquigarrow} \mathcal{U}_\sigma \tag{1.3}$$

where $\mathcal{U}_\sigma$ is the uniform law on the disc $\{z \in \mathbb{C}; |z| \leqslant \sigma\}$. This statement was established through a long sequence of partial results [39, 24, 26, 33, 21, 25, 4, 27, 5, 40, 28, 48, 50], the general case (1.3) being finally obtained by Tao and Vu [50] by using Girko's Hermitization with logarithmic potentials and uniform integrability, the convergence (1.1), and polynomial bounds on the extremal singular values.

1.1. **Main results.** The aim of this paper is to investigate what happens when $F$ does not have a finite second moment. We shall consider the following hypothesis:

(H1) there exists a slowly varying function $L$ (i.e. $\lim_{t\to\infty} L(x\,t)/L(t) = 1$ for any $x > 0$) and a real number $\alpha \in (0,2)$ such that for every $t \geqslant 1$

$$\mathbb{P}(|X_{11}| \geqslant t) = \int_{\{z\in\mathbb{C}; |z|\geqslant t\}} dF(z) = L(t)t^{-\alpha},$$

and there exists a probability measure $\theta$ on the unit circle $\mathbb{S}^1 := \{z \in \mathbb{C}; |z| = 1\}$ of the complex plane such that for every Borel set $D \subset \mathbb{S}^1$,

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{X_{11}}{|X_{11}|} \in D \,\middle|\, |X_{11}| \geqslant t\right) = \theta(D).$$

Assumption (H1) states a complex version of the classical criterion for the domain of attraction of a real $\alpha$-stable law, see e.g. Feller [23, Theorem IX.8.1a]. For instance, if $X_{11} = V_1 + iV_2$ with $i = \sqrt{-1}$ and where $V_1$ and $V_2$ are independent real random variables both belonging to the domain of attraction of an $\alpha$-stable law then (H1) holds. When (H1) holds, we define the sequence

$$a_n := \inf\{a > 0 \text{ s.t. } n\mathbb{P}(|X_{11}| \geqslant a) \leqslant 1\}$$

and (H1) implies that $\lim_{n\to\infty} n\mathbb{P}(|X_{11}| \geqslant a_n) = \lim_{n\to\infty} na_n^{-\alpha}L(a_n) = 1$. It follows then classically that $a_n = n^{1/\alpha}\ell(n)$ for every $n \geqslant 1$, for some slowly varying function $\ell$. The additional possible assumptions on $F$ to be considered in the sequel are the following:

(H2) $\mathbb{P}(|X_{11}| \geqslant t) \sim_{t\to\infty} c\,t^{-\alpha}$ for some $c > 0$ (this implies $a_n \sim_{n\to\infty} c^{1/\alpha}n^{1/\alpha}$)

(H3) $X_{11}$ has a bounded probability Lebesgue density on $\mathbb{R}$ or on $\mathbb{C}$.

One can check that (H1-H2-H3) hold e.g. when $|X_{11}|$ and $X_{11}/|X_{11}|$ are independent with $|X_{11}| = |S|$ where $S$ is real symmetric $\alpha$-stable. Another basic example is given by $X_{11} = \varepsilon W^{-1/\alpha}$ with $\varepsilon$ and $W$ independent such that $\varepsilon$ takes values in $S^1$ and $W$ is uniform on $[0,1]$.

For every $n \geqslant 1$, let us define the i.i.d. $n \times n$ complex matrix $A = A_n$ by

$$A_{ij} := a_n^{-1}X_{ij} \tag{1.4}$$

for every $1 \leqslant i, j \leqslant n$. Our first result concerns the singular values of $A - zI$, $z \in \mathbb{C}$.

**Theorem 1.1** (Singular values)**.** *If (H1) holds then for all $z \in \mathbb{C}$, there exists a probability measure $\nu_{\alpha,z}$ on $[0, \infty)$ depending only on $\alpha$ and $z$ such that a.s.*

$$\nu_{A-zI} \underset{n\to\infty}{\rightsquigarrow} \nu_{\alpha,z}.$$

The case $z = 0$ was already obtained by Belinschi, Dembo and Guionnet [6]. Theorem 1.1 is a heavy tailed version of the Dozier and Silverstein theorem (1.1). Our main results below give a non-Hermitian version of Wigner's theorem for Lévy matrices [14, 7, 6, 11], as well as a heavy tailed version of Girko's circular law theorem (1.3).

**Theorem 1.2** (Eigenvalues)**.** *If (H1-H2-H3) hold then there exists a probability measure $\mu_\alpha$ on $\mathbb{C}$ depending only on $\alpha$ such that a.s.*

$$\mu_A \underset{n\to\infty}{\rightsquigarrow} \mu_\alpha.$$

**Theorem 1.3** (Limiting law)**.** *The probability distribution $\mu_\alpha$ from theorem 1.2 is isotropic and has a continuous density. Its density at $z = 0$ equals*

$$\frac{\Gamma(1 + 2/\alpha)^2\Gamma(1 + \alpha/2)^{2/\alpha}}{\pi\Gamma(1 - \alpha/2)^{2/\alpha}}.$$

*Furthermore, up to a multiplicative constant, the density of $\mu_\alpha$ is equivalent to*

$$|z|^{2(\alpha-1)}e^{-\frac{\alpha}{2}|z|^\alpha} \text{ as } |z| \to \infty.$$

Recall that for a normal matrix (i.e. which commutes with its adjoint), the absolute value of the eigenvalues are equal to the singular values. Theorem 1.3 reveals a striking contrast between $\mu_\alpha$ and $\nu_{\alpha,0}$. The limiting law of the eigenvalues $\mu_\alpha$ has a stretched exponential tail while the limiting law $\nu_{\alpha,0}$ of the singular values is heavy tailed with power exponent $\alpha$, see e.g. [6]. This does not contradict the identity $\prod_{k=1}^n |\lambda_k(A)| = \prod_{k=1}^n s_k(A)$, but it does indicate that $A$ is typically far from being a normal matrix. A similar shrinking phenomenon appears already in the finite second moment case (1.1)-(1.3): the law of the absolute value under the circular law $\mathcal{U}_\sigma$ has density

$$r \mapsto 2\sigma^{-2}r\mathbb{1}_{[0,\sigma]}(r)$$

in contrast with the density (1.2) of the quarter-circular law $\mathcal{Q}_{\sigma,0}$, even the supports differ by a factor 2.

The proof of theorem 1.1 is given in section 2.8. It relies on an extension to non-Hermitian matrices of the "objective method" approach developed in [11]. More precisely, we build an explicit operator on Aldous' Poisson Weighted Infinite Tree (PWIT) and prove that it is the local limit of the matrices $A_n$ in an appropriate sense. While Poisson statistics arises naturally as in all heavy tailed phenomena, the fact that a tree structure appears in the limit is roughly explained by the observation that non vanishing entries of the rescaled matrix $A_n = a_n^{-1}X$ can be viewed as the adjacency matrix of a sparse random graph which locally looks like a tree. In particular, the convergence to PWIT is a weighted-graph version of familiar results on the local structure of Erdős-Rényi random graphs.

The proof of theorem 1.2 is given in section 3. It relies on Girko's Hermitization method with logarithmic potentials, on theorem 1.1, and on polynomial bounds on the extremal singular values needed to establish a uniform integrability property. This extends the Hermitization method to more general settings, by successfully mixing various arguments already developed in [11, 12, 50]. Following Tao and Vu, one of the key steps will be a lower bound on the distance of a row of the matrix $A$ to a subspace of dimension at most $n - n^{1-\gamma}$, for some small $\gamma > 0$.

Girko's Hermitization method gives a characterization of $\mu_\alpha$ in terms of its logarithmic potential (see appendix A). In our settings, however, this is not convenient to derive properties of the measure $\mu_\alpha$, and our proof of theorem 1.3 is based on an analysis of a self-adjoint operator on the PWIT and a recursive characterization of the spectral measure from the resolvent of this operator. This method is explained in section 2 while the actual computations on the PWIT are performed in section 4.

Let us conclude with some final remarks. Following [16], the derivation of a Markovian version of theorems 1.1 and 1.2 is an interesting problem, see [10, 11] for the symmetric case and [18, 12] for the light tailed non-symmetric case. In another direction, it is also tempting to seek for an interpretation of $\nu_{\alpha,z}$ and $\mu_\alpha$ in terms of a sort of graphical free probability theory. Indeed, our random operators are defined on trees and tree structures are closely related to freeness. Also, with a proper notion of trace, it is possible to define the spectral measure of an operator, see e.g. [15, 31, 36]. However these notions are usually defined on algebras of bounded operators and we will not pursue this goal here. Note finally that theorems 1.1 and 1.2 remain available for additive perturbations of finite rank, by following the methodology used in [17, 50, 47].

1.2. **Notation.** Throughout the paper, the notation $n \gg 1$ means *large enough* $n$. For any $c \in [0, \infty]$ and any couple $f, g$ of positive functions defined in a neighborhood of $c$ , we say that $f(t) \sim g(t)$ as $t$ goes to $c$, if $\lim_{t \to c} f(t)/g(t) = 1$. We denote by $\mathcal{D}'(\mathbb{C})$ the set of Schwartz-Sobolev distributions endowed with its usual convergence with respect to all infinitely differentiable functions with bounded support $C_0^\infty(\mathbb{C})$. We will consider the differential operators on $\mathbb{C} \simeq \mathbb{R}^2$, for $z = x + iy$ (here $i = \sqrt{-1}$)

$$\partial = \frac{1}{2}(\partial_x - i\partial_y) \quad \text{and} \quad \bar\partial = \frac{1}{2}(\partial_x + i\partial_y).$$

We have $\partial \bar z = \bar\partial z = 0$, $\partial z = \bar\partial \bar z = 1$ and the Laplace differential operator on $\mathbb{C}$ is given by

$$\Delta = 4\partial\bar\partial = \partial_x^2 + \partial_y^2.$$

We use sometimes the shortened notation $A - z$ instead of $A - zI$.

## 2. Bipartized resolvent matrix

The aim of this section is to develop an efficient machinery to analyze the spectral measures of a non-hermitian matrix which avoids a direct use of the logarithmic potential and the singular values. Our approach builds upon similar methods in the physics literature [22, 29, 44, 43].

2.1. **Bipartization of a matrix.** Let $n$ be an integer, and $A$ be a $n \times n$ complex matrix. We introduce the symmetrized version of $\nu_{A-z}$,

$$\check\nu_{A-z} = \frac{1}{2n} \sum_{k=1}^{n} \left( \delta_{\sigma_k(A-z)} + \delta_{-\sigma_k(A-z)} \right).$$

Let $\mathbb{C}_+ = \{z \in \mathbb{C} : \mathfrak{Im}(z) > 0\}$ and consider the quaternionic-type set

$$\mathbb{H}_+ = \left\{ U = \begin{pmatrix} \eta & z \\ \bar z & \eta \end{pmatrix}, \eta \in \mathbb{C}_+, z \in \mathbb{C} \right\} \subset \mathcal{M}_2(\mathbb{C}).$$

For $z \in \mathbb{C}, \eta \in \mathbb{C}_+$ and $1 \leqslant i, j \leqslant n$ integers, we define the elements of $\mathbb{H}_+$ and $\mathcal{M}_2(\mathbb{C})$ respectively,

$$U(z,\eta) = \begin{pmatrix} \eta & z \\ \bar z & \eta \end{pmatrix} \quad \text{and} \quad B_{ij} = \begin{pmatrix} 0 & A_{ij} \\ \bar A_{ji} & 0 \end{pmatrix}.$$

We define the matrix in $\mathcal{M}_n(\mathcal{M}_2(\mathbb{C})) \simeq \mathcal{M}_{2n}(\mathbb{C})$, $B = (B_{ij})_{1 \leqslant i,j \leqslant n}$. Since $B_{ji}^* = B_{ij}$, as an element of $\mathcal{M}_{2n}(\mathbb{C})$, $B$ is an Hermitian matrix. Graphically, the matrix $A$ can be identified with an oriented graph on the vertex set $\{1, \cdots, n\}$ with weight on the oriented edge $(i,j)$ equal to $A_{ij}$. Then, the matrix $B$ can be thought of as the bipartization of the matrix $A$, that is a non-oriented graph on the vertex set $\{1, -1, \cdots, -n, n\}$, for every integers $1 \leqslant i, j \leqslant n$ the weight on the non-oriented edge $\{i, -j\}$ is $A_{ij}$, and there is no edge between $i$ and $j$ or $-i$ and $-j$.

For $U \in \mathbb{H}_+$, let $U \otimes I_n \in \mathcal{M}_n(\mathcal{M}_2(\mathbb{C}))$ be the matrix given by $(U \otimes I_n)_{ij} = \delta_{ij} U$, $1 \leqslant i, j \leqslant n$. The resolvent matrix is defined in $\mathcal{M}_n(\mathcal{M}_2(\mathbb{C}))$ by

$$R(U) = (B - U \otimes I_n)^{-1},$$

so that for all $1 \leqslant i, j \leqslant n$, $R(U)_{ij} \in \mathcal{M}_2(\mathbb{C})$. For $1 \leqslant k \leqslant n$, we write, with $U = U(z, \eta)$,

$$R(U)_{kk} = \begin{pmatrix} a_k(z,\eta) & b_k(z,\eta) \\ b_k'(z,\eta) & c_k(z,\eta) \end{pmatrix}. \tag{2.1}$$

The modulus of the entries of the matrix $R(U)_{kk}$ are bounded by $(\mathfrak{Im}(\eta))^{-1}$ (see the forthcoming lemma 2.2).

As an element of $\mathcal{M}_{2n}(\mathbb{C})$, $R$ is the usual resolvent of the matrix

$$B(z) = B - U(z,0) \otimes I_n.$$

Indeed, with $U = U(z,\eta)$,

$$R(U) = (B(z) - \eta I_{2n})^{-1}. \tag{2.2}$$

In the next proposition, we shall check that the eigenvalues of $B(z)$ are $\pm\sigma_k(A-z)$, $1 \leqslant k \leqslant n$, and consequently

$$\mu_{B(z)} = \check{\nu}_{A-z}. \tag{2.3}$$

It will follow that the spectral measures $\mu_A$ and $\check{\nu}_{A-z}$ can be easily recovered from the resolvent matrix. Recall that the Cauchy-Stieltjes transform of a measure $\nu$ on $\mathbb{R}$ is defined, for $\eta \in \mathbb{C}_+$, as

$$m_\nu(\eta) = \int_{\mathbb{R}} \frac{1}{x-\eta}\nu(dx).$$

The Cauchy-Stieltjes transform characterizes the measure. For a probability measure on $\mathbb{C}$, it is possible to define a Cauchy-Stieltjes-like transform on quaternions, by setting for $U \in \mathbb{H}_+$,

$$M_\mu(U) = \int_{\mathbb{C}} \left( \begin{pmatrix} 0 & \lambda \\ \bar\lambda & 0 \end{pmatrix} - U \right)^{-1} \mu(d\lambda) \ \in \ \mathbb{H}_+.$$

This transform characterizes the measure : in $\mathcal{D}'(\mathbb{C})$, $\lim_{t\downarrow 0}(\partial M_\mu(U(z,it))_{12} = -\pi\mu$. If $A$ is normal, i.e. if $A^*A = AA^*$, then it can be checked that $R(U)_{kk} \in \mathbb{H}_+$ and

$$\frac{1}{n}\sum_{k=1}^n R(U)_{kk} = M_{\mu_A}(U).$$

However, if $A$ is not normal, the above formula fails to hold and the next proposition explains how to recover anyway $\mu_A$ from the resolvent.

**Theorem 2.1** (From resolvent to spectral measure). *Let $U = U(z,\eta) \in \mathbb{H}_+$, and $a_k, b_k, b'_k, c_k$ be as in* (2.1). *Then* (2.3) *holds,*

$$m_{\check{\nu}_{A-z}}(\eta) = \frac{1}{2n}\sum_{k=1}^n \left(a_k(z,\eta) + c_k(z,\eta)\right),$$

*and, in $\mathcal{D}'(\mathbb{C})$,*

$$\mu_A = -\frac{1}{\pi n}\sum_{k=1}^n \partial b_k(\cdot,0) = \lim_{t\downarrow 0} -\frac{1}{\pi n}\sum_{k=1}^n \partial b_k(\cdot,it).$$

*In particular, if $A$ is a random matrix with exchangeable entries, then by linearity we get*

$$m_{\mathbb{E}\check{\nu}_{A-z}}(\eta) = \mathbb{E}a_1(z,\eta),$$

*and, in $\mathcal{D}'(\mathbb{C})$,*

$$\mathbb{E}\mu_A = -\frac{1}{\pi}\partial\mathbb{E}b_1(\cdot,0) = \lim_{t\downarrow 0} -\frac{1}{\pi}\partial\mathbb{E}b_1(\cdot,it).$$

*Proof of theorem 2.1.* Through a permutation of the entries, the matrix $B(z)$ is similar to

$$\begin{pmatrix} 0 & (A-z) \\ (A-z)^* & 0 \end{pmatrix},$$

whose eigenvalues are easily seen to be $\pm\sigma_k(A-z)$, $1 \leqslant k \leqslant n$. We get

$$\mathrm{tr}R = \sum_{k=1}^n (a_k + c_k) = \sum_{k=1}^n (\sigma_k(A-z) - \eta)^{-1} + \sum_{k=1}^n (-\sigma_k(A-z) - \eta)^{-1}.$$

And the first statement and (2.3) follow. Also, from (A.3), in Appendix, for $z \notin \mathrm{supp}(\mu_A)$,

$$U_{\mu_A}(z) = \int \ln|x|\mu_{B(z)}(dx) = \frac{1}{2n}\ln|\det B(z)|, \tag{2.4}$$

where $U_\mu$ is the logarithmic potential of a measure $\mu$ on $\mathbb{C}$, see (A.1). Recall that the differential of $X \mapsto \det(X)$ at point $X$ (invertible) in the direction $Y$ is $\mathrm{tr}(X^{-1}Y)\det(X)$ (this is sometimes referred as the Jacobi formula). The sign of $\det B(z)$ is $(-1)^n$. We deduce that in $\mathcal{D}'(\mathbb{C})$,

$$\bar\partial \ln|\det B(z)| = \frac{\bar\partial \det B(z)}{\det B(z)} = \mathrm{tr}\left\{ B(z)^{-1}\bar\partial \begin{pmatrix} 0 & -z \\ -\bar z & 0 \end{pmatrix} \otimes I_n \right\}.$$

With $R_{kk} = R(U(z,0))_{kk} = (B(z)^{-1})_{kk}$, we get from $\bar\partial z = 0$, $\bar\partial \bar z = 1$,

$$\bar\partial \ln|\det B(z)| = \sum_{k=1}^n \operatorname{tr}\left\{ R_{kk} \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \right\} = -\sum_{k=1}^n b_k(z,0).$$

Now from Equation (A.2), in $\mathcal{D}'(\mathbb{C})$, using $\Delta = 4\partial\bar\partial$,

$$2\pi\mu_A = \Delta U_{\mu_A} = \Delta \frac{1}{2n}\ln|\det B(z)| = -\frac{2}{n}\sum_{k=1}^n \partial b_k.$$

To get the limit as $t \downarrow 0$, we note that for real $t > 0$,

$$\int \ln|x - it|\mu_{B(z)}(dx) = \frac{1}{2n}\ln|\det(B(z) - it)|.$$

Note that $\det(B(z) - it)$ is real and its sign is $(-1)^n$. As $t \downarrow 0$, the left hand side of the above identity converges in $\mathcal{D}'(\mathbb{C})$, to $U_{\mu_A}$. Taking the Laplacian, and arguing as above, we get

$$\Delta \int \ln|x - it|\mu_{B(z)}(dx) = -\frac{2}{n}\sum_{k=1}^n \partial b_k(z,it). \tag{2.5}$$

The conclusion follows. □

Note that even if $-\sum_k \partial b_k$ is a measure on $\mathbb{C}$, for each $1 \leqslant k \leqslant n$, $-\partial b_k$ is not in general a measure on $\mathbb{C}$ (default of positivity, this can be checked on $2 \times 2$ matrices).

2.2. **Bipartization of an operator.** We shall generalize the above finite dimensional construction. Let $V$ be a countable set and let $\ell^2(V)$ denote the Hilbert space defined by the scalar product

$$\langle \phi, \psi \rangle := \sum_{u \in V} \bar\phi_u \psi_u\,, \qquad \phi_u = \langle \delta_u, \phi \rangle,$$

where $\delta_u$ is the unit vector supported on $u \in V$. Let $\mathcal{D}(V)$ denote the dense subset of $\ell^2(V)$ of vectors with finite support. Let $(w_{uv})_{u,v\in V}$ be a collection of complex numbers such that for all $u \in V$,

$$\sum_{v \in V} |w_{uv}|^2 + |w_{vu}|^2 < \infty$$

We may then define a linear operator $A$ on $\mathcal{D}(V)$, by the formula,

$$\langle \delta_u, A\delta_v \rangle = w_{uv}. \tag{2.6}$$

Let $\hat V$ be a set in bijection with $V$, the image of $v \in V$ being denoted by $\hat v \in \hat V$. We set $V^b = V \cup \hat V$ and define the symmetric operator $B$ on $\mathcal{D}(V^b)$, by the formulas,

$$\begin{aligned} \langle \delta_u, B\delta_{\hat v} \rangle = \overline{\langle \delta_{\hat v}, B\delta_u \rangle} &= \quad w_{uv} \\ \langle \delta_u, B\delta_v \rangle = \langle \delta_{\hat u}, B\delta_{\hat v} \rangle &= \quad 0. \end{aligned} \tag{2.7}$$

In other words, if $\Pi_u : \ell^2(V^b) \to \mathbb{C}^2$ denotes the orthogonal projection on $(u, \hat u)$,

$$\Pi_u B \Pi_v^* = \begin{pmatrix} 0 & w_{uv} \\ \bar w_{vu} & 0 \end{pmatrix}.$$

For $z \in \mathbb{C}$, we also define on $\mathcal{D}(V^b)$, the symmetric operator $B(z)$: for all $u, v$ in $V$,

$$\begin{aligned} \langle \delta_u, B(z)\delta_{\hat v} \rangle = \overline{\langle \delta_{\hat v}, B(z)\delta_u \rangle} &= \quad w_{uv} - z\mathbb{1}(u = v) \\ \langle \delta_u, B(z)\delta_v \rangle = \langle \delta_{\hat u}, B(z)\delta_{\hat v} \rangle &= \quad 0. \end{aligned}$$

Hence, if we identify $V^b$ with $\{1,2\} \times V$, we have

$$B(z) = B - U(z,0) \otimes I_V. \tag{2.8}$$

The operator $B(z)$ is symmetric and it has a closure on a domain $D(B) \subset \ell^2(V^b)$. We also denote by $B(z)$ the closure of $B(z)$. If $B$ is self-adjoint then $B(z)$ is also self-adjoint (recall that the sum of a bounded self-adjoint operator and a self-adjoint operator is also a self-adjoint operator). Recall also that the spectrum of a self–adjoint operator is real. For all $U = U(z,\eta) \in \mathbb{H}_+$,

$B(z) - \eta I_{V^b} = B - U(z,\eta) \otimes I_V$ is invertible with bounded inverse and the resolvent operator is then well defined by

$$R(U) = (B(z) - \eta I_{V^b})^{-1}.$$

We may then define

$$R(U)_{vv} = \Pi_v R(U) \Pi_v^* = \begin{pmatrix} a_v(z,\eta) & b_v(z,\eta) \\ b'_v(z,\eta) & c_v(z,\eta) \end{pmatrix}.$$

In the sequel, we shall use some properties of resolvent operators.

**Lemma 2.2** (Properties of resolvent)**.** *Let $B$ be the above bipartized operator. Assume that $B$ is self-adjoint and let $U = U(z,\eta) \in \mathbb{H}_+$, $v \in V$. Then, $a_v, c_v \in \mathbb{C}_+$, for each $z \in \mathbb{C}$, the functions $a_v(z,\cdot), b_v(z,\cdot), b'_v(z,\cdot), c_v(z,\cdot)$ are analytic on $\mathbb{C}_+$, and*

$$|a_v| \leqslant (\mathfrak{Im}(\eta))^{-1}, \quad |c_v| \leqslant (\mathfrak{Im}(\eta))^{-1}, \quad |b_v| \leqslant (2\mathfrak{Im}(\eta))^{-1} \quad \text{and} \quad |b'_v| \leqslant (2\mathfrak{Im}(\eta))^{-1}.$$

*Moreover, if $\eta \in i\mathbb{R}_+$, then $a_v$ and $c_v$ are pure imaginary and $b'_v = \bar{b}_v$.*

*Proof.* For a proof of the first statements refer e.g. to Reed and Simon [42]. For the last statement concerning $\eta \in i\mathbb{R}_+$, we define the skeleton of $B(z)$ as the graph on $V^b$ obtained by putting an edge between two vertices $u, v$ in $V^b$, if $\langle \delta_u, B(z)\delta_v \rangle \neq 0$. Then since there is no edge between two vertices of $V$ or $\hat{V}$, the skeleton of $B(z)$ is a bipartite graph.

Assume first that $B(z)$ is bounded: for all $u \in V^b$, $\|B(z)\delta_u\| \leqslant C$. Then for $|\eta| > C$, the series expansion of the resolvent gives

$$R(U) = -\sum_{n=0}^{\infty} \frac{B(z)^n}{\eta^{n+1}}.$$

However since the skeleton is a bipartite graph, all cycles have an even length. It implies that for $n$ odd, $\langle \delta_u, B(z)^n \delta_u \rangle = 0$. Applied first to $v \in V$, we deduce that for $|\eta| > C$, $a(z, -\bar{\eta}) = -\bar{a}(z,\eta)$ and then applied to $\hat{v}$, we get $c(z,-\bar{\eta}) = -\bar{c}(z,\eta)$. We may then extend to $\mathbb{C}_+$ this last identity by analyticity. For $\eta = it \in i\mathbb{R}_+$, we deduce that $a_v$ and $c_v$ are pure imaginary. Similarly, since the skeleton is a bipartite graph, a path from a vertex $v \in V$ to a vertex $\hat{u} \in \hat{V}$ must of be of odd length. We get for $|\eta| > C$

$$\begin{aligned} \bar{b}'_v(z,-\bar{\eta}) = \overline{\langle \delta_{\hat{v}}, R(U(z,-\bar{\eta}))\delta_v \rangle} &= -\sum_{n=0}^{\infty} \frac{\overline{\langle \delta_{\hat{v}}, B(z)^{2n+1}\delta_v \rangle}}{\eta^{2n+2}} \\ &= \langle \delta_v, R(U)\delta_{\hat{v}} \rangle = b_v(z,\eta), \end{aligned}$$

where we have used the symmetry of $B(z)$. It follows that $b'_v(z,-\bar{\eta}) = \bar{b}_v(z,\eta)$. If $B(z)$ is not bounded, then $B(z)$ is limit of a sequence of bounded operators and we conclude by invoking Theorem VIII.25(a) in [42]. □

2.3. **Operator on a tree.** We keep the setting of the above paragraph and consider a (non-oriented) tree $T = (V, E)$ on the vertices $V$ with edge set $E$ (recall that a tree is a connected graph without cycles). For ease of notation, we note $u \sim v$ if $\{u, v\} \in E$. We assume that if $\{u, v\} \notin E$ then $w_{uv} = w_{vu} = 0$. In particular $w_{vv} = 0$ for all $v \in V$. We continue to consider the operator $A$ defined by (2.6).

In the special case when $w_{uv} = \overline{w}_{vu}$ for all $u, v$ in $V$, the operator $A$ is symmetric and we first look for sufficient conditions for $A$ to be essentially self-adjoint.

**Lemma 2.3** (Criterion of self-adjointness)**.** *Let $\kappa > 0$ and $T = (V, E)$ be a tree. Assume that for all $u, v \in V$, $w_{uv} = \overline{w}_{vu}$ and that if $\{u, v\} \notin E$ then $w_{uv} = w_{vu} = 0$. Assume also that there exists a sequence of connected finite subsets $(S_n)_{n \geqslant 1}$ in $V$, such that $S_n \subset S_{n+1}$, $\cup_n S_n = V$, and for every $n$ and $v \in S_n$,*

$$\sum_{u \notin S_n : u \sim v} |w_{uv}|^2 \leqslant \kappa.$$

*Then $A$ is essentially self-adjoint.*

For a proof, see [11, Lemma A.3]. The above lemma has an interesting corollary for the bipartized operator $B$ of $A$ defined by (2.7)-(2.8).

**Corollary 2.4** (Criterion of self-adjointness of bipartized operator)**.** *Let $\kappa > 0$ and $T = (V, E)$ be a tree. Assume that if $\{u, v\} \notin E$ then $w_{uv} = w_{vu} = 0$. Assume also that there exists a sequence of connected finite subsets $(S_n)_{n\geqslant 1}$ in $V$, such that $S_n \subset S_{n+1}$, $\cup_n S_n = V$, and for every $n$ and $v \in S_n$,*

$$\sum_{u\notin S_n : u\sim v} \left(|w_{uv}|^2 + |w_{vu}|^2\right) \leqslant \kappa.$$

*Then for all $z \in \mathbb{C}$, $B(z)$ is self-adjoint.*

*Proof.* From (2.8), it is sufficient to check that $B$ is self-adjoint. Let $\varnothing \in V$ be a distinguished vertex, we define two disjoint trees $G_\varnothing = (V_\varnothing, E_\varnothing)$ and $\hat G_\varnothing = (\hat V_\varnothing, \hat E_\varnothing)$ on a partition $(V_\varnothing, \hat V_\varnothing)$ of $V^b$ as follows. The trees $G_\varnothing$ and $\hat G_\varnothing$ are the unique trees such that $\varnothing \in V_\varnothing$, $\hat\varnothing \in \hat V_\varnothing$ and that satisfy the following properties

(i) if $\{u, v\} \in E$ and $u$ in $V_\varnothing$ (or $\hat V_\varnothing$) then $\hat v \in V_\varnothing$ (or $\hat V_\varnothing$) and $\{u, \hat v\} \in E_\varnothing$ (or $\hat E_\varnothing$),

(ii) if $\{u, v\} \in E$ and $\hat u$ in $V_\varnothing$ (or $\hat V_\varnothing$) then $v \in V_\varnothing$ (or $\hat V_\varnothing$) and $\{\hat u, v\} \in E_\varnothing$ (or $\hat E_\varnothing$).

We note that by construction if $u \in V_\varnothing$ and $v \in \hat V_\varnothing$ then $\langle \delta_u, B\delta_v\rangle = 0$. If follows that the operator $B$ decomposes orthogonally into two operators $B_\varnothing$ and $\hat B_\varnothing$ on domains in $\ell^2(V_\varnothing)$ and $\ell^2(\hat V_\varnothing)$ respectively: $B = B_\varnothing \oplus \hat B_\varnothing$. We may then safely apply lemma 2.3 to $B_\varnothing$ and $\hat B_\varnothing$. □

When the operator $B$ is self-adjoint, the resolvent operator has a nice recursive expression due to the tree structure. Let $\varnothing \in V$ be a distinguished vertex of $V$ (in graph language, we root the tree $T$ at $\varnothing$). For each $v \in V\backslash\{\varnothing\}$, we define $V_v \subset V$ as the set of vertices whose unique path to the root $\varnothing$ contains $v$. We define $T_v = (V_v, E_v)$ as the subtree of $T$ spanned by $V_v$. We finally consider $A_v$, the projection of $A$ on $V_v$, and $B_v$ the bipartized operator of $A_v$. The skeleton of $A_v$ is contained in $T_v$. Finally, we note that if $B$ is self-adjoint then so is $B_v(z)$ for every $z \in \mathbb{C}$. The next lemma can be interpreted as a Schur complement formula on trees.

**Lemma 2.5** (Resolvent on a tree)**.** *Assume that $B$ is self-adjoint and let $U = U(z, \eta) \in \mathbb{H}_+$. Then*

$$R(U)_{\varnothing\varnothing} \quad = \quad -\left(U + \sum_{v\sim\varnothing} \begin{pmatrix} 0 & w_{\varnothing v} \\ \overline{w}_{v\varnothing} & 0 \end{pmatrix} \widetilde R(U)_{vv} \begin{pmatrix} 0 & w_{v\varnothing} \\ \overline{w}_{\varnothing v} & 0 \end{pmatrix}\right)^{-1},$$

*where $\widetilde R(U)_{vv} = \Pi_v R_{B_v}(U)\Pi_v^*$ and $R_{B_v}(U) = (B_v(z) - \eta)^{-1}$ is the resolvent operator of $B_v$.*

*Proof.* Define the operator $C$ on $\mathcal{D}(V^b)$ by its matrix elements

$$C_\varnothing := \Pi_\varnothing C \Pi_\varnothing^* = -U(z, 0)\,, \qquad C_v := \Pi_\varnothing C \Pi_v^* = \Pi_v C \Pi_\varnothing^* = \begin{pmatrix} 0 & w_{\varnothing v} \\ \overline{w}_{v\varnothing} & 0 \end{pmatrix}$$

for all $v \in V$ such that $v \sim \varnothing$, and $\Pi_u C \Pi_v^* = 0$ otherwise. The operator $C$ is symmetric and bounded. Its extension to $\ell^2(V^b)$ is thus self-adjoint (also denoted by $C$). In this way, we have from $V = \{\varnothing\} \bigcup_{v\sim\varnothing} V_v$,

$$B(z) = C + \widetilde B \quad \text{ with } \quad \widetilde B = \bigoplus_{v\sim\varnothing} B_v(z).$$

We shall write $\widetilde R(U) = (\widetilde B - \eta I)^{-1}$ for the associated resolvent of $\widetilde B$. From the resolvent identity, these operators satisfy

$$\widetilde R(U) C R(U) = \widetilde R(U) - R(U)\,. \tag{2.9}$$

Set $\widetilde R_{uv} = \Pi_u \widetilde R(U)\Pi_v^*$ and $R_{uv} = \Pi_u R(U)\Pi_v^*$. Observe that $\widetilde R_{\varnothing\varnothing} = -\eta^{-1} I_2$. Also the direct sum decomposition $V = \{\varnothing\} \bigcup_{v\sim\varnothing} V_v$ implies $\widetilde R_{vv} = \Pi_v R_{B_v}(U)\Pi_v^*$ and $\widetilde R_{uv} = 0$ for every $u \neq v$ with $u \sim \varnothing$, $v \sim \varnothing$. Similarly we have that $\widetilde R_{\varnothing v} = 0 = \widetilde R_{v\varnothing}$ for every $v \in V\backslash\{\varnothing\}$. Using the identity $\sum_{u\in V} \Pi_u^*\Pi_u = I$, we get

$$\begin{aligned} \Pi_\varnothing \widetilde R(U) C R(U) \Pi_\varnothing^* \quad &= \quad \widetilde R_{\varnothing\varnothing} C_\varnothing R_{\varnothing\varnothing} + \sum_{v\sim\varnothing} \widetilde R_{\varnothing\varnothing} C_v R_{v\varnothing} \\ &= \quad \eta^{-1} U(z, 0) R_{\varnothing\varnothing} - \eta^{-1} \sum_{v\sim\varnothing} C_v R_{v\varnothing}. \end{aligned}$$

We compose the identity (2.9) on the left by $\Pi_v$ and on the right by $\Pi_\varnothing^*$, we obtain, for $v \sim \varnothing$,

$$\widetilde R_{vv} C_v^* R_{\varnothing\varnothing} = -R_{v\varnothing}\,.$$

We finally compose (2.9) on the left by $\Pi_\varnothing$ and on the right by $\Pi_\varnothing^*$,

$$\eta^{-1}U(z,0)R_{\varnothing\varnothing} + \eta^{-1}\sum_{v\sim\varnothing} C_v \widetilde{R}_{vv} C_v^* R_{\varnothing\varnothing} = -\eta^{-1}I_2 - R_{\varnothing\varnothing},$$

or equivalently $(U(z,\eta) + \sum_{v\sim\varnothing} C_v \widetilde{R}_{vv} C_v^*)R_{\varnothing\varnothing} = -I_2$. □

2.4. **Local operator convergence.** In the next paragraphs, we are going to prove that the sequence of random matrices $(A_n)$ converges to a limit random operator on an infinite tree. Let us recall a notion of convergence that we have already used in [11].

**Definition 2.6** (Local convergence)**.** *Suppose $(A_n)$ is a sequence of bounded operators on $\ell^2(V)$ and $A$ is a linear operator on $\ell^2(V)$ with domain $D(A) \supset \mathcal{D}(V)$. For any $u, v \in V$ we say that $(A_n, u)$ converges locally to $(A, v)$, and write*

$$(A_n, u) \to (A, v),$$

*if there exists a sequence of bijections $\sigma_n : V \to V$ such that $\sigma_n(v) = u$ and, for all $\phi \in \mathcal{D}(V)$,*

$$\sigma_n^{-1} A_n \sigma_n \phi \to A\phi,$$

*in $\ell^2(V)$, as $n \to \infty$.*

Assume in addition that $A$ is closed and $\mathcal{D}(V)$ is a core for $A$ (i.e. the closure of $A$ restricted to $\mathcal{D}(V)$ equals $A$). Then, the local convergence is the standard strong convergence of operators up to a re-indexing of $V$ which preserves a distinguished element. With a slight abuse of notation we have used the same symbol $\sigma_n$ for the linear isometry $\sigma_n : \ell^2(V) \to \ell^2(V)$ induced in the obvious way. As pointed out in [11], the point for using Definition 2.6 lies in the following theorem on strong resolvent convergence.

**Theorem 2.7** (From local convergence to resolvents)**.** *Assume that $(A_n)$ and $A$ satisfy the conditions of Definition 2.6 and $(A_n, u) \to (A, v)$ for some $u, v \in V$. Let $B_n$ be the self-adjoint bipartized operator of $A_n$. If the bipartized operator $B$ of $A$ is self-adjoint and $\mathcal{D}(V^b)$ is a core for $B$, then, for all $U \in \mathbb{H}_+$,*

$$R_{B_n}(U)_{uu} \to R_B(U)_{vv}. \tag{2.10}$$

*where $R_B(U)_{vv} = \Pi_v R_B(U) \Pi_v^*$ and $R_B(U) = (B(z) - \eta)^{-1}$ is the resolvent of $B(z)$.*

*Proof of theorem 2.7.* It is a special case of Reed and Simon [42, Theorem VIII.25(a)]. Indeed, we first fix $z \in \mathbb{C}$ and extend the bijection $\sigma_n$ to $V^b$ by the formula, for all $w \in V$, $\sigma_n(\hat{w}) = \hat{\sigma}_n(w)$. Then we define $\widetilde{B}_n(z) = \sigma_n^{-1} B_n(z) \sigma_n$, so that $\widetilde{B}_n(z)\phi \to B(z)\phi$ for all $\phi$ in a common core of the self–adjoint operators $\widetilde{B}_n(z), B(z)$. This implies the strong resolvent convergence, i.e. $(\widetilde{B}_n(z) - \eta I)^{-1}\psi \to (B(z) - \eta I)^{-1}\psi$ for any $\eta \in \mathbb{C}_+$, $\psi \in \ell^2(V)$. We conclude by using the identities : $\Pi_v(\widetilde{B}_n(z) - \eta I)^{-1}\delta_v = \Pi_u(B_n(z) - \eta I)^{-1}\delta_u$ and $\Pi_v(\widetilde{B}_n(z) - \eta I)^{-1}\delta_{\hat{v}} = \Pi_u(B_n(z) - \eta I)^{-1}\delta_{\hat{u}}$. □

We shall apply the above theorem in cases where the operators $A_n$ and $A$ are random operators on $\ell^2(V)$, which satisfy with probability one the conditions of theorem 2.7. In this case we say that $(A_n, u) \to (A, v)$ *in distribution* if there exists a random bijection $\sigma_n$ as in Definition 2.6 such that $\sigma_n^{-1} A_n \sigma_n \phi$ converges in distribution to $A\phi$, for all $\phi \in \mathcal{D}(V)$ (where a random vector $\psi_n \in \ell^2(V)$ converges in distribution to $\psi$ if $\lim_{n\to\infty} \mathbb{E} f(\psi_n) = \mathbb{E} f(\psi)$ for all bounded continuous functions $f : \ell^2(V) \to \mathbb{R}$). Under these assumptions then (2.10) becomes convergence in distribution of (bounded) complex random variables. Note that in order to prove theorems 1.1, 1.2, we will also need almost-sure convergence statements.

2.5. **Poisson Weighted Infinite Tree (PWIT).** We now define an operator on an infinite rooted tree with random edge–weights, the Poisson weighted infinite tree (PWIT) introduced by Aldous [1], see also [3].

Let $\rho$ be a positive Radon measure on $\mathbb{R}$ such that $\rho(\mathbb{R}) = \infty$. PWIT($\rho$) is the random weighted rooted tree defined as follows. The vertex set of the tree is identified with $\mathbb{N}^f := \cup_{k\in\mathbb{N}} \mathbb{N}^k$ by indexing the root as $\mathbb{N}^0 = \varnothing$, the offsprings of the root as $\mathbb{N}$ and, more generally, the offsprings of some $v \in \mathbb{N}^k$ as $(v1), (v2), \cdots \in \mathbb{N}^{k+1}$ (for short notation, we write $(v1)$ in place of $(v, 1)$). In this way the set of $v \in \mathbb{N}^n$ identifies the $n^{\text{th}}$ generation. We then define $T$ as the tree on $\mathbb{N}^f$ with (non-oriented) edges between the offsprings and their parents.

We denote by Be(1/2) the Bernoulli probability distribution $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Now assign marks to the edges of the tree $T$ according to a collection $\{\Xi_v\}_{v\in\mathbb{N}^f}$ of independent realizations of the Poisson point process with intensity measure $\rho \otimes \mathrm{Be}(1/2)$ on $\mathbb{R} \times \{0,1\}$. Namely, starting from the root $\varnothing$, let $\Xi_\varnothing = \{(y_1,\varepsilon_1),(y_2,\varepsilon_2),\dots\}$ be ordered in such a way that $|y_1| \leqslant |y_2| \leqslant \cdots$, and assign the mark $(y_i,\varepsilon_i)$ to the offspring of the root labeled $i$. Now, recursively, at each vertex $v$ of generation $k$, assign the mark $(y_{vi},\varepsilon_{vi})$ to the offspring labeled $vi$, where $\Xi_v = \{(y_{v1},\varepsilon_{v1}),(y_{v2},\varepsilon_{v2}),\dots\}$ satisfy $|y_{v1}| \leqslant |y_{v2}| \leqslant \cdots$. The Bernoulli mark $\varepsilon_{vi}$ should be understood as an orientation of the edge $\{v, vi\}$ : if $\varepsilon_{vi} = 1$, the edge is oriented from $vi$ to $v$ and from $v$ to $vi$ otherwise.

For a probability measure $\theta$ on $S^1$, we introduce the measure on $\mathbb{C}$, for all Borel $D$:

$$\ell_\theta(D) = \int_0^\infty \int_{S^1} \mathbb{1}_{\{\omega^{-\alpha} r \in D\}} \theta(d\omega) dr \tag{2.11}$$

Consider a realization of PWIT$(2\ell_\theta)$. We now define a random operator $A$ on $\mathcal{D}(\mathbb{N}^f)$ by the formula, for all $v \in \mathbb{N}^f$ and $k \in \mathbb{N}$,

$$\langle \delta_v, A\delta_{vk}\rangle = \varepsilon_{vk} y_{vk}^{-1/\alpha} \quad \text{and} \quad \langle \delta_{vk}, A\delta_v\rangle = (1-\varepsilon_{vk}) y_{vk}^{-1/\alpha} \tag{2.12}$$

and $\langle \delta_v, A\delta_u\rangle = 0$ otherwise. It is an operator as in §2.3. Indeed, if $u = vk$ is an offspring of $v$, we set

$$w_{vu} = \varepsilon_{vk} y_{vk}^{-1/\alpha} \quad \text{and} \quad w_{uv} = (1-\varepsilon_{vk}) y_{vk}^{-1/\alpha}, \tag{2.13}$$

otherwise, we set $w_{uv} = 0$. We may thus consider the bipartized operator $B$ of $A$.

**Proposition 2.8** (Self-adjointness of bipartized operator on PWIT)**.** *Let $A$ be the random operator defined by* (2.12)*. With probability one, for all $z \in \mathbb{C}$, $B(z)$ is self-adjoint.*

We shall use Corollary 2.4. We start with a technical lemma proved in [11, Lemma A.4].

**Lemma 2.9.** *Let $\kappa > 0$, $0 < \alpha < 2$ and let $0 < x_1 < x_2 < \cdots$ be a Poisson process of intensity* 1 *on $\mathbb{R}_+$. Define $\tau = \inf\{t \in \mathbb{N} : \sum_{k=t+1}^\infty x_k^{-2/\alpha} \leqslant \kappa\}$. Then $\mathbb{E}\tau$ is finite and goes to* 0 *as $\kappa$ goes to infinity.*

*Proof of proposition 2.8.* For $\kappa > 0$ and $v \in \mathbb{N}^f$, we define

$$\tau_v = \inf\{t \geqslant 0 : \sum_{k=t+1}^{\infty} |y_{vk}|^{-2/\alpha} \leqslant \kappa\}.$$

The variables $(\tau_v)$ are i.i.d. and by lemma 2.9, there exists $\kappa > 0$ such that $\mathbb{E}\tau_v < 1$. We fix such $\kappa$. Now, we put a green color to all vertices $v$ such that $\tau_v \geqslant 1$ and a red color otherwise. We consider an exploration procedure starting from the root which stops at red vertices and goes on at green vertices. More formally, define the sub-forest $T^g$ of $T$ where we put an edge between $v$ and $vk$ if $v$ is a green vertex and $1 \leqslant k \leqslant \tau_v$. Then, if the root $\varnothing$ is red, we set $S_1 = C^g(T) = \{\varnothing\}$. Otherwise, the root is green, and we consider $T^g_\varnothing = (V^g_\varnothing, E^g_\varnothing)$ the subtree of $T^g$ that contains the root. It is a Galton-Watson tree with offspring distribution $\tau_\varnothing$. Thanks to our choice of $\kappa$, $T^g_\varnothing$ is almost surely finite. Consider $L^g_\varnothing$ the leaves of this tree (i.e. the set of vertices $v$ in $V^g_\varnothing$ such that for all $1 \leqslant k \leqslant \tau_v$, $vk$ is red). We set $S_1 = V^g_\varnothing \bigcup_{v\in L^g_\varnothing} \{1 \leqslant k \leqslant \tau_v : vk\}$. Clearly, the set $S_1$ satisfies the condition of Lemma 2.3.

Now, we define the outer boundary of $\{\varnothing\}$ as $\partial_\tau\{\varnothing\} = \{1,\cdots,\tau_\varnothing\}$ and for $v = (i_1,\cdots i_k) \in \mathbb{N}^f\backslash\{\varnothing\}$ we set $\partial_\tau\{v\} = \{(i_1,\cdots,i_{k-1},i_k+1)\} \cup \{(i_1,\cdots,i_k,1),\cdots,(i_1,\cdots,i_k,\tau_v)\}$. For a connected set $S$, its outer boundary is

$$\partial_\tau S = \left(\bigcup_{v\in S} \partial_\tau\{v\}\right) \backslash S.$$

Now, for each vertex $u_1,\cdots,u_k \in \partial_\tau S_1$, we repeat the above procedure to the rooted subtrees $T_{u_1},\cdots,T_{u_k}$. We set $S_2 = S_1 \bigcup \cup_{1\leqslant i\leqslant k} C^b(T_{u_i})$. Iteratively, we may thus almost surely define an increasing connected sequence $(S_n)$ of vertices with the properties required for Corollary 2.4. □

2.6. **Local convergence to PWIT.** We may now come back to the random matrix $A_n$ defined by (1.4). We extend it as an operator on $\mathcal{D}(\mathbb{N}^f)$ by setting for $1 \leqslant i, j \leqslant n$, $\langle \delta_i, A\delta_j \rangle = A_{i,j}$ and otherwise, if either $i$ or $j$ is in $\mathbb{N}^f \backslash \{1, \cdots n\}$, $\langle \delta_i, A\delta_j \rangle = 0$.

The aim of this paragraph is to prove the following theorem.

**Theorem 2.10** (Local convergence to PWIT)**.** *Assume (H1). Let $A_n$ be as above and $A$ be the operator associated to* PWIT($2\ell_\theta$) *defined by* (2.12)*. Then in distribution $(A_n, 1) \to (A, \varnothing)$.*

Up to small differences, this theorem has already been proved in [11, Section 2]. We review here the method of proof and stress the differences. The method relies on the *local weak convergence*, a notion introduced by Benjamini and Schramm [8], Aldous and Steele [3], see also Aldous and Lyons [2].

We define a network as a graph with weights on its edges taking values in some metric space. Let $G_n$ be the complete network on $\{1, \ldots, n\}$ whose weight on edge $\{i, j\}$ equals $(\xi^n_{i,j})$, for some collection $(\xi^n_{i,j})_{1 \leqslant i \leqslant j \leqslant n}$ of i.i.d. complex random variables. We set $\xi^n_{j,i} = \xi^n_{i,j}$. We consider the rooted network $(G_n, 1)$ obtained by distinguishing the vertex labeled 1.

We follow Aldous [1, Section 3]. For every fixed realization of the marks $(\xi^n_{ij})$, and for any $B, H \in \mathbb{N}$, such that $(B^{H+1} - 1)/(B - 1) \leqslant n$, we define a finite rooted subnetwork $(G_n, 1)^{B,H}$ of $(G_n, 1)$, whose vertex set coincides with a $B$–ary tree of depth $H$ with root at 1. To this end we partially index the vertices of $(G_n, 1)$ as elements in

$$J_{B,H} = \cup_{\ell=0}^{H} \{1, \cdots, B\}^\ell \subset \mathbb{N}^f,$$

the indexing being given by an injective map $\sigma_n$ from $J_{B,H}$ to $V_n := \{1, \ldots, n\}$. We set $I_\varnothing = \{1\}$ and the index of the root 1 is $\sigma_n^{-1}(1) = \varnothing$. The vertex $v \in V_n \backslash I_\varnothing$ is given the index $(k) = \sigma_n^{-1}(v)$, $1 \leqslant k \leqslant B$, if $\xi^n_{1,v}$ has the $k^{\text{th}}$ smallest absolute value among $\{\xi^n_{1,j}, j \neq 1\}$, the marks of edges emanating from the root 1. We break ties by using the lexicographic order. This defines the first generation. Now let $I_1$ be the union of $I_\varnothing$ and the $B$ vertices that have been selected. If $H \geqslant 2$, we repeat the indexing procedure for the vertex indexed by (1) (the first child) on the set $V_n \backslash I_1$. We obtain a new set $\{11, \cdots, 1B\}$ of vertices sorted by their weights as before (for short notation, we concatenate the vector $(1, 1)$ into 11). Then we define $I_2$ as the union of $I_1$ and this new collection. We repeat the procedure for (2) on $V_n \backslash I_2$ and obtain a new set $\{21, \cdots, 2B\}$, and so on. When we have constructed $\{B1, \cdots, BB\}$, we have finished the second generation (depth 2) and we have indexed $(B^3 - 1)/(B - 1)$ vertices. The indexing procedure is then repeated until depth $H$ so that $(B^{H+1} - 1)/(B - 1)$ vertices are sorted. Call this set of vertices $V_n^{B,H} = \sigma_n J_{B,H}$. The subnetwork of $G_n$ generated by $V_n^{B,H}$ is denoted $(G_n, 1)^{B,H}$ (it can be identified with the original network $G_n$ where any edge $e$ touching the complement of $V_n^{B,H}$ is given a mark $x_e = \infty$). In $(G_n, 1)^{B,H}$, the set $\{u1, \cdots, uB\}$ is called the set of offsprings of the vertex $u$. Note that while the vertex set has been given a tree structure, $(G_n, 1)^{B,H}$ is still a complete network on $V_n^{B,H}$. The next proposition shows that it nevertheless converges to a tree (i.e. extra marks diverge to $\infty$) if the $\xi^n_{i,j}$ satisfy a suitable scaling assumption.

Let $\rho$ be a Radon measure on $\mathbb{C}$ and let $T$ be a realization of PWIT($\rho$) defined in §2.5. For the moment, we remove the Bernoulli marks $(\varepsilon_v)_{v \in \mathbb{N}^f}$ and, for $v \in \mathbb{N}^f$ and $k \in \mathbb{N}$, we define the weight on edge $\{v, vk\}$ to simply be $y_{vk}$. Then $(T, \varnothing)$ is a rooted network. We call $(T, \varnothing)^{B,H}$ the finite random network obtained by the same sorting procedure. Namely, $(T, \varnothing)^{B,H}$ consists of the subtree with vertices in $J_{B,H}$, with the marks inherited from the infinite tree. If an edge is not present in $(T, \varnothing)^{B,H}$, we assign to it the mark $+\infty$.

We say that the sequence of random finite networks $(G_n, 1)^{B,H}$ converges in distribution (as $n \to \infty$) to the random finite network $(T, \varnothing)^{B,H}$ if the joint distributions of the marks converge weakly. To make this precise we have to add the points $\{\pm\infty\}$ as possible values for each mark, and continuous functions on the space of marks have to be understood as functions such that the limit as any one of the marks diverges to $+\infty$ exists and coincides with the limit as the same mark diverges to $-\infty$. We may define $\overline{\mathbb{C}} = \mathbb{C} \cup \{\pm\infty\}$. The next proposition generalizes [1, Section 3], for a proof see [11, Proposition 2.6] (the proof there is stated for a measure $\rho$ on $\mathbb{R}$, the complex case extends verbatim).

**Proposition 2.11** (Local weak convergence to a tree)**.** *Let $(\xi^n_{i,j})_{1 \leqslant i \leqslant j \leqslant n}$ be a collection of i.i.d. random variables in $\overline{\mathbb{C}}$ and set $\xi^n_{j,i} = \xi^n_{i,j}$. Let $\rho$ be a Radon measure on $\mathbb{C}$ with no mass at* 0 *and* *assume that*

$$n\mathbb{P}(\xi_{12}^n \in \cdot) \underset{n\to\infty}{\rightsquigarrow} \rho. \tag{2.14}$$

*Let* $G_n$ *be the complete network on* $\{1,\dots,n\}$ *whose mark on edge* $\{i,j\}$ *equals* $\xi_{ij}^n$*, and* $T$ *a realization of* $\mathrm{PWIT}(\rho)$*. Then, for all integers* $B,H$*,*

$$(G_n,1)^{B,H} \underset{n\to\infty}{\rightsquigarrow} (T,\varnothing)^{B,H}.$$

Now, we shall extend the above statement to directed networks. More precisely, let $(\xi_{i,j}^n)_{1\leqslant i,j\leqslant n}$ be i.i.d. real random variables. We consider the complete graph $\bar{G}_n$ on $V_n$ whose weight on edge $\{i,j\}$ equals, if $i\leqslant j$, $(\xi_{i,j}^n,\xi_{j,i}^n)\in\mathbb{R}^2$. As above, we partially index the vertices of $(\bar{G}_n,1)$ as elements in

$$J_{B,H}=\cup_{\ell=0}^H\{1,\cdots,B\}^\ell\subset\mathbb{N}^f,$$

the indexing being given by an injective map $\sigma_n$ from $J_{B,H}$ to $V_n$ such that $\sigma_n^{-1}(1)=\varnothing$. The difference with the above construction, is that the vertex $v\in V_n\backslash\{1\}$ is given the index $(k)=\sigma_n^{-1}(v)$, $1\leqslant k\leqslant B$, if $\min(|\xi_{1,v}^n|,|\xi_{v,1}^n|)$ has the $k^{\text{th}}$ smallest value among $\{\min(|\xi_{1,j}^n|,|\xi_{j,1}^n|),\, j\neq 1\}$.

Similarly, let $(T,\varnothing)$ be the infinite random rooted network with distribution $\mathrm{PWIT}(\rho)$. This time we do not remove the Bernoulli marks $(\varepsilon_v)_{v\in\mathbb{N}^f}$ and define the weight on edge $\{v,vk\}$ as $(y_{vk},\infty)$ if $\varepsilon_{vk}=1$ and $(\infty,y_{vk})$ if $\varepsilon_{vk}=0$. Again, we call $(T,\varnothing)^{B,H}$ the finite random network obtained by the sorting procedure : $(T,\varnothing)^{B,H}$ consists of the subtree with vertices in $J_{B,H}$, with the marks inherited from the infinite tree.

We apply proposition 2.11 to the complete network $G_n$ with mark on edge $\{i,j\}$ equals, if $i\leqslant j$, to $\min(|\xi_{i,j}^n|,|\xi_{j,i}^n|)$. This network satisfies (2.14) with $2\rho$. We remark that if $u,v\in J_{B,H}$ then from (2.14), $\max(|\xi^n_{\sigma_n(u),\sigma_n(v)}|,|\xi^n_{\sigma_n(v),\sigma_n(u)}|)$ diverges weakly to infinity. We also notice that, given $(G_n,1)^{B,H}$, with equal probability $|\xi^n_{\sigma_n(u),\sigma_n(v)}|$ is larger or less than $|\xi^n_{\sigma_n(v),\sigma_n(u)}|$. We deduce the following.

**Corollary 2.12** (Local weak convergence to a tree)**.** *Let* $\rho$ *be a Radon measure on* $\mathbb{C}$ *with no mass at* $0$*. Let* $(\xi_{i,j}^n)_{1\leqslant i,j\leqslant n}$ *be a collection of i.i.d. random variables in* $\overline{\mathbb{C}}$ *such that (2.14) holds. Let* $\bar{G}_n$ *be the complete network on* $\{1,\dots,n\}$ *whose mark on edge* $\{i,j\}$ *equals, if* $i\leqslant j$*,* $(\xi_{i,j}^n,\xi_{j,i}^n)$*, and* $T$ *a realization of* $\mathrm{PWIT}(2\rho)$*. Then, for all integers* $B,H$*,*

$$(\bar{G}_n,1)^{B,H} \underset{n\to\infty}{\rightsquigarrow} (T,\varnothing)^{B,H}.$$

We may now prove theorem 2.10.

*Proof of theorem 2.10.* We argue as in the proof of theorem 2.3(i) in [11, Section 2]. We first define the weights $(\xi_{i,j}^n)_{i,j\in\mathbb{N}^f}$ as follows. For integers $1\leqslant i,j\leqslant n$, we set

$$\xi_{i,j}^n=A_{i,j}^{-\alpha}=a_n^\alpha X_{i,j}^{-\alpha},$$

with the convention that $\xi_{i,j}^n=\infty$ if $X_{i,j}=0$. For this choice, by assumption (H1), (2.14) holds with $\rho=\ell_\theta$ and $\ell_\theta$ in (2.11). If $i$ or $j$ is in $\mathbb{N}^f\backslash\{1,\cdots,n\}$, we set $\xi_{i,j}^n=\infty$.

Let $\bar{G}_n$ denote the complete network on $\{1,\cdots,n\}$ with marks $(\xi_{i,j}^n,\xi_{j,i}^n)$ on edge $\{i,j\}$, if $i\leqslant j$. From Corollary 2.12, for all $B,H$, $(\bar{G}_n,1)^{B,H}$ converges weakly to $(T,\varnothing)^{B,H}$, where $T$ has distribution $\mathrm{PWIT}(2\ell_\theta)$. Let $A$ be the random operator associated to $T$.

Let $\sigma_n^{B,H}$ be the map $\sigma_n$ associated to the network $(\bar{G}_n,1)^{B,H}$. The maps $\sigma_n$ are arbitrarily extended to a bijection $\mathbb{N}^f\to\mathbb{N}^f$. From the Skorokhod Representation Theorem we may assume that $(\bar{G}_n,1)^{B,H}$ converges a.s. to $(T,\varnothing)^{B,H}$ for all $B,H$. Thus we may find sequences $B_n,H_n$ tending to infinity and a sequence of bijections $\widetilde{\sigma}_n:=\sigma_n^{B_n,H_n}$ such that $(B_n^{H_n+1}-1)/(B_n-1)\leqslant n$ and such that for any pair $u,v\in\mathbb{N}^f$ we have $\xi^n_{\widetilde{\sigma}_n(u),\widetilde{\sigma}_n(v)}$ which converge a.s. to

$$\begin{cases} y_{uk} & \text{if for some integer } k,\ v=uk \text{ and } \varepsilon_{uk}=1\\ y_{vk} & \text{if for some integer } k,\ u=vk \text{ and } \varepsilon_{vk}=0\\ \infty & \text{otherwise}\end{cases}$$

It follows that a.s.

$$\langle\delta_u,\widetilde{\sigma}_n^{-1}A_n\widetilde{\sigma}_n\delta_v\rangle={\xi^n_{\widetilde{\sigma}_n(u),\widetilde{\sigma}_n(v)}}^{-1/\alpha}\to\langle\delta_u,A\delta_v\rangle\,.$$

For any $v$, set $\psi_n^v := \widetilde{\sigma}_n^{-1} A_n \widetilde{\sigma}_n \delta_v$. To prove theorem 2.10, it is sufficient to show that for any $v \in \mathbb{N}^f$, $\psi_n^v \to A\delta_v$ in $\ell^2(\mathbb{N}^f)$ almost surely as $n$ goes to infinity, i.e.,

$$\sum_u (\langle \delta_u, \psi_n^v \rangle - \langle \delta_u, A\delta_v \rangle)^2 \to 0\,.$$

From what precedes, we know that $\langle \delta_u, \psi_n^v \rangle \to \langle \delta_u, A\delta_v \rangle$ for every $u$. The claim follows if we have (almost surely) uniform (in $n$) square-integrability of $(\langle \delta_u, \psi_n^v \rangle)_u$. This in turn follows from Lemma 2.4(i) and Lemma 2.7 in [11]. □

2.7. **Convergence of the resolvent matrix.** Let $A_n$ and $A$ be as in theorem 2.10. From proposition 2.8, we may almost surely define the resolvent $R$ of the bipartized random operator of $A$. For $U = U(z,\eta) \in \mathbb{H}_+$, we set

$$R(U)_{\varnothing\varnothing} = \Pi_\varnothing R(U) \Pi_\varnothing^* = \begin{pmatrix} a(z,\eta) & b(z,\eta) \\ b'(z,\eta) & c(z,\eta) \end{pmatrix}. \tag{2.15}$$

We define similarly, $R_n(U) = (B_n(z) - \eta)^{-1}$, the resolvent of $B_n$, the bipartized operator of $A_n$. We set $R_n(U)_{11} = \Pi_1 R_n(U) \Pi_1^*$.

**Theorem 2.13** (Convergence of the Resolvent matrix)**.** *Let $A_n$ and $A$ be as in theorem 2.10. For all $U = U(z,\eta) \in \mathbb{H}_+$,*

$$R_n(U)_{11} \underset{n\to\infty}{\rightsquigarrow} R(U)_{\varnothing\varnothing}.$$

*Proof of theorem 2.13.* We apply proposition 2.8, theorem 2.10 and the"in distribution" version of theorem 2.7. □

2.8. **Proof of theorem 1.1.** Again, we consider the sequence of random $n \times n$ matrices $(A_n)$ defined in introduction by (1.4).

**Theorem 2.14.** *For all $z \in \mathbb{C}_+$, almost surely the measure $\check{\nu}_{A_n - z}(dx)$ converges weakly to a measure $\check{\nu}_{\alpha,z}(dx)$ whose Cauchy-Stieltjes transform is given, for $\eta \in \mathbb{C}_+$, by*

$$m_{\check{\nu}_{\alpha,z}}(\eta) = \mathbb{E}a(z,\eta),$$

*where $a(z,\eta)$ was defined in* (2.15).

*Proof.* For every $z \in \mathbb{C}$, by proposition 2.8, the operator $B(z)$ is a.s. self-adjoint. It implies that there exists a.s. a measure on $\mathbb{R}$, $\nu_{\varnothing,z}$, called the spectral measure with vector $\delta_\varnothing$, such that for all $\eta \in \mathbb{C}_+$,

$$a(z,\eta) = \langle \delta_\varnothing, R(U)\delta_\varnothing \rangle = \int \frac{\nu_{\varnothing,z}(dx)}{x - \eta} = m_{\nu_{\varnothing,z}}(\eta).$$

We define $R_n$ as the resolvent matrix of $B_n$, the bipartized matrix of $A_n$. For $U = U(z,\eta) \in \mathbb{H}_+$, we write $R_n(U)_{kk} = \begin{pmatrix} a_k & b_k \\ b'_k & c_k \end{pmatrix}$. By theorem 2.1,

$$m_{\mathbb{E}\check{\nu}_{A_n - z}}(\eta) = \mathbb{E}a_1(z,\eta).$$

By lemma 2.2, for $U \in \mathbb{H}_+$, the entries of the matrix $R_n(U)_{11}$ are bounded. It follows from theorem 2.13 that for all $U \in \mathbb{H}_+$,

$$\lim_{n\to\infty} \mathbb{E}R_n(U)_{11} = \mathbb{E}\begin{pmatrix} a & b \\ b' & c \end{pmatrix},$$

where the limit matrix was defined in (2.15). Hence, for all $z \in \mathbb{C}_+$,

$$\lim_{n\to\infty} m_{\mathbb{E}\check{\nu}_{A_n - z}}(\eta) = \mathbb{E}a(z,\eta).$$

We deduce that $\mathbb{E}\check{\nu}_{A_n - z}$ converges to the measure $\nu_{\alpha,z} = \mathbb{E}\nu_{\varnothing,z}$. This convergence can be improved to almost sure by showing that the random measure $\check{\nu}_{A_n - z}$ concentrates around its mean. This is done by applying Borel-Cantelli Lemma and lemma C.2 to the matrix $B_n(z)$ whose spectral measure equals $\check{\nu}_{A_n - z}$, see (2.3). □

Theorem 1.1 is a corollary of the above theorem up to the fact that $\mathbb{E}a(z,\eta)$ does not depend on the measure $\theta$ which appears in (H1). The latter will be a consequence of the forthcoming theorem 4.1.

## 3. Convergence of the spectral measure

3.1. **Tightness.** In this paragraph, we prove that the counting probability measures of the eigenvalues and singular values of the random matrices $(A_n)$ defined by (1.4) are a.s. tight. For ease of notation, we will often write $A$ in place of $A_n$.

**Lemma 3.1** (Tightness). *If (H1) holds, there exists $r > 0$ such that for all $z \in \mathbb{C}$, a.s.*

$$\varlimsup_{n\to\infty} \int_0^\infty t^r \, \nu_{A-zI}(dt) < \infty, \quad \textit{and thus} \quad (\nu_{A-zI})_{n\geqslant 1} \textit{ is tight.}$$

*Moreover, a.s.*

$$\varlimsup_{n\to\infty} \int_{\mathbb{C}} |z|^r \, \mu_A(dz) < \infty, \quad \textit{and thus} \quad (\mu_A)_{n\geqslant 1} \textit{ is tight.}$$

*Proof.* In both cases, the a.s. tightness follows from the moment bound and the Markov inequality. The moment bound on $\mu_A$ follows from the statement on $\nu_A$ (take $z = 0$) by using the Weyl inequality (B.6). It is therefore enough to establish the moment bound on $\nu_{A-zI}$ for every $\mathbb{C}$. Let us fix $z \in \mathbb{C}$ and $r > 0$. By definition of $\nu_{A-zI}$ we have

$$\int_0^\infty t^r \, \nu_{A-zI}(dt) = \frac{1}{n} \sum_{k=1}^n s_k(A - zI)^r.$$

From (B.2) we have $s_k(A - zI) \leqslant s_k(A) + |z|$ for every $1 \leqslant k \leqslant n$, and one can then safely assume that $z = 0$ for the proof. By using (B.7) we get for any $0 \leqslant r \leqslant 2$,

$$\int_0^\infty t^r \, \nu_A(dt) \leqslant Z_n := \frac{1}{n} \sum_{i=1}^n Y_{n,i} \quad \text{where} \quad Y_{n,i} := \left( \sum_{j=1}^n a_n^{-2} |X_{ij}|^2 \right)^{r/2}.$$

We need to show that $(Z_n)_{n\geqslant 1}$ is a.s. bounded. Assume for the moment that

$$\sup_{n\geqslant 1} \mathbb{E}(Y_{n,1}^4) < \infty \tag{3.1}$$

for some choice of $r$. Since $Y_{n,1}, \dots, Y_{n,n}$ are i.i.d. for every $n \geqslant 1$, we get from (3.1) that

$$\mathbb{E}((Z_n - \mathbb{E}Z_n)^4) = n^{-4} \mathbb{E}\left( \sum_{1\leqslant i,j\leqslant n} (Y_{n,i} - \mathbb{E}Y_{n,i})^2 (Y_{n,j} - \mathbb{E}Y_{n,j})^2 \right) = O(n^{-2}).$$

Therefore, by the monotone convergence theorem, we get $\mathbb{E}(\sum_{n\geqslant 1}(Z_n - \mathbb{E}Z_n)^4) < \infty$, which gives $\sum_{n\geqslant 1}(Z_n - \mathbb{E}Z_n)^4 < \infty$ a.s. and thus $Z_n - \mathbb{E}Z_n \to 0$ a.s. Now the sequence $(\mathbb{E}Z_n)_{n\geqslant 1} = (\mathbb{E}Y_{n,1})_{n\geqslant 1}$ is bounded by (3.1) and it follows that $(Z_n)_{n\geqslant 1}$ is a.s. bounded.

It remains to show that (3.1) holds, say if $0 < 4r < \alpha$. To this end, let us define

$$S_{n,a,b} := \sum_{j=1}^n a_n^{-2} |X_{1j}|^2 \mathbb{1}_{\{a_n^{-2}|X_{1j}|^2 \in [a,b)\}} \quad \text{for every } a < b.$$

Now $Y_{n,1}^4 = (S_{n,0,\infty})^{2r} = (S_{n,0,1} + S_{n,1,\infty})^{2r}$ and thus,

$$\mathbb{E}(Y_{n,1}^4) \leqslant 2^{2r-1} \{ \mathbb{E}(S_{n,0,1}^{2r}) + \mathbb{E}(S_{n,1,\infty}^{2r}) \}. \tag{3.2}$$

We have $\sup_{n\geqslant 1} \mathbb{E}(S_{n,0,1}^{2r}) < \infty$. Indeed, since $2r < 1$, by the Jensen inequality,

$$\mathbb{E}(S_{n,0,1}^{2r}) \leqslant (\mathbb{E}S_{n,0,1})^{2r}$$

and by lemma C.1,

$$\mathbb{E}S_{n,0,1} \sim_n \alpha/(2 - \alpha).$$

For the second term of the right hand side of (3.2), we set

$$M_n := \max_{1\leqslant j\leqslant n} a_n^{-1} |X_{1j}| \mathbb{1}_{\{a_n^{-1}|X_{1j}|>1\}} \text{ and } N_n := \#\{1 \leqslant j \leqslant n \text{ s.t. } a_n^{-1}|X_{1j}| > 1\}.$$

From Hölder inequality, if $1/p + 1/q = 1$, we have

$$\mathbb{E}(S_{n,1,\infty}^{2r}) \leqslant \mathbb{E}\left(N_n^{2r} M_n^{4r}\right) \leqslant \left(\mathbb{E}N_n^{2rp}\right)^{1/p} \left(\mathbb{E}M_n^{4rq}\right)^{1/q}. \tag{3.3}$$

Recall that $\mathbb{P}(|X_{12}| > a_n) = (1+o(1))/n \leqslant 2/n$ for $n \gg 1$. By the union bound, for $n \gg 1$,

$$\mathbb{P}(N_n \geqslant k) \leqslant \binom{n}{k} \mathbb{P}(|X_{12}| > a_n)^k \leqslant \frac{n^k}{k!}\frac{2^k}{n^k} = \frac{2^k}{k!}.$$

In particular, we have $\sup_{n\geqslant 1} \mathbb{E}N_n^\eta < \infty$ for any $\eta > 0$. Similarly, since the function $L$ is slowly varying, for $n \gg 1$ and all $t \geqslant 1$, we have

$$\mathbb{P}(M_n \geqslant t) \leqslant n\mathbb{P}(|X_{12}| > ta_n) = na_n^{-\alpha} t^{-\alpha} L(a_n t) \leqslant 2t^{-\alpha}.$$

It follows that if $\gamma < \alpha$, $\sup_{n\geqslant 1} \mathbb{E}M_n^\gamma < \infty$. Taking $p$ and $q$ so that $4rq < \alpha$, we thus conclude from (3.3) that $\sup_{n\geqslant 1} \mathbb{E}(S_{n,1,\infty}^{2r}) < \infty$. ☐

3.2. **Invertibility.** In this paragraph, we find a lower bound for the smallest singular value of the random matrix $A - zI$ where $A$ is defined by (1.4), in other words an upper bound on the operator norm of the resolvent of $A$. Such lower bounds on the smallest singular value of random matrices were developed in the recent years by using Littlewood-Offord type problems, as in [48, 49] and [45]. The available results require moments assumptions which are not satisfied when the entries have heavy tails. Here we circumvent the problem by requiring the bounded density hypothesis (H3). The removal of this hypothesis can be done by adapting the Rudelson and Vershynin approach already used by Götze and Tikhomirov [28].

**Lemma 3.2** (Invertibility)**.** *If (H3) holds then for some $r > 0$, every $z \in \mathbb{C}$, a.s.*

$$\varliminf_{n\to\infty} n^r s_n(A - zI) = +\infty.$$

*Proof.* For every $x, y \in \mathbb{C}^n$ and $S \subset \mathbb{C}^n$, we set $x \cdot y := x_1\overline{y_1} + \cdots + x_n\overline{y_n}$ and $\|x\|_2 := \sqrt{x\cdot x}$ and $\mathrm{dist}(x, S) := \min_{y\in S} \|x - y\|_2$. Let $R_1, \ldots, R_n$ be the rows of $A - zI$ and set

$$R_{-i} := \mathrm{span}\{R_j; j \neq i\}$$

for every $1 \leqslant i \leqslant n$. From lemma B.2 we have

$$\min_{1\leqslant i\leqslant n} \mathrm{dist}(R_i, R_{-i}) \leqslant \sqrt{n}\, s_n(A - zI)$$

and consequently, by the union bound, for any $u \geqslant 0$,

$$\mathbb{P}(\sqrt{n}\, s_n(A - zI) \leqslant u) \leqslant n \max_{1\leqslant i\leqslant n} \mathbb{P}(\mathrm{dist}(R_i, R_{-i}) \leqslant u).$$

Let us fix $1 \leqslant i \leqslant n$. Let $Y_i$ be a unit vector orthogonal to $R_{-i}$. Such a vector is not unique. We just pick one. This defines a random variable on the unit sphere $\mathbb{S}^{n-1} = \{x \in \mathbb{C}^n : \|x\|_2 = 1\}$. By the Cauchy–Schwarz inequality,

$$|R_i \cdot Y_i| \leqslant \|\pi_i(R_i)\|_2 \|Y_i\|_2 = \mathrm{dist}(R_i, R_{-i})$$

where $\pi_i(\cdot)$ is the orthogonal projection on the orthogonal complement of $R_{-i}$. Let $\nu_i$ be the distribution of $Y_i$ on $\mathbb{S}^{n-1}$. Since $Y_i$ and $R_i$ are independent, for any $u \geqslant 0$,

$$\mathbb{P}(\mathrm{dist}(R_i, R_{-i}) \leqslant u) \leqslant \mathbb{P}(|R_i \cdot Y_i| \leqslant u) = \int_{\mathbb{S}^{n-1}} \mathbb{P}(|R_i \cdot y| \leqslant u)\, d\nu_i(y)$$

Let us first consider the case where $X_{11}$ has a bounded density $\varphi$ on $\mathbb{C}$. Since $\|y\|_2 = 1$ there exists an index $j_0 \in \{1, \ldots, n\}$ such that $y_{j_0} \neq 0$ with $|y_{j_0}|^{-1} \leqslant \sqrt{n}$. The complex random variable $R_i \cdot y$ is a sum of independent complex random variables and one of them is $a_n^{-1} X_{ij_0}\, \overline{y_{j_0}}$, which is absolutely continuous with a density bounded above by $a_n\sqrt{n}\, \|\varphi\|_\infty$. Consequently, by a basic property of convolutions of probability measures, the complex random variable $R_i \cdot y$ is also absolutely continuous with a density $\varphi_i$ bounded above by $a_n\sqrt{n}\, \|\varphi\|_\infty$, and thus

$$\mathbb{P}(|R_i \cdot y| \leqslant u) = \int_{\mathbb{C}} \mathbb{1}_{|s|\leqslant u} \varphi_i(s)\, ds \leqslant \pi u^2\, a_n\sqrt{n}\, \|\varphi\|_\infty.$$

Therefore, for every $b > 0$,

$$\mathbb{P}(s_n(A - zI) \leqslant n^{-b-1/2}) = O(n^{3/2-2b} a_n)$$

where the $O$ does not depend on $z$. By taking $b$ large enough, the first Borel-Cantelli lemma implies that there exists $r > 0$ such that a.s. for every $z \in \mathbb{C}$ and $n \gg 1$,

$$s_n(A - zI) \geqslant n^{-r}.$$

It remains to consider the case where $X_{11}$ has a bounded density $\varphi$ on $\mathbb{R}$. As for the complex case, let us fix $y \in \mathbb{S}^{n-1}$. Since $\|y\|_2 = 1$ there exists an index $j_0 \in \{1, \ldots, n\}$ such that $|y_{j_0}|^{-1} \leqslant \sqrt{n}$. Also, either $|\mathfrak{Re}(y_{j_0})|^{-1} \leqslant \sqrt{2n}$ or $|\mathfrak{Im}(y_{j_0})|^{-1} \leqslant \sqrt{2n}$. Assume for instance that $|\mathfrak{Re}(y_{j_0})|^{-1} \leqslant \sqrt{2n}$. We observe that for every $u \geqslant 0$,

$$\mathbb{P}(|R_i \cdot y| \leqslant u) \leqslant \mathbb{P}(|\mathfrak{Re}(R_i \cdot y)| \leqslant u).$$

The real random variable $\mathfrak{Re}(R_i \cdot y)$ is a sum of independent real random variables and one of them is $a_n^{-1} X_{ij_0} \mathfrak{Re}(y_{j_0})$, which is absolutely continuous with a density bounded above by $a_n \sqrt{2n}\, \|\varphi\|_\infty$. Consequently, by a basic property of convolutions of probability measures, the real random variable $\mathfrak{Re}(R_i \cdot y)$ is also absolutely continuous with a density $\varphi_i$ bounded above by $a_n \sqrt{2n}\, \|\varphi\|_\infty$. Therefore, we have for every $u \geqslant 0$,

$$\mathbb{P}(|\mathfrak{Re}(R_i \cdot y)| \leqslant u) = \int_{[-u,u]} \varphi_i(s)\, ds \leqslant 2^{3/2} a_n \sqrt{n}\, u\, \|\varphi\|_\infty.$$

We skip the rest of the proof, which is identical to the complex case. □

3.3. **Distance from a row to a vector space.** In this paragraph, we give two lower bounds on the distance of a row of the random matrix $A - z$ defined by (1.4) to a vector space of not too large dimension. The first ingredient is an adaptation of Proposition 5.1 in Tao and Vu [50].

**Proposition 3.3** (Distance of a row to a subspace)**.** *Assume that (H1) holds. Let $0 < \gamma < 1/2$, and let $R$ be a row of $a_n(A - z)$. There exists $\delta > 0$ depending on $\alpha, \gamma$ such that for all $d$-dimensional subspaces $W$ of $\mathbb{C}^n$ with $n - d \geqslant n^{1-\gamma}$, one has*

$$\mathbb{P}\left(\operatorname{dist}(R, W) \leqslant n^{(1-2\gamma)/\alpha}\right) \leqslant e^{-n^\delta}.$$

The proof of proposition 3.3 is based on a concentration estimate for the truncated variables $X_{1i} \mathbb{1}_{\{|X_{1i}| \leqslant b_n\}}$ for suitable sequences $b_n$. We first recall a concentration inequality of Talagrand.

**Theorem 3.4** (Talagrand concentration inequality [46] and [34, Corollary 4.10])**.** *Let us denote by $\mathbb{D} := \{z \in \mathbb{C}; |z| \leqslant 1\}$ the complex unit disc and let $P$ be a product probability measure on the product space $\mathbb{D}^n$. Let $F : \mathbb{D}^n \to \mathbb{R}$ be a Lipschitz convex function on $\mathbb{D}^n$ with $\|F\|_{\mathrm{Lip}} \leqslant 1$. If $M(F)$ is a median of $F$ under $P$ then for every $r \geqslant 0$,*

$$P\left(|F - M(F)| \geqslant r\right) \leqslant 4e^{-r^2/4}.$$

*Proof of proposition 3.3.* We first perform some pre-processing of the vector $R$ as in Tao-Vu [50]. To fix ideas, we may assume that $R$ is the first row of $a_n(A - z)$. Then $R = X_1 - z a_n e_1$ where $X_1$ is the first row of $X = a_n A$. We then have

$$\operatorname{dist}(R, W) \geqslant \operatorname{dist}(X_1 - z a_n e_1, \operatorname{span}(W, e_1)) = \operatorname{dist}(X_1, W_1).$$

where we have set $W_1 = \operatorname{span}(W, e_1)$. Note that $d \leqslant \dim W_1 \leqslant d + 1$.

For any sequence $b_n$, from the Markov inequality,

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{\{|X_{1i}| \geqslant b_n\}} \geqslant \sqrt{n}\right) &\leqslant e^{-\sqrt{n}} \left(\mathbb{E} e^{\mathbb{1}_{|X_{11}| \geqslant b_n}}\right)^n \\
&\leqslant e^{-\sqrt{n}} \left(1 + eL(b_n) b_n^{-\alpha}\right)^n \\
&\leqslant e^{-\sqrt{n} + enL(b_n) b_n^{-\alpha}}. \qquad (3.4)
\end{aligned}$$

Choose $b_n = a_n n^{-2\gamma/\alpha}$. Clearly, $b_n / n^{(1-2\gamma)/\alpha} \in [n^{-\varepsilon}, n^{\varepsilon}]$ eventually for all $\varepsilon > 0$.

Let $\mathcal{J}$ denote the set of indexes $i$ such that $|X_{1i}| \leqslant b_n$. From (3.4) we see that, for some $\delta > 0$:

$$\mathbb{P}(|\mathcal{J}| < n - \sqrt{n}) \leqslant e^{-n^\delta}.$$

It follows that it is sufficient to prove the statement conditioned on the event $\{|\mathcal{J}| \geqslant n - \sqrt{n}\}$. In particular, we shall prove that for any fixed $I \subset \{1, \ldots, n\}$, such that $|I| \geqslant n - \sqrt{n}$,

$$\mathbb{P}\left(\operatorname{dist}(X_1, W_1) \leqslant n^{(1-2\gamma)/\alpha} \,|\, \mathcal{J} = I\right) \leqslant e^{-n^\delta}. \qquad (3.5)$$

Without loss of generality, we assume that $I = \{1, \cdots, n'\}$ with $n' \geqslant n - \sqrt{n}$. Let $\pi_I$ be the orthogonal projection on $\mathrm{span}(e_i : i \in I)$. If $W_2 = \pi_I(W_1)$, we find $d - \sqrt{n} \leqslant \dim(W_2) \leqslant \dim(W_1) \leqslant d + 1$ and

$$\mathrm{dist}(X_1, W_1) \geqslant \mathrm{dist}(\pi_I(X_1), W_2).$$

Note that $\pi_I(X_1)$ is simply the vector $X_{1i}$, $i = 1, \ldots, n'$. We set

$$W' = \mathrm{span}(W_2, \mathbb{E}[\pi_I(X_1) \,|\, \mathcal{J} = I])\,, \quad Y = \pi_I(X_1) - \mathbb{E}[\pi_I(X_1) \,|\, \mathcal{J} = I]\,,$$

so that $d - \sqrt{n} \leqslant \dim(W') \leqslant d + 2$ and

$$\mathrm{dist}(\pi_I(X_1), W_2) \geqslant \mathrm{dist}(Y, W').$$

Let $P$ denote the orthogonal projection matrix to the orthogonal complement of $W'$ in $\mathbb{C}^{n'}$. We have $\mathrm{dist}^2(Y, W') = \sum_{i,j} Y_i P_{ij} \bar{Y}_j$, and, since $Y = (Y_i)_{1 \leqslant i \leqslant n'}$ is a mean zero vector under $\mathbb{P}(\cdot \,|\, \mathcal{I} = I)$,

$$\begin{aligned}\mathbb{E}[\mathrm{dist}^2(Y, W') \,|\, \mathcal{J} = I] &= \mathbb{E}\Big[\sum_{i,j} Y_i P_{ij} \bar{Y}_j \,|\, \mathcal{J} = I\Big] \\ &= \sum_{i=1}^{n'} P_{ii} \mathbb{E}[|Y_i|^2 \,|\, \mathcal{J} = I] = \mathbb{E}[|Y_1|^2 \,|\, \mathcal{J} = I]\, \mathrm{tr} P.\end{aligned}$$

We have for any $\varepsilon > 0$ and for $n \gg 1$:

$$\mathbb{E}[|Y_1|^2 \,|\, \mathcal{J} = I] = \mathbb{E}[|X_{11}|^2 \,|\, \mathcal{J} = I] - (\mathbb{E}[|X_{11}| \,|\, \mathcal{J} = I])^2 \geqslant b_n^{2-\alpha}\, n^{-\varepsilon}\,,$$

where the last bound follows from lemma C.1, since by independence one has

$$\mathbb{E}[|X_{11}|^2 \,|\, \mathcal{J} = I] = \mathbb{E}[|X_{11}|^2 \,|\, |X_{11}| \leqslant b_n]\,,$$

and $|\mathbb{E}[X_{11} \,|\, \mathcal{J} = I]|^2 = |\mathbb{E}[X_{11} \,|\, |X_{11}| \leqslant b_n]|^2$ is $O(1)$ if $\alpha > 1$, while (by lemma C.1) it is $O(b_n^{2-2\alpha+\varepsilon})$ for any $\varepsilon > 0$, if $\alpha \in (0, 1]$.

Using $\mathrm{tr} P = n' - \dim(W') \geqslant \frac{1}{2}(n - d)$, it follows that, for any $\varepsilon > 0$, for $n \gg 1$:

$$\mathbb{E}[\mathrm{dist}^2(Y, W') \,|\, \mathcal{J} = I] \geqslant c L(b_n) b_n^{2-\alpha} (n - d) \geqslant n^{q(\varepsilon)}\,, \tag{3.6}$$

where $q := (1 - 2\gamma)\frac{2}{\alpha} + \gamma - \varepsilon$.

Under $\mathbb{P}(\cdot \,|\, \mathcal{J} = I)$, the vector $(Y_1/b_n, \cdots, Y_{n'}/b_n)$ is a vector of independent variables on $\mathbb{D}^{n'}$, where $\mathbb{D}$ be the unit complex ball. We consider the function $F : x \mapsto \mathrm{dist}(x, W')$. The mapping $F$ is 1-Lipschitz and convex. From theorem 3.4, we deduce that

$$\mathbb{P}(|\mathrm{dist}(Y, W') - M(\mathrm{dist}(Y, W'))| \geqslant r \,|\, \mathcal{J} = I) \leqslant 4 e^{-\frac{r^2}{8 b_n^2}} \tag{3.7}$$

where $M(\mathrm{dist}(Y, W'))$ is a median of $\mathrm{dist}(Y, W')$ under $\mathbb{P}(\cdot \,|\, \mathcal{J} = I)$.

It follows that, for e.g. $\delta = \gamma/2$, taking $\varepsilon = \gamma/4$ in (3.6), we obtain $q(\varepsilon) = (1 - 2\gamma)\frac{2}{\alpha} + \delta + \varepsilon$, and therefore there exists $c > 0$ such that $n \gg 1$,

$$b_n^{-2}\, \mathbb{E}[\mathrm{dist}^2(Y, W') \,|\, \mathcal{J} = I] \geqslant c\, \frac{n^{q(\varepsilon)}}{b_n^2} \geqslant c\, n^{\delta}\,. \tag{3.8}$$

From (3.7) it follows that

$$\mathbb{E}\left[|M(\mathrm{dist}(Y, W')) - \mathrm{dist}(Y, W')|^2 \,\,|\, \mathcal{J} = I\right] = O\left(b_n^2\right)$$

From the Cauchy-Schwarz inequality we then have

$$\begin{aligned}&\left| M(\mathrm{dist}(Y, W')) - \sqrt{\mathbb{E}[\mathrm{dist}^2(Y, W') \,|\, \mathcal{J} = I]} \right|^2 \\ &\qquad \leqslant \mathbb{E}\left[|M(\mathrm{dist}(Y, W')) - \mathrm{dist}(Y, W')|^2 \,\,|\, \mathcal{J} = I\right] = O\left(b_n^2\right)\,.\end{aligned}$$

The above estimates, with (3.6) and (3.8), imply that $M(\mathrm{dist}(Y, W')) \geqslant \frac{1}{2} n^{q(\varepsilon)/2}$ for $n \gg 1$. Therefore, for $n \gg 1$,

$$\begin{aligned}&\mathbb{P}\left(\mathrm{dist}(Y, W') \leqslant n^{(1-2\gamma)/\alpha} \,|\, \mathcal{J} = I\right) \\ &\qquad \leqslant \mathbb{P}\left(|M(\mathrm{dist}(Y, W')) - \mathrm{dist}(Y, W')| \geqslant \frac{1}{4}\, n^{q(\varepsilon)/2} \,|\, \mathcal{J} = I\right)\,.\end{aligned}$$

The desired conclusion (3.5) now follows from (3.7) and (3.8). □

So far we have shown that under assumption (H1), the distance of a row to a space with codimension $n-d \geqslant n^{1-\gamma}$ is at least $n^{(1-2\gamma)/\alpha}$ with large probability. We want a sharper estimate, namely at the order $n^{1/\alpha}$. We will obtain such a bound in a weak sense in the forthcoming proposition 3.7. Furthermore, we shall require assumption (H2) to do so. We start with some preliminary facts.

Below we write $Z = Z^{(\beta)}$, $\beta \in (0,1)$, for the one-sided $\beta$-stable distribution such that for all $s \geqslant 0$,

$$\mathbb{E}\exp(-sZ_i) = \exp(-s^\beta)\,.$$

From the standard inversion formula, for $m > 0$

$$y^{-m} = \Gamma(m)^{-1}\int_0^\infty x^{m-1}\,e^{-x\,y}\,dx\,,$$

we see that all moments

$$\mathbb{E}[Z^{-m}] = \Gamma(m)^{-1}\int_0^\infty x^{m-1}\,e^{-x^\beta}\,dx \tag{3.9}$$

are finite for $m > 0$. Also, recall that if $(Z_i)_{1\leqslant i\leqslant n}$ is an i.i.d. vector with distribution $Z$ then, for every $(w_i)_{1\leqslant i\leqslant n} \in \mathbb{R}_+^n$, in distribution

$$\sum_{i=1}^n w_i Z_i \stackrel{d}{=} \left(\sum_{i=1}^n w_i^\beta\right)^{1/\beta} Z_1\,. \tag{3.10}$$

Indeed, (3.10) follows from $\mathbb{E}\exp(-s\sum w_i Z_i) = \exp(-s^\beta \sum w_i^\beta)$ and a change of variables.

**Lemma 3.5.** *Assume (H2). There exists $\varepsilon > 0$ and $p \in (0,1)$ such that the random variable $|X_{11}|^2$ dominates stochastically the random variable $\varepsilon\,D\,Z$, where $\mathbb{P}(D=1) = 1-\mathbb{P}(D=0) = p$ is a random variable with law $\mathrm{Be}(p)$ , $Z = Z^{(\beta)}$ with $\beta = \frac{\alpha}{2}$, and $D$ and $Z$ are independent.*

*Proof.* From our assumptions, there exist $\delta > 0$ and $x_0 > 0$ such that

$$\mathbb{P}(|X_{11}|^2 > x) \geqslant \delta\,x^{-\beta} \geqslant \mathbb{P}(\delta^2\,Z > x)\,, \quad x > x_0\,.$$

Let $p$ be the probability that $|X_{11}|^2 > x_0$. If $x > x_0$ then $\mathbb{P}(|X_{11}|^2 > x) \geqslant p\,\mathbb{P}(\delta^2\,Z > x) = \mathbb{P}(\delta^2\,D\,Z > x)$. On the other hand, if $x \leqslant x_0$ then $\mathbb{P}(|X_{11}|^2 > x) \geqslant p \geqslant \mathbb{P}(\delta^2\,D\,Z > x)$. In any case, setting $\varepsilon = \delta^2$ we have

$$\mathbb{P}(|X_{11}|^2 > x) \geqslant \mathbb{P}(\varepsilon\,D\,Z > x)\,, \quad x > 0.$$

This implies the lemma. ☐

**Lemma 3.6.** *Assume (H2). Let $\omega_i \in [0,1]$ be numbers such that $\omega(n) := \sum_{i=1}^n \omega_i \geqslant n^{\frac12+\varepsilon}$ for some $\varepsilon > 0$. Let $X_1 = (X_{1i})_{1\leqslant i\leqslant n}$ be i.i.d. random variables distributed as $X_{11}$, and let $Z = Z^{(\beta)}$ with $\beta = \frac{\alpha}{2}$. There exist $\delta > 0$ and a coupling of $X_1$ and $Z$ such that*

$$\mathbb{P}\left(\sum_{i=1}^n \omega_i |X_{1i}|^2 \leqslant \delta\,\omega(n)^{\frac{1}{\beta}} Z\right) \leqslant e^{-n^\delta}\,. \tag{3.11}$$

*Proof.* Let $D = (D_i)_{1\leqslant i\leqslant n}$ denote an i.i.d. vector of Bernoulli variables with parameter $p$ given by lemma 3.5. From this latter lemma and (3.10) we know that there exist $\varepsilon > 0$ and a coupling of $X_1$, $D$ and $Z$ such that

$$\mathbb{P}\left(\sum_{i=1}^n \omega_i |X_{1i}|^2 \geqslant \varepsilon\Big(\sum_{i=1}^n \omega_i^\beta\,D_i\Big)^{\frac{1}{\beta}} Z\right) = 1\,.$$

It remains to show that for some $\varepsilon' > 0$:

$$\mathbb{P}\left(\sum_{i=1}^n \omega_i^\beta\,D_i \leqslant \varepsilon'\,\omega(n)\right) \leqslant e^{-n^{\varepsilon'}}\,.$$

Observe that $\omega_i^\beta \geqslant \omega_i$, so that $\mathbb{E}\sum_{i=1}^n \omega_i^\beta D_i \geqslant p\,\omega(n)$. Therefore, for $0<\varepsilon'<p$,

$$\mathbb{P}\left(\sum_{i=1}^n \omega_i^\beta D_i \leqslant \varepsilon'\,\omega(n)\right)$$

$$\leqslant \mathbb{P}\left(\Big|\sum_{i=1}^n \left(\omega_i^\beta D_i - \mathbb{E}\omega_i^\beta D_i\right)\Big| \geqslant (p-\varepsilon')\,\omega(n)\right) \leqslant 2e^{-2(p-\varepsilon')^2\omega(n)^2/n},$$

where we have used the Hoeffding inequality in the last bound. Since $\omega(n)\geqslant n^{\frac12+\varepsilon}$, this implies the lemma. □

**Proposition 3.7.** *Assume (H2) and take $0<\gamma\leqslant\alpha/4$. Let $R$ be the first row of the matrix $a_n(A-z)$. There exists a constant $c>0$ and an event $E$ such that for any $d$-dimensional subspace $W$ of $\mathbb{C}^n$ with codimension $n-d\geqslant n^{1-\gamma}$, we have*

$$\mathbb{E}[\mathrm{dist}^{-2}(R,W)\,;\,E]\leqslant c\,(n-d)^{-\frac{2}{\alpha}} \quad \text{and} \quad \mathbb{P}(E^c)\leqslant c\,n^{-(1-2\gamma)/\alpha}.$$

*Proof.* As in the proof of proposition 3.3, we have

$$\mathrm{dist}(R,W)\geqslant \mathrm{dist}(X_1,W_1),$$

where $W_1=\mathrm{span}(W,e_1)$, $d\leqslant\dim W_1\leqslant d+1$, and $X_1=(X_{1i})_{1\leqslant i\leqslant n}$ is the first row of $X=a_nA$. Let $\mathcal{I}$ denote the set of indexes $i$ such that $|X_{1i}|\leqslant a_n$. From (3.4) we know that

$$\mathbb{P}(|\mathcal{I}|<n-\sqrt{n})<e^{-n^\delta},$$

for some $\delta>0$. It is thus sufficient to prove that for any set $I\subset\{1,\dots,n\}$ such that $|I|\geqslant n-\sqrt{n}$,

$$\mathbb{E}[\mathrm{dist}^{-2}(R,W)\,;\,E_I\,|\,\mathcal{I}=I]\leqslant c\,(n-d)^{-\frac{2}{\alpha}},$$

for some event $E_I$ satisfying $\mathbb{P}((E_I)^c\,|\,\mathcal{I}=I)\leqslant n^{-(1-2\gamma)/\alpha}$. We will then simply set

$$E=E_{\mathcal{I}}\cap\{|\mathcal{I}|\geqslant n-\sqrt{n}\}.$$

Without loss of generality, we assume that $I=\{1,\cdots,n'\}$ with $n'\geqslant n-n^{1/2}$. Let $\pi_I$ be the orthogonal projection on $\mathrm{span}(e_i : i\in I)$. If $W_2=\pi_I(W_1)$, set

$$W'=\mathrm{span}\left(W_2,\mathbb{E}(\pi_I(X_1)\,|\,\mathcal{I}=I)\right).$$

Note that $d-\sqrt{n}\leqslant\dim(W')\leqslant\dim(W_1)+1\leqslant d+2$. Defining

$$Y=\pi_I(X_1)-\mathbb{E}(\pi_I(X_1)\,|\,\mathcal{I}=I),$$

we have

$$\mathrm{dist}(R,W)\geqslant\mathrm{dist}(X_1,W_1)\geqslant\mathrm{dist}(Y,W').$$

Thus, $Y=(Y_i)_{1\leqslant i\leqslant n'}$ is an i.i.d. mean zero vector under $\mathbb{P}(\cdot\,|\,\mathcal{I}=I)$. Let $P$ denote the orthogonal projection matrix to the orthogonal of $W'$ in $\mathbb{C}^{n'}$. By construction, we have

$$\mathbb{E}\left(\mathrm{dist}^2(Y,W')\,|\,\mathcal{I}=I\right)=\mathbb{E}\Big(\sum_{i,j=1}^{n'}Y_iP_{ij}\bar{Y}_j\,|\,\mathcal{I}=I\Big)=\mathbb{E}\left[|Y_1|^2\,|\,\mathcal{I}=I\right]\,\mathrm{tr}P.$$

Here $\mathrm{tr}P=\sum_{i=1}^{n'}P_{ii}$, where $P_{ii}=(e_i,Pe_i)\in[0,1]$ and $\mathrm{tr}P=n'-\dim(W')$ satisfies

$$2(n-d)\geqslant\mathrm{tr}P\geqslant\frac12(n-d). \tag{3.12}$$

Let $S=\sum_{i=1}^{n'}P_{ii}|Y_i|^2$. We have

$$\begin{aligned}
\mathbb{E}\left((\mathrm{dist}^2(Y,W')-S)^2\,|\,\mathcal{I}=I\right) &= \mathbb{E}\left(\Big(\sum_{i\neq j}Y_iP_{ij}\bar{Y}_j\Big)^2\,|\,\mathcal{I}=I\right)\\
&= \sum_{(i_1\neq j_1),(i_2\neq j_2)}P_{i_1j_1}P_{i_2j_2}\mathbb{E}\left(Y_{i_1}\bar{Y}_{j_1}Y_{i_2}\bar{Y}_{j_2}\,|\,\mathcal{I}=I\right)\\
&= 2\sum_{i_1\neq j_1}P_{i_1j_1}^2\mathbb{E}[|Y_1|^2\,|\,\mathcal{I}=I]\\
&\leqslant 2\mathbb{E}[|Y_1|^2\,|\,\mathcal{I}=I]\,\mathrm{tr}P^2.
\end{aligned}$$

Note that,

$$\begin{aligned}\mathbb{E}[|Y_1|^2 \,|\, \mathcal{I} = I] &\leqslant \mathbb{E}[|X_{11}|^2 \,|\, \mathcal{I} = I] \\ &= \mathbb{E}[|X_{11}|^2 \,|\, |X_{11}| \leqslant a_n] \\ &\leqslant \frac{\mathbb{E}[|X_{11}|^2 \,;\, |X_{11}| \leqslant a_n]}{\mathbb{P}(|X_{11}| \leqslant a_n)} = O(a_n^2/n)\,,\end{aligned}$$

where the last bound follows from lemma C.1. Since $P^2 = P$, we deduce that

$$\mathbb{E}\left[(\mathrm{dist}^2(Y, W') - S)^2 \,|\, \mathcal{I} = I\right] = O\left(a_n^2 \, \frac{n-d}{n}\right). \tag{3.13}$$

Next, let $Z = Z^{(\beta)}$ with $\beta = \frac{\alpha}{2}$, as in lemma 3.6. Set $\omega_i = P_{ii}$, $i = 1, \dots n'$, and for $\varepsilon > 0$, consider the event

$$\Gamma_I = \left\{ \sum_{i=1}^{n'} \omega_i |X_{1i}|^2 \geqslant \varepsilon\,(n-d)^{\frac{1}{\beta}} Z \right\}.$$

From lemma 3.6 (with $n$ replaced by $n' \geqslant n - n^{1/2}$) and using (3.12) there exists a coupling of the vector $X_{1i}, i = 1, \dots, n'$ and $Z$ such that

$$\mathbb{P}(\Gamma_I^c) \leqslant e^{-n^\delta}\,, \tag{3.14}$$

for some $\delta > 0$ and some choice of $\varepsilon > 0$. Also, since $(a-b)^2 \geqslant a^2/2 - b^2$ for all $a, b \in \mathbb{R}$, we have $S \geqslant \frac{1}{2}\, S_a - \, S_b$, where

$$S_a = \sum_{i=1}^{n'} \omega_i |X_{1i}|^2\,, \quad S_b = \sum_{i=1}^{n'} \omega_i \, \mathbb{E}\left[|X_{1i}| \,|\, |X_{1i}| \leqslant \, a_n\right]^2.$$

From Lemma C.1 and (3.12) we have

$$S_b = \mathbb{E}\left[|X_{11}| \,|\, |X_{11}| \leqslant a_n\right]^2 \, \mathrm{tr}P = h^{(\alpha)}(n, d) \tag{3.15}$$

where $h^{(\alpha)}(n,d) \sim (n-d)a_n^2/n^2$ if $\alpha \in (0,1]$ and $h^{(\alpha)}(n,d) \sim (n-d)$ if $\alpha \in (1,2)$. Let $G_I^1$ be the event that $S_a \geqslant 3\, S_b$. From (3.15) and the definition of $\Gamma_I$ we have, for some $c_0 > 0$

$$\mathbb{P}((G_I^1)^c \cap \Gamma_I \,|\, \mathcal{I} = I) \leqslant \mathbb{P}(Z \leqslant c_0 (n-d)^{-1/\beta} h^{(\alpha)}(n,d) \,|\, \mathcal{I} = I)\,.$$

Note that, thanks to the assumptions $n - d \geqslant n^{1-\gamma}$, $\gamma \leqslant \alpha/4$, we have $(n-d)^{-1/\beta} h^{(\alpha)}(n,d) \leqslant n^{-\varepsilon_0}$ for some $\varepsilon_0 = \varepsilon_0(\alpha) > 0$ for all $\alpha \in (0,2)$, for $n \gg 1$. Therefore, for $n \gg 1$,

$$\begin{aligned}\mathbb{P}((G_I^1)^c \cap \Gamma_I \,|\, \mathcal{I} = I) &\leqslant \mathbb{P}(Z \leqslant c_0 \, n^{-\varepsilon_0} \,|\, \mathcal{I} = I) \\ &= \frac{\mathbb{P}(Z \leqslant c_0 \, n^{-\varepsilon_0};\ |X_{1i}| \leqslant a_n\,,\ \forall i = 1, \dots, n')}{\mathbb{P}(|X_{1i}| \leqslant a_n\,,\ \forall i = 1, \dots, n')}\,,\end{aligned}$$

where the last identity follows from the independence of the $X_{1i}$. Observing that the probability for the event $\{|X_{1i}| \leqslant a_n\,,\ \forall i = 1, \dots, n'\}$ is lower bounded by $1/c > 0$ uniformly in $n$, we obtain

$$\mathbb{P}((G_I^1)^c \cap \Gamma_I \,|\, \mathcal{I} = I) \leqslant c \, \mathbb{P}(Z \leqslant c_0 \, n^{-\varepsilon_0})\,.$$

The latter probability can be estimated using Markov's inequality and the fact that $\mathbb{E}[Z^{-m}] = u_m$ is finite (cf. (3.9)). Indeed, for every $m > 0$, $\mathbb{P}(Z \leqslant t) \leqslant \, u_m t^{-m}$. Thus, we have shown that for every $p > 0$ there exists a constant $\kappa_p$ such that

$$\mathbb{P}((G_I^1)^c \cap \Gamma_I \,|\, \mathcal{I} = I) \leqslant \kappa_p \, n^{-p}\,. \tag{3.16}$$

Next, we set $\widetilde{\Gamma}_I = G_I^1 \cap \Gamma_I$ and we claim that

$$\mathbb{E}\left[S^{-2}\,;\, \widetilde{\Gamma}_I \,|\, \mathcal{I} = I\right] = O\left(\, (n-d)^{-4/\alpha} \right), \tag{3.17}$$

Indeed, on $\widetilde{\Gamma}_I$ we have $S \geqslant \frac{1}{6}\, S_a \geqslant \frac{\varepsilon}{6}(n-d)^{2/\alpha}\, Z$ and therefore, for some constant $c_1$,

$$\mathbb{E}\left[S^{-2}\,;\, \widetilde{\Gamma}_I \,|\, \mathcal{I} = I\right] \leqslant c_1 \, (n-d)^{-4/\alpha} \mathbb{E}\left[Z^{-2} \,|\, \mathcal{I} = I\right].$$

Using independence as before, and recalling that the event $\{|X_{1i}| \leqslant a_n\,,\ \forall i = 1, \dots, n'\}$ has uniformly positive probability we have

$$\mathbb{E}\left[Z^{-2} \,|\, \mathcal{I} = I\right] \leqslant c \, \mathbb{E}[Z^{-2}] = c\, u_2\,.$$

This proves (3.17).

Now, for the event Markov's and Cauchy-Schwarz' inequalities lead to

$$\begin{aligned}\mathbb{P}\left(\mathrm{dist}^2(Y,W')\leqslant S/2\,;\,\widetilde{\Gamma}_I\,|\,\mathcal{I}=I\right) &\leqslant \mathbb{P}\left(\frac{|\mathrm{dist}^2(Y,W')-S|}{S}\geqslant 1/2\,;\,\widetilde{\Gamma}_I\,|\,\mathcal{I}=I\right)\\ &\leqslant 2\mathbb{E}\left[\frac{|\mathrm{dist}^2(Y,W')-S|}{S}\,;\,\widetilde{\Gamma}_I\,|\,\mathcal{I}=I\right]\\ &\leqslant 2\sqrt{\mathbb{E}\left[|\mathrm{dist}^2(Y,W')-S|^2\,|\,\mathcal{I}=I\right]\mathbb{E}\left[S^{-2}\,;\,\widetilde{\Gamma}_I\,|\,\mathcal{I}=I\right]}.\end{aligned}$$

Hence, if $G_I^2$ denotes the event $\{\mathrm{dist}^2(Y,W')\geqslant S/2\}$, we deduce from (3.13) and (3.17)

$$\mathbb{P}\left((G_I^2)^c\cap\widetilde{\Gamma}_I\,|\,\mathcal{I}=I\right)=O\left(a_n n^{-\frac12}(n-d)^{\frac12-\frac2\alpha}\right). \tag{3.18}$$

Note that, using $n-d\geqslant n^{1-\gamma}$, the last expression is certainly $O(n^{-\frac1\alpha(1-2\gamma)})$. On the other hand, by (3.17) and Cauchy-Schwarz' inequality

$$\mathbb{E}\left[\mathrm{dist}^{-2}(X,W)\,;\,G_I^2\cap\widetilde{\Gamma}_I\,|\,\mathcal{I}=I\right]\leqslant 2\,\mathbb{E}\left[S^{-1}\,;\,\widetilde{\Gamma}_I\,|\,\mathcal{I}=I\right]=O\left((n-d)^{-2/\alpha}\right). \tag{3.19}$$

To conclude the proof we take $E_I=G_I^2\cap\widetilde{\Gamma}_I=G_I^1\cap G_I^2\cap\Gamma_I$. We have

$$\mathbb{P}((E_I)^c\,|\,\mathcal{I}=I)\leqslant\mathbb{P}\left((\Gamma_I)^c\,|\,\mathcal{I}=I\right)+\mathbb{P}\left((G_I^1)^c\cap\Gamma_I\,|\,\mathcal{I}=I\right)+\mathbb{P}\left((G_I^2)^c\cap G_I^1\cap\Gamma_I\,|\,\mathcal{I}=I\right).$$

From (3.16) and (3.18) we see that,

$$\mathbb{P}\left((G_I^1)^c\cap\Gamma_I\,|\,\mathcal{I}=I\right)+\mathbb{P}\left((G_I^2)^c\cap G_I^1\cap\Gamma_I\,|\,\mathcal{I}=I\right)=O\left(n^{-\frac1\alpha(1-2\gamma)}\right),$$

and all it remains to prove is an upper bound on $\mathbb{P}\left((\Gamma_I)^c\,|\,\mathcal{I}=I\right)$. By independence, as before

$$\mathbb{P}\left((\Gamma_I)^c\,|\,\mathcal{I}=I\right)\leqslant c\,\mathbb{P}\left((\Gamma_I)^c\,;\,|X_{1i}|\leqslant a_n\,,\ \forall i=1,\dots,n'\right).$$

From (3.14) we obtain $\mathbb{P}\left((\Gamma_I)^c\,|\,\mathcal{I}=I\right)\leqslant c\,e^{-n^\delta}$. This ends the proof. □

3.4. **Uniform integrability.** Let $z\in\mathbb{C}$ and $\sigma_n\leqslant\cdots\leqslant\sigma_1$ be the singular values of $A_n-z$ with $A_n$ defined by (1.4). For $0<\delta<1$, we define $K_\delta=[\delta,\delta^{-1}]$. In this paragraph, we prove the uniform integrability in probability, meaning that for all $\varepsilon>0$, there exists $\delta>0$ such that

$$\mathbb{P}\left(\int_{K_\delta^c}|\ln(x)|\nu_{A_n-z}(dx)>\varepsilon\right)\to 0. \tag{3.20}$$

From lemma 3.1, with probability 1 there exists $c_0>0$, such that for all $n$,

$$\int_1^\infty \ln^2(x)\nu_{A_n-z}(dx)<c_0.$$

It follows from Markov inequality that for all $t\geqslant 1$, $\int_t^\infty \ln(x)\nu_{A_n-z}(dx)<c_0/\ln t$. The upper part $(\delta^{-1},\infty)$ of (3.20) is thus not an issue. For the lower part $(0,\delta)$, it is sufficient to prove that

$$\frac1n\sum_{i=0}^{n-1}\mathbb{1}_{\{\sigma_{n-i}\leqslant\delta_n\}}\ln\sigma_{n-i}^{-2}$$

converges in probability to 0 for any sequence $(\delta_n)_n$ converging to 0. From lemma 3.2, we may a.s. lower bound $\sigma_{n-i}$ by $cn^{-r}$ for some constant $c$ and all integer $n\geqslant 1$. Take $0<\gamma<\alpha/4$ to be fixed later. Using this latter bound for every $1\leqslant i\leqslant n^{1-\gamma}$, it follows that it is sufficient to prove that

$$\frac1n\sum_{i=\lfloor n^{1-\gamma}\rfloor}^{n-1}\mathbb{1}_{\{\sigma_{n-i}\leqslant\delta_n\}}\ln\sigma_{n-i}^{-2}$$

converges in probability to 0. We are going to prove that there exists an event $F_n$ such that, for some $\delta>0$ and $c>0$,

$$\mathbb{P}((F_n)^c)\leqslant c\exp(-n^\delta), \tag{3.21}$$

and

$$\mathbb{E}\left[\sigma_{n-i}^{-2}\,|\,F_n\right]\leqslant c\left(\frac ni\right)^{\frac2\alpha+1}. \tag{3.22}$$

We first conclude the proof before proving (3.21)-(3.22). From Markov inequality, and (3.22), we deduce that

$$\mathbb{P}(\sigma_{n-i} \leqslant \delta_n) \leqslant \mathbb{P}((F_n)^c) + c\,\delta_n^2 \left(\frac{n}{i}\right)^{\frac{2}{\alpha}+1}.$$

If follows that there exists a sequence $\varepsilon_n = \delta_n^{1/(\frac{2}{\alpha}+1)}$ tending to 0 such that the probability that $\mathbb{P}(\sigma_{n-\lfloor n\varepsilon_n \rfloor} \leqslant \delta_n)$ converges to 0. We obtain that it is sufficient to prove that

$$\frac{1}{n}\sum_{i=\lfloor n^{1-\gamma}\rfloor}^{\lfloor \varepsilon_n n\rfloor} \ln \sigma_{n-i}^{-2}$$

given $F_n$ converges in probability to 0. However, using the concavity of the logarithm and (3.22) we have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i=\lfloor n^{1-\gamma}\rfloor}^{\lfloor \varepsilon_n n\rfloor} \ln \sigma_{n-i}^{-2} \,\middle|\, F_n\right] &\leqslant \frac{1}{n}\sum_{i=\lfloor n^{1-\gamma}\rfloor}^{\lfloor \varepsilon_n n\rfloor} \ln \mathbb{E}[\sigma_{n-i}^{-2}|F_n] \\
&\leqslant \frac{c_1}{n}\sum_{i=1}^{\lfloor \varepsilon_n n\rfloor} \ln\left(\frac{n}{i}\right) \\
&= c_1\left(-\varepsilon_n \ln \varepsilon_n + \varepsilon_n + O(n^{-1})\right).
\end{aligned}$$

It thus remain to prove (3.21)-(3.22). Let $B_n$ be the matrix formed by the first $n - \lfloor i/2 \rfloor$ rows of $a_n(A_n - zI)$. If $\sigma'_1 \geqslant \cdots \geqslant \sigma'_{n-\lfloor i/2\rfloor}$ are the singular values of $B_n$, then by the Cauchy interlacing Lemma B.4,

$$\sigma_{n-i} \geqslant \frac{\sigma'_{n-i}}{a_n}.$$

By the Tao-Vu negative second moment lemma B.3, we have

$$\sigma_1'^{-2} + \cdots + \sigma_{n-\lceil i/2\rceil}'^{-2} = \mathrm{dist}_1^{-2} + \cdots + \mathrm{dist}_{n-\lceil i/2\rceil}^{-2},$$

where $\mathrm{dist}_j$ is the distance from the $j$-th row of $B_n$ to the subspace spanned by the other rows of $B_n$. In particular,

$$\frac{i}{2}\sigma_{n-i}^{-2} \leqslant a_n^2 \sum_{j=1}^{n-\lfloor i/2\rfloor} \mathrm{dist}_j^{-2}.$$

Let $F_n$ be the event that for all $1 \leqslant j \leqslant n - \lfloor i/2 \rfloor$, $\mathrm{dist}_j \geqslant n^{(1-2\gamma)/\alpha}$. Since the dimension of the span of all but one rows of $B_n$ is at most $d \leqslant n - i/2$, we can use proposition 3.3, to obtain

$$\mathbb{P}((F_n)^c) \leqslant \exp(-n^\delta),$$

for some $\delta > 0$. Then we write

$$\frac{i}{2}\sigma_{n-i}^{-2}\,\mathbb{1}_{F_n} \leqslant a_n^2 \sum_{j=1}^{n-\lfloor i/2\rfloor} \mathrm{dist}_j^{-2}\,\mathbb{1}_{F_n},$$

Taking expectation, we get

$$\mathbb{E}\left[i\sigma_{n-i}^{-2}\,;\,F_n\right] \leqslant 2\,a_n^2 n \mathbb{E}\left[\mathrm{dist}_1^{-2}\,;\,F_n\right], \tag{3.23}$$

Since we are on $F_n$ we can always estimate $\mathrm{dist}_1 \geqslant n^{(1-2\gamma)/\alpha}$. By introducing a further decomposition we can strengthen this as follows. Recall that from proposition 3.7, there exists an event $E$ independent from the rows $j \neq 1$ such that $\mathbb{P}((E)^c) \leqslant n^{-(1-2\gamma)/\alpha}$ and for any $W \subset \mathbb{C}^n$ with dimension $d < n - n^{1-\gamma}$ one has

$$\mathbb{E}[\mathrm{dist}(R, W)^{-2}\,;\,E] \leqslant c\,(n-d)^{-2/\alpha}.$$

Here $R$ is the first row of the matrix $B_n$. By first conditioning on the value of the other rows of $B_n$ and recalling that the dimension $d$ of the span of these is at most $n - i/2 \leqslant n - 2n^{1-\gamma}$, we see that

$$\mathbb{E}[\mathrm{dist}_1^{-2}\,;\,E] = O\left(i^{-2/\alpha}\right).$$

Therefore

$$\mathbb{E}\left[\mathrm{dist}_1^{-2}\,;\,F_n\right] \leqslant \mathbb{E}(\mathrm{dist}_1^{-2}\,;\,E) + \mathbb{P}((E)^c)\,n^{-2(1-2\gamma)/\alpha}$$
$$\leqslant c_2\left(i^{-2/\alpha} + n^{-3(1-2\gamma)/\alpha}\right). \quad (3.24)$$

Now, if $\gamma < 1/6$ we have $3(1-2\gamma)/\alpha > 2/\alpha$ and therefore $n^{-3(1-2\gamma)/\alpha} \leqslant i^{-2/\alpha}$. Thus, (3.24) implies

$$\mathbb{E}\left[\mathrm{dist}_1^{-2}\,;\,F_n\right] \leqslant 2\,c_2\,i^{-2/\alpha}. \quad (3.25)$$

From (3.23) we obtain

$$\mathbb{E}\left[i\sigma_{n-i}^{-2}\,;\,F_n\right] \leqslant 2\,c_2\,a_n^2\,n\,i^{-2/\alpha}.$$

From (H2) it follows that (3.22) holds. This concludes the proof of (3.21)-(3.22).

3.5. **Proof of theorem 1.2.** We may now invoke theorem 1.1 and (3.20). From lemma A.2, $\mu_{A_n}$ converges in probability to $\mu_\alpha$, where for almost all $z \in \mathbb{C}$,

$$U_{\mu_\alpha}(z) = \int \ln(x)\nu_{\alpha,z}(dx).$$

Let us upgrade this convergence to an a.s. convergence. By lemmas 3.1 and A.1, it is sufficient to prove that for every $z \in \mathbb{C}$, a.s.

$$\lim_{n\to\infty} U_{\mu_{A_n}}(z) = U_{\mu_a}(z).$$

Let us fix $z \in \mathbb{C}$ from now on. Since $L = U_{\mu_a}(z)$ is *deterministic*, it actually suffices to show that there exists a *deterministic* sequence $L_n$ such that a.s.

$$\lim_{n\to\infty}\left(U_{\mu_{A_n}}(z) - L_n\right) = 0. \quad (3.26)$$

Now, by lemmas 3.1 and 3.2, there exists $b > 0$ such that a.s. for $n \gg 1$,

$$\mathrm{supp}(\nu_{A_n - zI}) \subset [s_n(A_n - zI), s_1(A_n - zI)] \subset [n^{-b}, n^b].$$

Denoting $f_n : x \in \mathbb{R}_+ \mapsto f_n(x) = \mathbf{1}_{[n^{-b},n^b]}(x)\log(x)$, we get that a.s. for $n \gg 1$,

$$U_{\mu_{A_n}}(z) = -\int_0^\infty \log(s)\,d\nu_{A_n-zI}(s) = -\int_0^\infty f_n(s)\,d\nu_{A_n-zI}(s). \quad (3.27)$$

The total variation of $f_n$ is bounded by $c\log n$ for some $c > 0$. Hence by lemma C.2, if

$$L_n := \mathbb{E}\int f_n(s)d\nu_{A_n-zI}(s),$$

then we have, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\int f_n(s)d\nu_{A_n-zI}(s) - L_n\right| \geqslant \varepsilon\right) \leqslant 2\exp\left(-2\frac{n\varepsilon^2}{(c\log n)^2}\right).$$

In particular, from the first Borel-Cantelli lemma, a.s.,

$$\lim_{n\to\infty}\left(\int f_n(s)\,d\nu_{A_n-zI}(s) - L_n\right) = 0.$$

Finally, using (3.27), we deduce that (3.26) holds almost surely, as required.

## 4. Limiting spectral measure

In this section, we give a close look to the resolvent of the random operator on the PWIT and we deduce some properties of the limiting spectral measure $\mu_\alpha$. For ease of notation we set

$$\beta = \frac{\alpha}{2}$$

and define the measure on $\mathbb{R}_+$,

$$\Lambda_\alpha = \frac{\alpha}{2}x^{-\frac{\alpha}{2}-1}dx.$$

4.1. **Resolvent operator on the Poisson Weighted Infinite Tree.** In this paragraph, we analyze the random variable

$$R(U)_{\varnothing\varnothing} = \begin{pmatrix} a(z,\eta) & b(z,\eta) \\ b'(z,\eta) & c(z,\eta) \end{pmatrix}.$$

By lemma 2.2, for $t \in \mathbb{R}_+$, $a(z,it)$ is pure imaginary and we set

$$h(z,t) = \mathfrak{Im}(a(z,it)) = -ia(z,it) \in [0, t^{-1}].$$

The random variables $a(z,\eta)$ and $h(z,t)$ solve a nice recursive distribution equation.

**Theorem 4.1** (Recursive Distributional Equation)**.** *Let $U = U(z,\eta) \in \mathbb{H}_+$, $t \in \mathbb{R}_+$. Let $L_U$ be the distribution on $\mathbb{C}_+$ of $a(z,\eta)$ and $L_{z,t}$ the distribution of $h(z,t)$.*

*(i) $L_U$ solves the equation in distribution*

$$a \overset{d}{=} \frac{\eta + \sum_{k\in\mathbb{N}} \xi_k a_k}{|z|^2 - \left(\eta + \sum_{k\in\mathbb{N}} \xi_k a_k\right)\left(\eta + \sum_{k\in\mathbb{N}} \xi'_k a'_k\right)}, \tag{4.1}$$

*where $a$, $(a_k)_{k\in\mathbb{N}}$ and $(a'_k)_{k\in\mathbb{N}}$ are i.i.d. with law $L_U$ independent of $\{\xi_k\}_{k\in\mathbb{N}}$, $\{\xi'_k\}_{k\in\mathbb{N}}$ two independent Poisson point processes on $\mathbb{R}_+$ with intensity $\Lambda_\alpha$.*

*(ii) $L_{z,t}$ is the unique probability distribution on $[0,\infty)$ such that*

$$h \overset{d}{=} \frac{t + \sum_{k\in\mathbb{N}} \xi_k h_k}{|z|^2 + \left(t + \sum_{k\in\mathbb{N}} \xi_k h_k\right)\left(t + \sum_{k\in\mathbb{N}} \xi'_k h'_k\right)} \tag{4.2}$$

*where $h$, $(h_k)_{k\in\mathbb{N}}$ and $(h'_k)_{k\in\mathbb{N}}$ are i.i.d. with law $L_{z,t}$, independent of $\{\xi_k\}_{k\in\mathbb{N}}$, $\{\xi'_k\}_{k\in\mathbb{N}}$ two independent Poisson point processes on $\mathbb{R}_+$ with intensity $\Lambda_\alpha$.*

*(iii) For $t = 0$ there are two probability distributions on $[0,\infty)$ solving (4.2) such that $\mathbb{E}h^{\alpha/2} < \infty$: $\delta_0$ and another denoted by $L_{z,0}$. Moreover, for the topology of weak convergence, $L_{z,t}$ converges to $L_{z,0}$ as $t$ goes to 0.*

We start with an important lemma.

**Lemma 4.2.** *For every $U = U(z,\eta) \in \mathbb{H}_+$, $\begin{pmatrix} a & b \\ b' & c \end{pmatrix}$ is equal in distribution to*

$$\frac{1}{|z|^2 - \left(\eta + \sum_{k\in\mathbb{N}} \xi_k a_k\right)\left(\eta + \sum_{k\in\mathbb{N}} \xi'_k a'_k\right)} \begin{pmatrix} \eta + \sum_{k\in\mathbb{N}} \xi_k a_k & -z \\ -\bar{z} & \eta + \sum_{k\in\mathbb{N}} \xi'_k a'_k \end{pmatrix}, \tag{4.3}$$

*where $a$, $(a_k)_{k\in\mathbb{N}}$ and $(a'_k)_{k\in\mathbb{N}}$ are i.i.d. with law $L_U$ independent of $\{\xi_k\}_{k\in\mathbb{N}}$, $\{\xi'_k\}_{k\in\mathbb{N}}$ two independent Poisson point processes on $\mathbb{R}_+$ with intensity $\Lambda_\alpha$.*

*Proof of lemma 4.2.* Consider a realization of PWIT$(2\ell_\theta)$ on the tree $T$. For $k \in \mathbb{N}$, we define $T_k$ as the subtree of $T$ spanned by $k\mathbb{N}^f$. With the notation of lemma 2.5, for $k \in \mathbb{N}$, $R_{B_k}(U) = (B_k(z) - \eta)^{-1}$ is the resolvent operator of $B_k$ and set

$$\widetilde{R}(U)_{kk} = \Pi_k R_{B_k}(U) \Pi_k^* = \begin{pmatrix} a_k & b_k \\ b'_k & c_k \end{pmatrix}.$$

Then, by lemma 2.5 and (2.13), we get

$$\begin{aligned} R(U)_{\varnothing\varnothing} &= -\left(U + \sum_{k\in\mathbb{N}} \begin{pmatrix} 0 & \varepsilon_k y_k^{-1/\alpha} \\ (1-\varepsilon_k) y_k^{-1/\alpha} & 0 \end{pmatrix} \begin{pmatrix} a_k & b_k \\ b'_k & c_k \end{pmatrix} \begin{pmatrix} 0 & (1-\varepsilon_k) y_k^{-1/\alpha} \\ \varepsilon_k y_k^{-1/\alpha} & 0 \end{pmatrix}\right)^{-1} \\ &= -\left(U + \begin{pmatrix} \sum_{k\in\mathbb{N}} (1-\varepsilon_k)|y_k|^{-2/\alpha} c_k & 0 \\ 0 & \sum_{k\in\mathbb{N}} \varepsilon_k |y_k|^{-2/\alpha} a_k \end{pmatrix}\right)^{-1} \\ &= D^{-1} \begin{pmatrix} \eta + \sum_{k\in\mathbb{N}} \varepsilon_k |y_k|^{-2/\alpha} a_k & -z \\ -\bar{z} & \eta + \sum_{k\in\mathbb{N}} (1-\varepsilon_k)|y_k|^{-2/\alpha} c_k \end{pmatrix}, \end{aligned}$$

with $D = |z|^2 - \left(\eta + \sum_{k\in\mathbb{N}} \varepsilon_k |y_k|^{-2/\alpha} a_k\right)\left(\eta + \sum_{k\in\mathbb{N}} (1-\varepsilon_k)|y_k|^{-2/\alpha} c_k\right)$.

Now the structure of the PWIT implies that (i) $a_k$ and $c_k$ have common distribution $L_U$; and (ii) the variables $(a_k, c_k)_{k\in\mathbb{N}}$ are i.i.d.. Also the thinning property of Poisson processes implies that (iii) $\{\varepsilon_k |y_k|^{-2/\alpha}\}_{k\in\mathbb{N}}$ and $\{(1-\varepsilon_k)|y_k|^{-2/\alpha}\}_{k\in\mathbb{N}}$ are independent Poisson point process with common intensity $\Lambda_\alpha$. □

The next well-known and beautiful lemma will be crucial in the computations that will follow. It is a consequence of the LePage-Woodroofe-Zinn representation of stable laws [35], see also Panchenko and Talagrand [41, Lemma 2.1].

**Lemma 4.3.** *Let $\{\xi_k\}_{k\in\mathbb{N}}$ be a Poisson process with intensity $\Lambda_\alpha$. If $(Y_k)$ is an i.i.d. sequence of non–negative random variables, independent of $\{\xi_k\}_{k\in\mathbb{N}}$, such that $\mathbb{E}[Y_1^\beta] < \infty$ then*

$$\sum_{k\in\mathbb{N}} \xi_k Y_k \stackrel{d}{=} \mathbb{E}[Y_1^\beta]^{\frac{1}{\beta}} \sum_{k\in\mathbb{N}} \xi_k \stackrel{d}{=} \mathbb{E}[Y_1^\beta]^{\frac{1}{\beta}} S,$$

*where $S$ is the positive $\beta$-stable random variable with Laplace transform for all $x \geqslant 0$,*

$$\mathbb{E}\exp(-xS) = \exp\left(-\Gamma(1-\beta)x^\beta\right). \tag{4.4}$$

*Proof of lemma 4.3.* Recall the formulas, for $y \geqslant 0$, $\eta > 0$ and $0 < \eta < 1$ respectively,

$$y^{-\eta} = \Gamma(\eta)^{-1}\int_0^\infty x^{\eta-1}e^{-xy}dx \quad \text{and} \quad y^\eta = \Gamma(1-\eta)^{-1}\eta\int_0^\infty x^{-\eta-1}(1-e^{-xy})dx. \tag{4.5}$$

From the Lévy-Khinchin formula we deduce that, with $s \geqslant 0$,

$$\begin{aligned}\mathbb{E}\exp\left(-s\sum_k \xi_k Y_k\right) &= \exp\left(\mathbb{E}\int_0^\infty (e^{-xsY_1}-1)\beta x^{-\beta-1}dx\right)\\ &= \exp\left(-\Gamma(1-\beta)s^\beta\mathbb{E}[Y_1^\beta]\right).\end{aligned}$$

□

*Proof of theorem 4.1.* Statement $(i)$ is contained in lemma 4.2. For $(ii)$, let $t > 0$ and $h$ a solution of (4.2). Then $h$ is positive and is upper bounded by $1/t$. By lemma 4.3, we may rewrite (4.2) as

$$h \stackrel{d}{=} \frac{t + \mathbb{E}[h^\beta]^{1/\beta}S}{|z|^2 + \left(t + \mathbb{E}[h^\beta]^{1/\beta}S\right)\left(t + \mathbb{E}[h^\beta]^{1/\beta}S'\right)} \tag{4.6}$$

where $S$ and $S'$ are i.i.d. variables with common Laplace transform (4.4). In particular, $\mathbb{E}[h^\beta]^{1/\beta}$ is solution of the equation in $y$:

$$y^\beta = \mathbb{E}\left(\frac{t+yS}{|z|^2 + (t+yS)(t+yS')}\right)^\beta$$

Since $t > 0$, $\mathbb{E}[h^\beta] > 0$, it follows that $\mathbb{E}[h^\beta]^{1/\beta}$ is solution of the equation in $y$:

$$1 = \mathbb{E}\left(\frac{ty^{-1}+S}{|z|^2 + (t+yS)(t+yS')}\right)^\beta. \tag{4.7}$$

For every $S, S' > 0$, the function $y \mapsto \frac{ty^{-1}+S}{|z|^2+(t+yS)(t+yS')}$ is decreasing in $y$. It follows that

$$y \mapsto \mathbb{E}\left(\frac{ty^{-1}+S}{|z|^2 + (t+yS)(t+yS')}\right)^\beta$$

is decreasing in $y$. As $y$ goes to 0 it converges to $\infty$ and as $y$ goes to infinity, it converges to 0. In particular, there is a unique point, $y_*(|z|^2, t)$ of such that (4.7) holds. This proves $(ii)$ since from (4.6), the law of $h$ is determined by $\mathbb{E}[h^\beta]^{1/\beta} = y_*(|z|^2, t)$.

For Statement $(iii)$ and $t = 0$, then $h = 0$ is a particular solution of (4.2). If $h$ is not a.s. equal to 0, then $\mathbb{E}[h^\beta]^{1/\beta} > 0$ and the argument above still works since, for every $s, s' > 0$, the function $y \mapsto \frac{s}{|z|^2+y^2ss'}$ is decreasing in $y$. We deduce the existence of a unique positive solution $y_*(|z|^2, 0)$ of (4.7). We also have the continuity of the function $t \mapsto y_*(|z|^2, t)$ on $[0, \infty)$. Finally

$$h \stackrel{d}{=} y_*(|z|^2, 0)S/(|z|^2 + y_*^2(|z|^2, 0)SS'),$$

and from (4.6), it implies the weak convergence of $L_{z,t}$ to $L_{z,0}$. □

4.2. **Density of the limiting measure.** In this paragraph, we analyze the RDE (4.3). For all $t > 0$, let $L_{z,t}$ be as in theorem 4.1. From Equation (4.6), $h$ may be expressed as

$$h \stackrel{d}{=} \frac{t + y_* S}{|z|^2 + (t + y_* S)(t + y_* S')}$$

where $S$ and $S'$ are i.i.d. variables with common Laplace transform (4.4) and $y_* := y_*(|z|^2, t)$ is the unique solution in $(0, \infty)$ of (4.7) (uniqueness is proved in theorem 4.1). We extend continuously the function $y_*(r, t)$ for $t = 0$ by defining $y_*(|z|^2, 0)$ as the unique solution in $(0, \infty)$:

$$1 = \mathbb{E}\left(\frac{S}{|z|^2 + y^2 SS'}\right)^{\beta}. \tag{4.8}$$

**Lemma 4.4.** *The function $y_* : [0, \infty)^2 \to (0, \infty)$ is $C^1$. For every $t \geqslant 0$, the mapping $r \mapsto y_*(r, t)$ is decreasing to* 0.

*Proof.* For every $t \geqslant 0$, the derivative in $y > 0$ of the function $\mathbb{E}\left(\frac{ty^{-1}+S}{|z|^2+(t+yS)(t+yS')}\right)^{\beta}$ is

$$-\beta t y^{-2} \mathbb{E}\frac{(ty^{-1} + S)^{\beta - 1}}{(|z|^2 + (t + yS)(t + yS'))^{\beta}} - \beta \mathbb{E}\frac{(ty^{-1} + S)^{\beta}(S(t + yS') + S'(t + yS))}{(|z|^2 + (t + yS)(t + yS'))^{\beta+1}}. \tag{4.9}$$

The last computation is justified since all terms are integrable, indeed we have

$$\frac{(ty^{-1} + S)^{\beta - 1}}{(|z|^2 + (t + yS)(t + yS'))^{\beta}} \leqslant \frac{y^{-\beta+1}}{(t + yS)(t + yS')^{\beta}} \leqslant \frac{y^{-2\beta}}{SS'^{\beta}}$$

and from (4.5), for all $\eta > 0$,

$$\mathbb{E}S^{-\eta} = \Gamma(\eta)^{-1} \int x^{\eta - 1} e^{-\Gamma(1-\beta)x^{\beta}} dx < \infty. \tag{4.10}$$

Similarly, for the second term of (4.9), we write

$$\begin{aligned}
\frac{(ty^{-1} + S)^{\beta}(S(t + yS') + S'(t + yS))}{(|z|^2 + (t + yS)(t + yS'))^{\beta+1}} &\leqslant y^{-1}\frac{S(t + yS') + S'(t + yS)}{(t + yS)(t + yS')^{\beta+1}} \\
&\leqslant y^{-1}\frac{S}{(t + yS)(t + yS')^{\beta}} + y^{-1}\frac{S'}{(t + yS')^{\beta+1}} \\
&\leqslant y^{-\beta-2}S'^{-\beta} + y^{-\beta-2}S'^{-\beta}
\end{aligned}$$

The expression (4.9) is finite and strictly negative for all $y > 0$. The statement follows from the implicit function theorem. □

From (4.3), for all $t > 0$,

$$b(z, it) \stackrel{d}{=} -\frac{z}{|z|^2 + (t + y_*(|z|^2, it)S)(t + y_*(|z|^2, it)S')}.$$

By lemma 4.4, we may also define

$$b(z, 0) = \lim_{t \downarrow 0} b(z, it) \stackrel{d}{=} -\frac{z}{|z|^2 + y_*^2(|z|^2, 0)SS'}.$$

For ease of notation, we set $y_*(r) = y_*(r, 0)$. Since $\partial z = 1$, $\partial |z|^2 = \bar{z}$, we deduce that

$$\begin{aligned}
-\mathbb{E}\partial b(z, 0) &= \mathbb{E}\partial \frac{z}{|z|^2 + y_*^2(|z|^2)SS'} \\
&= \mathbb{E}\left(|z|^2 + y_*^2(|z|^2)SS'\right)^{-1} - |z|^2 \mathbb{E}\left(|z|^2 + y_*^2(|z|^2)SS'\right)^{-2} \\
&\qquad -2|z|^2 y_*(|z|^2) y_*'(|z|^2) \mathbb{E}SS'\left(|z|^2 + y_*^2(|z|^2)SS'\right)^{-2} \\
&= \left(y_*^2(|z|^2) - 2|z|^2 y_*(|z|^2) y_*'(|z|^2)\right)\mathbb{E}\frac{SS'}{(|z|^2 + y_*^2(|z|^2)SS')^2}.
\end{aligned} \tag{4.11}$$

The latter is justified since

$$SS'\left(|z|^2 + y^2 SS'\right)^{-2} \leqslant y^{-4}(SS')^{-1}.$$

is integrable from (4.10). The next lemma is an important consequence of Theorems 2.13 and 1.2.

**Lemma 4.5.** *The following identity holds in $\mathcal{D}'(\mathbb{C})$:*

$$\mu_\alpha = -\frac{1}{\pi}\partial \mathbb{E} b(\cdot, 0).$$

*Therefore the measure $\mu_\alpha$ is isotropic and has a continuous density given by $1/\pi$ times the right hand side of* (4.11).

*Proof.* Let $R_n$ be the resolvent matrix of $B_n$, the bipartized matrix of $A_n$ defined by (1.4). By theorem 2.13 and lemma 2.2, for all $t > 0$ and $z \in \mathbb{C}$,

$$\lim_{n\to\infty} \mathbb{E}R_n(U(z,it))_{11} = \begin{pmatrix} i\mathbb{E}h(z,t) & \mathbb{E}b(z,it) \\ \mathbb{E}\bar{b}(z,it) & i\mathbb{E}h(z,t)\end{pmatrix},$$

From theorem 2.14, $\mathbb{E}\nu_{A_n - z}$ converge weakly to $\nu_{\alpha,z}$ and, by lemma 3.1, for all $t > 0$,

$$\lim_{n\to\infty} \frac{1}{2}\int \ln(x^2+t^2)\mathbb{E}\nu_{A_n-z}dx) = \frac{1}{2}\int \ln(x^2+t^2)\nu_{\alpha,z}(dx).$$

From Equation (3.20), $\int \ln(x)\nu_{\alpha,z}(dx)$ is integrable. We deduce that for all $z_0 \in \mathbb{C}$, there exists an open neighborhood of $z_0$ and a sequence $(t_n)_{n\geqslant 1}$ converging to 0 such that for all $z$ in the neighborhood,

$$\lim_{n\to\infty} \mathbb{E}R_n(U(z,it_n))_{11} = \begin{pmatrix} i\mathbb{E}h(z,0) & \mathbb{E}b(z,0) \\ \mathbb{E}\bar{b}(z,0) & i\mathbb{E}h(z,0)\end{pmatrix}, \tag{4.12}$$

and

$$\lim_{n\to\infty} \frac{1}{2}\int \ln(x^2+t_n^2)\mathbb{E}\nu_{A_n-z}(dx) = \int \ln(x)\nu_{\alpha,z}(dx). \tag{4.13}$$

Moreover from theorem 1.2, Equation (3.20), lemma A.2, in $\mathcal{D}'(\mathbb{C})$:

$$\Delta\int \ln(x)\nu_{\alpha,z}(dx) = 2\pi\mu_\alpha. \tag{4.14}$$

On the other hand, $\frac{1}{2}\int \ln(x^2+t^2)\nu_{A_n-z}(dx) = \frac{1}{2n}\ln|\det(B(z)-itI_{2n}))|$, and from (2.5),

$$\Delta\frac{1}{2}\int \ln(x^2+t^2)\mathbb{E}\nu_{A_n-z},(dx) = -2\partial\mathbb{E}b_1(z,it).$$

The conclusion follows from (4.12), (4.13) and (4.14). □

It is possible to compute explicitly the expression (4.11) at $z = 0$.

**Lemma 4.6.** *The density of $\mu_\alpha$ at $z = 0$ is*

$$\frac{1}{\pi}\frac{\Gamma(1+1/\beta)^2\Gamma(1+\beta)^{1/\beta}}{\Gamma(1-\beta)^{1/\beta}}.$$

*Proof.* By definition, the real $y_*(0)$ solves the equation

$$1 = \mathbb{E}\left(\frac{S}{y^2SS'}\right)^\beta = y^{-2\beta}\mathbb{E}S^{-\beta} = \frac{y^{-2\beta}}{\Gamma(\beta)}\int x^{\beta-1}e^{-\Gamma(1-\beta)x^\beta}dx.$$

With the change of variable $x \mapsto x^\beta$ and the identity $z\Gamma(z) = \Gamma(1+z)$, we find easily, $\mathbb{E}S^{-\beta} = (\Gamma(1-\beta)\Gamma(1+\beta))^{-1}$ and

$$y_*(0) = (\Gamma(1-\beta)\Gamma(1+\beta))^{-\frac{1}{2\beta}}.$$

We also have

$$\mathbb{E}S^{-1} = \int e^{-\Gamma(1-\beta)x^\beta}dx = \frac{1}{\beta\Gamma(1-\beta)^{1/\beta}}\int x^{1/\beta-1}e^{-x}dx = \frac{\Gamma(1+1/\beta)}{\Gamma(1-\beta)^{1/\beta}},$$

where we have used again the identity $z\Gamma(z) = \Gamma(1+z)$. Then the right hand side of (4.11) at $z = 0$ is equal to

$$y_*^2(0)y_*^{-4}(0)\mathbb{E}(SS')^{-1} = y_*^{-2}(0)\left(\mathbb{E}S^{-1}\right)^2.$$

□

4.3. **Proof of theorem 1.3.** In this subsection, we prove the last statement of theorem 1.3 (the first part of the theorem being contained in lemmas 4.5, 4.6). We start with a first technical lemma.

**Lemma 4.7.** *Let* $0 < \beta < 1$, $\delta > 0$, *and* $f$ *be a bounded measurable* $\mathbb{R}_+ \to \mathbb{R}$ *function such that* $f(y) = O(y^{\beta+\delta})$ *as* $y \downarrow 0$. *Let* $Y$ *be a random variable such that* $\mathbb{P}(Y \geqslant t) = L(t)t^{-\beta}$ *for some slowly varying function* $L$. *Then as* $t$ *goes to infinity*

$$\mathbb{E}f\left(\frac{Y}{t}\right) \sim \beta L(t)t^{-\beta}\int_0^\infty f(y)y^{-\beta-1}dy.$$

*Proof.* Define $Y_t = Y/t$. We fix $\varepsilon > 0$ and consider the distribution $\mathbb{P}(Y_t \in \cdot | Y_t \geqslant \varepsilon)$. By assumption, for $s > \varepsilon$,

$$\mathbb{P}(Y_t \geqslant s | Y_t \geqslant \varepsilon) \sim (s/\varepsilon)^{-\beta}.$$

In particular, the distribution of $Y_t$ given $\{Y_t \geqslant \varepsilon\}$ converges weakly as $t$ goes to infinity to the distribution with density $\beta x^{-\beta-1}\varepsilon^\beta dx$. Since $f$ is bounded and $L$ slowly varying, we get

$$\begin{aligned}
\mathbb{E}\left[f\left(\frac{Y}{t}\right)\mathbb{1}_{\{Y \geqslant \varepsilon t\}}\right] &= \mathbb{P}(Y_t \geqslant \varepsilon)\mathbb{E}\left[f(Y_t) \,\Big|\, Y_t \geqslant \varepsilon\right] \\
&\sim L(\varepsilon t)\varepsilon^{-\beta}t^{-\beta}\int_\varepsilon^\infty f(y)\beta y^{-\beta-1}\varepsilon^\beta dy \\
&\sim \beta L(t)t^{-\beta}\int_\varepsilon^\infty f(y)y^{-\beta-1}dy.
\end{aligned}$$

Finally, by assumption, for some constant, $c > 0$,

$$\mathbb{E}\left[f\left(\frac{Y}{t}\right)\mathbb{1}_{\{Y \leqslant \varepsilon t\}}\right] \leqslant ct^{-\beta-\delta}\mathbb{E}[Y^{\beta+\delta}\mathbb{1}_{\{Y \leqslant \varepsilon t\}}].$$

Thus by lemma C.1, for some new constant $c > 0$ and all $t \geqslant 1/\varepsilon$,

$$\mathbb{E}\left[f\left(\frac{Y}{t}\right)\mathbb{1}_{\{Y \leqslant \varepsilon t\}}\right] \leqslant ct^{-\beta-\delta}L(\varepsilon t)(\varepsilon t)^\delta = ct^{-\beta}L(t)\varepsilon^\delta\frac{L(\varepsilon t)}{L(t)}.$$

We may thus conclude by letting $t$ tend to infinity and then $\varepsilon$ to 0. □

**Lemma 4.8.** *Let* $S$ *be a random variable with Laplace transform* (4.4). *There exists a constant* $c_0 > 0$ *such that as* $t$ *goes to infinity,*

$$\mathbb{E}S^\beta\mathbb{1}_{\{S \leqslant t\}} = \ln t + c_0 + o(1).$$

*Proof.* Let $g_\beta$ be the density function of $S$. From Equation (2.4.8) in Zolotarev [56], $g_\beta$ has a convergent power series representation

$$g_\beta(x) = \frac{1}{\pi}\sum_{n=1}^\infty(-1)^{n-1}\frac{\Gamma(n\beta+1)}{\Gamma(n+1)\Gamma(1-\beta)^n}\sin(\pi n\beta)x^{-n\beta-1}.$$

The Stirling formula $\Gamma(x) \sim_{x\to\infty} \sqrt{\frac{2\pi}{x}}\,\left(\frac{x}{e}\right)^x$ implies that the convergence radius of the series is $+\infty$. Recall that $\Gamma(\beta+1) = \beta\Gamma(\beta)$, and the Euler reflection formula, $\Gamma(1-\beta)\sin(\pi\beta)/\pi = \Gamma(\beta)$. Thus, as $x$ goes to infinity,

$$g_\beta(x) = \beta x^{-\beta-1} + O(x^{-2\beta-1}).$$

□

The next lemma is a consequence of the Karamata Tauberian theorem.

**Lemma 4.9.** *As* $t$ *goes to infinity,*

$$\mathbb{P}(SS' \geqslant t) \sim \beta t^{-\beta}\ln t,$$

*and, with* $c_1 = \beta^2\int_0^\infty(x+1)^{-2}x^{-\beta}dx$,

$$\mathbb{E}\frac{SS'}{(t+SS')^2} \sim c_1t^{-1-\beta}\ln t.$$

*Proof.* Let $x > 0$, since $S$ and $S'$ are independent we have

$$\mathbb{E}\exp(-xSS') = \mathbb{E}\exp\left(-\Gamma(1-\beta)x^\beta S^\beta\right).$$

From Corollary 8.1.7 in [9], we have as $t$ goes to infinity, $\mathbb{P}(S > t) \sim t^{-\beta}$. In particular, we have $\mathbb{P}(S^\beta > t) \sim t^{-1}$ and a new application of Corollary 8.1.7 in [9] gives as $x \downarrow 0$,

$$1 - \mathbb{E}\exp(-xS^\beta) \sim x \ln x^{-1}.$$

We obtain

$$1 - \mathbb{E}\exp(-xSS') \sim \Gamma(1-\beta)x^\beta \ln(\Gamma(1-\beta)x^{-\beta}) \sim \beta\Gamma(1-\beta)x^\beta \ln(x^{-1}).$$

We then conclude by a third application of Corollary 8.1.7 in [9]. The second statement is a consequence of lemma 4.7. □

The next lemma gives the asymptotic behavior of $y_*(r)$ as $r$ goes to infinity.

**Lemma 4.10.** *There exists a constant $c_2 > 0$ such that as $r$ goes to infinity,*

$$y_*(r) \sim c_2\sqrt{r}e^{-r^\beta/2}.$$

*Proof.* From Equations (4.5), (4.8), we have with $y_* = y_*(r)$,

$$\begin{aligned} 1 &= \frac{1}{\Gamma(\beta)}\int x^{\beta-1}\mathbb{E}\exp\left(-\frac{xr}{S} - xy_*^2 S'\right)dx \\ &= \frac{1}{\Gamma(\beta)}\int x^{\beta-1}e^{-x^\beta y_*^{2\beta}\Gamma(1-\beta)}\mathbb{E}e^{-\frac{xr}{S}}dx \\ &= \frac{1}{\Gamma(1+\beta)\Gamma(1-\beta)y_*^{2\beta}}\int e^{-x}\mathbb{E}e^{-\frac{x^{1/\beta}ry_*^{-2}}{S\Gamma(1-\beta)^{1/\beta}}}dx. \end{aligned} \quad (4.15)$$

By lemma 4.4, $\lim_{r\to\infty} y_*(r) = 0$. Hence, from the above expression, we deduce that the term $ry_*^{-2}$ goes to infinity as $r$ goes to infinity. Define

$$I(y) = \frac{1}{\Gamma(1+\beta)\Gamma(1-\beta)}\int e^{-x}e^{-\frac{x^{1/\beta}}{y\Gamma(1-\beta)^{1/\beta}}}dx = I_0(y) + I_1(y) + I_2(y),$$

with $I_0(y) = I(y)\mathbb{1}_{\{y\geqslant 1\}}$,

$$\begin{aligned} I_1(y) &= \frac{\mathbb{1}_{\{y\leqslant 1\}}}{\Gamma(1+\beta)\Gamma(1-\beta)}\int e^{-\frac{x^{1/\beta}}{y\Gamma(1-\beta)^{1/\beta}}}dx = y^\beta \mathbb{1}_{\{y\leqslant 1\}}, \\ I_2(y) &= \frac{\mathbb{1}_{\{y\leqslant 1\}}}{\Gamma(1+\beta)\Gamma(1-\beta)}\int (e^{-x}-1)e^{-\frac{x^{1/\beta}}{y\Gamma(1-\beta)^{1/\beta}}}dx. \end{aligned}$$

The function $I$ is increasing and $\lim_{y\to\infty} I(y) < \infty$. Also, the function $I_0$ is equal to 0 in a neighborhood of 0. By lemma 4.7, we get as $t$ goes to infinity,

$$\mathbb{E}I_0(S/t) \sim a_0 t^{-\beta},$$

for some positive constant $a_0 = \frac{1}{\Gamma(1+\beta)\Gamma(1-\beta)}\int_1^\infty \int e^{-x}e^{-\frac{x^{1/\beta}}{y\Gamma(1-\beta)^{1/\beta}}}\beta y^{-\beta-1}dxdy$. By lemma 4.8,

$$\mathbb{E}[I_1(S/t)] = t^{-\beta}\ln t + c_0 t^{-\beta} + o(1).$$

Also, from Laplace method, $I_2(y) \sim -\Gamma(2\beta)\Gamma(1-\beta)^2 y^{2\beta}$ as $y$ goes to 0. By lemma 4.7,

$$\mathbb{E}I_2(S/t) \sim a_2 t^{-\beta},$$

with $a_2 = \frac{1}{\Gamma(1+\beta)\Gamma(1-\beta)}\int_0^1 \int (e^{-x}-1)e^{-\frac{x^{1/\beta}}{y\Gamma(1-\beta)^{1/\beta}}}\beta y^{-\beta-1}dxdy$. Hence, for $t = ry_*^{-2}$, we get from (4.15)

$$y_*^{2\beta} = (ry_*^{-2})^{-\beta}\ln(ry_*^{-2}) + (c_0 + a_0 + a_2)(ry_*^{-2})^{-\beta} + o((ry_*^{-2})^{-\beta}).$$

In other words,

$$r^\beta = \ln(ry_*^{-2}) + (c_0 + a_0 + a_2) + o(1).$$

We conclude by setting $c_2 = \exp((c_0 + a_0 + a_2)/2)$. □

**Lemma 4.11.** *As $r$ goes to infinity,*

$$y'(r) \sim -c_3^{-1} y_*(r) r^{\beta-1},$$

*where* $c_3 = 2\int_0^\infty \int_0^\infty x e^{-x} e^{-\frac{x^{1/\beta}}{s\Gamma(1-\beta)^{1/\beta}}} \beta s^{-\beta-1} dx ds / (\Gamma(1+\beta)\Gamma(1-\beta))$.

*Proof.* We define

$$G(y,r) = \mathbb{E}\left(\frac{S}{r + y^2 SS'}\right)^\beta = \frac{1}{\Gamma(\beta)} \int x^{\beta-1} e^{-x^\beta y^{2\beta}\Gamma(1-\beta)} \mathbb{E} e^{-\frac{xr}{S}} dx.$$

From the implicit function theorem

$$y_*'(r) = -\frac{\partial_r G(y_*, r)}{\partial_y G(y_*, r)}.$$

We have

$$\begin{aligned} \partial_y G(y,r) &= -\frac{2\beta\Gamma(1-\beta) y^{2\beta-1}}{\Gamma(\beta)} \int x^{2\beta-1} e^{-x^\beta y^{2\beta}\Gamma(1-\beta)} \mathbb{E} e^{-\frac{xr}{S}} dx \\ &= -\frac{2}{y^{2\beta+1}\Gamma(1+\beta)\Gamma(1-\beta)} \int x e^{-x} \mathbb{E} e^{-\frac{x^{1/\beta} r y^{-2}}{S\Gamma(1-\beta)^{1/\beta}}} dx \end{aligned}$$

The Laplace method implies that, as $t$ goes infinity,

$$\int x e^{-x} e^{-\frac{x^{1/\beta} t}{\Gamma(1-\beta)^{1/\beta}}} dx \sim \Gamma(2\beta)\Gamma(1-\beta)^2 t^{-2\beta}.$$

Thus by lemma 4.7, we deduce that

$$\int x e^{-x} \mathbb{E} e^{-\frac{x^{1/\beta} t}{S\Gamma(1-\beta)^{1/\beta}}} dx \sim t^{-\beta} \int\int x e^{-x} e^{-\frac{x^{1/\beta}}{s\Gamma(1-\beta)^{1/\beta}}} \beta s^{-\beta-1} dx ds \sim c t^{-\beta}.$$

Applying the above to $t = r y_*^{-2}(r)$ we deduce, with $c_3 = 2c/(\Gamma(1+\beta)\Gamma(1-\beta))$,

$$\partial_y G(y_*, r) \sim -c_3 r^{-\beta} y_*^{-1}(r).$$

Similarly, the derivative of $G$ with respect to $r$ is

$$\partial_r G(y,r) = -\frac{1}{y^{2\beta+2}\Gamma(1-\beta)^{1/\beta+1}\Gamma(1+\beta)} \int x^{1/\beta} e^{-x} \mathbb{E} e^{-\frac{x^{1/\beta} r y^{-2}}{S\Gamma(1-\beta)^{1/\beta}}} S^{-1} dx.$$

Once again, Laplace method implies that, as $t$ goes infinity,

$$\int x^{1/\beta} e^{-x} e^{-\frac{x^{1/\beta} t}{\Gamma(1-\beta)^{1/\beta}}} dx \sim \Gamma(\beta+1)\Gamma(1-\beta)^{1/\beta+1} t^{-\beta-1}.$$

In particular, for all $\varepsilon > 0$ there exists $t_0$ such that

$$\begin{aligned} &(1-\varepsilon) t^{-\beta-1} \mathbb{E} S^\beta \mathbb{1}_{\{S \leqslant t/t_0\}} \\ &\leqslant \frac{1}{\Gamma(1-\beta)^{1/\beta+1}\Gamma(1+\beta)} \int x^{1/\beta} e^{-x} \mathbb{E} e^{-\frac{x^{1/\beta} t}{S\Gamma(1-\beta)^{1/\beta}}} S^{-1} \mathbb{1}_{\{S \leqslant t/t_0\}} dx \\ &\leqslant (1+\varepsilon) t^{-\beta-1} \mathbb{E} S^\beta \mathbb{1}_{\{S \leqslant t/t_0\}}. \end{aligned}$$

By lemma 4.8,

$$\mathbb{E} S^\beta \mathbb{1}_{\{S \leqslant t/t_0\}} \sim \ln t.$$

It follows that for some $t_1 > t_0$ and all $t \geqslant t_1$,

$$(1-2\varepsilon) t^{-\beta-1} \ln t \leqslant \frac{1}{\Gamma(1-\beta)^{1/\beta+1}\Gamma(1+\beta)} \int x^{1/\beta} e^{-x} \mathbb{E} e^{-\frac{x^{1/\beta} t}{S\Gamma(1-\beta)^{1/\beta}}} S^{-1} \mathbb{1}_{\{S \leqslant t/t_0\}} dx \leqslant (1+2\varepsilon) t^{-\beta-1} \ln t.$$

On the other hand, for some constant $c > 0$ and all $t \geqslant 1$,

$$\int x^{1/\beta} e^{-x} \mathbb{E} e^{-\frac{x^{1/\beta} t}{S\Gamma(1-\beta)^{1/\beta}}} S^{-1} \mathbb{1}_{\{S \geqslant t/t_0\}} dx \leqslant \int x^{1/\beta} e^{-x} dx \mathbb{P}(S \geqslant t/t_0) \leqslant c t^{-\beta-1} t_0^{\beta+1}.$$

We thus have proved that

$$\frac{1}{\Gamma(1-\beta)^{1/\beta+1}\Gamma(1+\beta)} \int x^{1/\beta} e^{-x} \mathbb{E} e^{-\frac{x^{1/\beta} t}{S\Gamma(1-\beta)^{1/\beta}}} S^{-1} dx \sim t^{-\beta-1} \ln t,$$

and

$$\partial_r G(y_*(r), r) \sim -r^{-\beta-1} \ln(ry^{-2}) \sim -r^{-1}.$$

The statement follows. □

*Proof of theorem 1.3.* From Equation (4.11) and lemma 4.9, the density at $r = |z|^2$ is equivalent to $1/\pi$ times

$$\left(1 - 2r\frac{y_*'(r)}{y_*(r)}\right) y_*^{-2}(r) c_1 (r y_*^{-2})^{-1-\beta} \ln(r y_*^{-2}).$$

It remains to apply lemmas 4.10 and 4.11, and set the multiplicative constant to be $c = 2\pi^{-1} c_3^{-1} c_1 c_2^{2\beta}$. □

## Appendix A. Logarithmic potentials and Hermitization

Let $\mathcal{P}(\mathbb{C})$ be the set of probability measures on $\mathbb{C}$ which integrate $\ln|\cdot|$ in a neighborhood of infinity. For every $\mu \in \mathcal{P}(\mathbb{C})$, the *logarithmic potential* $U_\mu$ of $\mu$ on $\mathbb{C}$ is the function $U_\mu : \mathbb{C} \to [-\infty, +\infty)$ defined for every $z \in \mathbb{C}$ by

$$U_\mu(z) = \int_{\mathbb{C}} \ln|z - z'| \, \mu(dz') = (\ln|\cdot| * \mu)(z). \tag{A.1}$$

Note that in classical potential theory, the definition is opposite in sign, but ours turns out to be more convenient (lightweight) for our purposes. Since $\ln|\cdot|$ is Lebesgue locally integrable on $\mathbb{C}$, one can check by using the Fubini theorem that $U_\mu$ is Lebesgue locally integrable on $\mathbb{C}$. In particular, $U_\mu < \infty$ a.e. (Lebesgue almost everywhere) and $U_\mu \in \mathcal{D}'(\mathbb{C})$. Since $\ln|\cdot|$ is the fundamental solution of the Laplace equation in $\mathbb{C}$, we have, in $\mathcal{D}'(\mathbb{C})$,

$$\Delta U_\mu = 2\pi\mu. \tag{A.2}$$

**Lemma A.1** (Unicity)**.** *For every $\mu, \nu \in \mathcal{P}(\mathbb{C})$, if $U_\mu = U_\nu$ a.e. then $\mu = \nu$.*

*Proof.* Since $U_\mu = U_\nu$ in $\mathcal{D}'(\mathbb{C})$, we get $\Delta U_\mu = \Delta U_\nu$ in $\mathcal{D}'(\mathbb{C})$. Now (A.2) gives $\mu = \nu$ in $\mathcal{D}'(\mathbb{C})$, and thus $\mu = \nu$ as measures since $\mu$ and $\nu$ are Radon measures. □

If $A$ is an $n \times n$ complex matrix and $P_A(z) := \det(A - zI)$ is its characteristic polynomial,

$$U_{\mu_A}(z) = \int_{\mathbb{C}} \ln|z' - z| \, \mu_A(dz') = \frac{1}{n} \ln|\det(A - zI)| = \frac{1}{n} \ln|P_A(z)|$$

for every $z \in \mathbb{C} \setminus \{\lambda_1(A), \ldots, \lambda_n(A)\}$. We have also the alternative expression

$$U_{\mu_A}(z) = \frac{1}{n} \ln\det(\sqrt{(A - zI)(A - zI)^*}) = \int_0^\infty \ln(t) \, \nu_{A-zI}(dt). \tag{A.3}$$

The identity above bridges the eigenvalues with the singular values, and is at the heart of the following lemma, which allows to deduce the convergence of $\mu_A$ from the one of $\nu_{A-zI}$. The strength of this Hermitization lies in the fact that contrary to the eigenvalues, one can control the singular values with the entries of the matrix. The price paid here is the introduction of the auxiliary variable $z$ and the uniform integrability. We recall that on a Borel measurable space $(E, \mathcal{E})$, we say that a Borel function $f : E \to \mathbb{R}$ is uniformly integrable for a sequence of probability measures $(\eta_n)_{n\geqslant 1}$ on $E$ when

$$\lim_{t\to\infty} \varlimsup_{n\to\infty} \int_{\{|f|>t\}} |f| \, d\eta_n = 0.$$

We will use this property as follows: if $\eta_n \rightsquigarrow \eta$ and $f$ is continuous and uniformly integrable for $(\eta_n)_{n\geqslant 1}$ then $f$ is $\eta$-integrable and $\lim_{n\to\infty} \int f \, d\eta_n = \int f \, \eta$. Similarly for a sequence random probability measures $(\eta_n)_{n\geqslant 1}$ we will say that $f$ is uniformly integrable for $(\eta_n)_{n\geqslant 1}$ in probability, if for all $\varepsilon > 0$

$$\lim_{t\to\infty} \varlimsup_{n\to\infty} \mathbb{P}\left(\int_{\{|f|>t\}} |f| \, d\eta_n > \varepsilon\right) = 0.$$

A proof of lemma A.2 below can be found in [12] which covers the "a.s." case, the "in probability" case being similar. It relies only on the unicity lemma A.1, the classical Prohorov theorem, and the Weyl inequalities of Lemma B.5 linking eigenvalues and singular values.

**Lemma A.2** (Girko's Hermitization method). *Let $(A_n)_{n\geqslant 1}$ be a sequence of complex random matrices where $A_n$ is $n\times n$ for every $n\geqslant 1$. Suppose that for Lebesgue almost all $z\in\mathbb{C}$, there exists a probability measure $\nu_z$ on $[0,\infty)$ such that*

(i) *a.s. $(\nu_{A_n-zI})_{n\geqslant 1}$ tends weakly to $\nu_z$*
(ii) *a.s. (resp. in probability) $\ln(\cdot)$ is uniformly integrable for $(\nu_{A_n-zI})_{n\geqslant 1}$*

*Then there exists a probability measure $\mu\in\mathcal{P}(\mathbb{C})$ such that*

(j) *a.s. (resp. in probability) $(\mu_{A_n})_{n\geqslant 1}$ converges weakly to $\mu$*
(jj) *for a.a. $z\in\mathbb{C}$,*

$$U_\mu(z)=\int_0^\infty \ln(t)\,\nu_z(dt).$$

## Appendix B. General spectral estimates

**Lemma B.1** (Basic inequalities [32]). *If $A$ and $B$ are $n\times n$ complex matrices then*

$$s_1(AB)\leqslant s_1(A)s_1(B)\quad\text{and}\quad s_1(A+B)\leqslant s_1(A)+s_1(B) \tag{B.1}$$

*and*

$$\max_{1\leqslant i\leqslant n}|s_i(A)-s_i(B)|\leqslant s_1(A-B). \tag{B.2}$$

**Lemma B.2** (Rudelson-Vershynin row bound [45, 12]). *Let $A$ be a complex $n\times n$ matrix with rows $R_1,\dots,R_n$. Define the vector space $R_{-i}:=\mathrm{span}\{R_j;j\neq i\}$. We have then*

$$n^{-1/2}\min_{1\leqslant i\leqslant n}\mathrm{dist}(R_i,R_{-i})\leqslant s_n(A)\leqslant \min_{1\leqslant i\leqslant n}\mathrm{dist}(R_i,R_{-i}).$$

Recall that the singular values $s_1(A),\dots,s_{n'}(A)$ of a rectangular $n'\times n$ complex matrix $A$ with $n'\leqslant n$ are defined by $s_i(A):=\lambda_i(\sqrt{AA^*})$ for every $1\leqslant i\leqslant n'$.

**Lemma B.3** (Tao-Vu negative second moment [50, Lemma A4]). *If $A$ is a full rank $n'\times n$ complex matrix ($n'\leqslant n$) with rows $R_1,\dots,R_{n'}$, and $R_{-i}:=\mathrm{span}\{R_j;j\neq i\}$, then*

$$\sum_{i=1}^{n'}s_i(A)^{-2}=\sum_{i=1}^{n'}\mathrm{dist}(R_i,R_{-i})^{-2}.$$

**Lemma B.4** (Cauchy interlacing by rows deletion [32]). *Let $A$ be an $n\times n$ complex matrix. If $B$ is $n'\times n$, obtained from $A$ by deleting $n-n'$ rows, then for every $1\leqslant i\leqslant n'$,*

$$s_i(A)\geqslant s_i(B)\geqslant s_{i+n-n'}(A).$$

**Lemma B.5** (Weyl inequalities [53]). *For every $n\times n$ complex matrix $A$, we have*

$$\prod_{i=1}^k|\lambda_i(A)|\leqslant\prod_{i=1}^k s_i(A)\quad\text{and}\quad\prod_{i=k}^n s_i(A)\leqslant\prod_{i=k}^n|\lambda_i(A)| \tag{B.3}$$

*for all $1\leqslant k\leqslant n$. In particular, by viewing $|\det(A)|$ as a volume,*

$$|\det(A)|=\prod_{k=1}^n|\lambda_k(A)|=\prod_{k=1}^n s_k(A)=\prod_{k=1}^n\mathrm{dist}(R_k,\mathrm{span}\{R_1,\dots,R_{k-1}\}) \tag{B.4}$$

*where $R_1,\dots,R_n$ are the rows of $A$. Moreover, for every increasing function $\varphi$ from $(0,\infty)$ to $(0,\infty)$ such that $t\mapsto\varphi(e^t)$ is convex on $(0,\infty)$ and $\varphi(0):=\lim_{t\to0^+}\varphi(t)=0$, we have*

$$\sum_{i=1}^k\varphi(|\lambda_i(A)|^2)\leqslant\sum_{i=1}^k\varphi(s_i(A)^2) \tag{B.5}$$

*for every $1\leqslant k\leqslant n$. In particular, with $\varphi(t)=t^{r/2}$, $r>0$, and $k=n$, we obtain*

$$\sum_{k=1}^n|\lambda_k(A)|^r\leqslant\sum_{k=1}^n s_k(A)^r. \tag{B.6}$$

**Lemma B.6** (Schatten bound [55, proof of Theorem 3.32]). *Let $A$ be an $n \times n$ complex matrix with rows $R_1, \ldots, R_n$. Then for every $0 < r \leqslant 2$,*

$$\sum_{k=1}^{n} s_k(A)^r \leqslant \sum_{k=1}^{n} \|R_k\|_2^r. \tag{B.7}$$

## Appendix C. Additional lemmas

We begin with a lemma on truncated moments. We skip the proof since it follows from an adaptation of the proof in the real case given by e.g. Feller [23, Theorem VIII.9.2].

**Lemma C.1** (Truncated moments). *If (H1) holds then for every $p > \alpha$,*

$$\mathbb{E}\big[|X_{11}|^p \mathbb{1}_{\{|X_{11}| \leqslant t\}}\big] \sim c(p) L(t) t^{p-\alpha}$$

*where $c(p) := \alpha/(p-\alpha)$. In particular, we have*

$$\mathbb{E}\big[|X_{11}|^p \mathbb{1}_{\{|X_{11}| \leqslant a_n\}}\big] \sim c(p) \frac{a_n^p}{n}.$$

We end up this section by a result on the concentration of the spectral measure of Hermitian or Hermitized random matrices, mentioned in [13]. The total variation norm of $f : \mathbb{R} \to \mathbb{R}$ is

$$\|f\|_{\mathrm{TV}} := \sup \sum_{k \in \mathbb{Z}} |f(x_{k+1}) - f(x_k)|,$$

where the supremum runs over all sequences $(x_k)_{k \in \mathbb{Z}}$ such that $x_{k+1} \geqslant x_k$ for any $k \in \mathbb{Z}$. If $f = \mathbb{1}_{(-\infty, s]}$ for some real $s$ then $\|f\|_{\mathrm{TV}} = 1$, while if $f$ has a derivative in $\mathrm{L}^1(\mathbb{R})$, we get

$$\|f\|_{\mathrm{TV}} = \int_{\mathbb{R}} |f'(t)| \, dt.$$

The following lemma comes with remarkably weak assumptions, and allows to deduce the almost sure weak convergence of empirical spectral measures of random matrices without any moment assumptions on the entries. We discovered that this lemma was obtained independently by Guntuboyina and Leeb in [30], where they discuss the relationships with more classical results.

**Lemma C.2** (Concentration for spectral measures). *Let $H$ be an $n \times n$ random Hermitian matrix. Let us assume that the vectors $(H_i)_{1 \leqslant i \leqslant n}$, where $H_i := (H_{ij})_{1 \leqslant j \leqslant i} \in \mathbb{C}^i$, are independent. Then for any $f : \mathbb{R} \to \mathbb{R}$ going to $0$ at $\pm\infty$ and such that $\|f\|_{\mathrm{TV}} \leqslant 1$ and every $t \geqslant 0$,*

$$\mathbb{P}\left(\left|\int f \, d\mu_H - \mathbb{E} \int f \, d\mu_H\right| \geqslant t\right) \leqslant 2 \exp\left(-\frac{nt^2}{2}\right).$$

*Similarly, if $M$ is an $n \times n$ complex random matrix with independent rows (or with independent columns) then for any $f : \mathbb{R}_+ \to \mathbb{R}$ going to $0$ at $+\infty$ with $\|f\|_{\mathrm{TV}} \leqslant 1$ and every $t \geqslant 0$,*

$$\mathbb{P}\left(\left|\int f \, d\nu_M - \mathbb{E} \int f \, d\nu_M\right| \geqslant t\right) \leqslant 2 \exp\left(-2nt^2\right).$$

*Proof.* We prove only the Hermitian version, the non-Hermitian version being entirely similar. Let us start by showing that for every $n \times n$ deterministic Hermitian matrices $A$ and $B$ and any measurable function $f$ with $\|f\|_{\mathrm{TV}} = 1$,

$$\left|\int f \, d\mu_A - \int f \, d\mu_B\right| \leqslant \frac{\mathrm{rank}(A-B)}{n}. \tag{C.1}$$

Indeed, it is well known (follows from interlacing, see e.g. [51] or [5, Theorem 11.42]) that

$$\|F_A - F_B\|_\infty \leqslant \frac{\mathrm{rank}(A-B)}{n}$$

where $F_A$ and $F_B$ are the cumulative distribution functions of $\mu_A$ and $\mu_B$ respectively. Now if $f$ is smooth, we get, by integrating by parts,

$$\left|\int f \, d\mu_A - \int f \, d\mu_B\right| = \left|\int_{\mathbb{R}} f'(t) F_A(t) \, dt - \int_{\mathbb{R}} f'(t) F_B(t) \, dt\right| \leqslant \frac{\mathrm{rank}(A-B)}{n} \int_{\mathbb{R}} |f'(t)| \, dt,$$

and since the left hand side depends on at most $2n$ points, we get (C.1) by approximating $f$ by smooth functions. Next, for any $x = (x_1, \ldots, x_n) \in \mathcal{X} := \{(x_i)_{1 \leqslant i \leqslant n} : x_i \in \mathbb{C}^{i-1} \times \mathbb{R}\}$, let $H(x)$ be the $n\times n$ Hermitian matrix given by $H(x)_{ij} := x_{i,j}$ for $1 \leqslant j \leqslant i \leqslant n$. We have $\mu_H = \mu_{H(H_1,\dots,H_n)}$. For all $x\in\mathcal{X}$ and $x_i' \in \mathbb{C}^{i-1}\times\mathbb{R}$, the matrix

$$H(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n) - H(x_1,\dots,x_{i-1},x_i',x_{i+1},\dots,x_n)$$

has only the $i$-th row and column possibly different from 0, and thus

$$\mathrm{rank}(H(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n) - H(x_1,\dots,x_{i-1},x_i',x_{i+1},\dots,x_n)) \leqslant 2.$$

Therefore from C.1, we obtain, for every $f:\mathbb{R}\to\mathbb{R}$ with $\|f\|_{\mathrm{TV}}\leqslant 1$,

$$\left|\int f\,d\mu_{H(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n)} - \int f\,d\mu_{H(x_1,\dots,x_{i-1},x_i',x_{i+1},\dots,x_n)}\right| \leqslant \frac{2}{n}.$$

The desired result follows now from the Azuma–Hoeffding inequality, see e.g. [38, Lemma 1.2]. □

## References


[1] D. Aldous, *Asymptotics in the random assignment problem*, Probab. Theory Related Fields **93** (1992), no. 4, 507–534. 9, 11

[2] D. Aldous and R. Lyons, *Processes on unimodular random networks*, Electron. J. Probab. **12** (2007), no. 54, 1454–1508 (electronic). 11

[3] D. Aldous and J. M. Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 1–72. 9, 11

[4] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), no. 1, 494–529. 2

[5] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Mathematics Monograph Series 2, Science Press, Beijing, 2006. 2, 33

[6] S. Belinschi, A. Dembo, and A. Guionnet, *Spectral measure of heavy tailed band and covariance random matrices*, Comm. Math. Phys. **289** (2009), no. 3, 1023–1055. 3

[7] G. Ben Arous and A. Guionnet, *The spectrum of heavy tailed random matrices*, Comm. Math. Phys. **278** (2008), no. 3, 715–751. 3

[8] I. Benjamini and O. Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13 pp. (electronic). 11

[9] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*, Encyclopedia of Mathematics and its Applications, vol. 27, Cambridge University Press, Cambridge, 1989. 29

[10] Ch. Bordenave, P. Caputo, and D. Chafaï, *Spectrum of large random reversible Markov chains: two examples*, ALEA Lat. Am. J. Probab. Math. Stat. 7 (2010), 41–64. 3

[11] ———, *Spectrum of large random reversible Markov chains: heavy tailed weigths on the complete graph*, The Annals of Probability 2011, Vol. 39, No. 4, 1544–1590. 3, 7, 9, 10, 11, 12, 13

[12] ———, *Circular Law Theorem for Random Markov Matrices*, preprint `arXiv:0808.1502 [math.PR]`, accepted in Probability Theory and Related Fields, 2010. 3, 31, 32

[13] Ch. Bordenave, M. Lelarge, and J. Salez, *The rank of diluted random graphs*, Annals of Probability 2011, Vol. 39, No. 3, 1097–1121. 33

[14] J. Bouchaud and P. Cizeau, *Theory of Lévy matrices*, Phys. Rev. E **3** (1994), 1810–1822. 3

[15] L. G. Brown, *Lidskiĭ's theorem in the type* II *case*, Geometric methods in operator algebras (Kyoto, 1983), Pitman Res. Notes Math. Ser., vol. 123, Longman Sci. Tech., Harlow, 1986, pp. 1–35. 4

[16] D. Chafaï, *Aspects of large random Markov kernels*, Stochastics 81 (2009), no. 3-4, 415–429. 3

[17] ———, *Circular law for noncentral random matrices*, J. Theoret. Probab. 23 (2010), no. 4, 945–950. 4

[18] ———, *The Dirichlet Markov ensemble*, J. Multivariate Anal. 101 (2010), no. 3, 555–567. 3

[19] R. B. Dozier and J. W. Silverstein, *Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices*, J. Multivariate Anal. **98** (2007), no. 6, 1099–1122. 2

[20] ———, *On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices*, J. Multivariate Anal. **98** (2007), no. 4, 678–694. 2

[21] A. Edelman, *The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law*, J. Multivariate Anal. **60** (1997), no. 2, 203–232. 2

[22] J. Feinberg and A. Zee, *Non-Hermitian random matrix theory: Method of Hermitian reduction*, Nucl. Phys. B (1997), no. 3, 579–608. 4

[23] W. Feller, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971. 2, 33

[24] V. L. Girko, *The circular law*, Teor. Veroyatnost. i Primenen. **29** (1984), no. 4, 669–679. 2

[25] ———, *Strong circular law*, Random Oper. Stochastic Equations **5** (1997), no. 2, 173–196. 2

[26] ———, *The circular law. Twenty years later. III*, Random Oper. Stochastic Equations **13** (2005), no. 1, 53–109. 2

[27] I. Y. Goldsheid and B. A. Khoruzhenko, *The Thouless formula for random non-Hermitian Jacobi matrices*, Israel J. Math. **148** (2005), 331–346, Probability in mathematics. MR MR2191234 (2006k:47082) 2

[28] F. Götze and A. Tikhomirov, *The Circular Law for Random Matrices*, Ann. Probab. 38 (2010), no. 4, 1444–1491. 2, 15

[29] E. Gudowska-Nowak, A. Jarosz, M. Nowak, and G. Pappe, *Towards non-Hermitian random Lévy matrices*, Acta Physica Polonica B **38** (2007), no. 13, 4089–4104. 4

[30] A. Guntuboyina and H. Leeb, *Concentration of the spectral measure of large Wishart matrices with dependent entries*, Electron. Commun. Probab. 14 (2009), 334–342. 33

[31] U. Haagerup and H. Schultz, *Brown measures of unbounded operators affiliated with a finite von Neumann algebra*, Math. Scand. **100** (2007), no. 2, 209–263. 4

[32] R. A. Horn and Ch. R. Johnson, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1994, Corrected reprint of the 1991 original. 32

[33] C.-R. Hwang, *A brief survey on the spectral radius and the spectral distribution of large random matrices with i.i.d. entries*, Random matrices and their applications (Brunswick, Maine, 1984), Contemp. Math., vol. 50, Amer. Math. Soc., Providence, RI, 1986, pp. 145–152. 2

[34] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. 16

[35] R. LePage, M. Woodroofe and J. Zinn, *Convergence to a stable distribution via order statistics*, Ann. Probab. 9 (1981), no. 4. 624–632 25

[36] R. Lyons, *Identities and Inequalities for Tree Entropy*, Combin. Probab. Comput. **19** (2010), no. 2, 303–313. 4

[37] V.A. Marchenko and L.A. Pastur, *The distribution of eigenvalues in sertain sets of random matrices*, Mat. Sb. **72** (1967), 507–536. 2

[38] C. McDiarmid, *On the method of bounded differences*, Surveys in combinatorics, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, pp. 148–188. 34

[39] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Academic Press, New York, 1967. 2

[40] G.M. Pan and W. Zhou, *Circular law, extreme singular values and potential theory*, J. Multivar. Anal. **101** (2010), no. 3, 645–656. 2

[41] D. Panchenko and M. Talagrand, *On one property of Derrida-Ruelle cascades*, C. R. Math. Acad. Sci. Paris 345 (2007), no. 11, 653–656. 25

[42] M. Reed and B. Simon, *Methods of modern mathematical physics. I*, second ed., Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980, Functional analysis. 7, 9

[43] T. Rogers, *Universal sum and product rules for random matrices*, J. Math. Phys. **51**, (2010), 093304, 15 . 4

[44] T. Rogers and I.P. Castillo, *Cavity approach to the spectral density of non-Hermitian sparse matrices*, Phys. Rev. E 79, 012101 (2009). 4

[45] M. Rudelson and R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633. 15, 32

[46] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205. MR MR1361756 (97h:60016) 16

[47] T. Tao, *Outliers in the spectrum of iid matrices with bounded rank perturbations*, preprint `arXiv:1012.4818 [math.PR]` 4

[48] T. Tao and V. Vu, *Random matrices: the circular law*, Commun. Contemp. Math. **10** (2008), no. 2, 261–307. 2, 15

[49] ———, *Smooth analysis of the condition number and the least singular value*, Math. Comp. 79 (2010), no. 272, 23332352. 15

[50] ———, *Random matrices: universality of ESDs and the circular law*, with an appendix by Manjunath Krishnapur. Ann. Probab. 38 (2010), no. 5, 2023–2065. 2, 3, 4, 16, 32

[51] R. C. Thompson, *The behavior of eigenvalues and singular values under perturbations of restricted rank*, Linear Algebra and Appl. **13** (1976), no. 1/2, 69–78, Collection of articles dedicated to Olga Taussky Todd. 33

[52] K. W. Wachter, *The strong limits of random matrix spectra for sample matrices of independent elements*, Ann. Probability **6** (1978), no. 1, 1–18. 2

[53] H. Weyl, *Inequalities between the two kinds of eigenvalues of a linear transformation*, Proc. Nat. Acad. Sci. U. S. A. **35** (1949), 408–411. 32

[54] Y. Q. Yin, *Limiting spectral distribution for a class of random matrices*, J. Multivariate Anal. **20** (1986), no. 1, 50–68. 2

[55] X. Zhan, *Matrix inequalities*, Lecture Notes in Mathematics, vol. 1790, Springer-Verlag, Berlin, 2002. 33

[56] V. M. Zolotarev, *One-dimensional stable distributions*, Translations of Mathematical Monographs, vol. 65, American Mathematical Society, Providence, RI, 1986, Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver. 28



(Ch. Bordenave) IMT UMR 5219 CNRS and Université Paul-Sabatier Toulouse III, France
*E-mail address*: `charles.bordenave(at)math.univ-toulouse.fr`
*URL*: `http://www.math.univ-toulouse.fr/~bordenave/`

(P. Caputo) Dipartimento di Matematica, Università Roma Tre, Italy
*E-mail address*: `caputo(at)mat.uniroma3.it`
*URL*: `http://www.mat.uniroma3.it/users/caputo/`

(D. Chafaï) LAMA UMR 8050 CNRS and Université Paris-Est Marne-la-Vallée, France
*E-mail address*: `djalil(at)chafai.net`
*URL*: `http://djalil.chafai.net/`