Around the circular law

Charles Bordenave, Djalil Chafai

AROUND THE CIRCULAR LAW

CHARLES BORDENAVE AND DJALIL CHAFAÏ

Abstract. These expository notes are centered around the circular law theorem, which states that the empirical spectral distribution of a $n \times n$ random matrix with i.i.d. entries of variance $1/n$ tends to the uniform law on the unit disc of the complex plane as the dimension n tends to infinity. This phenomenon is the non-Hermitian counterpart of the semi circular limit for Wigner random Hermitian matrices, and the quarter circular limit for Marchenko-Pastur random covariance matrices. We present a proof in a Gaussian case, due to Silverstein, based on a formula by Ginibre, and a proof of the universal case by revisiting the approach of Tao and Vu, based on the Hermitization of Girko, the logarithmic potential, and the control of the small singular values. Beyond the finite variance model, we also consider the case where the entries have heavy tails, by using the objective method of Aldous and Steele borrowed from randomized combinatorial optimization. The limiting law is then no longer the circular law and is related to the Poisson weighted infinite tree. We provide a weak control of the smallest singular value under weak assumptions, using asymptotic geometric analysis tools. We also develop a quaternionic Cauchy-Stieltjes transform borrowed from the Physics literature.

Contents

These expository notes are split in seven sections and an appendix. Section 1 introduces the notion of eigenvalues and singular values and discusses their relationships. Section 2

Date: September 2011, draft version compiled October 5, 2011.
2000 *Mathematics Subject Classification.* Random matrices; Spectrum; Circular law; Singular values.
Key words and phrases. 60B20; 60F15.

states the circular law theorem. Section 3 is devoted to the Gaussian model known as the Complex Ginibre Ensemble, for which the law of the spectrum is known and leads to the circular law. Section 4 provides the proof of the circular law theorem in the universal case, using the approach of Tao and Vu based on the Hermitization of Girko and the logarithmic potential. Section 5 presents some models related to the circular law and discusses an algebraic-analytic interpretation via free probability. Section 6 is devoted to the heavy tailed counterpart of the circular law theorem, using the objective method of Aldous and Steele and the Poisson Weighted Infinite Tree. Finally, section 7 lists some open problems. The notes end with appendix A devoted to a novel general weak control of the smallest singular value of random matrices with i.i.d. entries, with weak assumptions, well suited for the proof of the circular law theorem and its heavy tailed analogue.

1. Two kinds of spectra

The *eigenvalues* of $A \in \mathcal{M}_n(\mathbb{C})$ are the roots in $\mathbb{C}$ of its characteristic polynomial $P_A(z) := \det(A - zI)$. We label them $\lambda_1(A), \dots, \lambda_n(A)$ so that $|\lambda_1(A)| \geq \cdots \geq |\lambda_n(A)|$ with growing phases. The *spectral radius* is $|\lambda_1(A)|$. The eigenvalues form the algebraic spectrum of A. The *singular values of* A are defined by

$$s_k(A) := \lambda_k(\sqrt{AA^*})$$

for all $1 \leq k \leq n$, where $A^* = \bar{A}^\top$ is the conjugate-transpose. We have

$$s_1(A) \geq \cdots \geq s_n(A) \geq 0.$$

The matrices $A, A^\top, A^*$ have the same singular values. The $2n \times 2n$ Hermitian matrix

$$H(A) := \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}$$

has eigenvalues $s_1(A), -s_1(A), \dots, s_n(A), -s_n(A)$. This turns out to be useful because the mapping $A \mapsto H(A)$ is linear in A, in contrast with the mapping $A \mapsto \sqrt{AA^*}$. Geometrically, the matrix A maps the unit sphere to an ellipsoid, the half-lengths of its principal axes being exactly the singular values of A. The *operator norm* or *spectral norm* of A is

$$\|A\|_{2\to 2} := \max_{\|x\|_2=1} \|Ax\|_2 = s_1(A) \quad \text{while} \quad s_n(A) = \min_{\|x\|_2=1} \|Ax\|_2.$$

The rank of A is equal to the number of non-zero singular values. If A is non-singular then $s_i(A^{-1}) = s_{n-i}(A)^{-1}$ for all $1 \leq i \leq n$ and $s_n(A) = s_1(A^{-1})^{-1} = \|A^{-1}\|_{2\to 2}^{-1}$.

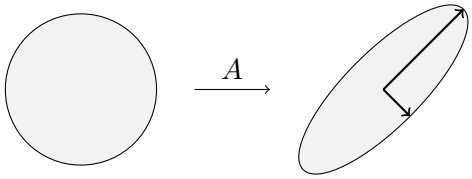

Figure 1. Largest and smallest singular values of $A \in \mathcal{M}_2(\mathbb{R})$. Taken from [33].

Since the singular values are the eigenvalues of a Hermitian matrix, we have variational formulas for all of them, often called the Courant-Fischer variational formulas [76, Th. 3.1.2]. Namely, denoting $\mathcal{G}_{n,i}$ the Grassmannian of all i-dimensional subspaces, we have

$$s_i(A) = \max_{E \in \mathcal{G}_{n,i}} \min_{\substack{x \in E \\ \|x\|_2=1}} \|Ax\|_2 = \max_{E,F \in \mathcal{G}_{n,i}} \min_{\substack{(x,y) \in E \times F \\ \|x\|_2=\|y\|_2=1}} \langle Ax, y \rangle.$$

Most useful properties of the singular values are consequences of their Hermitian nature via these variational formulas, which are valid on $\mathbb{R}^n$ and on $\mathbb{C}^n$. In contrast, there are no

such variational formulas for the eigenvalues in great generality, beyond the case of normal matrices. If the matrix A is normal[1] (i.e. $A^*A = A^*A$) then for every $1 \leq i \leq n$,

$$s_i(A) = |\lambda_i(A)|.$$

Beyond normal matrices, the relationships between the eigenvalues and the singular values are captured by a set of inequalities due to Weyl [136][2], which can be obtained by using the Schur unitary triangularization[3], see for instance [76, Theorem 3.3.2 page 171].

Theorem 1.1 (Weyl inequalities)**.** *For every $n \times n$ complex matrix A and $1 \leq k \leq n$,*

$$\prod_{i=1}^{k} |\lambda_i(A)| \leq \prod_{i=1}^{k} s_i(A). \tag{1.1}$$

The reversed form $\prod_{i=n-k+1}^{n} s_i(A) \leq \prod_{i=n-k+1}^{n} |\lambda_i(A)|$ for every $1 \leq k \leq n$ can be deduced without much difficulty (exercise!). Equality is achieved for $k = n$ and we have

$$\prod_{k=1}^{n} |\lambda_k(A)| = |\det(A)| = \sqrt{|\det(A)||\det(A^*)|} = |\det(\sqrt{AA^*})| = \prod_{k=1}^{n} s_k(A). \tag{1.2}$$

By using *majorization techniques* which can be found for instance in [76, Section 3.3], one may deduce from Weyl's inequalities that for every real valued function φ such that $t \mapsto \varphi(e^t)$ is increasing and convex on $[s_n(A), s_1(A)]$, we have, for every $1 \leq k \leq n$,

$$\sum_{i=1}^{k} \varphi(|\lambda_i(A)|) \leq \sum_{i=1}^{k} \varphi(s_i(A)), \tag{1.3}$$

see [76, Theorem 3.3.13]. In particular, taking $k = n$ and $\varphi(t) = t^2$ gives

$$\sum_{i=1}^{n} |\lambda_i(A)|^2 \leq \sum_{i=1}^{n} s_i(A)^2 = \mathrm{Tr}(AA^*) = \sum_{i,j=1}^{n} |A_{i,j}|^2. \tag{1.4}$$

We define the empirical eigenvalues and singular values measures by

$$\mu_A := \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k(A)} \quad \text{and} \quad \nu_A := \frac{1}{n} \sum_{k=1}^{n} \delta_{s_k(A)}.$$

Note that μ_A and ν_A are supported respectively in $\mathbb{C}$ and $\mathbb{R}_+$. There is a rigid determinantal relationship between μ_A and ν_A, namely from (1.2) we get

$$\begin{aligned} \int \log|\lambda| \, d\mu_A(\lambda) &= \frac{1}{n} \sum_{i=1}^{n} \log|\lambda_i(A)| \\ &= \frac{1}{n} \log|\det(A)| \\ &= \frac{1}{n} \sum_{i=1}^{n} \log(s_i(A)) \\ &= \int \log(s) \, d\nu_A(s). \end{aligned}$$

This identity is at the heart of the Hermitization technique used in sections 4 and 6.

[1] In these notes, the word *normal* is always used in this way, and never as a synonym for *Gaussian*.

[2] Horn [74] showed a remarkable converse to Weyl's theorem: if a sequence $s_1 \geq \cdots \geq s_n$ of non-negative real numbers and a sequence $\lambda_1, \ldots, \lambda_n$ of complex numbers of non increasing modulus satisfy to all Weyl's inequalities (1.1) then there exists $A \in \mathcal{M}_n(\mathbb{C})$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ and singular values $s_1, \ldots, s_n$.

[3] If $A \in \mathcal{M}_n(\mathbb{C})$ then there exists a unitary matrix U such that UAU^* is upper triangular.

In the sequel, we say that a sequence of probability measures $(\eta_n)_{n\geq 1}$ on $\mathbb{C}$ (respectively on $\mathbb{R}$) tends weakly to a probability measure η, and we denote $\eta_n \rightsquigarrow \eta$, when for all continuous and bounded function $f : \mathbb{C} \to \mathbb{R}$ (respectively $f : \mathbb{R} \to \mathbb{R}$) we have

$$\lim_{n\to\infty} \int f \, d\eta_n = \int f \, d\eta.$$

This type of convergence does not capture the behavior of the support and of the moments.

Since $s_1(\cdot) = \|\cdot\|_{2\to 2}$ we have for any $A, B \in \mathcal{M}_n(\mathbb{C})$ that

$$s_1(AB) \leq s_1(A)s_1(B) \quad \text{and} \quad s_1(A+B) \leq s_1(A) + s_1(B). \tag{1.5}$$

The singular values are quite regular functions of the matrix entries. For instance, the Courant-Fischer formulas imply that the mapping $A \mapsto (s_1(A), \ldots, s_n(A))$ is 1-Lipschitz for the operator norm and the ℓ^∞ norm in the sense that for any $A, B \in \mathcal{M}_n(\mathbb{C})$,

$$\max_{1\leq i\leq n} |s_i(A) - s_i(B)| \leq s_1(A - B). \tag{1.6}$$

Recall that $\mathcal{M}_n(\mathbb{C})$ or $\mathcal{M}_n(\mathbb{R})$ are Hilbert spaces for the scalar product $A \cdot B = \mathrm{Tr}(AB^*)$. The norm $\|\cdot\|_2$ associated to this scalar product, called the trace norm[4], satisfies to

$$\|A\|_2^2 = \mathrm{Tr}(AA^*) = \sum_{i=1}^n s_i(A)^2 = n \int s^2 \, d\nu_A(s). \tag{1.7}$$

The Hoffman-Wielandt inequality for the singular values states that for all $A, B \in \mathcal{M}_n(\mathbb{C})$,

$$\sum_{i=1}^n (s_i(A) - s_i(B))^2 \leq \|A - B\|_2^2. \tag{1.8}$$

In other words the mapping $A \mapsto (s_1(A), \ldots, s_n(A))$ is 1-Lipschitz for the trace norm and the Euclidean 2-norm. See [76, equation (3.3.32)] and [75, Theorem 6.3.5] for a proof.

Example 1.2 (Spectra of non-normal matrices)**.** *The eigenvalues depend continuously on the entries of the matrix. It turns out that for non-normal matrices, the eigenvalues are more sensitive to perturbations than the singular values. Among non-normal matrices, we find non-diagonalizable matrices, including nilpotent matrices. Let us recall a striking example taken from* [120] *and* [11, Chapter 10]*. Let us consider $A, B \in \mathcal{M}_n(\mathbb{R})$ given by*

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \quad \textit{and} \quad B = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \kappa_n & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where (κ_n) is a sequence of positive real numbers. The matrix A is nilpotent, and B is a perturbation with small norm (and rank one!):

$$\mathrm{rank}(A - B) = 1 \quad \textit{and} \quad \|A - B\|_{2\to 2} = \kappa_n.$$

We have $\lambda_1(A) = \cdots = \lambda_{\kappa_n}(A) = 0$ and thus

$$\mu_A = \delta_0.$$

In contrast, $B^n = \kappa_n I$ and thus $\lambda_k(B) = \kappa_n^{1/n} e^{2k\pi i/n}$ for all $1 \leq k \leq n$ which gives

$$\mu_B \rightsquigarrow \mathrm{Uniform}\{z \in \mathbb{C} : |z| = 1\}$$

as soon as $\kappa_n^{1/n} \to 1$ (this allows $\kappa_n \to 0$). On the other hand, from the identities

$$AA^* = \mathrm{diag}(1, \ldots, 1, 0) \quad \textit{and} \quad BB^* = \mathrm{diag}(1, \ldots, 1, \kappa_n^2)$$

[4]Also known as the Hilbert-Schmidt norm, the Schur norm, or the Frobenius norm.

we get $s_1(A) = \cdots = s_{n-1}(A) = 1, s_n(A) = 0$ *and* $s_1(B) = \cdots = s_{n-1}(B) = 1, s_n(B) = \kappa_n$ *for large enough* n*, and therefore, for any choice of* κ_n*, since the atom* κ_n *has weight* $1/n$*,*

$$\nu_A \rightsquigarrow \delta_1 \quad \textit{and} \quad \nu_B \rightsquigarrow \delta_1.$$

This example shows the stability of the limiting distribution of singular values under an additive perturbation of rank 1 *of arbitrary large norm, and the instability of the limiting eigenvalues distribution under an additive perturbation of rank* 1 *of arbitrary small norm.*

Beyond square matrices, one may define the singular values $s_1(A), \ldots, s_m(A)$ of a rectangular matrix $A \in \mathcal{M}_{m,n}(\mathbb{C})$ with $m \leq n$ by $s_i(A) := \lambda_i(\sqrt{AA^*})$ for every $1 \leq i \leq m$. The famous Singular Value Decomposition (SVD, see [76, Th. 3.3.1]) states then that

$$A = UDV^*$$

where U and V are the unitary matrices of the eigenvectors of AA^* and A^*A and where $D = \mathrm{diag}(s_1(A), \ldots, s_n(A))$ is a $m \times n$ diagonal matrix. The SVD is at the heart of many numerical techniques in concrete applied mathematics (pseudo-inversion, regularization, low dimensional approximation, principal component analysis, etc). Note that if A is square then the Hermitian matrix $H := VDV^*$ and the unitary matrix $W := UV^*$ form the polar decomposition $A = WH$ of A. Note also that if W_1 and W_2 are unitary matrices then W_1AW_2 and A have the same singular values.

We refer to the books [76] and [62] for more details on basic properties of the singular values and eigenvalues of deterministic matrices. The sensitivity of the spectrum to perturbations of small norm is captured by the notion of *pseudo-spectrum*. Namely, for a matrix norm $\|\cdot\|$ and a positive real ε, the $(\|\cdot\|, \varepsilon)$-pseudo-spectrum of A is defined by

$$\Lambda_{\|\cdot\|,\varepsilon}(A) := \bigcup_{\|A-B\| \leq \varepsilon} \{\lambda_1(B), \ldots, \lambda_n(B)\}.$$

If A is normal then its pseudo-spectrum for the operator norm $\|\cdot\|_{2\to 2}$ coincides with the ε-neighborhood of its spectrum. The pseudo-spectrum can be much larger for non-normal matrices. For instance, if A is the nilpotent matrix considered earlier, then the asymptotic (as $n \to \infty$) pseudo-spectrum for the operator norm contains the unit disc if κ_n is well chosen. For more, see the book [133].

2. Quarter circular and circular laws

The variance of a random variable Z on $\mathbb{C}$ is $\mathrm{Var}(Z) = \mathbb{E}(|Z|^2) - |\mathbb{E}(Z)|^2$. Let $(X_{ij})_{i,j\geq 1}$ be an infinite table of i.i.d. random variables on $\mathbb{C}$ with variance 1. We consider the square random matrix $X := (X_{ij})_{1\leq i,j\leq n}$ as a random variable in $\mathcal{M}_n(\mathbb{C})$. We write *a.s.*, *a.a.*, and *a.e.* for *almost surely*, *Lebesgue almost all*, and *Lebesgue almost everywhere* respectively.

We start with a reformulation in terms of singular values of the classical Marchenko-Pastur theorem for the "empirical covariance matrix" $\frac{1}{n}XX^*$. This theorem is universal in the sense that the limiting distribution does not depend on the law of X_{11}.

Theorem 2.1 (Marchenko-Pastur quarter circular law)**.** *a.s.* $\nu_{n^{-1/2}X} \rightsquigarrow \mathcal{Q}_2$ *as* $n \to \infty$*, where* $\mathcal{Q}_2$ *is the quarter circular law*[5] *on* $[0, 2] \subset \mathbb{R}_+$ *with density* $x \mapsto \pi^{-1}\sqrt{4 - x^2}\mathbf{1}_{[0,2]}(x)$*.*

The $n^{-1/2}$ normalization factor is easily understood from the law of large numbers:

$$\int s^2 \, d\nu_{n^{-1/2}X}(s) = \frac{1}{n^2}\sum_{i=1}^{n} s_i(X)^2 = \frac{1}{n^2}\mathrm{Tr}(XX^*) = \frac{1}{n^2}\sum_{i,j=1}^{n} |X_{i,j}|^2 \to \mathbb{E}(|X_{1,1}|^2). \quad (2.1)$$

The central subject of these notes is the following counterpart for the eigenvalues.

[5]Actually, it is a quarter ellipse rather than a quarter circle, due to the normalizing factor $1/\pi$. However, one may use different scales on the horizontal and vertical axes to see a true quarter circle, as in figure 2.

Theorem 2.2 (Girko circular law)**.** *a.s.* $\mu_{n^{-1/2}X} \rightsquigarrow \mathcal{C}_1$ *as* $n \to \infty$*, where* $\mathcal{C}_1$ *is the circular law*[6] *which is the uniform law on the unit disc of* $\mathbb{C}$ *with density* $z \mapsto \pi^{-1}\mathbf{1}_{\{z\in\mathbb{C}:|z|\leq 1\}}$.

Note that if Z is a complex random variable following the uniform law on the unit disc $\{z \in \mathbb{C} : |z| \leq 1\}$ then the random variables $\mathfrak{Re}(Z)$ and $\mathfrak{Im}(Z)$ follow the semi circular law on $[-1, 1]$, but are not independent. Additionally, the random variables $|\mathfrak{Re}(Z)|$ and $|\mathfrak{Im}(Z)|$ follow the quarter circular law on $[0, 1]$, and $|Z|$ follows the law with density $\rho \mapsto \frac{1}{2}\rho\mathbf{1}_{[0,1]}(\rho)$. We will see in section 5 that the notion of freeness developed in free probability is the key to understand these relationships. An extension of theorem 2.1 is the key to deduce theorem 2.2 via a Hermitization technique, as we will see in section 4.

The circular law theorem 2.2 has a long history. It was established through a sequence of partial results during the period 1965–2009, the general case being finally obtained by Tao and Vu [132]. Indeed Mehta [98] was the first to obtain a circular law theorem for the expected empirical spectral distribution in the complex Gaussian case, by using the explicit formula for the spectrum due to Ginibre [51]. Edelman was able to prove the same kind of result for the far more delicate real Gaussian case [40]. Silverstein provided an argument to pass from the expected to the almost sure convergence in the complex Gaussian case [78]. Girko worked on the universal version and came with very good ideas such as the Hermitization technique [52, 54, 56, 57, 58]. Unfortunately, his work was controversial due to a lack of clarity and rigor[7]. In particular, his approach relies implicitly on an unproved uniform integrability related to the behavior of the smallest singular values of random matrices. Bai [10] was the first to circumvent the problem in the approach of Girko, at the price of bounded density assumptions and moments assumptions[8]. Bai improved his approach in his book written with Silverstein [11]. His approach involves the control of the speed of convergence of the singular values distribution. Śniady considered a universal version beyond random matrices and the circular law, using the notion of $*$-moments and Brown measure of operators in free probability, and a regularization by adding an independent Gaussian Ginibre noise [120]. Goldsheid and Khoruzhenko [15] used successfully the logarithmic potential to derive the analogue of the circular law theorem for random non-Hermitian tridiagonal matrices. The smallest singular value of random matrices was the subject of an impressive activity culminating with the works of Tao and Vu [127] and of Rudelson and Vershynin [113], using tools from asymptotic geometric analysis and additive combinatorics (Littlewood-Offord problems). These achievements allowed Götze and Tikhomirov [63] to obtain the expected circular law theorem up to a small loss in the moment assumption, by using the logarithmic potential. Similar ingredients are present in the work of Pan and Zhou [101]. At the same time, Tao and Vu, using a refined bound on the smallest singular value and the approach of Bai, deduced the circular law theorem up to a small loss in the moment assumption [128]. As in the works of Girko, Bai and their followers, the loss was due to a sub-optimal usage of the Hermitization approach. In [132], Tao and Vu finally obtained the full circular law theorem 2.2 by using the full strength of the logarithmic potential, and a new control of the count of the small singular values which replaces the speed of convergence estimates of Bai. See also their synthetic paper [129]. We will follow essentially their approach in section 4 to prove theorem 2.2.

[6] It is not customary to call it instead the "disc law". The terminology corresponds to what we actually draw: a circle for the circular law, a quarter circle (actually a quarter ellipse) for the quarter circular law, even if it is the boundary of the support in the first case, and the density in the second case. See figure 2.

[7] Girko's writing style is also quite original, see for instance the recent paper [59].

[8] *. . . I worked for 13 years from 1984 to 1997, which was eventually published in Annals of Probability. It was the hardest problem I have ever worked on.* Zhidong Bai, interview with Atanu Biswas in 2006 [36].

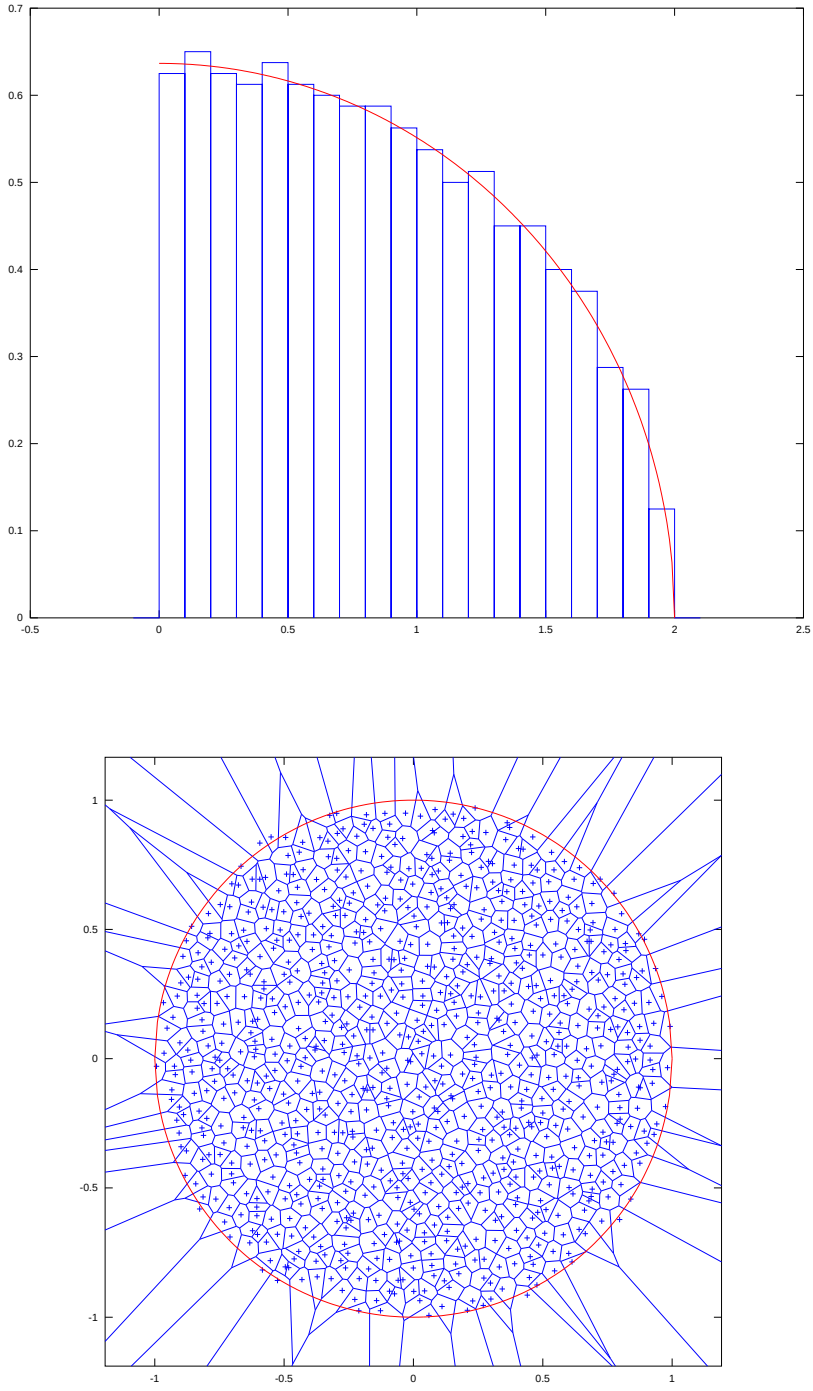

FIGURE 2. Simulation of the quarter circular law and circular law theorems 2.1 and 2.2 in the complex Gaussian case (Complex Ginibre Ensemble) for a single matrix with $n = 800$. The bottom graphic shows the Voronoï tessellation of the spectrum. One of the striking fact behind this graphic is the remarkable stability of the numerical algorithms despite the sensitivity of the spectrum of non-normal matrices (numerical self regularization). The Śniady regularization of Brown measure theorem [120] at work?

The a.s. tightness of $\mu_{n^{-1/2}X}$ is easily understood since by Weyl's inequality we obtain

$$\int |\lambda|^2 \, d\mu_{n^{-1/2}X}(\lambda) = \frac{1}{n^2}\sum_{i=1}^{n} |\lambda_i(X)|^2 \leq \frac{1}{n^2}\sum_{i=1}^{n} s_i(X)^2 = \int s^2 \, d\nu_{n^{-1/2}X}(s).$$

The convergence in the couple of theorems above is the weak convergence of probability measures with respect to continuous bounded functions. We recall that this mode of convergence does not capture the convergence of the support. It implies only that a.s.

$$\varliminf_{n\to\infty} s_1(n^{-1/2}X) \geq 2 \quad \text{and} \quad \varliminf_{n\to\infty} |\lambda_1(n^{-1/2}X)| \geq 1.$$

However, following [14, 12, 11] and [13, 101], if $\mathbb{E}(X_{1,1}) = 0$ and $\mathbb{E}(|X_{1,1}|^4) < \infty$ then a.s.[9]

$$\lim_{n\to\infty} s_1(n^{-1/2}X) = 2 \quad \text{and} \quad \lim_{n\to\infty} |\lambda_1(n^{-1/2}X)| = 1.$$

The asymptotic factor 2 between the operator norm and the spectral radius indicates in a sense that X is a non-normal matrix asymptotically as $n \to \infty$ (note that if X_{11} is absolutely continuous then X is absolutely continuous and thus $XX^* \neq X^*X$ a.s. which means that X is non-normal a.s.). The law of the modulus under the circular law has density $\rho \mapsto 2\rho \mathbf{1}_{[0,1]}(\rho)$ which differs completely from the shape of the quarter circular law $s \mapsto \pi^{-1}\sqrt{4-s^2}\,\mathbf{1}_{[0,2]}(s)$, see figure 2. The integral of "log" for both laws is the same.

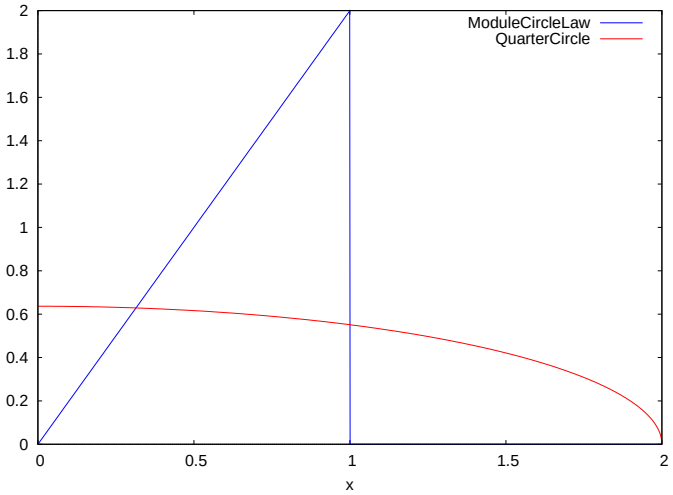

FIGURE 3. Comparison between the quarter circular distribution of theorem 2.1 for the singular values, and the modulus under the circular law of theorem 2.2 for the eigenvalues. The supports and the shapes are different. This difference indicates the asymptotic non-normality of these matrices. The integral of the function $t \mapsto \log(t)$ is the same for both distributions.

3. GAUSSIAN CASE

This section is devoted to the case where $X_{11} \sim \mathcal{N}(0, \frac{1}{2}I_2)$. From now on, we denote G instead of X in order to distinguish the Gaussian case from the general case. We say that G belongs to the *Complex Ginibre Ensemble*. The Lebesgue density of the $n \times n$ random matrix $G = (G_{i,j})_{1\leq i,j\leq n}$ in $\mathcal{M}_n(\mathbb{C}) \equiv \mathbb{C}^{n\times n}$ is

$$A \in \mathcal{M}_n(\mathbb{C}) \mapsto \pi^{-n^2} e^{-\sum_{i,j=1}^{n} |A_{ij}|^2} = \pi^{-n^2} e^{-\mathrm{Tr}(AA^*)} \tag{3.1}$$

[9]The argument is based on Gelfand's formula: if $A \in \mathcal{M}_n(\mathbb{C})$ then $|\lambda_1(A)| = \lim_{k\to\infty} \|A^k\|^{1/k}$ for any sub-multiplicative matrix norm $\|\cdot\|$ (e.g. operator or trace norm). In the same spirit, the Yamamoto theorem states that $\lim_{k\to\infty} s_i(A^k)^{1/k} = |\lambda_i(A)|$ for every $1 \leq i \leq n$, see [76, Theorem 3.3.21].

where A^* the conjugate-transpose of A. This law is unitary invariant, in the sense that if U and V are $n \times n$ unitary matrices then UGV and G are equally distributed. If H_1 and H_2 are two independent copies of GUE[10] then $(H_1 + iH_2)/\sqrt{2}$ has the law of G. Conversely, the matrices $(G + G^*)/\sqrt{2}$ and $(G - G^*)/\sqrt{2i}$ are independent and belong to the GUE.

The singular values of G are the square root of the eigenvalues of the positive semidefinite Hermitian matrix GG^*. The matrix GG^* is a complex Wishart matrix, and belongs to the complex Laguerre Ensemble ($\beta = 2$). The empirical distribution of the singular values of $n^{-1/2}G$ tends to the Marchenko-Pastur quarter circular distribution (Gaussian case in theorem 2.1). This section is rather devoted to the study of the eigenvalues of G, and in particular to the proof of the circular law theorem 2.2 in this Gaussian settings.

Lemma 3.1 (Diagonalizability)**.** *For every $n \geq 1$, the set of elements of $\mathcal{M}_n(\mathbb{C})$ with multiple eigenvalues has zero Lebesgue measure in $\mathbb{C}^{n\times n}$. In particular, the set of non-diagonalizable elements of $\mathcal{M}_n(\mathbb{C})$ has zero Lebesgue measure in $\mathbb{C}^{n\times n}$.*

Proof. If $A \in \mathcal{M}_n(\mathbb{C})$ has characteristic polynomial

$$P_A(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0,$$

then $a_0, \ldots, a_{n-1}$ are polynomials of the entries of A. The *resultant* $R(P_A, P'_A)$ of P_A, P'_A, called the *discriminant* of P_A, is the determinant of the $(2n-1)\times(2n-1)$ Sylvester matrix of P_A, P'_A. It is a polynomial in $a_0, \ldots, a_{n-1}$. We have also the Vandermonde formula

$$|R(P_A, P'_A)| = \prod_{i<j} |\lambda_i(A) - \lambda_j(A)|^2.$$

Consequently, A has all eigenvalues distinct if and only if A lies outside the proper polynomial hyper-surface $\{A \in \mathbb{C}^{n\times n} : R(P_A, P'_A) = 0\}$. □

Since G is absolutely continuous, we have a.s. $GG^* \neq G^*G$ (non-normality). Additionally, lemma 3.1 gives that a.s. G is diagonalizable with distinct eigenvalues. Following Ginibre [51] – see also [99, Chapter 15] and [47, Chapter 15] – one may then compute the joint density of the eigenvalues $\lambda_1(G), \ldots, \lambda_n(G)$ of G by integrating (3.1) over the eigenvectors matrix. The result is stated in theorem 3.2 below. The law of G is invariant by the multiplication of the entries with a common phase, and thus the law of the spectrum of G has also the same property. In the sequel we set

$$\Delta_n := \{(z_1, \ldots, z_n) \in \mathbb{C}^n : |z_1| \geq \cdots \geq |z_n|\}.$$

Theorem 3.2 (Spectrum law)**.** *$(\lambda_1(G), \ldots, \lambda_n(G))$ has density $n!\varphi_n \mathbf{1}_{\Delta_n}$ where*

$$\varphi_n(z_1, \ldots, z_n) = \frac{\pi^{-n^2}}{1!2!\cdots n!} \exp\left(-\sum_{k=1}^n |z_k|^2\right) \prod_{1\leq i<j\leq n} |z_i - z_j|^2.$$

In particular, for every symmetric Borel function $F : \mathbb{C}^n \to \mathbb{R}$,

$$\mathbb{E}[F(\lambda_1(G), \ldots, \lambda_n(G))] = \int_{\mathbb{C}^n} F(z_1, \ldots, z_n)\varphi_n(z_1, \ldots, z_n)\, dz_1 \cdots dz_n.$$

We will use theorem 3.2 with symmetric functions of the form

$$F(z_1, \ldots, z_n) = \sum_{i_1,\ldots,i_k \text{ distinct}} f(z_{i_1}) \cdots f(z_{i_k}).$$

The Vandermonde determinant comes from the Jacobian of the diagonalization, and can be interpreted as an electrostatic repulsion. The spectrum is a Gaussian determinantal point process, see [77, Chapter 4].

[10] Up to a scaling factor, a random $n \times n$ Hermitian matrix H belongs to the Gaussian Unitary Ensemble (GUE) when its density is proportional to $H \mapsto \exp(-\frac{1}{2}\mathrm{Tr}(H^2)) = \exp(-\frac{1}{2}\sum_{i=1}^n |H_{ii}|^2 - \sum_{1\leq i<j\leq n} |H_{ij}|^2)$. Equivalently $(H_{ii}, H_{ij})_{1\leq i\leq n, i<j\leq n}$ are independent with $H_{ii} \sim \mathcal{N}(0, 1)$ and $H_{ij} \sim \mathcal{N}(0, \frac{1}{2}I_2)$ for $i \neq j$.

Theorem 3.3 (k-points correlations)**.** *Let $z \in \mathbb{C} \mapsto \gamma(z) = \pi^{-1} e^{-|z|^2}$ be the density of the standard Gaussian $\mathcal{N}(0, \frac{1}{2} I_2)$ on $\mathbb{C}$. Then for every $1 \leq k \leq n$, the "k-point correlation"*

$$\varphi_{n,k}(z_1, \ldots, z_k) := \int_{\mathbb{C}^{n-k}} \varphi_n(z_1, \ldots, z_k)\, dz_{k+1} \cdots dz_n$$

satisfies

$$\varphi_{n,k}(z_1, \ldots, z_k) = \frac{(n-k)!}{n!} \pi^{-k^2} \gamma(z_1) \cdots \gamma(z_k) \det\left[K(z_i, z_j)\right]_{1 \leq i,j \leq k}$$

where

$$K(z_i, z_j) := \sum_{\ell=0}^{n-1} \frac{(z_i z_j^*)^\ell}{\ell!} = \sum_{\ell=0}^{n-1} H_\ell(z_i) H_\ell(z_j)^* \quad \textit{with} \quad H_\ell(z) := \frac{1}{\sqrt{\ell!}} z^\ell.$$

In particular, by taking $k = n$ we get

$$\varphi_{n,n}(z_1, \ldots, z_n) = \varphi_n(z_1, \ldots, z_n) = \frac{1}{n!} \pi^{-n^2} \gamma(z_1) \cdots \gamma(z_n) \det\left[K(z_i, z_j)\right]_{1 \leq i,j \leq n}.$$

Proof. Calculations made by [99, Chapter 15 page 271 equation 15.1.29] using

$$\prod_{1 \leq i < j \leq n} |z_i - z_j|^2 = \prod_{1 \leq i < j \leq n} (z_i - z_j) \prod_{1 \leq i < j \leq n} (z_i - z_j)^*$$

and

$$\det\left[z_j^{i-1}\right]_{1 \leq i,j \leq n} \det\left[(z_j^*)^{i-1}\right]_{1 \leq i,j \leq n} = \frac{1}{n!} \det\left[K(z_i, z_j)\right]_{1 \leq i,j \leq n}.$$

□

Recall that if μ is a random probability measure on $\mathbb{C}$ then $\mathbb{E}\mu$ is the deterministic probability measure defined for every bounded measurable f by

$$\int f\, d\mathbb{E}\mu := \mathbb{E} \int f\, d\mu.$$

Theorem 3.4 (Mean circular Law)**.** *$\mathbb{E}\mu_{n^{-1/2}G} \rightsquigarrow \mathcal{C}_1$ as $n \to \infty$, where $\mathcal{C}_1$ is the circular law i.e. the uniform law on the unit disc of $\mathbb{C}$ with density $z \mapsto \pi^{-1} \mathbf{1}_{\{z \in \mathbb{C} : |z| \leq 1\}}$.*

Proof. From theorem 3.3, with $k = 1$, we get that the density of $\mathbb{E}\mu_G$ is

$$\varphi_{n,1} : z \mapsto \gamma(z) \left(\frac{1}{n} \sum_{\ell=0}^{n-1} |H_\ell|^2(z) \right) = \frac{1}{n\pi} e^{-|z|^2} \sum_{\ell=0}^{n-1} \frac{|z|^{2\ell}}{\ell!}.$$

Following Mehta [99, Chapter 15 page 272], some calculus gives for every compact $C \subset \mathbb{C}$

$$\lim_{n \to \infty} \sup_{z \in C} \left| n \varphi_{n,1}(\sqrt{n} z) - \pi^{-1} \mathbf{1}_{[0,1]}(|z|) \right| = 0.$$

The n in front of $\varphi_{n,1}$ is due to the fact that we are on the complex plane $\mathbb{C} = \mathbb{R}^2$ and thus $d\sqrt{n}x d\sqrt{n}y = n dx dy$. Here is the start of the elementary calculus: for $r^2 < n$,

$$e^{r^2} - \sum_{\ell=0}^{n-1} \frac{r^{2\ell}}{\ell!} = \sum_{\ell=n}^{\infty} \frac{r^{2\ell}}{\ell!} \leq \frac{r^{2n}}{n!} \sum_{\ell=0}^{\infty} \frac{r^{2\ell}}{(n+1)^\ell} = \frac{r^{2n}}{n!} \frac{n+1}{n+1-r^2}$$

while for $r^2 > n$,

$$\sum_{\ell=0}^{n-1} \frac{r^{2\ell}}{\ell!} \leq \frac{r^{2(n-1)}}{(n-1)!} \sum_{\ell=0}^{n-1} \left(\frac{n-1}{r^2} \right)^\ell \leq \frac{r^{2(n-1)}}{(n-1)!} \frac{r^2}{r^2 - n + 1}.$$

By taking $r^2 = |\sqrt{n} z|^2$ we obtain the convergence of the density uniformly on compact subsets, which implies in particular the weak convergence. □

The sequence $(H_k)_{k\in\mathbb{N}}$ forms an orthonormal basis (orthogonal polynomials) of square integrable analytic functions on $\mathbb{C}$ for the standard Gaussian on $\mathbb{C}$. The uniform law on the unit disc (known as the circular law) is the law of $\sqrt{V}e^{2i\pi W}$ where V and W are i.i.d. uniform random variables on the interval $[0,1]$. This point of view can be used to interpolate between complex Ginibre and GUE via Girko's elliptic laws, see [89, 80, 20].

We are ready to prove the complex Gaussian version of the circular law theorem 2.2.

Theorem 3.5 (Circular law)**.** *a.s. $\mu_{n^{-1/2}G} \rightsquigarrow \mathcal{C}_1$ as $n \to \infty$, where $\mathcal{C}_1$ is the circular law i.e. the uniform law on the unit disc of $\mathbb{C}$ with density $z \mapsto \pi^{-1}\mathbf{1}_{\{z\in\mathbb{C}:|z|\leq 1\}}$.*

Proof. We reproduce Silverstein's argument, published by Hwang [78]. The argument is similar to the proof of the strong law of large numbers for i.i.d. random variables with finite fourth moment. It suffices to establish the result for compactly supported continuous bounded functions. Let us pick such a function f and set

$$S_n := \int_{\mathbb{C}} f\, d\mu_{n^{-1/2}G} \quad \text{and} \quad S_\infty := \pi^{-1}\int_{|z|\leq 1} f(z)\, dxdy.$$

Suppose for now that we have

$$\mathbb{E}[(S_n - \mathbb{E}S_n)^4] = O(n^{-2}). \tag{3.2}$$

By monotone convergence (or by the Fubini-Tonelli theorem),

$$\mathbb{E}\sum_{n=1}^{\infty}(S_n - \mathbb{E}S_n)^4 = \sum_{n=1}^{\infty}\mathbb{E}[(S_n - \mathbb{E}S_n)^4] < \infty$$

and consequently $\sum_{n=1}^{\infty}(S_n - \mathbb{E}S_n)^4 < \infty$ a.s. which implies $\lim_{n\to\infty} S_n - \mathbb{E}S_n = 0$ a.s. Since $\lim_{n\to\infty}\mathbb{E}S_n = S_\infty$ by theorem 3.4, we get that a.s.

$$\lim_{n\to\infty} S_n = S_\infty.$$

Finally, one can swap the universal quantifiers on ω and f thanks to the separability of $(\mathcal{C}_c(\mathbb{C},\mathbb{R}), \|\cdot\|_\infty)$. To establish (3.2), we set

$$S_n - \mathbb{E}S_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \quad \text{with} \quad Z_i := f\Big(\lambda_i\Big(n^{-1/2}G\Big)\Big).$$

Next, we obtain, with $\sum_{i_1,\ldots}$ running over distinct indices in $1,\ldots,n$,

$$\begin{aligned}\mathbb{E}\Big[(S_n - \mathbb{E}S_n)^4\Big] = \frac{1}{n^4}\sum_{i_1}\mathbb{E}[Z_{i_1}^4] &\\ + \frac{4}{n^4}\sum_{i_1,i_2}\mathbb{E}[Z_{i_1}Z_{i_2}^3] &\\ + \frac{3}{n^4}\sum_{i_1,i_2}\mathbb{E}[Z_{i_1}^2Z_{i_2}^2] &\\ + \frac{6}{n^4}\sum_{i_1,i_2,i_3}\mathbb{E}[Z_{i_1}Z_{i_2}Z_{i_3}^2] &\\ + \frac{1}{n^4}\sum_{i_1,i_2,i_3,i_3,i_4}\mathbb{E}[Z_{i_1}Z_{i_3}Z_{i_3}Z_{i_4}]. &\end{aligned}$$

The first three terms of the right hand side are $O(n^{-2})$ since $\max_{1\leq i\leq n}|Z_i| \leq \|f\|_\infty$. Finally, some calculus using the expressions of $\varphi_{n,3}$ and $\varphi_{n,4}$ provided by theorem 3.3 allows to show that the remaining two terms are also $O(n^{-2})$. See Hwang [78, p. 151]. □

Following Kostlan [86] (see also Rider [107] and [77]) the integration of the phases in the joint density of the spectrum given by theorem 3.2 leads to theorem 3.6 below.

Theorem 3.6 (Layers)**.** *If $Z_1, \dots, Z_n$ are independent with $Z_k^2 \sim \Gamma(k,1)$ for every k, then*

$$(|\lambda_1(G)|, \dots, |\lambda_n(G)|) \stackrel{d}{=} (Z_{(1)}, \dots, Z_{(n)})$$

where $Z_{(1)}, \dots, Z_{(n)}$ is the non-increasing reordering of the sequence $Z_1, \dots, Z_n$.

Note that $(\sqrt{2}Z_k)^2 \sim \chi^2(2k)$ which is useful for $\sqrt{2}G$. Since $Z_k^2 \stackrel{d}{=} E_1 + \cdots + E_k$ where $E_1, \dots, E_k$ are i.i.d. exponential random variables of unit mean, we get, for every $r > 0$,

$$\mathbb{P}\big(|\lambda_1(G)| \le \sqrt{n}r\big) = \prod_{1 \le k \le n} \mathbb{P}\bigg(\frac{E_1 + \cdots + E_k}{n} \le r^2\bigg)$$

The law of large numbers suggests that $r = 1$ is a critical value. The central limit theorem suggests that $\big|\lambda_1(n^{-1/2}G)\big|$ behaves when $n \gg 1$ as the maximum of i.i.d. Gaussians, for which the fluctuations follow the Gumbel law. A quantitative central limit theorem and the Borel-Cantelli lemma provides the follow result. The full proof is in Rider [107].

Theorem 3.7 (Convergence and fluctuation of the spectral radius)**.**

$$\mathbb{P}\Big(\lim_{n\to\infty} |\lambda_1(n^{-1/2}G)| = 1\Big) = 1.$$

Moreover, if $\gamma_n := \log(n/2\pi) - 2\log(\log(n))$ then

$$\sqrt{4n\gamma_n}\bigg(|\lambda_1(n^{-1/2}G)| - 1 - \sqrt{\frac{\gamma_n}{4n}}\bigg) \xrightarrow[n\to\infty]{d} \mathcal{G}$$

where $\mathcal{G}$ is the Gumbel law with cumulative distribution function $x \mapsto e^{-e^{-x}}$ on $\mathbb{R}$.

The convergence of the spectral radius was obtained by Mehta [99, chapter 15 page 271 equation 15.1.27] by integrating the joint density of the spectrum of theorem 3.2 over the set $\bigcap_{1\le i\le n}\{|\lambda_i| > r\}$. The same argument is reproduced by Hwang [78, pages 149–150]. Let us give now an alternative derivation of theorem 3.4. From theorem 3.7, the sequence $(\mathbb{E}\mu_{n^{-1/2}G})_{n\ge1}$ is tight and every adherence value μ is supported in the unit disc. From theorem 3.2, such a μ is rotationally invariant, and from theorem 3.6, the image of μ by $z \in \mathbb{C} \mapsto |z|$ has density $r \mapsto 2r\mathbf{1}_{[0,1]}(r)$ (use moments!). Theorem 3.4 follows immediately.

It is remarkable that the large eigenvalues in modulus of the complex Ginibre ensemble are asymptotically independent, which gives rise to a Gumbel fluctuation, in contrast with the GUE and its delicate Tracy-Widom fluctuation, see [80] for an interpolation.

Remark 3.8 (Real Ginibre Ensemble)**.** *Ginibre considered also in his paper* [51] *the case where $\mathbb{C}$ is replaced by $\mathbb{R}$ or by the quaternions. These cases are less understood than the complex case due to their peculiarities. Let us focus on the Real Ginibre Ensemble, studied by Edelman and his collaborators. The expected number of real eigenvalues is equivalent to $\sqrt{2n/\pi}$ as $n \to \infty$, see* [41]*, while the probability that all the eigenvalues are real is exactly $2^{-n(n-1)/4}$, see* [40, Corollary 7.1]*. The expected counting measure of the real eigenvalues, scaled by $\sqrt{n}$, tends to the uniform law on the interval $[-1,1]$, see* [41, Theorem 4.1] *and figures 5. The eigenvalues do not have a density in $\mathbb{C}^n$, except if we condition on the real eigenvalues, see* [40]*. The analogue of the weak circular law theorem 3.4 was proved by Edelman* [41, Theorem 6.3]*. More information on the structure of the Real Ginibre Ensemble can be found in* [2]*,* [27]*, and* [47, Chapter 15] *and references therein.*

On overall, one can remember that the Complex Ginibre Ensemble is in a way "simpler" than the GUE while the Real Ginibre Ensemble is "harder" than the GOE.

$$\boxed{\text{Real Ginibre} \ge \text{GOE} \ge \text{GUE} \ge \text{Complex Ginibre}}$$

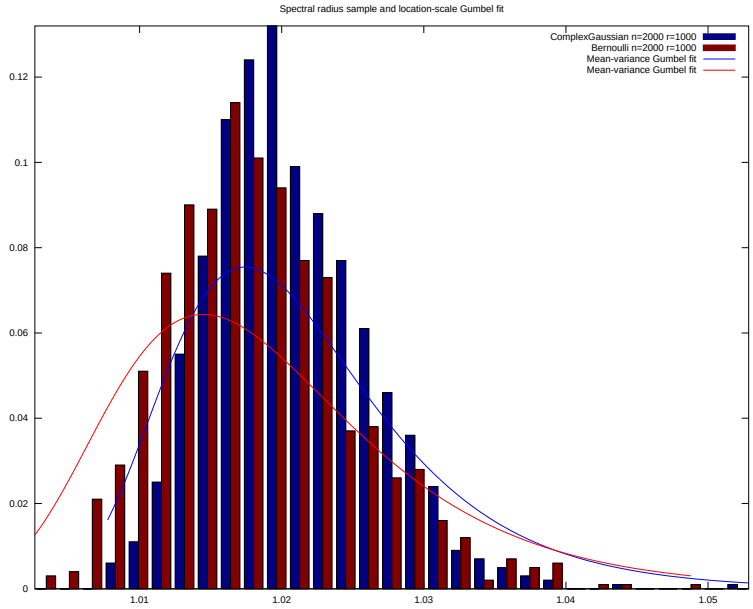

FIGURE 4. Spectral radius empirical distribution and mean-variance Gumbel fit, in the Complex Gaussian case $X_{11} \sim \mathcal{N}(0, \frac{1}{2}I_2)$, and in the Bernoulli case $X_{11} \sim \frac{1}{2}(\delta_{-1} + \delta_1)$, for a spectral radius sample. The convergence seems slow even for the Complex Gaussian case, making the Gumbel fits not convincing. However, one can point out that the scale is quite tight.

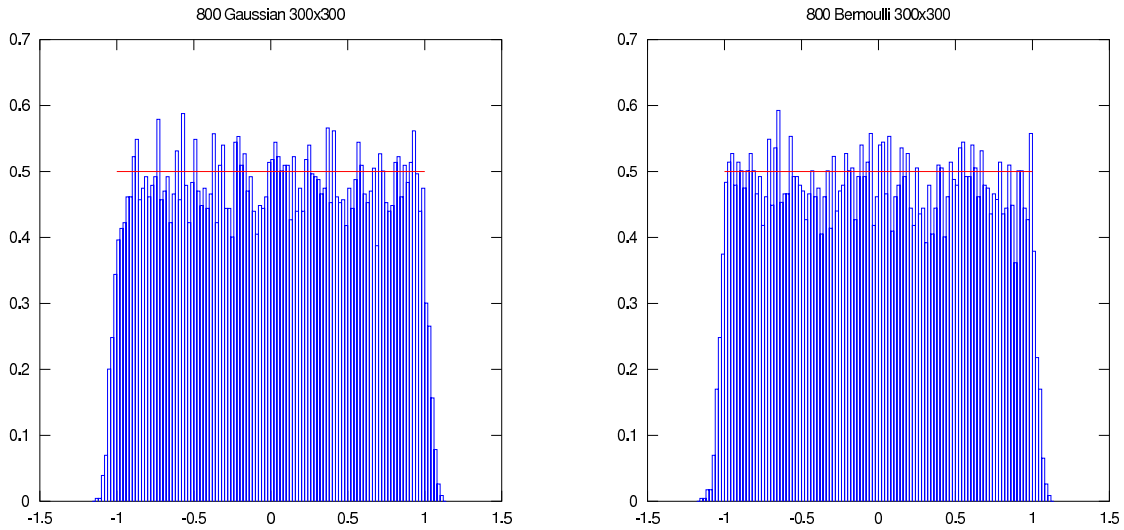

FIGURE 5. Histograms of real eigenvalues of 800 i.i.d. copies of $n^{-1/2}X$ with $n = 300$. The left hand side graphic corresponds to the standard real Gaussian case $X_{11} \sim \mathcal{N}(0, 1)$, while the right hand side graphic corresponds to the symmetric Bernoulli case $X_{11} \sim \frac{1}{2}(\delta_{-1} + \delta_1)$. See remark 3.8.

Remark 3.9 (Quaternionic Ginibre Ensemble)**.** *The quaternionic Ginibre Ensemble was considered at the origin by Ginibre* [51]*. It has been recently shown* [19] *by using the logarithmic potential that there exists an analogue of the circular law theorem for this ensemble, in which the limiting law is supported in the unit ball of the quaternions field.*

4. Universal case

This section is devoted to the proof of the circular law theorem 2.2 following [132]. The universal Marchenko-Pastur theorem 2.1 can be proved by using powerful Hermitian techniques such as truncation, centralization, the method of moments, or the Cauchy-Stieltjes trace-resolvent transform. It turns out that all these techniques fail for the eigenvalues of non-normal random matrices. Indeed, the key to prove the circular law theorem 2.2 is to use a bridge pulling back the problem to the Hermitian world. This is called *Hermitization*.

Actually, and as we will see in sections 5 and 6, there is a non-Hermitian analogue of the method of moments called the $*$-moments, and there is an analogue of the Cauchy-Stieltjes trace-resolvent in which the complex variable is replaced by a quaternionic type variable.

4.1. **Logarithmic potential and Hermitization.** Let $\mathcal{P}(\mathbb{C})$ be the set of probability measures on $\mathbb{C}$ which integrate $\log|\cdot|$ in a neighborhood of infinity. The *logarithmic potential* U_μ of $\mu \in \mathcal{P}(\mathbb{C})$ is the function $U_\mu : \mathbb{C} \to (-\infty, +\infty]$ defined for all $z \in \mathbb{C}$ by

$$U_\mu(z) = -\int_{\mathbb{C}} \log|z-\lambda|\, d\mu(\lambda) = -(\log|\cdot| * \mu)(z). \tag{4.1}$$

For instance, for the circular law $\mathcal{C}_1$ of density $\pi^{-1}\mathbf{1}_{\{z\in\mathbb{C}:|z|\leq 1\}}$, we have, for every $z \in \mathbb{C}$,

$$U_{\mathcal{C}_1}(z) = \begin{cases} -\log|z| & \text{if } |z| > 1, \\ \frac{1}{2}(1-|z|^2) & \text{if } |z| \leq 1, \end{cases} \tag{4.2}$$

see e.g. [115]. Let $\mathcal{D}'(\mathbb{C})$ be the set of Schwartz-Sobolev distributions. We have $\mathcal{P}(\mathbb{C}) \subset \mathcal{D}'(\mathbb{C})$. Since $\log|\cdot|$ is Lebesgue locally integrable on $\mathbb{C}$, the Fubini-Tonelli theorem implies that U_μ is Lebesgue locally integrable on $\mathbb{C}$. In particular, $U_\mu < \infty$ a.e. and $U_\mu \in \mathcal{D}'(\mathbb{C})$.

Let us define the first order linear differential operators in $\mathcal{D}'(\mathbb{C})$

$$\partial := \frac{1}{2}(\partial_x - i\partial_y) \quad \text{and} \quad \overline{\partial} := \frac{1}{2}(\partial_x + i\partial_y) \tag{4.3}$$

and the Laplace operator $\Delta = 4\partial\overline{\partial} = 4\overline{\partial}\partial = \partial_x^2 + \partial_y^2$. Each of these operators coincide on smooth functions with the usual differential operator acting on smooth functions. By using Green's or Stockes' theorems, one may show, for instance via the Cauchy-Pompeiu formula, that for any smooth and compactly supported function $\varphi : \mathbb{C} \to \mathbb{R}$,

$$\int_{\mathbb{C}} \Delta\varphi(z) \log|z|\, dxdy = -2\pi\varphi(0) \tag{4.4}$$

where $z = x + iy$. Now (4.4) can be written, in $\mathcal{D}'(\mathbb{C})$,

$$\Delta \log|\cdot| = -2\pi\delta_0$$

In other words, $-\frac{1}{2\pi}\log|\cdot|$ is the fundamental solution of the Laplace equation on $\mathbb{R}^2$. Note that $\log|\cdot|$ is harmonic on $\mathbb{C}\setminus\{0\}$. It follows that in $\mathcal{D}'(\mathbb{C})$,

$$\Delta U_\mu = -2\pi\mu. \tag{4.5}$$

This means that for every smooth and compactly supported "test function" $\varphi : \mathbb{C} \to \mathbb{R}$,

$$\int_{\mathbb{C}} \Delta\varphi(z) U_\mu(z)\, dxdy = -2\pi \int_{\mathbb{C}} \varphi(z)\, d\mu(z)$$

where $z = x+iy$. This means that $-\frac{1}{2\pi}U_\cdot$ is the Green operator on $\mathbb{R}^2$ (Laplacian inverse).

Lemma 4.1 (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 (4.5) gives $\mu = \nu$ in $\mathcal{D}'(\mathbb{C})$, and thus $\mu = \nu$ as measures since μ and ν are Radon measures. Note more generally that the lemma remains valid if $U_\mu = U_\nu + h$ for some harmonic $h \in \mathcal{D}'(\mathbb{C})$. □

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

$$U_{\mu_A}(z) = -\int_{\mathbb{C}} \log|\lambda - z| \, d\mu_A(\lambda) = -\frac{1}{n} \log|\det(A - zI)| = -\frac{1}{n} \log|P_A(z)|$$

for every $z \in \mathbb{C} \setminus \{\lambda_1(A), \ldots, \lambda_n(A)\}$. We have also the alternative expression[11]

$$U_{\mu_A}(z) = -\frac{1}{n} \log\det(\sqrt{(A - zI)(A - zI)^*}) = -\int_0^\infty \log(t) \, d\nu_{A-zI}(t). \tag{4.6}$$

One may retain from this determinantal Hermitization that for any $A \in \mathcal{M}_n(\mathbb{C})$,

$$\boxed{\text{knowledge of } \nu_{A-zI} \text{ for a.a. } z \in \mathbb{C} \quad \Rightarrow \quad \text{knowledge of } \mu_A}$$

Note that from (4.5), for every smooth compactly supported function $\varphi : \mathbb{C} \to \mathbb{R}$,

$$2\pi \int \varphi \, d\mu_A = \int_{\mathbb{C}} (\Delta\varphi) \log|P_A| \, dxdy.$$

The identity (4.6) bridges the eigenvalues with the singular values, and is at the heart of the following lemma, which allows to deduce the convergence of μ_A from the one of ν_{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 using powerful methods such as the method of moments or the trace-resolvent Cauchy-Stieltjes transform. The price paid here is the introduction of the auxiliary variable z. Moreover, we cannot simply deduce the convergence of the integral from the weak convergence of ν_{A-zI} since the logarithm is unbounded on $\mathbb{R}_+$. We circumvent this problem by requiring uniform integrability. We recall that on a Borel measurable space $(E, \mathcal{E})$, a Borel function $f : E \to \mathbb{R}$ is *uniformly integrable* for a sequence of probability measures $(\eta_n)_{n\geq 1}$ on E when

$$\lim_{t\to\infty} \overline{\lim_{n\to\infty}} \int_{\{|f|>t\}} |f| \, d\eta_n = 0. \tag{4.7}$$

We will use this property as follows: if $\eta_n \rightsquigarrow \eta$ as $n \to \infty$ for some probability measure η and if f is continuous and uniformly integrable for $(\eta_n)_{n\geq 1}$ then f is η-integrable and

$$\lim_{n\to\infty} \int f \, d\eta_n = \int f \, d\eta.$$

Remark 4.2 (Uniform integrability in probability)**.** *When $(\eta_n)_{n\geq 1}$ are random probability measures, we say that f is* uniformly integrable in probability *when for any $\varepsilon > 0$,*

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

This will be helpful for the heavy tailed case of section 6 together with lemma 4.4 below, in order to circumvent the lack of almost sure bounds on small singular values.

Remark 4.3 (Weak convergence in probability)**.** *Let E be a topological space such as $\mathbb{R}$ or $\mathbb{C}$ equipped with its Borel σ-field $\mathcal{E}$. Let $(\eta_n)_{n\geq 1}$ be a sequence of random probability measures on $(E, \mathcal{E})$ and η be a probability measure on $(E, \mathcal{E})$. We say that $\eta_n \rightsquigarrow \eta$ as $n \to \infty$* in probability *if for all bounded continuous function $f : E \to \mathbb{R}$, and for all $\varepsilon > 0$,*

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

This mode of convergence is implied by the a.s. weak convergence.

[11] Girko uses the name "V-transform of μ_A", where V stands for "Victory".

The idea of using Hermitization goes back at least to Girko [53]. However, the proofs of lemmas 4.4 and 4.5 below are inspired from the approach of Tao and Vu [132].

Lemma 4.4 (Girko Hermitization)**.** *Let $(A_n)_{n\geq 1}$ be a sequence of complex random matrices where A_n is $n\times n$ for every $n\geq 1$, defined on a common probability space. Suppose that for a.a. $z\in\mathbb{C}$, there exists a probability measure ν_z on $[0,\infty)$ such that*

(i) *a.s. $\nu_{A_n-zI}\rightsquigarrow\nu_z$ as $n\to\infty$*
(ii) *a.s. (respectively in probability)* log *is uniformly integrable for $(\nu_{A_n-zI})_{n\geq 1}$*

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

(j) *a.s. (respectively in probability) $\mu_{A_n}\rightsquigarrow\mu$ as $n\to\infty$*
(jj) *for a.a. $z\in\mathbb{C}$,*

$$U_\mu(z)=-\int_0^\infty \log(s)\,d\nu_z(s).$$

Proof. Let us give first the proof when (ii) holds a.s. Let z and ω be such that (i-ii) hold. For every $1\leq k\leq n$, define

$$a_{n,k}:=|\lambda_k(A_n-zI)|\quad\text{and}\quad b_{n,k}:=s_k(A_n-zI)$$

and set $\nu:=\nu_z$. Note that $\mu_{A_n-zI}=\mu_{A_n}*\delta_{-z}$. Thanks to the Weyl inequalities (1.1) and to the assumptions (i-ii), one can use lemma 4.5 below, which gives that $(\mu_{A_n})_{n\geq 1}$ is tight, that $\log|z-\cdot|$ is uniformly integrable for $(\mu_{A_n})_{n\geq 1}$, and that

$$\lim_{n\to\infty}U_{\mu_{A_n}}(z)=-\int_0^\infty\log(s)\,d\nu_z(s)=:U(z).$$

Consequently, a.s. if μ is an adherence value of $(\mu_{A_n})_{n\geq 1}$ then $\mu\in\mathcal{P}(\mathbb{C})$ and $U_\mu=U$ a.e. Now, since U does not depend on μ, by lemma 4.1, a.s. $(\mu_{A_n})_{n\geq 1}$ has a unique adherence value μ, and since $(\mu_n)_{n\geq 1}$ is tight, $\mu_{A_n}\rightsquigarrow\mu$ as $n\to\infty$ by the Prohorov theorem. Finally, by (4.5), μ is deterministic since U is deterministic, and (j-jj) hold. This achieves the proof when (ii) holds a.s. The "in probability" version follows the same lines. □

Note the remarkable swap of quantifiers on z and ω in lemma 4.4 allowed by the deterministic nature of the limit. The following lemma is in a way the skeleton of proof of lemma 4.4. It states essentially a propagation of a uniform logarithmic integrability for a couple of triangular arrays, provided that a logarithmic majorization holds between the arrays. We skip the elementary yet lengthy proof which can be found in [23, Lemma C2].

Lemma 4.5 (Logarithmic majorization and uniform integrability)**.** *Let $(a_{n,k})_{1\leq k\leq n}$ and $(b_{n,k})_{1\leq k\leq n}$ be two triangular arrays in $[0,\infty)$. Define the discrete probability measures*

$$\mu_n:=\frac{1}{n}\sum_{k=1}^n\delta_{a_{n,k}}\quad\text{and}\quad\nu_n:=\frac{1}{n}\sum_{k=1}^n\delta_{b_{n,k}}.$$

If the following properties hold

(i) *$a_{n,1}\geq\cdots\geq a_{n,n}$ and $b_{n,1}\geq\cdots\geq b_{n,n}$ for $n\gg 1$,*
(ii) *$\prod_{i=1}^k a_{n,i}\leq\prod_{i=1}^k b_{n,i}$ for every $1\leq k\leq n$ for $n\gg 1$,*
(iii) *$\prod_{i=k}^n b_{n,i}\leq\prod_{i=k}^n a_{n,i}$ for every $1\leq k\leq n$ for $n\gg 1$,*
(iv) *$\nu_n\rightsquigarrow\nu$ as $n\to\infty$ for some probability measure ν,*
(v) log *is uniformly integrable for $(\nu_n)_{n\geq 1}$,*

then

(j) log *is uniformly integrable for $(\mu_n)_{n\geq 1}$ (in particular, $(\mu_n)_{n\geq 1}$ is tight),*
(jj) *we have, as $n\to\infty$,*

$$\int_0^\infty\log(t)\,d\mu_n(t)=\int_0^\infty\log(t)\,d\nu_n(t)\to\int_0^\infty\log(t)\,d\nu(t),$$

and in particular, for every adherence value μ of $(\mu_n)_{n\geq 1}$,

$$\int_0^\infty \log(t)\, d\mu(t) = \int_0^\infty \log(t)\, d\nu(t).$$

Remark 4.6 (Logarithmic potential and Cauchy-Stieltjes transform)**.** *We may define the Cauchy-Stieltjes transform $m_\mu : \mathbb{C} \to \mathbb{C} \cup \{\infty\}$ of a probability measure μ on $\mathbb{C}$ by*

$$m_\mu(z) := \int_{\mathbb{C}} \frac{1}{\lambda - z}\, d\mu(\lambda).$$

Since $1/|\cdot|$ is Lebesgue locally integrable on $\mathbb{C}$, the Fubini-Tonelli theorem implies that $m_\mu(z)$ is finite for a.a. $z \in \mathbb{C}$, and moreover m_μ is locally Lebesgue integrable on $\mathbb{C}$ and thus belongs to $\mathcal{D}'(\mathbb{C})$. Suppose now that $\mu \in \mathcal{P}(\mathbb{C})$. The logarithmic potential is related to the Cauchy-Stieltjes transform via the identity

$$m_\mu = 2\partial U_\mu$$

in $\mathcal{D}'(\mathbb{C})$. In particular, since $4\partial\overline{\partial} = 4\overline{\partial}\partial = \Delta$ as operators on $\mathcal{D}'(\mathbb{C})$, we obtain, in $\mathcal{D}'(\mathbb{C})$,

$$2\overline{\partial} m_\mu = -\Delta U_\mu = -2\pi\mu.$$

Thus we can recover μ from m_μ. Note that for any $\varepsilon > 0$, m_μ is bounded on

$$D_\varepsilon = \{z \in \mathbb{C} : \mathrm{dist}(z, \mathrm{supp}(\mu)) > \varepsilon\}.$$

If $\mathrm{supp}(\mu)$ is one-dimensional then one may completely recover μ from the knowledge of m_μ on D_ε as $\varepsilon \to 0$. If $\mathrm{supp}(\mu)$ is not one-dimensional then one needs the knowledge of m_μ inside the support to recover μ. If $A \in \mathcal{M}_n(\mathbb{C})$ then m_{μ_A} is the trace of the resolvent

$$m_{\mu_A}(z) = \mathrm{Tr}((A - zI)^{-1})$$

for every $z \in \mathbb{C} \setminus \{\lambda_1(A), \ldots, \lambda_n(A)\}$. The lack of a Hermitization identity expressing the Cauchy-Stieltjes transform m_{μ_A} in terms of singular values explains the advantage of U_{μ_A} over m_{μ_A} for non-Hermitian matrices.

Remark 4.7 (Logarithmic potential and logarithmic energy)**.** *The term "logarithmic potential" comes from the fact that U_μ is the electrostatic potential of μ viewed as a distribution of charged particles in the plane $\mathbb{C} = \mathbb{R}^2$* [115]*. The so called logarithmic energy of this distribution of charged particles, defined by*

$$\mathcal{E}(\mu) := \int_{\mathbb{C}} U_\mu(z)\, d\mu(z) = -\int_{\mathbb{C}}\int_{\mathbb{C}} \log|z - \lambda|\, d\mu(z) d\mu(\lambda), \tag{4.8}$$

is equal, up to a sign, to the Voiculescu free entropy of μ in free probability theory [134]*. The circular law minimizes $\mu \mapsto \mathcal{E}(\mu)$ under a second moment constraint* [115]*.*

4.2. **Proof of the circular law.** The proof of theorem 2.2 is based on the Hermitization lemma 4.4. The part (i) of lemma 4.4 is obtained from corollary 4.9 below.

Theorem 4.8 (Convergence of singular values with additive perturbation)**.** *Let $(M_n)_{n\geq 1}$ be a deterministic sequence such that $M_n \in \mathcal{M}_n(\mathbb{C})$ for every n. If $\nu_{M_n} \rightsquigarrow \rho$ as $n \to \infty$ for some probability measure ρ on $\mathbb{C}$ then there exists a probability measure ν_ρ on $\mathbb{R}_+$ which depends only on ρ and such that a.s. $\nu_{n^{-1/2}X + M_n} \rightsquigarrow \nu_\rho$ as $n \to \infty$.*

Theorem 4.8 appears as a special case of the work of Dozier and Silverstein for information plus noise random matrices [38]. Their proof relies on powerful Hermitian techniques such as truncation, centralization, trace-resolvent recursion via Schur block inversion, leading to a fixed point equation for the Cauchy-Stieltjes transform of ν_ρ. It is important to stress that ν_ρ does not depend on the law of X_{11} (recall that X_{11} has unit variance). One may possibly produce an alternative proof of theorem 4.8 using free probability theory.

Corollary 4.9 (Convergence of singular values)**.** *For all $z \in \mathbb{C}$, there exists a probability measure ν_z depending only on z such that a.s. $\nu_{n^{-1/2}X - zI} \rightsquigarrow \nu_z$ as $n \to \infty$.*

For completeness, we will give in subsection 4.5 a proof of corollary 4.9. Note also that for $z = 0$, we recover the quarter circular Marchenko-Pastur theorem 2.1.

It remains to check the uniform integrability assumption (ii) of lemma 4.4. From Markov's inequality, it suffices to show that for all $z \in \mathbb{C}$, there exists $p > 0$ such that a.s.

$$\overline{\lim_{n\to\infty}} \int s^{-p}\, d\nu_{n^{-1/2}X - zI}(s) < \infty \quad \text{and} \quad \overline{\lim_{n\to\infty}} \int s^{p}\, d\nu_{n^{-1/2}X - zI}(s) < \infty. \tag{4.9}$$

The second statement in (4.9) with $p \leq 2$ follows from the strong law of large numbers (2.1) together with (1.6), which gives $s_i(n^{-1/2}X - zI) \leq s_i(n^{-1/2}X) + |z|$ for all $1 \leq i \leq n$.

The first statement in (4.9) concentrates most of the difficulty behind theorem 2.2. In the next two subsections, we will prove and comment the following couple of key lemmas taken from [132] and [128] respectively.

Lemma 4.10 (Count of small singular values)**.** *There exists $c_0 > 0$ and $0 < \gamma < 1$ such a.s. for $n \gg 1$ and $n^{1-\gamma} \leq i \leq n-1$ and all $M \in \mathcal{M}_n(\mathbb{C})$,*

$$s_{n-i}(n^{-1/2}X + M) \geq c_0 \frac{i}{n}.$$

Lemma 4.10 is more meaningful when i is close to $n^{1-\gamma}$. For $i = n-1$, it gives only a lower bound on s_1. The linearity in i corresponds to what we can expect on spacing.

Lemma 4.11 (Polynomial lower bound on least singular value)**.** *For every $a, d > 0$, there exists $b > 0$ such that if M is a deterministic complex $n \times n$ matrix with $s_1(M) \leq n^d$ then*

$$\mathbb{P}(s_n(X + M) \leq n^{-b}) \leq n^{-a}.$$

In particular there exists $b > 0$ which may depend on d such that a.s. for $n \gg 1$,

$$s_n(X + M) \geq n^{-b}.$$

For ease of notation, we write s_i in place of $s_i(n^{-1/2}X - zI)$. Applying lemmas 4.10-4.11 with $M = -zI$ and $M = -z\sqrt{n}I$ respectively, we get, for any $c > 0$, $z \in \mathbb{C}$, a.s. for $n \gg 1$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} s_i^{-p} &\leq \frac{1}{n}\sum_{i=1}^{n-\lfloor n^{1-\gamma}\rfloor} s_i^{-p} + \frac{1}{n}\sum_{i=n-\lfloor n^{1-\gamma}\rfloor+1}^{n} s_i^{-p} \\ &\leq c_0^{-p}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{i}\right)^{p} + 2n^{-\gamma}n^{bp}.\end{aligned}$$

The first term of the right hand side is a Riemann sum for $\int_0^1 s^{-p}\, ds$ which converges as soon as $0 < p < 1$. We finally obtain the first statement in (4.9) as soon as $0 < p < \min(\gamma/b, 1)$. Now the Hermitization lemma 4.4 ensures that there exists a probability measure $\mu \in \mathcal{P}(\mathbb{C})$ such that a.s. $\mu_Y \rightsquigarrow \mu$ as $n \to \infty$ and for all $z \in \mathbb{C}$,

$$U_\mu(z) = -\int_0^\infty \log(s)\, d\nu_z(s).$$

Since ν_z does not depend on the law of X_{11} (we say that it is then universal), it follows that μ also does not depend on the law of X_{11}, and therefore, by using the circular law theorem 3.5 for the Complex Ginibre Ensemble we obtain that μ is the uniform law on the unit disc. Alternatively, following Pan and Zhou [101, Lemma 3], it is possible to avoid the knowledge of the Gaussian case by computing the integral of $\int_0^\infty \log(s)\, d\nu_z(s)$ which should match the formula for the logarithmic potential of the uniform law on the unit disc.

4.3. **Count of small singular values.** This subsection is devoted to lemma 4.10 used in the proof of theorem 2.2 to check the uniform integrability assumption in lemma 4.4.

Proof of lemma 4.10. We follow the original proof of Tao and Vu [132]. Up to increasing γ, it is enough to prove the statement for all $2n^{1-\gamma} \leq i \leq n-1$ for some $\gamma \in (0,1)$ to be chosen later. To lighten the notations, we denote by $s_1 \geq \cdots \geq s_n$ the singular values of $Y := n^{-1/2}X + M$. We fix $2n^{1-\gamma} \leq i \leq n-1$ and we consider the matrix Y' formed by the first $m := n - \lceil i/2 \rceil$ rows of $\sqrt{n}Y$. Let $s'_1 \geq \cdots \geq s'_m$ be the singular values of Y'. By the Cauchy-Poincaré interlacing[12], we get

$$n^{-1/2} s'_{n-i} \leq s_{n-i}$$

Next, by lemma 4.13 we obtain

$$s_1'^{-2} + \cdots + s_{n-\lceil i/2 \rceil}'^{-2} = \mathrm{dist}_1^{-2} + \cdots + \mathrm{dist}_{n-\lceil i/2 \rceil}^{-2},$$

where $\mathrm{dist}_j := \mathrm{dist}(R_j, H_j)$ is the distance from the j^{th} row R_j of Y' to H_j, the subspace spanned by the other rows of Y'. In particular, we have

$$\frac{i}{2n} s_{n-i}^{-2} \leq \frac{i}{2n} s_{n-i}'^{-2} \leq \sum_{j=n-\lceil i \rceil}^{n-\lceil i/2 \rceil} s_j'^{-2} \leq \sum_{j=1}^{n-\lceil i/2 \rceil} \mathrm{dist}_j^{-2}. \tag{4.10}$$

Now H_j is independent of R_j and $\dim(H_j) \leq n - \frac{i}{2} \leq n - n^{1-\gamma}$, and thus, for the choice of γ given in the forthcoming lemma 4.12,

$$\sum_{n \gg 1} \mathbb{P}\left(\bigcup_{i=2n^{1-\gamma}}^{n-1} \bigcup_{j=1}^{n-\lceil i/2 \rceil} \left\{ \mathrm{dist}_j \leq \frac{\sqrt{i}}{2\sqrt{2}} \right\} \right) < \infty$$

(note that the exponential bound in lemma 4.12 kills the polynomial factor due to the union bound over i, j). Consequently, by the first Borel-Cantelli lemma, we obtain that a.s. for $n \gg 1$, all $2n^{1-\gamma} \leq i \leq n-1$, and all $1 \leq j \leq n - \lceil i/2 \rceil$,

$$\mathrm{dist}_j \geq \frac{\sqrt{i}}{2\sqrt{2}} \geq \frac{\sqrt{i}}{4}$$

Finally, (4.10) gives $s_{n-i}^2 \geq (i^2)/(32n^2)$, i.e. the desired result with $c_0 := 1/(4\sqrt{2})$. □

Lemma 4.12 (Distance of a random vector to a subspace)**.** *There exist $\gamma > 0$ and $\delta > 0$ such that for all $n \gg 1$, $1 \leq i \leq n$, any deterministic vector $v \in \mathbb{C}^n$ and any subspace H of $\mathbb{C}^n$ with $1 \leq \dim(H) \leq n - n^{1-\gamma}$, we have, denoting $R := (X_{i1}, \ldots, X_{in}) + v$,*

$$\mathbb{P}\left(\mathrm{dist}(R, H) \leq \frac{1}{2}\sqrt{n - \dim(H)} \right) \leq \exp(-n^\delta).$$

The exponential bound above is obviously not optimal, but is more than enough for our purposes: in the proof of lemma 4.10, a large enough polynomial bound on the probability suffices.

Proof. The argument is due to Tao and Vu [132, Proposition 5.1]. We first note that if H' is the vector space spanned by H, v and $\mathbb{E}R$, then $\dim(H') \leq \dim(H) + 2$ and

$$\mathrm{dist}(R, H) \geq \mathrm{dist}(R, H') = \mathrm{dist}(R', H'),$$

where $R' := R - \mathbb{E}(R)$. We may thus directly suppose without loss of generality that $v = 0$ and that $\mathbb{E}(X_{ik}) = 0$. Then, it is easy to check that

$$\mathbb{E}(\mathrm{dist}(R, H)^2) = n - \dim(H)$$

[12] If $A \in \mathcal{M}_n(\mathbb{C})$ and $1 \leq m \leq n$ and if $B \in \mathcal{M}_{m,n}(\mathbb{C})$ is obtained from A by deleting $r := n - m$ rows, then $s_i(A) \geq s_i(B) \geq s_{i+r}(A)$ for every $1 \leq i \leq m$. In particular, $[s_m(B), s_1(B)] \subset [s_n(A), s_1(A)]$, i.e. the smallest singular value increases while the largest singular value is diminished. See [76, Corollary 3.1.3]

(see computation below). The lemma is thus a statement on the deviation probability of $\mathrm{dist}(R, H)$. We first perform a truncation. Let $0 < \varepsilon < 1/3$. Markov's inequality gives

$$\mathbb{P}(|X_{ik}| \geq n^{\varepsilon}) \leq n^{-2\varepsilon}.$$

Hence, from Hoeffding's deviation inequality[13], for $n \gg 1$,

$$\mathbb{P}\left(\sum_{k=1}^{n} \mathbf{1}_{\{|X_{ik}| \leq n^{\varepsilon}\}} < n - n^{1-\varepsilon}\right) \leq \exp(-2n^{1-2\varepsilon}(1-n^{-\varepsilon})^2) \leq \exp(-n^{1-2\varepsilon}).$$

It is thus sufficient to prove that the result holds by conditioning on

$$E_m := \{|X_{i1}| \leq n^{\varepsilon}, \ldots, |X_{im}| \leq n^{\varepsilon}\} \quad \text{with} \quad m := \lceil n - n^{1-\varepsilon} \rceil.$$

Let $\mathbb{E}_m[\,\cdot\,] := \mathbb{E}[\,\cdot\,|E_m; \mathcal{F}_m]$ denote the conditional expectation given E_m and the filtration $\mathcal{F}_m$ generated by $X_{i,m+1}, \ldots, X_{i,n}$. Let W be the subspace spanned by

$$H, \quad u = (0, \ldots, 0, X_{i,m+1}, \ldots, X_{i,n}), \quad w = (\mathbb{E}_m[X_{i1}], \ldots, \mathbb{E}_m[X_{im}], 0, \ldots, 0).$$

Then, by construction $\dim(W) \leq \dim(H) + 2$ and W is $\mathcal{F}_m$-measurable. We also have

$$\mathrm{dist}(R, H) \geq \mathrm{dist}(R, W) = \mathrm{dist}(Y, W),$$

where $Y = (X_{i1} - \lambda, \ldots, X_{im} - \lambda, 0, \ldots, 0) = R - u - w$ and $\lambda = \mathbb{E}_m[X_{i1}]$. Next we have

$$\sigma^2 := \mathbb{E}_m\big[Y_1^2\big] = \mathbb{E}\Big[\big(X_{i1} - \mathbb{E}\big[X_{i1} \;\big|\; |X_{i1}| \leq n^{\varepsilon}\big]\big)^2 \;\Big|\; |X_{i1}| \leq n^{\varepsilon}\Big] = 1 - o(1).$$

Now, let us consider the disc $D := \{z \in \mathbb{C} : |z| \leq n^{\varepsilon}\}$ and define the $D^m \to \mathbb{R}_+$ convex function $f : x \mapsto \mathrm{dist}((x, 0, \ldots, 0), W)$. From the triangle inequality, f is 1-Lipschitz:

$$|f(x) - f(x')| \leq \mathrm{dist}(x, x').$$

We deduce from Talagrand's concentration inequality[14] that

$$\mathbb{P}_m(|\mathrm{dist}(Y, W) - M_m| \geq t) \leq 4 \exp\left(-\frac{t^2}{16 n^{2\varepsilon}}\right), \tag{4.11}$$

where M_m is the median of $\mathrm{dist}(Y, W)$ under $\mathbb{E}_m$. In particular,

$$M_m \geq \sqrt{\mathbb{E}_m \mathrm{dist}^2(Y, W)} - c n^{\varepsilon}.$$

Also, if P denotes the orthogonal projection on the orthogonal of W, we find

$$\begin{aligned}
\mathbb{E}_m \mathrm{dist}^2(Y, W) &= \sum_{k=1}^{m} \mathbb{E}_m\big[Y_k^2\big] P_{kk} \\
&= \sigma^2 \left(\sum_{k=1}^{n} P_{kk} - \sum_{k=m+1}^{n} P_{kk}\right) \\
&\geq \sigma^2 (n - \dim(W) - (n - m)) \\
&\geq \sigma^2 \big(n - \dim(H) - n^{1-\varepsilon} - 2\big)
\end{aligned}$$

We choose some $0 < \gamma < \varepsilon$. Then, from the above expression for any $1/2 < c < 1$ and $n \gg 1$, $M_m \geq c\sqrt{n - \dim(H)}$. We choose finally $t = (c - 1/2)\sqrt{n - \dim(H)}$ in (4.11). □

The following lemma, taken from [132, Lemma A4], is used in the proof of lemma 4.10.

[13] If $X_1, \ldots, X_n$ are independent and bounded real r.v. with $d_i := \max(X_i) - \min(X_i)$, and if $S_n := X_1 + \cdots + X_n$, then $\mathbb{P}(S_n - \mathbb{E}S_n \leq tn) \leq \exp(-2n^2t^2/(d_1^2 + \cdots + d_n^2))$ for any $t \geq 0$. See [95, Th. 5.7].

[14] If $X_1, \ldots, X_n$ are i.i.d. r.v. on $D := \{z \in \mathbb{C} : |z| \leq r\}$ and if $f : D^n \to \mathbb{R}$ is convex, 1-Lipschitz, with median M, then $\mathbb{P}(|f(X_1, \ldots, X_n) - M| \geq t) \leq 4\exp(-\frac{t^2}{16r^2})$ for any $t \geq 0$. See [124] and [88, Cor. 4.9].

Lemma 4.13 (Rows and trace norm of the inverse)**.** *Let $1 \leq m \leq n$. If $A \in \mathcal{M}_{m,n}(\mathbb{C})$ has full rank, with rows $R_1, \ldots, R_m$ and $R_{-i} := \mathrm{span}\{R_j : j \neq i\}$, then*

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

Proof. The orthogonal projection of R_i^* on the subspace R_{-i} is $B^*(BB^*)^{-1}BR_i^*$ where B is the $(m-1)\times n$ matrix obtained from A by removing the row R_i. In particular, we have

$$|R_i|_2^2 - \mathrm{dist}_2(R_i, R_{-i})^2 = \left|B^*(BB^*)^{-1}BR_i^*\right|_2^2 = (BR_i^*)^*(BB^*)^{-1}BR_i^*$$

by the Pythagoras theorem. On the other hand, the Schur block inversion formula states that if M is a $m \times m$ matrix then for every partition $\{1, \ldots, m\} = I \cup I^c$,

$$(M^{-1})_{I,I} = (M_{I,I} - M_{I,I^c}(M_{I^c,I^c})^{-1}M_{I^c,I})^{-1}. \tag{4.12}$$

Now we take $M = AA^*$ and $I = \{i\}$, and we note that $(AA^*)_{i,j} = R_iR_j^*$, which gives

$$((AA^*)^{-1})_{i,i} = (R_iR_i^* - (BR_i^*)^*(BB^*)^{-1}BR_i^*)^{-1} = \mathrm{dist}_2(R_i, R_{-i})^{-2}.$$

The desired formula follows by taking the sum over $i \in \{1, \ldots, m\}$. □

Remark 4.14 (Local Wegner estimates)**.** *Lemma 4.10 says that $\nu_{n^{-1/2}X - zI}([0,\eta]) \leq \eta/C$ for every $\eta \geq 2Cn^{-\gamma}$. In this form, we see that lemma 4.10 is an upper bound on the counting measure $n\nu_{n^{-1/2}X-zI}$ on a small interval $[0,\eta]$. This type of estimate has already been studied. Notably, an alternative proof of lemma 4.10 can be obtained following the work of* [45] *on the resolvent of Wigner matrices.*

4.4. **Smallest singular value.** This subsection is devoted to lemma 4.11 which was used in the proof of theorem 2.2 to get the uniform integrability in lemma 4.4. We start with a comment. Let $(Y_n)_{n\geq 1}$ be any sequence of random matrices where Y_n takes its values in $\mathcal{M}_n(\mathbb{C})$, defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For every $\omega \in \Omega$, the set $\cup_{n\geq 1}\{\lambda_1(Y_n(\omega)), \ldots, \lambda_n(Y_n(\omega))\}$ is at most countable and has thus zero Lebesgue measure. Therefore, for **all** $\omega \in \Omega$ and **a.a.** $z \in \mathbb{C}$, we have $s_n(Y_n(\omega) - zI) > 0$ for **all** $n \geq 1$. Lemma 4.11 implies that for $Y_n = X + M$ this holds for **all** $z \in \mathbb{C}$, **a.s.** on ω, and $n \gg 1$.

The full proof of lemma 4.11 by Tao and Vu in [128] is based on Littlewood-Offord type problems. The main difficulty is the possible presence of atoms in the law of the entries (in this case X is non-invertible with positive probability). Regarding the assumptions, the finite second moment hypothesis on X_{11} is not crucial and can be considerably weakened. For the sake of simplicity, we give here a simplified proof when the law of X_{11} has a bounded density on $\mathbb{C}$ or on $\mathbb{R}$ (which implies that $X + M$ is invertible with probability one). In lemma A.1 in Appendix A, we prove a general statement of this type at the price of a weaker probabilistic estimate which is still good enough to obtain the uniform integrability "in probability" required by lemma 4.4.

Proof of lemma 4.11 with bounded density assumption. It suffices to show the first statement since the last statement follows from the first Borel-Cantelli lemma used with $a > 1$.

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 $X + M$ and set

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

for every $1 \leq i \leq n$. The lower bound in lemma 4.15 gives

$$\min_{1\leq i\leq n} \mathrm{dist}(R_i, R_{-i}) \leq \sqrt{n}\, s_n(X + M)$$

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

$$\mathbb{P}(\sqrt{n}\, s_n(X + M) \leq u) \leq n \max_{1\leq i\leq n} \mathbb{P}(\mathrm{dist}(R_i, R_{-i}) \leq u).$$

Let us fix $1 \leq i \leq n$. Let Y_i be a unit vector orthogonal to R_{-i}. Such a vector is not unique, but we may just pick one which is independent of R_i. 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,

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

where π_i is the orthogonal projection on the orthogonal complement of R_{-i}. Let ν_i be the distribution of Y_i on $\mathbb{S}^{n-1}$. Since Y_i and R_i are independent, for any $u \geq 0$,

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

Let us assume that X_{11} has a bounded density φ 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 $|y_{j_0}|^{-1} \leq \sqrt{n}$. The complex random variable $R_i \cdot y$ is a sum of independent complex random variables and one of them is $X_{ij_0}\, \overline{y_{j_0}}$, which is absolutely continuous with a density bounded above by $\sqrt{n}\, \|\varphi\|_\infty$. 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 φ_i bounded above by $\sqrt{n}\, \|\varphi\|_\infty$, and thus

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

Therefore, for every $b > 0$, we obtain the desired result (the O does not depend on M)

$$\mathbb{P}(s_n(X + M) \leq n^{-b-1/2}) = O(n^{3/2-2b}).$$

This scheme remains indeed valid in the case where X_{11} has a bounded density on $\mathbb{R}$. □

Lemma 4.15 (Rows and operator norm of the inverse)**.** *Let A be a complex $n \times n$ matrix with rows $R_1, \ldots, R_n$. Define the vector space $R_{-i} := \mathrm{span}\{R_j : j \neq i\}$. We have then*

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

Proof of lemma 4.15. The argument, due to Rudelson and Vershynin, is buried in [113]. Since A and $A^\top$ have same singular values, one can consider the columns $C_1, \ldots, C_n$ of A instead of the rows. For every column vector $x \in \mathbb{C}^n$ and $1 \leq i \leq n$, the triangle inequality and the identity $Ax = x_1 C_1 + \cdots + x_n C_n$ give

$$\|Ax\|_2 \geq \mathrm{dist}(Ax, C_{-i}) = \min_{y \in C_{-i}} \|Ax - y\|_2 = \min_{y \in C_{-i}} \|x_i C_i - y\|_2 = |x_i| \mathrm{dist}(C_i, C_{-i}).$$

If $\|x\|_2 = 1$ then necessarily $|x_i| \geq n^{-1/2}$ for some $1 \leq i \leq n$ and therefore

$$s_n(A) = \min_{\|x\|_2 = 1} \|Ax\|_2 \geq n^{-1/2} \min_{1 \leq i \leq n} \mathrm{dist}(C_i, C_{-i}).$$

Conversely, for every $1 \leq i \leq n$, there exists a vector y with $y_i = 1$ such that

$$\mathrm{dist}(C_i, C_{-i}) = \|y_1 C_1 + \cdots + y_n C_n\|_2 = \|Ay\|_2 \geq \|y\|_2 \min_{\|x\|_2 = 1} \|Ax\|_2 \geq s_n(A)$$

where we used the fact that $\|y\|_2^2 = |y_1|^2 + \cdots + |y_n|^2 \geq |y_i|^2 = 1$. □

Remark 4.16 (Assumptions for the control of the smallest singular value)**.** *In the proof of lemma 4.11 with the bounded density assumption, we have not used the assumption on the second moment of X_{11} nor the assumption on the norm of M.*

4.5. **Convergence of singular values measure.** This subsection is devoted to corollary 4.9. The proof is divided into five steps.

Step One: Concentration of singular values measure. First, it turns out that it is sufficient to prove the convergence to ν_z of $\mathbb{E}\nu_{n^{-1/2}X-z}$. Indeed, for matrices with independent rows, there is a remarkable concentration of measure phenomenon. More precisely, recall that the total variation norm of $f : \mathbb{R} \to \mathbb{R}$ is defined as

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

where the supremum runs over all sequences $(x_k)_{k\in\mathbb{Z}}$ such that $x_{k+1} \geq x_k$ for any $k \in \mathbb{Z}$. If $f = \mathbf{1}_{(-\infty,s]}$ for some $s \in \mathbb{R}$ then $\|f\|_{\mathrm{TV}} = 1$, while if f has a derivative in $L^1(\mathbb{R})$, $\|f\|_{\mathrm{TV}} = \int |f'(t)|\, dt$. The following lemma is extracted from [26], see also [69].

Lemma 4.17 (Concentration for the singular values empirical measure)**.** *If M is a $n \times n$ complex random matrix with independent rows (or with independent columns) then for any $f : \mathbb{R} \to \mathbb{R}$ going to 0 at $\pm\infty$ with $\|f\|_{\mathrm{TV}} \leq 1$ and every $t \geq 0$,*

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

Proof. If $A, B \in \mathcal{M}_n(\mathbb{C})$ and if $F_A(\cdot) := \nu_A((-\infty, \cdot))$ and $F_B(\cdot) := \nu_B((-\infty, \cdot))$ are the cumulative distribution functions of the probability measures ν_A and ν_B then it is easily seen from the Lidskii inequality for singular values[15] that

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

Now for a smooth $f : \mathbb{R} \to \mathbb{R}$, we get, by integrating by parts,

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

Since the left hand side depends on at most $2n$ points, we get, by approximation, for every measurable function $f : \mathbb{R} \to \mathbb{R}$ with $\|f\|_{\mathrm{TV}} \leq 1$,

$$\left|\int f\, d\nu_A - \int f\, d\nu_B\right| \leq \frac{\mathrm{rank}(A-B)}{n}. \tag{4.13}$$

From now on, $f : \mathbb{R} \to \mathbb{R}$ is a fixed measurable function with $\|f\|_{\mathrm{TV}} \leq 1$. For every row vectors $x_1, \ldots, x_n$ in $\mathbb{C}^n$, we denote by $A(x_1, \ldots, x_n)$ the $n \times n$ matrix with rows $x_1, \ldots, x_n$ and we define $F : (\mathbb{C}^n)^n \to \mathbb{R}$ by

$$F(x_1, \ldots, x_n) := \int f\, d\mu_{A(x_1,\ldots,x_n)}.$$

For any $i \in \{1, \ldots, n\}$ and any row vectors $x_1, \ldots, x_n, x_i'$ of $\mathbb{C}^n$, we have

$$\mathrm{rank}(A(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - A(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)) \leq 1$$

and thus

$$|F(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - F(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq \frac{1}{n}.$$

Finally, the desired result follows from the McDiarmid-Azuma-Hoeffding concentration inequality for bounded differences[16] applied to the function F and to the random variables $R_1, \ldots, R_n$ (the rows of M). □

[15] If $A, B \in \mathcal{M}_n(\mathbb{C})$ with $\mathrm{rank}(A-B) \leq k$, then $s_{i-k}(A) \geq s_i(B) \geq s_{i+k}(A)$ for any $1 \leq i \leq n$ with the convention $s_i \equiv \infty$ if $i < 1$ and $s_i \equiv 0$ if $i > n$. This allows the extremes to blow. See [76, Th. 3.3.16].

[16] If $X_1, \ldots, X_n$ are independent r.v. in $\mathcal{X}_1, \ldots, \mathcal{X}_n$ and if $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ is a measurable function then $\mathbb{P}(|f(X_1, \ldots, X_n) - \mathbb{E}f(X_1, \ldots, X_n)| \geq t) \leq 2\exp(-2t^2/(c_1^2 + \cdots + c_n^2))$ for any $t \geq 0$, where $c_k := \sup_{x,x'\in\mathcal{D}_k} |f(x) - f(x')|$ and $\mathcal{D}_k := \{(x, x') : x_i = x_i' \text{ for all } i \neq k\}$. We refer to McDiarmid [95].

Step Two: Truncation and centralization. In the second step, we prove that it is sufficient to prove the convergence for entries with bounded support. More precisely, we define

$$Y_{ij} = X_{ij}\mathbf{1}_{\{|X_{ij}|\leq\kappa\}},$$

where $\kappa = \kappa_n$ is a sequence growing to infinity. Then if $Y = (Y_{ij})_{1\leq i,j\leq n}$, we have from Hoffman-Wielandt inequality (1.8),

$$\frac{1}{n}\sum_{k=1}^{n}\left|s_k(n^{-1/2}Y - zI) - s_k(n^{-1/2}X - zI)\right|^2 \leq \frac{1}{n^2}\sum_{1\leq i,j\leq n}|X_{ij}|^2\mathbf{1}_{\{|X_{ij}|>\kappa\}}.$$

By assumption $\mathbb{E}|X_{ij}|^2\mathbf{1}_{\{|X_{ij}|>\kappa\}}$ goes to 0 as κ goes to infinity. Hence, by the law of large numbers, the right hand side of the above inequality converges a.s. to 0. On the left hand side we recognize the square of the Wasserstein W_2 coupling distance[17] between $\nu_{n^{-1/2}Y-zI}$ and $\nu_{n^{-1/2}X-zI}$. Since the convergence in W_2 distance implies weak convergence, we deduce that it is sufficient to prove the convergence of $\mathbb{E}\nu_{n^{-1/2}Y-zI}$ to ν_z.

Next, we turn to the centralization by setting

$$Z_{ij} = Y_{ij} - \mathbb{E}Y_{ij} = Y_{ij} - \mathbb{E}X_{11}\mathbf{1}_{\{|X_{11}|\leq\kappa\}}.$$

Then if $Z = (Z_{ij})_{1\leq i,j\leq n}$, we have from the Lidskii inequality for singular values,

$$\max_{t>0}\left|\nu_{n^{-1/2}Y-zI}([0,t]) - \nu_{n^{-1/2}Z-zI}([0,t])\right| \leq \frac{\mathrm{rank}(Y-Z)}{n} \leq \frac{1}{n}.$$

In particular, it is sufficient to prove the convergence of $\mathbb{E}\nu_{n^{-1/2}Z-zI}$ to ν_z.

In summary, in the remainder of this subsection, we will allow the law of X_{11} to depend on n but we will assume that

$$\mathbb{E}X_{11} = 0\,, \quad \mathbb{P}(|X_{11}| \geq \kappa_n) = 0 \quad \text{and} \quad \mathbb{E}|X_{11}|^2 = \sigma_n^2, \tag{4.14}$$

where $\kappa = \kappa_n$ is a sequence growing to infinity and $\sigma = \sigma_n$ goes to 1 as n goes to infinity.

Step Three: Linearization. We use a popular linearization technique: we remark the identity of the Cauchy-Stieltjes transform, for $\eta \in \mathbb{C}_+$,

$$m_{\check{\nu}_{n^{-1/2}X-zI}}(\eta) = \frac{1}{2n}\mathrm{Tr}(H(z) - \eta I)^{-1}, \tag{4.15}$$

where $\check{\nu}(\cdot) = (\nu(\cdot) + \nu(-\cdot))/2$ is the symmetrized version of a measure ν, and

$$H(z) := \begin{pmatrix} 0 & n^{-1/2}X - z \\ (n^{-1/2}X - z)^* & 0 \end{pmatrix}.$$

Through a permutation of the entries, this matrix $H(z)$ is equivalent to the matrix

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

where

$$q(z,\eta) := \begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix}$$

and for every $1 \leq i,j \leq n$,

$$B_{ij} := \frac{1}{\sqrt{n}}\begin{pmatrix} 0 & X_{ij} \\ \bar{X}_{ji} & 0 \end{pmatrix}.$$

Note that $B(z) \in \mathcal{M}_n(\mathcal{M}_2(\mathbb{C})) \simeq \mathcal{M}_{2n}(\mathbb{C})$ is Hermitian and its resolvent is denoted by

$$R(q) = (B(z) - \eta I_{2n})^{-1} = (B - q(z,\eta)\otimes I_n)^{-1}.$$

[17]The W_2 distance between two probability measures η_1, η_2 on $\mathbb{R}$ is $W_2(\eta_1,\eta_2) := \min \mathbb{E}(|X_1 - X_2|^2)^{1/2}$ where the inf runs over the set of r.v. (X_1, X_2) on $\mathbb{R}\times\mathbb{R}$ with $X_1 \sim \eta_1$ and $X_2 \sim \eta_2$. In the case where $\eta_1 = \frac{1}{n}\sum_{i=1}^n \delta_{a_i}$ with $0 \leq a_i \nearrow$ and $\eta_2 = \frac{1}{n}\sum_{i=1}^n \delta_{b_i}$ with $0 \leq b_i \nearrow$ then $W_2(\eta_1,\eta_2)^2 = \frac{1}{n}\sum_{i=1}^n (a_i - b_i)^2$.

Then $R(q) \in \mathcal{M}_n(\mathcal{M}_2(\mathbb{C}))$ and, by (4.15), we deduce that

$$m_{\check{\nu}_{n^{-1/2}X-zI}}(\eta) = \frac{1}{2n}\mathrm{Tr}R(q).$$

We set

$$R(q)_{kk} = \begin{pmatrix} a_k(q) & b_k(q) \\ c_k(q) & d_k(q) \end{pmatrix} \in \mathcal{M}_2(\mathbb{C}).$$

It is easy to check that

$$a(q) := \frac{1}{n}\sum_{k=1}^{n} a_k(q) = \frac{1}{n}\sum_{k=1}^{n} d_k(q) \quad \text{and} \quad b(q) := \frac{1}{n}\sum_{k=1}^{n} b_k(q) = \frac{1}{n}\sum_{k=1}^{n} \bar{c}_k(q), \tag{4.16}$$

(see the forthcoming lemma 4.18). So finally,

$$m_{\check{\nu}_{n^{-1/2}X-zI}}(\eta) = a(q) \tag{4.17}$$

Hence, in order to prove that $\mathbb{E}\nu_{n^{-1/2}X-zI}$ converges, it is sufficient to prove that $\mathbb{E}a(q)$ converges to, say, $\alpha(q)$ which, by tightness, will necessarily be the Cauchy-Stieltjes transform of a symmetric measure.

Step Four: Approximate fixed point equation. We use a resolvent method to deduce an approximate fixed point equation satisfied by $a(q)$. Schur's block inversion (4.12) gives

$$R_{nn} = \left(\frac{1}{\sqrt{n}} \begin{pmatrix} 0 & X_{nn} \\ \bar{X}_{nn} & 0 \end{pmatrix} - q - Q^* \widetilde{R} Q \right)^{-1},$$

where $Q \in \mathcal{M}_{n-1,1}(\mathcal{M}_2(\mathbb{C}))$,

$$Q_i = \frac{1}{\sqrt{n}} \begin{pmatrix} 0 & X_{ni} \\ \bar{X}_{in} & 0 \end{pmatrix}$$

and, with $\widetilde{B} = (B_{ij})_{1\le i,j\le n-1}$, $\widetilde{B}(z) = \widetilde{B} - q(z,0) \otimes I_{n-1}$,

$$\widetilde{R} = (\widetilde{B} - q \otimes I_{n-1})^{-1} = (\widetilde{B}(z) - \eta I_{2(n-1)})^{-1}$$

is the resolvent of a minor. We denote by $\mathcal{F}_{n-1}$ the smallest σ-algebra spanned by the variables $(X_{ij})_{1\le i,j\le n-1}$. We notice that $\widetilde{R}$ is $\mathcal{F}_{n-1}$-measurable and is independent of Q. If $\mathbb{E}_n[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathcal{F}_{n-1}]$, we get, using (4.14) and (4.16)

$$\begin{aligned} \mathbb{E}_n\Big[Q^*\widetilde{R}Q\Big] &= \sum_{1\le k,\ell\le n-1} \mathbb{E}_n\Big[Q_k^*\widetilde{R}_{k\ell}Q_\ell\Big] \\ &= \frac{\sigma^2}{n}\sum_{k=1}^{n-1} \begin{pmatrix} \widetilde{a}_k & 0 \\ 0 & \widetilde{d}_k \end{pmatrix} \\ &= \frac{\sigma^2}{n}\sum_{k=1}^{n-1} \begin{pmatrix} \widetilde{a}_k & 0 \\ 0 & \widetilde{a}_k \end{pmatrix}, \end{aligned}$$

where

$$\widetilde{R}_{kk} = \begin{pmatrix} \widetilde{a}_k & \widetilde{b}_k \\ \widetilde{c}_k & \widetilde{d}_k \end{pmatrix}.$$

Recall that $\widetilde{B}(z)$ is a minor of $B(z)$. We may thus use interlacing as in (4.13) for the function $f = (\cdot - \eta)^{-1}$, and we find

$$\left| \sum_{k=1}^{n-1} \widetilde{a}_k - \sum_{k=1}^{n} a_k \right| \le 2 \int_{\mathbb{R}} \frac{1}{|x-\eta|^2} dx = O\Big(\frac{1}{\mathfrak{Im}(\eta)}\Big).$$

Hence, we have checked that

$$\mathbb{E}_n\Big[Q^*\widetilde{R}Q\Big] = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \varepsilon_1,$$

with $\|\varepsilon_1\|_2 = o(1)$ (note here that q is fixed). Moreover, if we define

$$\varepsilon_2 := \mathbb{E}_n\Big[\Big(Q^*\widetilde{R}Q - \mathbb{E}_n\Big[Q^*\widetilde{R}Q\Big]\Big)^*\Big(Q^*\widetilde{R}Q - \mathbb{E}_n\Big[Q^*\widetilde{R}Q\Big]\Big)\Big]$$

then, since $\left\|\widetilde{R}\right\|_2 \leq \mathfrak{Im}(z)^{-1}$, an elementary computation gives

$$\|\varepsilon_2\|_2 = O\left(\frac{1}{n\mathfrak{Im}(z)^2}\right) = o(1).$$

Also, we note by lemma 4.17 that $a(q)$ is close to its expectation:

$$\mathbb{E}|a(q) - \mathbb{E}a(q)|^2 = O\left(\frac{1}{n\mathfrak{Im}(z)^2}\right) = o(1).$$

Thus, the matrix

$$D = \frac{1}{\sqrt{n}}\begin{pmatrix} 0 & X_{nn} \\ \bar{X}_{nn} & 0 \end{pmatrix} - Q^*\widetilde{R}Q - \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

has a norm which converges to 0 in expectation as $n \to \infty$. Now, we use the identity

$$R_{nn} + \left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}\right)^{-1} = R_{nn}\, D \left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}\right)^{-1}.$$

Hence, since the norms of $\left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}\right)^{-1}$ and R_{nn} are at most $\mathfrak{Im}(\eta)^{-1}$, we get

$$\mathbb{E}R_{nn} = -\left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}\right)^{-1} + \varepsilon$$

with $\|\varepsilon\|_2 = o(1)$. In other words, using exchangeability,

$$\mathbb{E}\begin{pmatrix} a & b \\ \bar{b} & a \end{pmatrix} = -\left(q + \mathbb{E}\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}\right)^{-1} + \varepsilon.$$

Step Five: Unicity of the fixed point equation. From what precedes, any accumulation point of $\mathbb{E}\begin{pmatrix} a & b \\ \bar{b} & a \end{pmatrix}$ is solution of the fixed point equation

$$\begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \alpha \end{pmatrix} = -\left(q + \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}\right)^{-1}, \tag{4.18}$$

with $\alpha = \alpha(q) \in \mathbb{C}_+$. We find

$$\alpha = \frac{\alpha + \eta}{|z|^2 - (\alpha + \eta)^2}.$$

Hence, α is a root of a polynomial of degree 3. For any $z \in \mathbb{C}_+$, it is simple to check that this equation has a unique solution in $\mathbb{C}_+$ which can be explicitly computed. Alternatively, we know from (4.17) and Montel's theorem that $\eta \in \mathbb{C}_+ \mapsto \alpha(q(z,\eta)) \in \mathbb{C}_+$ is analytic. In particular, it is sufficient to check that there is a unique solution in $\mathbb{C}_+$ for $\eta = it$, with $t > 0$. To this end, we also notice from (4.17) that $\alpha(q) \in i\mathbb{R}_+$ for $q = q(z, it)$. Hence, if $h(z,t) = \mathfrak{Im}(\alpha(q))$, we find

$$h = \frac{h + t}{|z|^2 + (h + t)^2}.$$

Thus, $h \neq 0$ and

$$1 = \frac{1 + th^{-1}}{|z|^2 + (h + t)^2}.$$

The right hand side in a decreasing function in h on $(0,\infty)$ with limits equal to $+\infty$ and 0 at $h \to 0$ and $h \to \infty$. Thus, there is a unique solution of the above equation.

We have thus proved that $\mathbb{E}\begin{pmatrix} a & b \\ \bar{b} & a \end{pmatrix}$ converges. The proof of corollary 4.9 is over.

4.6. **The quaternionic resolvent: an alternative look at the circular law.**

Motivation. The aim of this subsection 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. This approach is built upon methods in the physics literature, e.g. [46, 66, 112, 111]. As we will see, it is a refinement of the linearization procedure used in the proof of corollary 4.9. Recall that the Cauchy-Stieltjes transform of a measure ν on $\mathbb{R}$ is defined, for $\eta \in \mathbb{C}_+$, as

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

The Cauchy-Stieltjes transform characterizes every probability measure on $\mathbb{R}$, and actually, following Remark 4.6, every probability measure on $\mathbb{C}$. However, if the support of the measure is not one-dimensional, then one needs the knowledge of the Cauchy-Stieltjes transform inside the support, which is not convenient. For a probability measure on $\mathbb{C}$, it is tempting to define a quaternionic Cauchy-Stieltjes transform. For $q \in \mathbb{H}_+$, where

$$\mathbb{H}_+ := \left\{ \begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix}, z \in \mathbb{C}, \eta \in \mathbb{C}_+ \right\},$$

we would define

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

This transform characterizes the measure: in $\mathcal{D}'(\mathbb{C})$,

$$\lim_{t \downarrow 0} (\partial M_\mu(q(z,it))_{12} = -\pi\mu,$$

where ∂ is as in (4.3) and

$$q(z,\eta) := \begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix}.$$

If $A \in \mathcal{M}_n(\mathbb{C})$ is normal then M_{μ_A} can be recovered from the trace of a properly defined quaternionic resolvent. If A is not normal, the situation is however more delicate and needs a more careful treatment.

Definition of quaternionic resolvent. For further needs, we will define this quaternionic resolvent in any Hilbert space. Let H be an Hilbert space with inner product $\langle \cdot, \cdot \rangle$. We define the Hilbert space $H_2 = H \times \mathbb{Z}/2\mathbb{Z}$. For $x = (y, \varepsilon) \in H_2$, we set $\hat{x} = (y, \varepsilon + 1)$. In particular, this transform is an involution $\hat{\hat{x}} = x$. There is the direct sum decomposition $H_2 = H_0 \oplus H_1$ with $H_\varepsilon = \{x = (y, \varepsilon) : y \in H\}$.

For A be an operator defined on a dense domain $D(A) \subset H$. This operator can be extended to an operator on $D(A) \otimes \mathbb{Z}/2\mathbb{Z}$ by setting $Ax = (ay, \varepsilon)$, for all $x = (y, \varepsilon) \in D(A) \otimes \mathbb{Z}/2\mathbb{Z}$ (in other word we extend A by $A \otimes I_2$). We define the operator B in $D(B) = D(A) \otimes \mathbb{Z}/2\mathbb{Z}$ by

$$Bx = \begin{cases} A^* \hat{x} & \text{if } x \in H_0 \\ A\hat{x} & \text{if } x \in H_1. \end{cases}$$

For $x \in H$, if $\Pi_x : H_2 \to \mathbb{C}^2$ denotes the orthogonal projection on $((x,0),(x,1))$, for $x, y \in D(A)$, we find

$$B_{xy} := \Pi_x B \Pi_y^* = \begin{pmatrix} 0 & \langle x, Ay \rangle \\ \langle x, A^* y \rangle & 0 \end{pmatrix} \in \mathcal{M}_2(\mathbb{C}).$$

The operator B will be called the *bipartized operator* of A, it is an Hermitian operator (i.e. for all $x, y \in D(B)$, $\langle Bx, y\rangle = \langle x, By\rangle$). If B is essentially self-adjoint (i.e. it has a unique self-adjoint extension), we may define the quaternionic resolvent of A for all $q \in \mathbb{H}_+$ as

$$R_A(q) = (B - I_H \otimes q)^{-1}$$

Indeed, if $q = q(z, \eta)$, we note that R_A is the usual resolvent at η of the essentially self-adjoint operator $B(z) = B - I_H \otimes q(z, 0)$. Hence R_A inherits the usual properties of resolvent operators (analyticity in η, bounded norm). We define

$$R_A(q)_{xy} := \Pi_x R_A(q) \Pi_y^*.$$

If H is separable and $(e_i)_{i\geq 1}$ is a canonical orthonormal basis of H, we simply write R_{ij} instead of $R_A(q)_{e_i e_j}$, $i, j \in V$. Finally, if $A \in \mathcal{M}_n(\mathbb{C})$, we set

$$\Gamma_A(q) = \frac{1}{n} \sum_{k=1}^{n} R_A(q)_{kk}.$$

If A is normal then it can be checked that $R(q)_{kk} \in \mathbb{H}_+$ and $\Gamma_A(q) = M_{\mu_A}(q)$. However, if A is not normal, this formula fails to hold. However, the next lemma explains how to recover anyway μ_A from the resolvent.

Lemma 4.18 (From quaternionic transform to spectral measures)**.** *Let $A \in \mathcal{M}_n(\mathbb{C})$ and $q = q(z, \eta) \in \mathbb{H}_+$. Then,*

$$\Gamma_A(q) = \begin{pmatrix} a(q) & b(q) \\ \bar{b}(q) & a(q) \end{pmatrix} \in \mathbb{H}_+.$$

Moreover,

$$m_{\check{\nu}_{A-z}}(\eta) = a(q)$$

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

$$\mu_A = -\frac{1}{\pi} \lim_{q(z,it):t\downarrow 0} \partial b(q).$$

Proof. For ease of notation, assume that $z = 0$ and set $\tau(\cdot) = \frac{1}{n}\mathrm{Tr}(\cdot)$. If P is the permutation matrix associated to the permutation $\sigma(2k-1) = k$, $\sigma(2k) = n + k$, we get

$$(B - I_H \otimes q)^{-1} = P^* \begin{pmatrix} -\eta & A \\ A^* & -\eta \end{pmatrix}^{-1} P = -P \begin{pmatrix} \eta(\eta^2 - AA^*)^{-1} & A(\eta^2 - A^*A)^{-1} \\ A^*(\eta^2 - AA^*)^{-1} & \eta(\eta^2 - A^*A)^{-1} \end{pmatrix} P.$$

Hence,

$$\Gamma_A(q) = -\begin{pmatrix} \eta\tau(\eta^2 - AA^*)^{-1} & \tau\big(A(\eta^2 - A^*A)^{-1}\big) \\ \tau\big(A^*(\eta^2 - AA^*)^{-1}\big) & \eta\tau(\eta^2 - A^*A)^{-1} \end{pmatrix}.$$

Notice that

$$\begin{aligned} m_{\check{\nu}_A}(\eta) &= \frac{1}{2} \int \frac{1}{x - \eta} - \frac{1}{x + \eta}\, d\nu_A(x) \\ &= \int \frac{\eta}{x^2 - \eta^2}\, d\nu_A(x) \\ &= \int \frac{\eta}{x - \eta^2}\, d\mu_{AA^*}(x) \\ &= \eta\tau(AA^* - \eta^2)^{-1}. \end{aligned}$$

Note also that $\mu_{A^*A} = \mu_{AA^*}$ implies that

$$\tau(\eta^2 - AA^*)^{-1} = \tau(\eta^2 - A^*A)^{-1}.$$

Finally, since τ is a trace,

$$\begin{aligned}\tau\big(A(\eta^2 - A^*A)^{-1}\big) &= \tau\big((\eta^2 - A^*A)^{-1}A\big)\\ &= \overline{\tau((\eta^2 - A^*A)^{-1}A)^*}\\ &= \overline{\tau(A^*(\eta^2 - AA^*)^{-1})}.\end{aligned}$$

Applying the above to $A - z$, we deduce the first two statements.

For the last statement, we write

$$\int \log|s+it|\, d\nu_{A-z}(s) = \frac{1}{2}\int \log(s^2+t^2)\, d\nu_{A-z}(s) = \frac{1}{2}\tau\log((A-z)(A-z)^* + t^2).$$

Hence, from Jacobi formula (see remark 4.6 for the definition of ∂ and $\overline{\partial}$)

$$\begin{aligned}\overline{\partial}\int \log|s+it|\, d\nu_{A-z}(s) &= \frac{1}{2}\tau\big(((A-z)(A-z)^* + t^2)^{-1}\overline{\partial}((A-z)(A-z)^* + t^2)\big)\\ &= -\frac{1}{2}\tau\big(((A-z)(A-z)^* + t^2)^{-1}(a-z)\big)\\ &= -\frac{1}{2}b(q(z,it)).\end{aligned}$$

The function $\int \log|s+it|\, d\nu_{A-z}(s)$ decreases monotonically to

$$\int \log(s)\, d\nu_{A-z}(s) = -U_{\mu_A}(z)$$

as $t \downarrow 0$. Hence, in distribution,

$$\mu_A = \lim_{t\downarrow 0} \frac{2}{\pi}\partial\overline{\partial}\int \log|s+it|\, d\nu_{A-z}(s).$$

The conclusion follows. □

Girko's Hermitization lemma revisited. There is a straightforward extension of Girko's lemma 4.4 that uses the quaternionic resolvent.

Lemma 4.19 (Girko Hermitization)**.** *Let $(A_n)_{n\geq 1}$ be a sequence of complex random matrices defined on a common probability space where A_n takes its values in $\mathcal{M}_n(\mathbb{C})$. Assume that for all $q \in \mathbb{H}_+$, there exists*

$$\Gamma(q) = \begin{pmatrix} a(q) & b(q) \\ \bar{b}(q) & a(q)\end{pmatrix} \in \mathbb{H}_+$$

such that for a.a. $z \in \mathbb{C}$, $\eta \in \mathbb{C}_+$, with $q = q(z,\eta)$,

(i') *a.s. $\Gamma_{A_n}(q)$ converges to $\Gamma(q)$ as $n \to \infty$*
(ii) *a.s. (respectively in probability) $\log$ is uniformly integrable for $(\nu_{A_n - zI})_{n\geq 1}$*

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

(j) *a.s. (respectively in probability) $\mu_{A_n} \rightsquigarrow \mu$ as $n \to \infty$*
(jj') *In $\mathcal{D}'(\mathbb{C})$,*

$$\mu = -\frac{1}{\pi}\lim_{q(z,it):t\downarrow 0} \partial b(q).$$

Note that using lemma 4.18, assumption (i') implies assumption (i) of lemma 4.4 : the limit probability measure ν_z is characterized by

$$m_{\check{\nu}_z}(\eta) = a(q).$$

The potential interest of lemma 4.19 lies in the formula for μ. It avoids any use the logarithmic potential.

Concentration. The quaternionic resolvent enjoys a simple concentration inequality, exactly as for the empirical singular values measure.

Lemma 4.20 (Concentration for the quaternionic resolvent)**.** *If A is a random matrix in $\mathcal{M}_n(\mathbb{C})$ with independent rows (or columns) then for any $q = q(z,\eta) \in \mathbb{H}_+$ and $t \geq 0$,*

$$\mathbb{P}(\|\Gamma_A(q) - \mathbb{E}\Gamma_A(q)\|_2 \geq t) \leq 2\exp\left(-\frac{n\mathfrak{Im}(\eta)^2 t^2}{8}\right).$$

Proof. Let $M, N \in \mathcal{M}_n(\mathbb{C})$ with bipartized matrices $B, C \in \mathcal{M}_{2n}(\mathbb{C})$. We have

$$\|\Gamma_M(q) - \Gamma_N(q)\|_2 \leq \frac{4\,\mathrm{rank}(M-N)}{n\mathfrak{Im}(\eta)}. \tag{4.19}$$

Indeed, from the resolvent identity, for any $q \in \mathbb{H}_+$,

$$D = R_M(q) - R_N(q) = R_M(q)(C-B)R_N(q).$$

It follows that D has rank $r \leq \mathrm{rank}(B-C) = 2\,\mathrm{rank}(M-N)$. Also, recall that the operator norm of D is at most $2\mathfrak{Im}(\eta)^{-1}$. Hence, in the singular values decomposition

$$D = \sum_{i=1}^{r} s_i u_i v_i^*$$

we have $s_i \leq 2\mathfrak{Im}(\eta)^{-1}$. If $\Pi_k : \mathbb{C}^{2n} \to \mathbb{C}^2$ is the orthogonal projection on $\mathrm{span}\{e_{2k-1}, e_{2k}\}$, then

$$\Gamma_M(q) - \Gamma_N(q) = \frac{1}{n}\sum_{k=1}^{n} \Pi_k D \Pi_k^* = \frac{1}{n}\sum_{i=1}^{r} s_i \sum_{k=1}^{n} (\Pi_k u_i)(\Pi_k v_i)^*.$$

Using Cauchy-Schwartz inequality,

$$\|\Gamma_M(q) - \Gamma_N(q)\|_2 \leq \frac{1}{n}\sum_{i=1}^{r} s_i \sqrt{\left(\sum_{k=1}^{n} \|\Pi_k u_i\|_2^2\right)\left(\sum_{k=1}^{n} \|\Pi_k v_i\|_2^2\right)} = \frac{1}{n}\sum_{i=1}^{r} s_i.$$

We obtain precisely (4.19). The remainder of the proof is now identical to the proof of lemma 4.17: we express $\Gamma_A(q) - \mathbb{E}\Gamma_A(q)$ has a sum of bounded martingales difference. □

Computation for the circular law. As pointed out in [112], the circular law is easily found from the quaternionic resolvent. Indeed, using lemma 4.20 and the proof of corollary 4.9, we get, for all $q \in \mathbb{H}_+$, a.s.

$$\lim_{n\to\infty} \Gamma_{n^{-1/2}X}(q) = \Gamma(q) = \begin{pmatrix} \alpha(q) & \beta(q) \\ \bar{\beta}(q) & \alpha(q) \end{pmatrix},$$

where, from (4.18),

$$\Gamma = -(q + \mathrm{diag}(\Gamma))^{-1}\ {}^{18}.$$

In the proof of corollary 4.9, we have checked that for $\eta = it$, $\alpha(q) = ih(z,t) \in i\mathbb{R}_+$ where

$$1 = \frac{1 + th^{-1}}{|z|^2 + (h+t)^2}.$$

We deduce easily that

$$\lim_{t\downarrow 0} h(z,t) = \begin{cases} \sqrt{1-|z|^2} & \text{if } |z| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then, from

$$\beta(q) = \frac{-z}{|z|^2 - (a(q)+\eta)^2},$$

[18] This equation is the analog of the fixed point equation satisfied by the Cauchy-Stieltjes transform m of the semi circular law: $m(\eta) = -(\eta + m(\eta))^{-1}$.

we find

$$\lim_{q(z,it):t\downarrow 0} \beta(q) = \begin{cases} -z & \text{if } |z| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

As lemma 4.19 dictates, if we compose by $-\pi^{-1}\partial$ we retrieve the circular law.

5. Related models and free probability interpretation

It is worthwhile to mention that as for the Fourier transform, the pointwise convergence of logarithmic potentials along a sequence of probability measures to the logarithmic potential of a probability measure implies the weak convergence of the sequence. We need however some tightness. This is captured by the following lemma.

Lemma 5.1 (From converging potentials to weak convergence)**.** *If $(\mu_n)_{n\geq 1}$ is a sequence in $\mathcal{P}(\mathbb{C})$ and if $\mu \in \mathcal{P}(\mathbb{C})$ are such that $U_{\mu_n}(z) \to U_\mu(z)$ as $n \to \infty$ for a.a. $z \in \mathbb{C}$ and $\overline{\lim}_{n\to\infty} \int_{\{z\in\mathbb{C}:|z|>r\}} \log|z|\, d\mu_n(z) < \infty$ for some $r \geq 1$, then $\mu_n \rightsquigarrow \mu$ as $n \to \infty$.*

Proof. The function $\log|\cdot|$ is uniformly locally integrable for the sequence $(\mu_n)_{n\geq 1}$. It follows then by dominated convergence that ΔU_{μ_n} tends to ΔU_μ as $n \to \infty$, in $\mathcal{D}'(\mathbb{C})$. □

It is worthwhile to state the following lemma, which can be seen as a variant of the Hermitization lemma 4.4. It is also a slight generalization of [132, Lemma 3.1].

Lemma 5.2 (Replacement principle)**.** *Let $(A_n)_{n\geq 1}$ and $(B_n)_{n\geq 1}$ be two sequences where A_n and B_n are random variables in $\mathcal{M}_n(\mathbb{C})$. If for a.a. $z \in \mathbb{C}$, a.s.*

(k) $\lim_{n\to\infty} U_{\mu_{A_n}}(z) - U_{\mu_{B_n}}(z) = 0$

(kk) $\log$ *is uniformly integrable for $\nu_{A_n - zI_n}$ and $\nu_{B_n - zI_n}$*

then a.s. $\mu_{A_n} - \mu_{B_n} \rightsquigarrow 0$ as $n \to \infty$.

Using the replacement principle, Tao and Vu have proved in [132] that the universality of the limit spectral measures of random matrices goes far beyond the circular law. It is stated here in a stronger form than the original version, see [22].

Theorem 5.3 (Universality principle for sums of matrices)**.** *Let X and G be the random matrices considered in sections 3 and 4 obtained from infinite tables with i.i.d. entries. Consider a deterministic sequence $(M_n)_{n\geq 1}$ such that $M_n \in \mathcal{M}_n(\mathbb{C})$ and for some $p > 0$,*

$$\overline{\lim_{n\to\infty}} \int x^p \, d\nu_{M_n}(s) < \infty.$$

Then a.s. $\mu_{n^{-1/2}X + M_n} - \mu_{n^{-1/2}G + M_n} \rightsquigarrow 0$ as $n \to \infty$.

5.1. **Related models.** We give a list of models related to the circular law theorem 2.2.

Sparsity. The circular law theorem 2.2 may remain valid if one allows the entries law to depend on n. This extension contains for instance sparse models in which the law has an atom at 0 with mass $p_n \to 1$ at a certain speed, see [63, 128, 137].

Outliers. The circular law theorem 2.2 allows the blow up of an arbitrary (asymptotically negligible) fraction of the extremal eigenvalues. Indeed, it was shown by Silverstein [119] that if $\mathbb{E}(|X_{11}|^4) < \infty$ and $\mathbb{E}(X_{11}) \neq 0$ then the spectral radius $|\lambda_1(n^{-1/2}X)|$ tends to infinity at speed $\sqrt{n}$ and has a Gaussian fluctuation. This observation of Silverstein is the base of [31], see also the ideas of Andrew [8]. More recently, Tao studied in [125] the outliers produced by various types of perturbations including general additive perturbations.

Sum and products. The scheme of proof of theorem 2.2 (based on Hermitization, logarithmic potential, and uniform integrability) turns out to be quite robust. It allows for instance to study the limit of the empirical distribution of the eigenvalues of sums and products of random matrices, see [22], and also [64] in relation with Fuss-Catalan laws. We may also mention [100]. The crucial step lies in the control of the small singular values.

Cauchy and the sphere. It is well known that the ratio of two independent standard real Gaussian variables is a Cauchy random variable, which has heavy tails. The complex analogue of this phenomenon leads to a complex Cauchy random variable, which is also the image law by the stereographical projection of the uniform law on the sphere. The matrix analogue consists in starting from two independent copies G_1 and G_2 of the Complex Ginibre Ensemble, and to consider the random matrix $Y = G_1^{-1}G_2$. The limit of μ_Y was analyzed by Forrester and Krishnapur [48]. Note that Y does not have i.i.d. entries.

Random circulant matrices. The eigenvalues of a non-Hermitian circulant matrix are linear functionals of the matrix entries. Meckes [97] used this fact together with the central limit theorem in order to show that if the entries are i.i.d. with finite positive variance then the scaled empirical spectral distribution of the eigenvalues tends to a Gaussian law. We can easily imagine a heavy tailed version of this phenomenon with α-stable limiting laws.

Single ring theorem. Let $D \in \mathcal{M}_n(\mathbb{R}_+)$ be a random diagonal matrix and $U, V \in \mathcal{M}_n(\mathbb{C})$ be two independent Haar unitary matrices, independent of D. The law of $X := UDV^*$ is unitary invariant by construction, and $\nu_X = \mu_D$ (it is a random SVD). Assume that μ_D tends to some limiting law ν as $n \to \infty$. It was conjectured by physicists that μ_X tends to a limiting law which is supported in a centered ring of the complex plane, i.e. a set of the form $\{z \in \mathbb{C} : r \leq |z| \leq R\}$. Under some additional assumptions, this was proved by Guionnet, Krishnapur, and Zeitouni [67] by using the Hermitization technique and specific aspects such as the Schwinger-Dyson non-commutative integration by parts. Guionnet and Zeitouni have also obtained the convergence of the support in a more recent work [68]. The Complex Ginibre Ensemble is a special case of this unitary invariant model.

Large deviations and logarithmic potential with external field. The circular law theorem 3.5 for the Complex Ginibre Ensemble can be seen as a special case of the circular law theorem for unitary invariant random matrices with eigenvalues density proportional to

$$(\lambda_1, \ldots, \lambda_n) \mapsto \exp\left(-\frac{1}{2n}\sum_{i=1}^{n} V(\lambda_i)\right) \prod_{i<j} |\lambda_i - \lambda_j|^2$$

where $V : \mathbb{C} \mapsto \mathbb{R}$ is a smooth potential growing enough at infinity. Since

$$\exp\left(-\frac{1}{2n}\sum_{i=1}^{n} V(\lambda_i)\right) \prod_{i<j} |\lambda_i - \lambda_j|^2 = \exp\left(-\frac{1}{2n}\sum_{i=1}^{n} V(\lambda_i) + \frac{1}{2}\sum_{i<j} \log|\lambda_i - \lambda_j|\right)$$

we discover an empirical version of the logarithmic energy functional $\mathcal{E}(\cdot)$ defined in (4.8) penalized by the "external" potential V. Indeed, it has been shown by Hiai and Petz [104] (see also Ben Arous and Zeitouni [18]) that the Complex Ginibre Ensemble satisfies a large deviation principle at speed n^2 for the weak topology on the set of symmetric probability measures (with respect to conjugacy), with good rate function given by

$$\mu \mapsto \frac{1}{2}\left(\mathcal{E}(\mu) + \int V\, d\mu\right) - \frac{3}{8} = \frac{1}{4}\iint (V(z) + V(\lambda) - 2\log|\lambda - z|)\, d\mu(z) d\mu(\lambda) - \frac{3}{8}.$$

This rate function achieves its minimum 0 at point $\mu = \mathcal{C}_1$. This is coherent with the fact that the circular law $\mathcal{C}_1$ is the minimum of the logarithmic energy among the probability measures on $\mathbb{C}$ with fixed variance, see the book of Saff and Totik [115]. Note that this

large deviation principle gives an alternative proof of the circular law for the Ginibre Ensemble thanks to the first Borel-Cantelli lemma.

Dependent entries. According to Girko, in relation to his "canonical equation K20", the circular law theorem 2.2 remains valid for random matrices with independent rows provided some natural hypotheses [55]. A circular law theorem is available for random Markov matrices including the Dirichlet Markov Ensemble [23], and random matrices with i.i.d. log-concave isotropic rows[19] [1]. Another Markovian model consists in a non-Hermitian random Markov generator with i.i.d. off-diagonal entries, which gives rise to a new limiting spectral distribution, possibly not rotationally invariant, which can be interpreted using free probability theory, see [24]. Yet another model related to projections in which each row has a zero sum is studied in [125]. To end up this tour, let us mention another kind of dependence which comes form truncations of random matrices with depend entries such as Haar unitary matrices. Namely, let U be distributed according to the uniform law on the unitary group $\mathbb{U}_n$ (we say that U is Haar unitary). Dong, Jiang, and Li have shown in [37] that the empirical spectral distribution of the diagonal sub-matrix $(U_{ij})_{1\leq i,j\leq m}$ tends to the circular law if $m/n \to 0$, while it tends to the arc law (uniform law on the unit circle $\{z \in \mathbb{C} : |z| = 1\}$) if $m/n \to 1$. Other results of the same flavor can be found in [79].

Tridiagonal matrices. The limiting spectral distributions of random tridiagonal Hermitian matrices with i.i.d. entries are not universal and depend on the law of the entries, see [105] for an approach based on the method of moments. The non-Hermitian version of this model was studied by Goldsheid and Khoruzhenko [61] by using the logarithmic potential. Indeed, the tridiagonal structure produces a three terms recursion on characteristic polynomials which can be written as a product of random 2×2 matrices, leading to the usage of a multiplicative ergodic theorem to show the convergence of the logarithmic potential (which appears as a Lyapunov exponent). In particular, neither the Hermitization nor the control the smallest and small singular values are needed here. Indeed the approach relies on a version of the converging potential lemma 5.1. Despite this apparent simplicity, the structure of the limiting distributions may be incredibly complicated and mathematically mysterious, as shown on the Bernoulli case by the physicists Holz, Orland, and Zee [73].

5.2. **Free probability interpretation.** As we shall see, the circular law and its extensions have an interpretation in free probability theory, a sub-domain of operator algebra theory. Before going further, we should recall briefly certain classical notions of operator algebra. We refer to Voiculescu, Dykema and Nica[135] for a complete treatment of free non-commutative variables, see also the book by Anderson, Guionnet, and Zeitouni for the link with random matrices [7]. In the sequel, H is an Hilbert space and we consider a pair $(\mathcal{M}, \tau)$ where $\mathcal{M}$ is an algebra of bounded operators on H, stable by the adjoint operation $*$, and where $\tau : \mathcal{M} \to \mathbb{C}$ is a linear map such that $\tau(1) = 1$, $\tau(aa^*) = \tau(a^*a) \geq 0$.

Definition of Brown measure. For $a \in \mathcal{M}$, define $|a| = \sqrt{aa^*}$. For b self-adjoint element in $\mathcal{M}$, we denote by μ_b the spectral measure of b: it is the unique probability measure on the real line satisfying, for any integer $k \in \mathbb{N}$,

$$\tau(b^k) = \int t^k d\mu_b(t).$$

Also, if $a \in \mathcal{M}$, we define

$$\nu_a = \mu_{|a|}.$$

[19] An absolutely continuous probability measure on $\mathbb{R}^n$ is log-concave if its density is e^{-V} with V convex.

Then, in the spirit of (4.6), the Brown measure [30] of $a \in \mathcal{M}$ is the unique probability measure μ_a on $\mathbb{C}$, which satisfies for almost all $z \in \mathbb{C}$,

$$\int \log|z-\lambda|\, d\mu_a(\lambda) = \int \log(s)\, d\nu_{a-z}(s).$$

In distribution, it is given by the formula[20]

$$\mu_a = \frac{1}{2\pi}\Delta \int \log(s)\, d\nu_{a-z}(s). \tag{5.1}$$

The fact that the above definition is indeed a probability measure requires a proof, which can be found in [71]. Our notation is consistent: first, if a is self-adjoint, then the Brown (spectral) measure coincides with the spectral measure. Secondly, if $\mathcal{M} = \mathcal{M}_n(\mathbb{C})$ and $\tau := \frac{1}{n}\mathrm{Tr}$ is the normalized trace on $\mathcal{M}_n(\mathbb{C})$, then we retrieve our usual definition for ν_A and μ_A. It is interesting to point out that the identity (5.1) which is a consequence of the definition of the eigenvalues when $\mathcal{M} = \mathcal{M}_n(\mathbb{C})$ serves as a definition in the general setting of von Neumann algebras.

Beyond bounded operators, and as explained in Brown [30] and in Haagerup and Schultz [71], it is possible to define, for a class $\bar{\mathcal{M}} \supset \mathcal{M}$ of closed densely defined operators affiliated with $\mathcal{M}$, a probability measure on $\mathbb{C}$ called the Brown spectral measure of $a \in \bar{\mathcal{M}}$.

Failure of the method of moments. For non-Hermitian matrices, the spectrum does not necessarily belong to the real line, and in general, the limiting spectral distribution is not supported in the real line. The problem here is that the moments are not enough to characterize laws on $\mathbb{C}$. For instance, if Z is a complex random variable following the uniform law $\mathcal{C}_\kappa$ on the centered disc $\{z \in \mathbb{C}; |z| \le \kappa\}$ of radius κ then for every $r \ge 0$, $\mathbb{E}(Z^r) = 0$ and thus $\mathcal{C}_\kappa$ is not characterized by its moments. Any rotational invariant law on $\mathbb{C}$ with light tails shares with $\mathcal{C}_\kappa$ the same sequence of null moments. One can try to circumvent the problem by using "mixed moments" which uniquely determine μ by the Weierstrass theorem. Namely, for every $A \in \mathcal{M}_n(\mathbb{C})$, if $A = UTU^*$ is the Schur unitary triangularization of A then for every integers $r, r' \ge 0$ and with $z = x + iy$ and $\tau = \frac{1}{n}\mathrm{Tr}$,

$$\int_{\mathbb{C}} z^r \overline{z}^{r'}\, d\mu_A(z) = \sum_{i=1}^n \lambda_i^r(A)\overline{\lambda_i(A)}^{r'} = \tau(T^r \overline{T}^{r'}) \neq \tau(T^r T^{*r'}) = \tau(A^r A^{*r'}).$$

Indeed equality holds true when $\overline{T} = T^*$, i.e. when T is diagonal, i.e. when A is normal. This explains why the method of moments looses its strength for non-normal operators. To circumvent the problem, one may think about using the notion of $\star$-moments. Note that if A is normal then for every word $A^{\varepsilon_1} \cdots A^{\varepsilon_k}$ where $\varepsilon_1, \ldots, \varepsilon_n \in \{1, *\}$, we have $\tau(A^{\varepsilon_1} \cdots A^{\varepsilon_k}) = \tau(A^{k_1} A^{*k_2})$ where k_1, k_2 are the number of occurrence of A and A^*.

$\star$-distribution. The $\star$-distribution of $a \in \mathcal{M}$ is the collection of all its $\star$-moments:

$$\tau(a^{\varepsilon_1} a^{\varepsilon_2} \cdots a^{\varepsilon_n}),$$

where $n \ge 1$ and $\varepsilon_1, \ldots, \varepsilon_n \in \{1, *\}$. The element $c \in \mathcal{M}$ is *circular* when it has the $\star$-distribution of $(s_1 + is_2)/\sqrt{2}$ where s_1 and s_2 are free semi circular variables with spectral measure of Lebesgue density $x \mapsto \frac{1}{\pi}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}(x)$.

The $\star$-distribution of $a \in \mathcal{M}$ allows to recover the moments of $|a-z|^2 = (a-z)(a-z)^*$ for all $z \in \mathbb{C}$, and thus ν_{a-z} for all $z \in \mathbb{C}$, and thus the Brown measure μ_a of a. Actually, for a random matrix, the $\star$-distribution contains, in addition to the spectral measure, an information on the eigenvectors of the matrix.

We say that a sequence of matrices $(A_n)_{n\ge1}$ where A takes it values in $\mathcal{M}_n(\mathbb{C})$ converges in $\star$-moments to $a \in \mathcal{M}$, if all $\star$-moments converge to the $\star$-moments of $a \in \mathcal{M}$. For

[20]The quantity $\exp \int \log(t)\, d\mu_{|a|}(t)$ is known as the Fuglede-Kadison determinant of $a \in \mathcal{M}$, see [49].

example, if $G \in \mathcal{M}_n(\mathbb{C})$ is our complex Ginibre matrix, then a.s. as $n \to \infty$, $n^{-1/2}G$ converges in $\star$-moments to a circular element.

Discontinuity of the Brown measure. Due to the unboundedness of the logarithm, the Brown measure μ_a depends discontinuously on the $\star$-moments of a [21, 120]. The limiting measures are perturbations by "balayage". A simple counter example is given by the matrices of example 1.2. For random matrices, this discontinuity is circumvented in the Girko Hermitization by requiring a uniform integrability, which turns out to be a.s. satisfied the random matrices $n^{-1/2}X$ in the circular law theorem 2.2.

However, Śniady [120, Theorem 4.1] has shown that it is always possible to regularize the Brown measure by adding an additive noise. More precisely, if G is as above and $(A_n)_{n\geq 1}$ is a sequence of matrices where A_n takes its values in $\mathcal{M}_n(\mathbb{C})$, and if the $\star$-moments of A_n converge to the $\star$-moments of $a \in \mathcal{M}$ as $n \to \infty$, then a.s. $n \to \infty$ $\mu_{A_n+tn^{-1/2}G}$ converges to μ_{a+tc}, c is circular element free of a. In particular, by choosing a sequence t_n going to 0 sufficiently slowly, it possible to regularize the Brown measure: a.s. $\mu_{A_n+t_n n^{-1/2}G}$ converges to μ_a. Note that the universality theorem 5.3 shows that the same result holds if we replace G by our matrix X. We refer to Ryan [114] and references therein for the analysis of the convergence in $\star$-moments. See also the forthcoming book of Tao [126].

6. Heavy tailed entries and new limiting spectral distributions

This section is devoted to the study of the analogues of the quarter circular and circular law theorems 2.1-2.2 when X_{11} has an infinite variance (and thus heavy tails). The approach taken from [26] involves many ingredients including the Hermitization of section 4. To lighten the notations, we often abridge $A - zI$ into $A - z$ for an operator or matrix A and a complex number z.

6.1. **Heavy tailed analogs of quarter circular and circular laws.** We now come back to an array $X := (X_{ij})_{1\leq i,j\leq n}$ of i.i.d. random variables on $\mathbb{C}$. We lift the hypothesis that the entries have a finite second moment: we will assume that,

- for some $0 < \alpha < 2$,
$$\lim_{t\to\infty} t^\alpha \mathbb{P}(|X_{11}| \geq t) = 1, \tag{6.1}$$
- as $t \to \infty$, the conditional probability
$$\mathbb{P}\left(\frac{X_{11}}{|X_{11}|} \in \cdot \;\middle|\; |X_{11}| \geq t\right)$$
converges to a probability measure on the unit circle $\mathbb{S}^1 := \{z \in \mathbb{C} : |z| = 1\}$.

The law of the entries belongs then to the domain of attraction of an α-stable law. An example is obtained when $|X_{11}|$ and $X_{11}/|X_{11}|$ are independent with $|X_{11}| = |S|$ where S is real symmetric α-stable. Another example is given by $X_{11} = \varepsilon W^{-1/\alpha}$ with ε and W independent such that ε is supported in $\mathbb{S}^1$ while W is uniform on $[0, 1]$.

The interest on this type of random matrices has started with the work of the physicists Bouchaud and Cizeau [28]. One might think that the analog of the Ginibre ensemble is a matrix with i.i.d. α-stable entries. It turns out that this random matrix ensemble is not unitary invariant and there is no explicit expression for the distribution of its eigenvalues. This lack of comparison with a canonical ensemble makes the analysis of the limit spectral measures more delicate. We may first wonder what is the analog of the quarter circular law theorem 2.1. This question has been settled by Belinschi, Dembo and Guionnet [16].

Theorem 6.1 (Singular values of heavy tailed random matrices)**.** *There exists a probability measure ν_α on $\mathbb{R}_+$ such that a.s. $\nu_{n^{-1/\alpha}X} \rightsquigarrow \nu_\alpha$ as $n \to \infty$.*

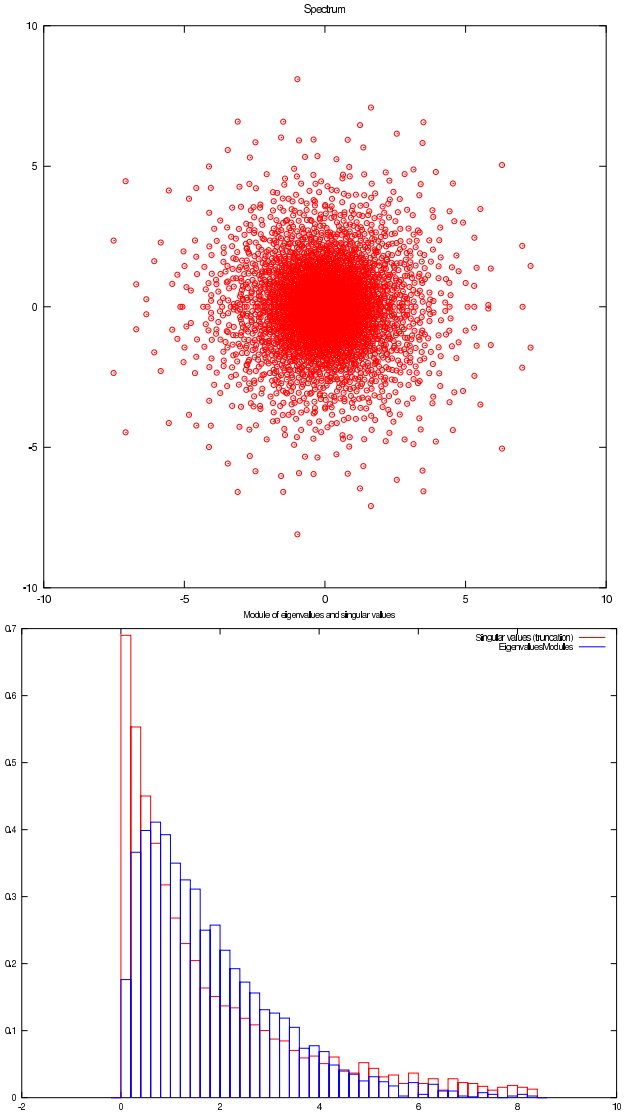

FIGURE 6. Spectrum of a single $n \times n$ matrix $n^{-1/\alpha}X$ with $n = 4000$ and i.i.d. heavy tailed entries with $X_{11} \stackrel{d}{=} \varepsilon U^{-1/\alpha}$ with $\alpha = 1$ and U uniform on $[0,1]$ and ε uniform on $\{-1,1\}$ independent of U.

This probability measure ν_α depends only on α. It does not have a known explicit closed form but has been studied in [17, 25, 16]. We know that ν_α has a bounded continuous density f_α on $\mathbb{R}_+$, which is analytic on some neighborhood of ∞. The explicit value of $f_\alpha(x)$ is only known for $x = 0$. But, more importantly, we have

$$\lim_{t\to\infty} t^{\alpha+1} f_\alpha(t) = \alpha.$$

In particular, ν_α inherits the tail behavior of the entries:

$$\lim_{t\to\infty} t^{\alpha} \nu_\alpha([t,\infty)) = 1.$$

The measure ν_α is a perturbation of the quarter circular law: it can be proved that ν_α converges weakly to the quarter circular law as α converges to 2. Contrary to the finite

variance case, the $n^{-1/\alpha}$ normalization cannot be understood from the computation of

$$\int s^2 \, d\nu_{n^{-1/\alpha}X}(s) = \frac{1}{n^{1+1/\alpha}} \sum_{i,j=1}^{n} |X_{ij}|^2$$

since the later diverges. A proof of the tightness of $\nu_{n^{-1/\alpha}X}$ requires some extra care that we will explain later on. However, at a heuristic level, we may remark that if $R_1, \ldots, R_n$ denotes the rows of $n^{-1/\alpha}X$ then for each k,

$$\|R_k\|_2^2 = \frac{1}{n^{2/\alpha}} \sum_{i=1}^{n} |X_{ki}|^2$$

converges weakly to a non-negative $\frac{\alpha}{2}$-stable random variable. Hence the $n^{-1/\alpha}$ normalization stabilizes the norm of each row of X.

Following [26], we may also investigate the behavior of the eigenvalues of X. Here is the analogue of the circular law theorem 2.2 for our heavy tailed entries matrix model.

Theorem 6.2 (Eigenvalues of heavy tailed random matrices)**.** *There exists a probability measure μ_α on $\mathbb{C}$ such that in probability $\mu_{n^{-1/\alpha}X} \rightsquigarrow \mu_\alpha$ as $n \to \infty$. Moreover, if X_{12} has a bounded density, then the convergence is almost sure.*

We believe that theorem 6.2 can be upgraded to an a.s. weak convergence, but our method does not catch this due to slow "in probability" controls on small singular values.

Again, the measure μ_α depends only on α and is not known explicitly. However, it is isotropic and has a bounded continuous density with respect to Lebesgue measure $dxdy$ on $\mathbb{C}$: $d\mu_\alpha(z) = g_\alpha(|z|)dxdy$. The value of $g_\alpha(r)$ is explicit for $r = 0$. As $r \to \infty$, the tail behavior of g_α is up to multiplicative constant equivalent to

$$r^{2(\alpha-1)} e^{-\frac{\alpha}{2} r^\alpha}.$$

This exponential decay is quite surprising and contrasts with the power tail behavior of f_α. It indicates that X is typically far from being a normal matrix. Also, we see that the eigenvalues limit spectrum is more concentrated than the singular values limit spectrum. In fact, in the finite variance case, the phenomenon is already present: the quarter circular law has support $[0, 2]$ while the circular law has support the unit disc. Again, the measure is μ_α is perturbation of the circular law: μ_α converges weakly to the circular law as α converges to 2.

The proof of theorem 6.2 will follow the general strategy of Girko's Hermitization. Lemma 4.4 gives a characterization of the limit measure in terms of its logarithmic potential. Here, it turns out to be not so convenient in order to analyze the measure μ_α. We will rather use the quaternionic version of Girko's Hermitization, i.e. lemma 4.19. For statement (i') in lemma 4.19, we will prove a generalized version of theorem 6.1.

Theorem 6.3 (Singular values of heavy tailed random matrices)**.** *For all $z \in \mathbb{C}$ there exists a probability measure $\nu_{\alpha,z}$ on $\mathbb{R}_+$ such that a.s. $\nu_{n^{-1/\alpha}X-z} \rightsquigarrow \nu_{\alpha,z}$ as $n \to \infty$. Moreover, with the notations used in lemma 4.19, for all $q = q(z, \eta) \in \mathbb{H}_+$, there exists $\Gamma(q) \in \mathbb{H}_+$, such that a.s. $\Gamma_{n^{-1/\alpha}X}(q)$ converges to $\Gamma(q)$ and $\Gamma(q)_{11} = m_{\check{\nu}_{\alpha,z}}(\eta)$.*

Objective method - sparse random graphs and trees. The strategy for proving theorem 6.3, borrowed from [26], will differ significantly from the proof of theorem 2.1. We will prove that $n^{-1/\alpha}X$ converges in some sense, as $n \to \infty$, to a limit random operator A defined in the Hilbert space $\ell^2(\mathbb{N})$. This will be done by using the "objective method" initially developed by Aldous and Steele in the context of randomized combinatorial optimization, see [6]. We build an explicit operator on Aldous' Poisson Weighted Infinite Tree (PWIT) and prove that it is the local limit of the matrices $n^{-1/\alpha}X$ 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 $n^{-1/\alpha}X$ 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 tree structure of Erdős-Rényi random graphs.

Free probability. We note finally that it is possible to associate to the PWIT a natural operator algebra $\mathcal{M}$ with a tracial state τ. Then for some operator a affiliated to $\mathcal{M}$, the probability measure μ_α is equal to the Brown measure μ_a of a, and $\nu_\alpha = \mu_{|a|} = \nu_a$ is the singular value measure of a. See the work of Aldous and Lyons [5, 93, Example 9.7 and Subsection 5], and the recent work of Male [94].

6.2. **Tightness and uniform integrability.**

Large singular values. We first prove the a.s. tightness of the sequences $(\mu_{n^{-1/\alpha}X})_{n\geq 1}$ and $(\nu_{n^{-1/\alpha}X-z})_{n\geq 1}$ with $z \in \mathbb{C}$. It is sufficient to prove that for some $p > 0$, for all $z \in \mathbb{C}$ a.s.

$$\overline{\lim_{n\to\infty}} \int s^p \, d\nu_{n^{-1/\alpha}X-z}(s) < \infty. \tag{6.2}$$

From (1.6), for any $A \in \mathcal{M}_n(\mathbb{C})$, with have $s_i(A-z) \leq s_i(A) + |z|$ and thus

$$\int s^p \, d\nu_{n^{-1/\alpha}X-z}(s) \leq \int (x+|z|)^p \, d\nu_{n^{-1/\alpha}X}(s).$$

Moreover, from 1.3 we get, for any $p > 0$,

$$\int |\lambda|^p \, d\mu_{n^{-1/\alpha}X}(\lambda) \leq \int s^p \, d\nu_{n^{-1/\alpha}X}(s).$$

In summary, it it is sufficient to prove that for some $p > 0$, a.s.

$$\overline{\lim_{n\to\infty}} \int s^p \, d\nu_{n^{-1/\alpha}X}(s) < \infty. \tag{6.3}$$

and (6.2) will follow. We shall use the following Schatten bound: for all $0 < p \leq 2$,

$$\int s^p \, d\nu_A(s) \leq \frac{1}{n} \sum_{k=1}^n \|R_k\|_2^p.$$

for every $A \in \mathcal{M}_n(\mathbb{C})$, where $R_1, \ldots, R_n$ are the rows of A (for a proof, see Zhan [138, proof of Theorem 3.32]). The above inequality is an equality if $p = 2$ (for $p > 2$, the inequality is reversed). For our matrix, $A = n^{-1/\alpha}X$, we find

$$\int |s|^p \, d\nu_{n^{-1/\alpha}X}(s) \leq \frac{1}{n} \sum_{k=1}^n \left(\frac{1}{n^{2/\alpha}} \sum_{i=1}^n |X_{ki}|^2 \right)^{\frac{p}{2}}.$$

The strategy of proof of (6.3) is now clear: the right hand side is a sum of i.i.d. variables, and from (6.1), $Y_{k,n} = n^{-2/\alpha} \sum_{i=1}^n |X_{ki}|^2$ is the domain attraction of a non-negative $\alpha/2$-stable law. We may thus expect, and it is possible to prove, that for p small enough,

$$\overline{\lim_{n\to\infty}} \, \mathbb{E} Y_{k,n}^{4p} < \infty.$$

Then, the classical proof of the strong law of large numbers for independent random variables bounded in L^4 implies (6.3).

Uniform integrability. We will prove statement (ii) of lemma 4.19 in probability. Fix $z \in \mathbb{C}$. Using (6.2), we shall prove the uniform integrability in probability of $\min(0, \log)$ for $(\nu_{n^{-1/\alpha}X-z})_{n\geq 1}$. From Markov's inequality, it is sufficient to prove that for some $c > 0$,

$$\lim_{t\to\infty} \overline{\lim_{n\to\infty}} \mathbb{P}\left(\int s^{-c}\, d\nu_{n^{-1/\alpha}X-z}(s) > t \right) = 0. \tag{6.4}$$

Arguing as in the finite variance case, the latter will in turn follow from two lemmas:

Lemma 6.4 (Lower bound on least singular value)**.** *For all $d \geq 0$, there exist $b, c \geq 0$ such that if $M \in \mathcal{M}_n(\mathbb{C})$ is deterministic with $\|M\|_2 \leq n^d$, then for $n \gg 1$,*

$$\mathbb{P}(s_n(X + M) \leq n^{-b}) \leq c\sqrt{\frac{\log n}{n}}.$$

The next lemma asserts that the i-th smallest singular of the random matrix $n^{-1/\alpha}X+M$ is at least of order $(i/n)^{2\alpha/(\alpha+2)}$ in a weak sense. This is not optimal but enough.

Lemma 6.5 (Count of small singular values)**.** *There exist $0 < \gamma < 1$ and $c_0 > 0$ such that for all $M \in \mathcal{M}_n(\mathbb{C})$, there exists an event F_n such that $\lim_{n\to\infty} \mathbb{P}(F_n) = 1$ and for all $n^{1-\gamma} \leq i \leq n-1$ and $n \gg 1$,*

$$\mathbb{E}\Big[s_{n-i}^{-2}(n^{-1/\alpha}X + M) \;\Big|\; F_n \Big] \leq c_0 \Big(\frac{n}{i}\Big)^{\frac{2}{\alpha}+1}.$$

Let us first check that these two lemmas imply (6.4) (and thus statement (ii) of lemma 4.19). Let us define the event $E_n := F_n \cap \{s_n(n^{-1/\alpha}X - z) \geq n^{-b}\}$. Let us define also

$$\mathbb{E}_n[\,\cdot\,] := \mathbb{E}[\,\cdot\,|E_n].$$

Since the event E_n has probability tending 1, the proof of (6.4) would follow from

$$\overline{\lim_{n\to\infty}} \mathbb{E}_n\left[\int x^{-p}\, d\nu_{n^{-1/\alpha}X-z}(s) \right] < \infty.$$

For simplicity, we write s_i instead $s_i(n^{-1/\alpha}X - zI)$. Since $s_n \geq n^{-b}$ has probability tending to 1, by lemma 6.5, for all $n^{1-\gamma} \leq i \leq n-1$,

$$\mathbb{E}_n\big[s_{n-i}^{-2}\big] \leq \frac{\mathbb{E}\Big[s_{n-i}^{-2} \;\Big|\; F_n\Big]}{\mathbb{P}(s_n \geq n^{-b})} \leq c_1 \Big(\frac{n}{i}\Big)^{\frac{2}{\alpha}+1}.$$

Then, for $0 < p \leq 2$, using Jensen inequality, we find

$$\begin{aligned}
\mathbb{E}_n\left[\int s^{-p}\, d\nu_{n^{-1/\alpha}X-z}(s) \right] &= \frac{1}{n} \sum_{i=0}^{\lfloor n^{1-\gamma} \rfloor} \mathbb{E}_n\Big[s_{n-i}^{-p}\Big] + \frac{1}{n} \sum_{i=\lfloor n^{1-\gamma} \rfloor + 1}^{n-1} \mathbb{E}_n\Big[s_{n-i}^{-p}\Big] \\
&\leq n^{-\gamma} n^{pb} + \frac{1}{n} \sum_{i=\lfloor n^{1-\gamma} \rfloor + 1}^{n-1} \mathbb{E}_n\big[s_{n-i}^{-2}\big]^{\frac{p}{2}} \\
&\leq n^{-\gamma+pb} + \frac{1}{n} \sum_{i=1}^{n} c_1^{-p} \Big(\frac{n}{i}\Big)^{(\frac{2}{\alpha}+1)(\frac{p}{2})}.
\end{aligned}$$

In this last expression we discover a Riemann sum. It is uniformly bounded if $p < \gamma/b$ and $p < 2\alpha/(\alpha+2)$. The uniform bound (6.4) follows.

Proof of lemma 6.4. The probability that $s_1(X) \geq n^{1+p}$ is upper bounded by the probability that one of the entries of X is larger that n^p. From Markov's inequality and the union bound, for p large enough, this event has probability at most $1/n$. In particular, $s_1(X + M) \leq s_1(X) + s_1(M)$ is at most $2n^q$ for $q = \max(p, d)$ with probability at least $1 - 1/n$. The statement is then a corollary of lemma A.1. Note: a simplified proof in the bounded density case may be obtained by adapting the proof of lemma 4.11 (see [26]). □

Sketch of proof of lemma 6.5. We now comment the proof of lemma 6.5, the detailed argument is quite technical and is omitted here. It can be found in extenso in [26]. First, as in the finite variance case, the proof reduces to derive a good lower bound on

$$\mathrm{dist}^2(X_1, W) = \langle X_1, PX_1 \rangle,$$

where X_1 is the first row of X, W is a vector space of Co-dimension $n - d \geq n^{1-\gamma}$ (in $\mathbb{R}^n$ or $\mathbb{C}^n$) and P is the orthogonal projection on the orthogonal of W. However, in the finite variance case, $\mathrm{dist}^2(X_1, W)$ concentrates sharply around its average: $n - d$. Here, the situation is quite different, for instance if $W = \mathrm{vect}(e_{n-d+1}, \ldots, e_n)$, we have

$$(n-d)^{-\frac{2}{\alpha}} \mathrm{dist}^2(X_1, W) = (n-d)^{-\frac{2}{\alpha}} \sum_{i=1}^{n-d} |X_{1i}|^2.$$

and thus $(n-d)^{-\frac{2}{\alpha}} \mathrm{dist}^2(X_1, W)$ is close in distribution to a non-negative $\alpha/2$-stable random variable, say S.

On the other hand, if U is a $n \times n$ unitary matrix uniformly distributed on the unitary group (normalized Haar measure), and if W is the span of the last d row vectors, then it can be argued than $\mathrm{dist}^2(X_1, W)$ is close in distribution to $c(n-d)n^{\frac{2}{\alpha}-1}S$. Hence, contrary to the finite variance case, the order of magnitude of the distance of X_1 to the vector space W depends on the geometry of W with respect to the coordinate basis. We have proved some lower bound on this distance which are universal on W. More precisely, for any $0 < \gamma < \alpha/4$, there exists $c_1 > 0$, such that for some event G_n with $\mathbb{P}(G_n^c) \leq c_1 n^{-(1-2\gamma)/\alpha}$,

$$\mathbb{E}\big[\mathrm{dist}^{-2}(X_1, W) \;\big|\; G_n\big] \leq c_1 (n-d)^{-\frac{2}{\alpha}}.$$

The above holds for $n - d \geq n^{1-\gamma}$. We have crucially used the fact that for all $p > 0$, $\mathbb{E}S^{-p}$ is finite, i.e. the non-negative $\alpha/2$-stable law is flat in the neighborhood of 0.

Note: the result implies that the vector space $W = \mathrm{vect}(e_{n-d+1}, \ldots, e_n)$ reaches the worst possible order of magnitude. Unfortunately, the upper bound on the probability of the event G_n^c is not good enough, and we also have to define the proper event F_n given in lemma 6.5. This event F_n satisfies $\mathbb{P}(F_n^c) \leq c \exp(-n^\delta)$ for some $\delta > 0$ and $c > 0$. □

6.3. **The objective method and the Poisson Weighted Infinite Tree (PWIT).**

Local convergence. We now describe our strategy to obtain the convergence of $\mathbb{E}\Gamma_{n^{-1/\alpha}X}$. It is an instance of the objective method : we prove that our sequence of random matrices converges locally to a limit random operator. To do this, we first notice that a $n \times n$ complex matrix M can be identified with a bounded operator in $\ell^2(\mathbb{N}) = \{(x_k)_{k\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}} : \sum_k |x_k|^2 < \infty\}$, by setting,

$$Me_i = \begin{cases} \sum_{j=1}^n M_{ji} e_j & \text{if } 1 \leq i \leq n \\ 0 & \text{otherwise.} \end{cases}$$

With an abuse of notation, without further notice, we will identify our matrices with their associated bounded operator in $\ell^2(\mathbb{N})$. The precise notion of convergence that we will use is the following.

Definition 6.6 (Local convergence)**.** *Let $\mathcal{D}(\mathbb{N})$ be the set of compactly supported elements of $\ell^2(\mathbb{N})$. Suppose (A_n) is a sequence of bounded operators on $\ell^2(\mathbb{N})$ and A is a linear operator on $\ell^2(\mathbb{N})$ with domain $D(A) = \mathcal{D}(\mathbb{N})$. For any $u, v \in \mathbb{N}$ 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 $\sigma_n : \mathbb{N} \to \mathbb{N}$ such that

- $\sigma_n(v) = u$
- *for all $\phi \in \mathcal{D}(\mathbb{N})$, $\lim_{n\to\infty} \sigma_n^{-1} A_n \sigma_n \phi = A\phi$ in $\ell^2(\mathbb{N})$.*

With a slight abuse of notation we have used the same symbol σ_n for the linear isometry $\sigma_n : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ induced in the obvious way. Note that the local convergence is the standard strong convergence of the operator $\sigma_n^{-1} A_n \sigma_n$ to A. This re-indexing of $\mathbb{N}$ preserves a distinguished element. It is a local convergence in the following way, if $P(x, y)$ is a non-commutative polynomial in $\mathbb{C}$, then the definition implies

$$\langle e_u, P(A_n, A_n^*) e_u \rangle \to \langle e_v, P(A, A^*) e_v \rangle.$$

We shall apply this definition to random operators A_n and A on $\ell^2(\mathbb{N})$: to be precise, in this case we say that $(A_n, u) \to (A, v)$ *in distribution* if there exists a random bijection σ_n as in definition 6.6 such that $\sigma_n^{-1} A_n \sigma_n \phi$ converges in distribution to $A\phi$, for all $\phi \in \mathcal{D}(\mathbb{N})$, where a random vector $\psi_n \in \ell^2(\mathbb{N})$ converges in distribution to ψ if

$$\lim_{n \to \infty} \mathbb{E} f(\psi_n) = \mathbb{E} f(\psi)$$

for all bounded continuous functions $f : \ell^2(\mathbb{N}) \to \mathbb{R}$. Finally, we may without harm replace $\mathbb{N}$ by an infinite countable set V. All definitions carry over by considering any bijection from $\mathbb{N}$ to V: namely $\ell^2(V)$, for $v \in V$, the unit vector e_v, $\mathcal{D}(V)$ and so on.

The Poisson Weighted Infinite Tree (PWIT). We now define our limit operator on an infinite rooted tree with random edge-weights, the Poisson weighted infinite tree (PWIT) introduced by Aldous [4], see also [6].

The PWIT is the random weighted rooted tree defined as follows. The vertex set of the tree is identified with $\mathbb{N}^f := \cup_{k \geq 1} \mathbb{N}^k$ by indexing the root as $\mathbb{N}^0 = \emptyset$, the offsprings of the root as $\mathbb{N}$ and, more generally, the offsprings of some $v \in \mathbb{N}^k$ as $(v1), (v2), \ldots \in \mathbb{N}^{k+1}$ (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^{th} generation. We then define T as the tree on $\mathbb{N}^f$ with (non-oriented) edges between the offsprings and their parents (see figure 7).

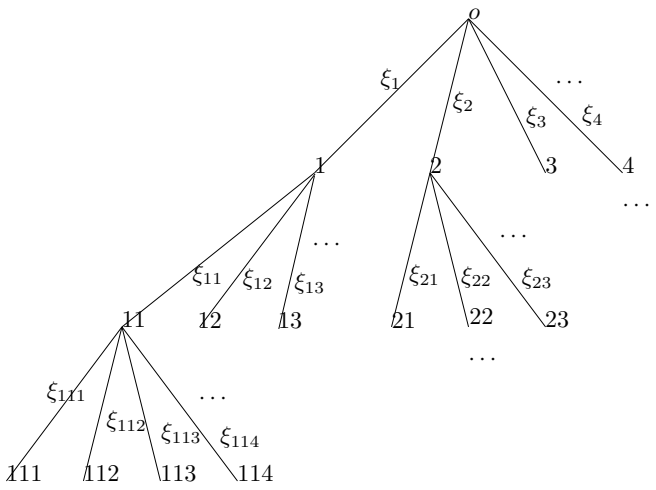

FIGURE 7. Representation of the PWIT.

We denote by Be(1/2) the Bernoulli probability distribution $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Also, recall that by assumption $\lim_{t \to \infty} \mathbb{P}(X_{11}/|X_{11}| \in \cdot \,||X_{11}| \geq t) = \theta(\cdot)$, a probability measure on the unit circle $\mathbb{S}^1$. 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 $(2\ell) \otimes \theta \otimes \text{Be}(1/2)$ on $\mathbb{R}_+ \times \mathbb{S}^1 \times \{0, 1\}$, where ℓ denotes the Lebesgue measure on $\mathbb{R}_+$. Namely, starting from the root $\emptyset$, let $\Xi_\emptyset = \{(y_1, \omega_1, \varepsilon_1), (y_2, \omega_2, \varepsilon_2), \ldots\}$ be ordered in such a way that $0 \leq y_1 \leq y_2 \leq \cdots$, and assign the mark $(y_i, \omega_i, \varepsilon_i)$ to the offspring of the root labeled i. Now, recursively, at each vertex v of generation k, assign the mark

$(y_{vi}, \omega_{vi}, \varepsilon_{vi})$ to the offspring labeled vi, where $\Xi_v = \{(y_{v1}, \omega_{v1}, \varepsilon_{v1}), (y_{v2}, \omega_{v2}, \varepsilon_{v2}), \dots\}$ satisfy $0 \leq y_{v1} \leq y_{v2} \leq \cdots$. The Bernoulli mark ε_{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.

We may define a random operator A on $\mathcal{D}(\mathbb{N}^f)$ by the formula, for all $v \in \mathbb{N}^f \backslash \{\emptyset\}$

$$Ae_v = \sum_{k \geq 1} (1 - \varepsilon_{vk}) \omega_{vk} y_{vk}^{-1/\alpha} e_{vk} + \varepsilon_v \omega_{a(v)} y_v^{-1/\alpha} e_{a(v)} \tag{6.5}$$

where $a(v)$ denotes the ancestor of v, while

$$Ae_\emptyset = \sum_{k \geq 1} (1 - \varepsilon_k) \omega_{vk} y_k^{-1/\alpha} e_k.$$

This defines a proper operator on $\mathcal{D}(\mathbb{N}^f)$. Indeed, since $\{y_{v1}, y_{v2}, \dots\}$ is an homogeneous Poisson point process of intensity 2 on $\mathbb{R}_+$: we have a.s. $\lim_{k \to \infty} y_{vk}/k = 2$. We thus find for $v \in \mathbb{N}^f \backslash \{\emptyset\}$

$$\|Ae_v\|_2^2 = \sum_{k \geq 1} (1 - \varepsilon_{vk}) y_{vk}^{-2/\alpha} + \varepsilon_{a(v)} y_v^{-2/\alpha} < \infty,$$

and similarly with $\|Ae_\emptyset\|_2$.

Theorem 6.7 (Local convergence to PWIT)**.** *In distribution* $(n^{-1/\alpha} X, 1) \to (A, \emptyset)$.

Sketch of proof. We start with some intuition behind theorem 6.7. The presence of Poisson point processes is an instance of the Poisson behavior of extreme ordered statistics. If $V_{11} \geq V_{12} \geq \cdots \geq V_{1n}$ is the ordered statistics of vector $(|X_{11}|, \dots, |X_{1n}|)$ then, it is well-known that the random variable in the space of non-increasing infinite sequences

$$n^{-1/\alpha}(V_{11}, V_{12}, \dots, V_{1n}, 0, \dots)$$

converges weakly, for the finite dimensional convergence, to

$$\left(x_1^{-1/\alpha}, x_2^{-1/\alpha}, \dots\right) \tag{6.6}$$

where $x_1 \leq x_2 \leq \dots$ are the points of an homogeneous Poisson point process of intensity 1 on $\mathbb{R}_+$. As observed by LePage, Woodroofe and Zinn [90], this fact follows easily from a beautiful representation for the order statistics of i.i.d. random variables. Namely, if $G(u) = \mathbb{P}(|X_{11}| > u)$ is (one minus) the distribution function of $|X_{11}|$, then

$$(V_{11}, \dots, V_{1n}) \stackrel{d}{=} \big(G^{-1}(x_1/x_{n+1}), \dots, G^{-1}(x_n/x_{n+1})\big),$$

where $G^{-1}(u) = \inf\{y > 0 : G(y) \leq u\}$, $u \in (0,1)$. To obtain the convergence to (6.6), it remains to notice that $G^{-1}(u) \sim u^{-1/\alpha}$ as $u \to 0$, and $x_n \sim n$ a.s. as $n \to \infty$.

More generally, we may reorder non-increasingly the vector

$$((X_{11}, X_{11}), (X_{12}, X_{21}), \dots, (X_{1n}, X_{n1})),$$

and find a permutation $\pi \in \mathcal{S}_n$ such that

$$\big\|(X_{1\pi(1)}, X_{\pi(1)1})\big\|_2 \geq \big\|(X_{1\pi(2)}, X_{\pi(2)1})\big\|_2 \geq \cdots \geq \big\|(X_{1\pi(n)}, X_{\pi(n)1})\big\|_2.$$

Then, the random variable (in the space of infinite sequences in $\mathbb{C}^2$ of non-increasing norm)

$$n^{-1/\alpha}\big((X_{1\pi(1)}, X_{\pi(1)1}), (X_{1\pi(2)}, X_{\pi(2)1}), \dots, (X_{1\pi(n)}, X_{\pi(n)1}), (0,0), \dots\big)$$

converges weakly, for the finite dimensional convergence, to

$$\Big((\varepsilon_1 w_1 y_1^{-1/\alpha}, (1 - \varepsilon_1) w_1 y_1^{-1/\alpha}), (\varepsilon_2 w_2 y_2^{-1/\alpha}, (1 - \varepsilon_2) w_2 y_2^{-1/\alpha}), \dots\Big). \tag{6.7}$$

In particular, we may define a bijection σ_n in $\mathbb{N}^f$, such that $\sigma_n(\emptyset) = 1$, $\sigma_n(k) = \pi(k)$ if $k \neq \pi^{-1}(1)$, and σ_n arbitrary otherwise. Then, for this sequence σ_n, we may check that $n^{-1/\alpha} \sigma_n^{-1} X \sigma_n e_\emptyset$ converges weakly to $Ae_\emptyset$ in $\ell^2(\mathbb{N}^f)$.

This is not good enough since we aim at the convergence for all $\phi \in D(\mathbb{N}^f)$, not only $e_\emptyset$. In particular, the above argument does not explain the presence of a tree in the limit operator. We notice however that from what precedes, only the entries such that $|X_{ij}| \geq \delta n^{1/\alpha}$ will matter for the operator convergence (for some small $\delta > 0$). By assumption,

$$\mathbb{P}(|X_{ij}| \geq \delta n^{1/\alpha}) = \frac{c}{n},$$

where $c = c(n) \sim \delta^{-1/\alpha}$. In other words, if we define G as the oriented graph on $\{1, \dots, n\}$ such that the oriented edge (i, j) is present if $|X_{ij}| \geq \delta n^{1/\alpha}$ then G is an oriented Erdős-Rényi graph (each oriented edge is present independently of the other and with equal probability). An elementary computation shows that the expected number of oriented cycles in G containing 1 and of length k is equivalent to c^k/n. This implies that there is no short cycles in G around a typical vertex. At a heuristic level, this locally tree-like structure of random graphs is thus responsible for the presence of the infinite tree T in the limit.

We are not going to give the full proof of theorem 6.7. For details, we refer to [25, 26]. The strategy is as follows. For integer m, define $J_m = \cup_{k=0}^m \{1, \cdots, m\}^k \subset \mathbb{N}^f$ and consider the matrix $A_{|m}$ obtained as the projection of the random operator A on J_m. We prove that for all integer m, there exists an injection π_m from J_m to $\{1, \dots, n\}$ such that $\pi_m(\emptyset) = 1$ and the projection of $n^{-1/\alpha}X$ on $\pi_m(J_m)$ converges weakly to $A_{|m}$. The conclusion of theorem 6.7 follows by extracting a sequence m_n going to infinity such that the latter holds.

To construct such injection π_m, we explore the entries of X: we first consider the m largest entries of the vector in $(\mathbb{C}^2)^m$, $((X_{12}, X_{21}), \dots, (X_{1n}, X_{n1}))$, whose indices are denoted by $i_1, \dots, i_m$. We then look at the m-largest entries of $((X_{i_1 j}, X_{j i_1}))_{j \neq (1, i_1, \dots, i_k)}$, whose indices are $i_{1,1}, \dots, i_{1,m}$. We repeat this procedure iteratively until we have discovered $|J_m|$ indices, and we define the injection π_m as $\pi_m(v) = i_v$. The fact that the restriction of $n^{-1/\alpha}X$ to $(i_v)_{v \in J_m}$ converges weakly to $A_{|m}$ can be proved by developing the ideas presented above. □

Continuity of quaternionic resolvent for local convergence. Note that theorem 6.7 will have a potential interest for us, only if we know how to link the local convergence of definition 6.6 to the convergence of the quaternionic resolvent introduced in subsection 4.6.

Recall that an operator B on a dense domain $D(B)$ is Hermitian if for all $x, y \in D(B)$, $\langle x, By \rangle = \langle Bx, y \rangle$. This operator will be *essentially self-adjoint* if there is a unique self-adjoint operator B_1 on $D(B_1) \supset D(B)$ such that for all $x \in D(B)$, $B_1 x = Bx$ (i.e. B_1 is an extension of B).

Lemma 6.8 (From local convergence to resolvents)**.** *Assume that (A_n) and A satisfy the conditions of definition 6.6 and $(A_n, u) \to (A, v)$ for some $u, v \in \mathbb{N}$. If the bipartized operator B of A is essentially self-adjoint, then, for all $q \in \mathbb{H}_+$,*

$$R_{A_n}(q)_{uu} \to R_A(q)_{vv}.$$

Proof. Fix $z \in \mathbb{C}$ and let $B_n(z) = B_n - q(z, 0) \otimes I$, where B_n is bipartized operator of A_n. By construction, for all $\phi \in D(B) = \mathcal{D}(\mathbb{N} \times \mathbb{Z}/2\mathbb{Z})$, $\sigma_n^{-1} B_n(z) \sigma_n \phi$ converges to $B(z)\phi$ (this is the strong operator convergence). The proof is then a direct consequence of [106, Theorem VIII.25(a)]: in this framework, the strong operator convergence implies the strong resolvent convergence. Namely, for all $\phi, \psi \in D(B)$ and $\eta \in \mathbb{C}_+$,

$$\langle \phi, (\sigma_n^{-1} B_n(z) \sigma_n - \eta I)^{-1} \psi \rangle \to \langle \phi, (B(z) - \eta I)^{-1} \psi \rangle.$$

We conclude by applying this to $\phi, \psi \in \{e_v, e_{\hat{v}}\}$. □

Remark 6.9 (A non-self-adjoint Hermitian operator)**.** *A key assumption in the above lemma is the essential self-adjointness of the bipartized limit operator. A local limit of Hermitian matrices will necessary be Hermitian. It may not however be always the case that the limit is essentially self-adjoint. Since any bounded Hermitian operator is essentially self-adjoint, for an example, we should look for an unbounded operator. Let* $(a_k)_{k\in\mathbb{N}}$ *be a sequence on* $\mathbb{R}_+$ *and define on* $D(\mathbb{N})$*, for* $k \geq 2$*,*

$$Be_k = a_k e_{k+1} + a_{k-1} e_{k-1}.$$

while $Be_1 = a_1 e_2$*. In matrix form,* B *is a tridiagonal symmetric infinite matrix. The work of Stieltjes* [123] *implies that* B *will be essentially self-adjoint if and only if*

$$\overline{\lim_{n\to\infty}} \sum_{k\geq n} a_k^{-1} = \infty.$$

6.4. **Skeleton of the main proofs.** All ingredients have finally been gathered. The skeleton of proof of theorems 6.2, 6.3 and the characterization of μ_α and $\nu_{\alpha,z}$ is as follows:

(1) By lemma 4.20, for all $q \in \mathbb{H}_+$, a.s. , in norm,
$$\Gamma_{n^{-1/\alpha}X}(q) - \mathbb{E}\Gamma_{n^{-1/\alpha}X}(q) \to 0$$
(2) Since X has exchangeable rows, for all $q \in \mathbb{H}_+$,
$$\mathbb{E}\Gamma_{n^{-1/\alpha}X}(q) = \mathbb{E}R_{n^{-1/\alpha}X}(q)_{11}$$
(3) We prove in subsection 6.5 that the bipartized operator B of the random operator A of subsection 6.3 is a.s. essentially self-adjoint
(4) It follows by theorem 6.7 and lemma 6.8,
$$\lim_{n\to\infty} \mathbb{E}\Gamma_{n^{-1/\alpha}X}(q) = \mathbb{E}R_A(q)_{\emptyset\emptyset} = \begin{pmatrix} a(q) & b(q) \\ \bar{b}(q) & a(q) \end{pmatrix}$$
(5) By lemma 4.18, a.s. $\nu_{n^{-1/\alpha}X - z} \rightsquigarrow \nu_{\alpha,z}$ as $n \to \infty$, where $\nu_{\alpha,z}$ is characterized by
$$m_{\check{\nu}_{\alpha,z}}(\eta) = a(q)$$
(6) We know from subsection 6.2 that statement (ii) of lemma 4.19 holds for $n^{-1/\alpha}X$ in probability. Then, in probability, $\mu_{n^{-1/\alpha}X} \rightsquigarrow \mu_\alpha$ as $n \to \infty$, where μ_α is characterized by, in $\mathcal{D}'(\mathbb{C})$,
$$\mu_\alpha = -\frac{1}{\pi} \lim_{q(z,it):t\downarrow 0} \partial b(q)$$
(7) We analyze in subsection 6.5 $R_A(q)_{\emptyset\emptyset}$ to obtain the properties of $\nu_{\alpha,z}$ and μ_α.
(8) Finally, when X_{12} has a bounded density we improve the convergence to almost sure (in subsection 6.6).

6.5. **Analysis of the limit operator.** This subsection is devoted to items 3 and 7 which appear above in the skeleton of proof of theorems 6.2 and 6.3.

Self-adjointness. Here we check the self-adjointness of the bipartized operator B of A.

Proposition 6.10 (Self-adjointness of bipartized operator on PWIT)**.** *Let* A *be the random operator defined by* (6.5)*. With probability one,* B *is essentially self-adjoint.*

This proposition relies on the following criterion of self-adjointness (see [26] for a proof).

Lemma 6.11 (Criterion of self-adjointness of the bipartized operator)**.** *Let* $\kappa > 0$ *and* $T = (V, E)$ *be an infinite tree on* $\mathbb{N}^f$ *and* $(w_{uv}, w_{vu})_{\{v,u\}\in E}$ *be a collection of pairs of weight in* $\mathbb{C}$ *such that for all* $u \in \mathbb{C}$*,*

$$\sum_{v:\{u,v\}\in E} |w_{uv}|^2 + |w_{vu}|^2 < \infty.$$

Define the operator on $\mathcal{D}(V)$ *as*

$$Ae_u = \sum_{v:\{u,v\}\in E} w_{vu} e_v.$$

Assume also that there exists a sequence of connected finite subsets $(S_n)_{n\geq 1}$ *in* V*, such that* $S_n \subset S_{n+1}$*,* $\cup_n S_n = V$*, and for every* n *and* $v \in S_n$*,*

$$\sum_{u\notin S_n:\{u,v\}\in E} \left(|w_{uv}|^2 + |w_{vu}|^2\right) \leq \kappa.$$

Then the bipartized operator B *of* A *is essentially self-adjoint.*

We will use a simple lemma on Poisson processes (for a proof [25, Lemma A.4]).

Lemma 6.12 (Poisson process tail)**.** *Let* $\kappa > 0$*,* $0 < \alpha < 2$ *and let* $0 < x_1 < x_2 < \cdots$ *be a Poisson process of intensity* 1 *on* $\mathbb{R}_+$*. If we define*

$$\tau := \inf\left\{ t \in \mathbb{N} : \sum_{k=t+1}^{\infty} x_k^{-2/\alpha} \leq \kappa \right\}$$

then $\mathbb{E}\tau$ *is finite and goes to* 0 *as* κ *goes to infinity.*

Proof of proposition 6.10. For $\kappa > 0$ and $v \in \mathbb{N}^f$, we define

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

The variables (τ_v) are i.i.d. and by lemma 6.12, there exists $\kappa > 0$ such that $\mathbb{E}\tau_v < 1$. We fix such κ. Now, we put a green color to all vertices v such that $\tau_v \geq 1$ 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 $1 \leq k \leq \tau_v$. Then, if the root $\emptyset$ is red, we set $S_1 = C^g(T) = \{\emptyset\}$. Otherwise, the root is green, and we consider $T^g_\emptyset = (V^g_\emptyset, E^g_\emptyset)$ the subtree of T^g that contains the root. It is a Galton-Watson tree with offspring distribution $\tau_\emptyset$. Thanks to our choice of κ, $T^g_\emptyset$ is almost surely finite. Consider $L^g_\emptyset$ the leaves of this tree (i.e. the set of vertices v in $V^g_\emptyset$ such that for all $1 \leq k \leq \tau_v$, vk is red). The following set satisfies the condition of Lemma 6.11:

$$S_1 = V^g_\emptyset \bigcup_{v\in L^g_\emptyset} \{1 \leq k \leq \tau_v : vk\}.$$

We define the outer boundary of $\{\emptyset\}$ as

$$\partial_\tau\{\emptyset\} = \{1, \ldots, \tau_\emptyset\}$$

and for $v = (i_1, \ldots, i_k) \in \mathbb{N}^f \backslash \{\emptyset\}$ we set

$$\partial_\tau\{v\} = \{(i_1, \ldots, i_{k-1}, i_k + 1)\} \cup \{(i_1, \ldots, i_k, 1), \ldots, (i_1, \ldots, i_k, \tau_v)\}.$$

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 $u_1, \ldots, u_k \in \partial_\tau S_1$, we repeat the above procedure to the rooted subtrees $T_{u_1}, \ldots, T_{u_k}$. We set

$$S_2 = S_1 \bigcup \cup_{1\leq i\leq k} C^b(T_{u_i}).$$

Iteratively, we may thus almost surely define an increasing connected sequence (S_n) of vertices with the properties required for lemma 6.11. □

Computation of resolvent. As explained in section 6.4, the properties of the measures μ_α and $\nu_{\alpha,z}$ can be deduced from the analysis of the limit resolvent operator. Resolvent are notoriously easy to compute on trees. More precisely, let $T = (V, E)$ be a tree and A, B be as in lemma 6.11 and let $\emptyset \in V$ be a distinguished vertex of V (in graph language, we root the tree T at $\emptyset$). For each $v \in V \backslash \{\emptyset\}$, we define $V_v \subset V$ as the set of vertices whose unique path to the root $\emptyset$ contains v. We define $T_v = (V_v, E_v)$ as the subtree of T spanned by V_v. We may consider A_v, the projection of A on V_v, and B_v the bipartized operator of A_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 is an operator analog of the Schur inversion by block formula (4.12).

Lemma 6.13 (Resolvent on a tree)**.** *Let A, B be as in lemma 6.11. If B is self-adjoint then for any $q = q(z, \eta) \in \mathbb{H}_+$,*

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

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

We come back to our random operator A defined on the PWIT and its quaternionic resolvent $R_A(q)$. We analyze the random variable

$$R_A(q)_{\emptyset\emptyset} := \begin{pmatrix} a(z,\eta) & b(z,\eta) \\ b'(z,\eta) & c(z,\eta) \end{pmatrix}.$$

The random variables $a(z, \eta)$ solves a nice *recursive distribution equation* (RDE). This type of recursion equation is typical of combinatorial observable defined on random rooted trees. More precisely, we define the measure on $\mathbb{R}_+$,

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

Lemma 6.14 (Recursive distribution equation)**.** *For all $q = q(z, \eta) \in \mathbb{H}_+$, if L_q is the distribution on $\mathbb{C}_+$ of $a(z, \eta)$ then L_q solves the equation in distribution:*

$$a \stackrel{d}{=} \frac{\eta + \sum_{k \geq 1} \xi_k a_k}{|z|^2 - \left(\eta + \sum_{k \geq 1} \xi_k a_k\right)\left(\eta + \sum_{k \geq 1} \xi'_k a'_k\right)}, \tag{6.8}$$

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

Moreover, with the same notation,

$$b \stackrel{d}{=} \frac{-z}{|z|^2 - \left(\eta + \sum_{k \geq 1} \xi_k a_k\right)\left(\eta + \sum_{k \geq 1} \xi'_k a'_k\right)}. \tag{6.9}$$

Proof. This is sa simple consequence of lemma 6.13. Indeed, 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 6.13, for $k \in \mathbb{N}$, $R_{B_k}(q) = (B_k(z) - \eta)^{-1}$ is the resolvent operator of B_k and set

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

Then, by lemma 6.13, we get

$$\begin{aligned} R(q)_{\emptyset\emptyset} &= -\left(q + \sum_{k\geq 1} \begin{pmatrix} 0 & \varepsilon_k w_k y_k^{-1/\alpha} \\ (1-\varepsilon_k) w_k y_k^{-1/\alpha} & 0 \end{pmatrix} \begin{pmatrix} a_k & b_k \\ b_k' & c_k \end{pmatrix} \begin{pmatrix} 0 & (1-\varepsilon_k) w_k y_k^{-1/\alpha} \\ \varepsilon_k w_k y_k^{-1/\alpha} & 0 \end{pmatrix} \right)^{-1} \\ &= -\left(U + \begin{pmatrix} \sum_k (1-\varepsilon_k) y_k^{-2/\alpha} c_k & 0 \\ 0 & \sum_k \varepsilon_k y_k^{-2/\alpha} a_k \end{pmatrix} \right)^{-1} \\ &= D^{-1} \begin{pmatrix} \eta + \sum_k \varepsilon_k y_k^{-2/\alpha} a_k & -z \\ -\bar{z} & \eta + \sum_k (1-\varepsilon_k) y_k^{-2/\alpha} c_k \end{pmatrix}, \end{aligned}$$

with

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

Now the structure of the PWIT implies that

(j) a_k and c_k have common distribution L_q
(jj) the variables $(a_k, c_k)_{k\in\mathbb{N}}$ are i.i.d.

Also the thinning property of Poisson point processes implies that

(jjj) $\{\varepsilon_k y_k^{-2/\alpha}\}_{k\in\mathbb{N}}$ and $\{(1-\varepsilon_k) y_k^{-2/\alpha}\}_{k\in\mathbb{N}}$ are independent Poisson point process with common intensity Λ_α.

□

Even if (6.8) looks complicated at first sight, for $\eta = it$, it is possible to solve it explicitly. First, 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 crucial ingredient, is a well-known and beautiful lemma. It can be derived form a representation of stable laws, see e.g. LePage, Woodroofe, and Zinn [90] and also Panchenko and Talagrand [102, Lemma 2.1].

Lemma 6.15 (Poisson-stable magic formula)**.** *Let $\{\xi_k\}_{k\in\mathbb{N}}$ be a Poisson process with intensity Λ_α. If (Y_k) is an i.i.d. sequence of non-negative random variables, independent of $\{\xi_k\}_{k\in\mathbb{N}}$, such that $\mathbb{E}[Y_1^{\frac{\alpha}{2}}] < \infty$ then*

$$\sum_{k\in\mathbb{N}} \xi_k Y_k \stackrel{d}{=} \mathbb{E}[Y_1^{\frac{\alpha}{2}}]^{\frac{2}{\alpha}} \sum_{k\in\mathbb{N}} \xi_k \stackrel{d}{=} \mathbb{E}[Y_1^{\frac{\alpha}{2}}]^{\frac{2}{\alpha}} S,$$

where S is the positive $\frac{\alpha}{2}$-stable random variable with Laplace transform for all $x \geq 0$,

$$\mathbb{E}\exp(-xS) = \exp\left(-\Gamma\Big(1-\frac{\alpha}{2}\Big) x^{\frac{\alpha}{2}}\right). \tag{6.10}$$

Proof of lemma 6.15. Recall the formulas, for $y \geq 0$, $\eta > 0$ and $0 < \eta < 1$ respectively,

$$y^{-\eta} = \Gamma(\eta)^{-1} \int_0^\infty x^{\eta-1} e^{-xy} dx \quad \text{and} \quad y^\eta = \Gamma(1-\eta)^{-1} \eta \int_0^\infty x^{-\eta-1} (1-e^{-xy}) dx. \tag{6.11}$$

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

$$\begin{aligned} \mathbb{E}\exp\left(-s\sum_k \xi_k Y_k\right) &= \exp\left(\mathbb{E}\int_0^\infty (e^{-xsY_1} - 1)\beta x^{-\frac{\alpha}{2}-1} dx\right) \\ &= \exp\left(-\Gamma\Big(1-\frac{\alpha}{2}\Big) s^{\frac{\alpha}{2}} \mathbb{E}[Y_1^{\frac{\alpha}{2}}]\right). \end{aligned}$$

□

Hence, by lemma 6.15, we may rewrite (6.8) as

$$h \stackrel{d}{=} \frac{t+yS}{|z|^2+(t+yS)(t+yS')}, \tag{6.12}$$

where S and S' are i.i.d. variables with common Laplace transform (6.10) and the function $y = y(|z|^2,t) = \mathbb{E}[h^{\alpha/2}]^{2/\alpha}$ is the unique solution of the equation in y:

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

(since the left hand side is decreasing in y, the solution is unique). In the above equations, it is also possible to consider the limits as $t \downarrow 0$.

As explained in section 6.4, this implies that, in $\mathcal{D}'(\mathbb{C})$, μ_α is equal to

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

Using (6.9), after a simple computation, we find that the density g_α of μ_α at z is

$$\frac{1}{\pi}\big(y_*^2(|z|^2)-2|z|^2y_*(|z|^2)y_*'(|z|^2)\big)\mathbb{E}\frac{SS'}{\left(|z|^2+y_*^2(|z|^2)SS'\right)^2}$$

where $y_*(r) = y(r)$ is the unique solution

$$1 = \mathbb{E}\left(\frac{S}{r+y^2SS'}\right)^{\frac{\alpha}{2}}.$$

After more computations, it is even possible to study the regularity of y_*, find the explicit solution at 0, and an asymptotic equivalent as $r \to \infty$. All these results can then be translated into properties of μ_α. We will not pursue here these computation which are done in [26]. We may simply point out that μ_α converges weakly to the circular law as $\alpha \to 2$, is a consequence of the fact that the non-negative $\alpha/2$-stable random variable $S/\Gamma(1-\alpha/2)^{2/\alpha}$ converges to a Dirac mass as $\alpha \to 2$ (see (6.10)).

6.6. **Improvement to almost sure convergence.** Let $\nu_{\alpha,z}$ be as in theorem 6.3. In order to improve the convergence to a.s., it is sufficient to prove that for all $z \in \mathbb{C}$, a.s.

$$\lim_{n\to\infty} U_{\mu_{n^{-1/\alpha}X}}(z) = L \quad \text{where} \quad L := -\int_0^\infty \log(s)\, d\nu_{\alpha,z}(s).$$

We have already proved that this convergence holds in probability. It is thus sufficient to prove that there exists a *deterministic* sequence L_n such that a.s.

$$\lim_{n\to\infty}\Big(U_{\mu_{n^{-1/\alpha}X}}(z) - L_n\Big) = 0. \tag{6.13}$$

Now, thanks to the bounded density assumption and remark 4.16, one may use lemma 4.11 for the matrix $X - n^{1/\alpha}zI$ in order to show that that there exists a number $b > 0$ such that a.s. for $n \gg 1$,

$$s_n(n^{-1/\alpha}X - zI) \geq n^{-b}.$$

Similarly, up to an increase of b if needed, we also get from (6.2) that a.s. for $n \gg 1$,

$$s_1(n^{-1/\alpha}X - zI) \leq n^b.$$

Now, we consider the function

$$f_n(x) = \mathbf{1}_{\{n^{-b}\leq|x|\leq n^b\}}\log(x).$$

From what precedes, a.s. for $n \gg 1$,

$$U_{\mu_{n^{-1/\alpha}X}}(z) = -\int_0^\infty \log(s)\, d\nu_{n^{-1/\alpha}X-zI}(s) = -\int_0^\infty f_n(s)\, d\nu_{n^{-1/\alpha}X-zI}(s). \tag{6.14}$$

The total variation of f_n is bounded by $c \log n$ for some $c > 0$. Hence by lemma 4.17, if

$$L_n := \mathbb{E} \int f_n(s) d\nu_{n^{-1/\alpha}X - zI}(s),$$

then we have,

$$\mathbb{P}\left(\left| \int f_n(s) d\nu_{n^{-1/\alpha}X - zI}(s) - L_n \right| \geq t \right) \leq 2 \exp\left(-2 \frac{nt^2}{(c \log n)^2} \right).$$

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

$$\lim_{n \to \infty} \left(\int f_n(s)\, d\nu_{n^{-1/\alpha}X - zI}(s) - L_n \right) = 0.$$

Finally, using (6.14), we deduce that (6.13) holds almost surely.

7. Open problems

We list in this section some open problems related to the circular law theorem.

Universality of Gaussian Ensembles. The universality dogma states that if a real or complex functional of X is enough symmetric and depend on enough entries then it is likely that this functional behaves asymptotically ($n \to \infty$) like in the Gaussian case (Ginibre Ensemble here) as soon as the first moments of X_{11} match certain Gaussian moments (depends on the functional). This can be understood as a sort of non-linear central limit theorem. Among interesting functionals, we find for instance the following:

- spectral radius (has Gumbel fluctuations for the Complex Ginibre Ensemble)
- argument of $\lambda_1(X)$ (is uniform on $[0, 2\pi]$ for the Complex Ginibre Ensemble)
- law of $\lambda_n(n^{-1/2}X)$ (see [47, Chapter 15] for the Complex Ginibre Ensemble). The square of the smallest singular value $s_n(n^{-1/2}G)^2$ of the Complex Ginibre Ensemble follows an exponential law [39] and this result is asymptotically universal [131]
- gap probabilities and Voronoï cells (see [3] and [60] for the Ginibre Ensemble)
- linear statistics of μ_X (some results are available such as [109, 110, 27, 108])
- empirical distribution of the real eigenvalues of $n^{-1/2}X$ when X_{11} is real (tends to uniform law on $[-1, 1]$ for the Real Ginibre Ensemble)
- unitary matrix in the polar decomposition (Haar unitary for the Complex Ginibre).
- if X_{11} has infinite fourth moment then the eigenvalues of largest modulus blow up and are asymptotically independent (Poisson statistics at some scale);
- a large deviation principle for μ_X at speed n^2 which includes as a special case the one obtained for the Complex Ginibre Ensemble by Hiai and Petz [104] (see also Ben Arous and Zeitouni [18]) and references therein. The analogous question for Hermitian models (Wigner and GUE) is also open. The answer depends on the chosen scale, the class of deviations, and the topology.

One may group most of these functionals by considering the spectrum as a point process.

It is also possible to consider universality beyond i.i.d. entries models. For instance, if X has exchangeable entries as a random vector of $\mathbb{C}^{n^2}$ and if X satisfies to suitable mean-variance normalizations, then we expect that $\mathbb{E}\mu_X$ tends to the circular law due to a Lindeberg type phenomenon, see [34] for the Hermitian case (Wigner). Similarly, if X, as a random vector of $\mathbb{C}^{n^2}$, is log-concave (see footnote 19) and isotropic (i.e. its covariance matrix is identity) then we expect that $\mathbb{E}\mu_X$ tends to the circular law, see [1] for i.i.d. log-concave rows. Since the indicator of a convex set is a log-concave measure, one may think about the Birkhoff polytope (convex envelope of permutation matrices) and ask if the circular law holds for random uniform doubly stochastic matrices, see [32] and [35].

Variance profile. We may consider the matrix Y defined as $Y_{ij} = X_{ij}\sigma(i/N, j/N)$ where $\sigma : [0,1]^2 \to [0,1]$ is a measurable function. The measure $\mu_{n^{-1/2}Y}$ should converge a.s. to a limit probability measure μ_σ on $\mathbb{C}$. For finite variance Hermitian matrices, this question has been settled by Khorunzhy, Khoruzhenko, Pastur and Shcherbina [85], for heavy tailed Hermitian matrices, by Belinschi, Dembo, Guionnet [16]. Girko has also results on the singular values of random matrices with variance profile.

Elliptic laws. We add some dependence in the array $(X_{ij})_{i,j\geq 1}$: we consider an infinite array $(X_{ij}, X_{ji})_{1\leq i<j\leq n}$ of i.i.d. pairs of complex random variables, independent of $(X_{ii})_{i\geq 1}$ an i.i.d. sequence of random variables. Assume that $\mathrm{Var}(X_{12}) = \mathrm{Var}(X_{21}) = 1$ and $\mathrm{Cor}(X_{12}, X_{21}) = t \in \{z \in \mathbb{C} : |z| \leq 1\}$. There is a conjectured universal limit for $\mu_{n^{-1/2}X}$ computed by Girko [53], called the elliptic law. This model interpolates between Hermitian and non-Hermitian random matrices. When $X = \sqrt{(1+\tau)/2}H_1 + i\sqrt{(1-\tau)/2}H_2$, with $0 \leq \tau \leq 1$ and H_1, H_2 two independent GUE, this model has been carefully analyzed by Bender in [20], see also Ledoux [89] and Johansson [80].

Oriented r-regular graphs and Kesten-McKay measure. Random oriented graphs are host of many open problems. For example, for integers $n \geq r \geq 3$, an oriented r-regular graph is a graph on n vertices such that all vertices have r incoming and r outgoing oriented edges. Consider the adjacency matrix A of a random oriented r-regular graph sampled from the uniform measure (there exists suitable simulation algorithms using matchings of half edges). It is conjectured that as $n \to \infty$, a.s. μ_A converges to the probability measure

$$\frac{1}{\pi}\frac{r^2(r-1)}{(r^2-|z|^2)^2}\mathbf{1}_{\{|z|<\sqrt{r}\}}\,dxdy.$$

It turns out that this probability measure is also the Brown measure of the free sum of r unitary, see Haagerup and Larsen [70]. The Hermitian (actually symmetric) version of this measure is known as the Kesten-McKay distribution for random non-oriented r-regular graphs, see [84, 96]. We recover the circular law when $r \to \infty$ up to renormalization.

Invertibility of random matrices. The invertibility of random matrices is one of the keys behind the circular law theorem 2.2. Let us consider the case were X_{11} is Bernoulli $\frac{1}{2}(\delta_{-1}+\delta_1)$. A famous conjecture by Spielman and Teng (related to their work on smoothed analysis of algorithms [122, 121]) states that there exists a constant $0 < c < 1$ such that

$$\mathbb{P}(\sqrt{n}\, s_n(X) \leq t) \leq t + c^n$$

for $n \gg 1$ and any small enough $t \geq 0$. This was almost solved by Rudelson and Vershynin [113] and Tao and Vu [131]. In particular, taking $t = 0$ gives $\mathbb{P}(s_n(X) = 0) = c^n$. This positive probability of being singular does not contradict the asymptotic invertibility since by the first Borel-Cantelli lemma, a.s. $s_n(X) > 0$ for $n \gg 1$. Regarding the constant c above, it has been conjectured years ago that

$$\mathbb{P}(s_n(X) = 0) = \left(\frac{1}{2} + o(1)\right)^n.$$

This intuition comes from the probability of equality of two rows, which implies that $\mathbb{P}(s_n(X) = 0) \geq (1/2)^n$. Many authors contributed to the analysis of this difficult nonlinear discrete problem, starting from Komlós, Kahn, and Szemerédi. The best result to date is due to Bourgain, Vu, and Wood [29] who proved that $\mathbb{P}(s_n(X) = 0) \leq \left(1/\sqrt{2} + o(1)\right)^n$.

Roots of random polynomials. The random matrix X has i.i.d. entries and its eigenvalues are the roots of its characteristic polynomial. The coefficients of this random polynomial are neither independent nor identically distributed. Beyond random matrices, let us consider a random polynomial $P(z) = a_0 + a_1 z + \cdots + a_n z^n$ where $a_0, \ldots, a_n$ are independent random variables. By analogy with random matrices, one may ask about the behavior as

$n \to \infty$ of the roots $\lambda_1(P), \dots, \lambda_n(P)$ of P in $\mathbb{C}$ and in particular the behavior of their empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(P)}$. The literature on this subject is quite rich and takes its roots in the works of Littlewood and Offord, Rice, and Kac. We refer to Shub and Smale [118], Azaïs and Wschebor [9], and Edelman and Kostlan [42, 43] for (partial) reviews. As for random matrices, the case where the coefficients are real is more subtle due to the presence of real roots. Regarding the complex case, the zeros of Gaussian analytic functions is the subject of a recent monograph [77] in connection with determinantal processes. Various cases are considered in the literature, including the following three families:

- Kac polynomials, for which $(a_i)_{0\leq i\leq n}$ are i.i.d.
- Binomial polynomials, for which $a_i = \sqrt{\binom{n}{i}} b_i$ for all i and $(b_i)_{0\leq i\leq n}$ are i.i.d.
- Weyl polynomials, for which $a_i = \frac{1}{\sqrt{i!}} b_i$ for all i and $(b_i)_{0\leq i\leq n}$ are i.i.d.

Geometrically, the complex number z is a root of P if and only if the vectors $(1, z, \dots, z^n)$ and $(\overline{a_0}, \overline{a_1}, \dots, \overline{a_n})$ are orthogonal in $\mathbb{C}^{n+1}$, and this connects the problem to Littlewood-Offord type problems [91] and small balls probabilities. Regarding Kac polynomials, Kac [82, 81] has shown in the real Gaussian case that the asymptotic number of real roots is about $\frac{2}{\pi}\log(n)$ as $n \to \infty$. Kac obtained the same result when the coefficients are uniformly distributed [83]. Hammersley [72] derived an explicit formula for the k-point correlation of the roots of Kac polynomials. Shparo and Shur [117] have shown that the empirical measure of the roots of Kac polynomials with light tailed coefficients tends as $n \to \infty$ to the uniform law one the unit circle $\{z \in \mathbb{C} : |z| = 1\}$ (the arc law). If the coefficients are heavy tailed then the limiting law concentrates on the union of two centered circles, see [65] and references therein. Regarding Weyl polynomials, various simulations and conjectures have been made [50, 44]. For instance, if $(b_i)_{0\leq i\leq n}$ are i.i.d. standard Gaussian, it was conjectured that the asymptotic behavior of the roots of the Weyl polynomials is analogous to the Ginibre Ensemble. Namely, the empirical distribution of the roots tends as $n \to \infty$ to the uniform law on the centered disc of the complex plane (circular law), and moreover, in the real Gaussian case, there are about $\frac{2}{\pi}\sqrt{n}$ real roots as $n \to \infty$ and their empirical distribution tends as $n \to \infty$ to a uniform law on an interval, as for the real Ginibre Ensemble, see Remark 3.8. The complex Gaussian case was considered by Leboeuf [87] and by Peres and Virág [103], while the real roots of the real Gaussian case were studied by Schehr and Majumdar [116]. Beyond the Gaussian case, one may try to use the companion matrix[21] of P and the logarithmic potential approach. Numerical simulations reveal strange phenomena depending on the law of the coefficients but we ignore it they are purely numerical. Note that if the coefficients are all real positive then the roots cannot be real positive. The heavy tailed case is also of interest (rings?).

Appendix A. Invertibility of random matrices with independent entries

This appendix is devoted to the proof of a general statement (lemma A.1 below) on the smallest singular value of random matrix models with independent entries. It follows form lemma A.1 below that if $X = (X_{ij})_{1\leq i,j\leq n}$ is a random matrix with i.i.d. entries such as X_{11} is not constant and $\mathbb{E}(|X_{11}|^\kappa) < \infty$ for some arbitrarily small real number $\kappa > 0$, then for any $\gamma > 0$ there exists are real number $\beta > 0$ such that for any $n \gg 1$ and any deterministic matrix $M \in \mathcal{M}_n(\mathbb{C})$ with $s_1(M) \leq n^\gamma$,

$$\lim_{n\to\infty} \mathbb{P}(s_n(X + M) \leq n^{-\beta}) = 0.$$

Both the assumptions and the conclusion are strictly weaker than the result of Tao and Vu. It is enough for the proof of the circular law in probability and its heavy tailed analogue.

[21] The companion matrix M of $Q(z) := c_0 + c_1 z + \cdots + c_{n-1}z^{n-1} + z^n$ is the $n \times n$ matrix with null entries except $M_{i,i+1} = 1$ and $M_{n,i} = c_{i-1}$ for every licit i. The characteristic polynomial of M is Q.

Lemma A.1 (Smallest singular value of matrices with independent entries)**.** *Let $(X_{ij})_{1\leq i,j\leq n}$ be a random matrix with independent entries in $\mathbb{C}$ (without any integrability assumptions) such that for some $a,b>0$ and all $1\leq i,j\leq n$,*

$$\mathbb{P}(|X_{ij}|\leq a)>b.$$

Then, there exists $c=c(a,b)\geq a$ such that for any $M\in\mathcal{M}_n(\mathbb{C})$, $s\gg 1$, $t\leq 1$,

$$\mathbb{P}\bigg(s_n(X+M)\leq\frac{t}{\sqrt{n}}\ ;\ s_1(X+M)\leq s\sqrt{n}\bigg)\leq\frac{c\sqrt{\log(cs/\sigma)}}{\sigma}\bigg(\frac{ts^2}{\sigma^2}+\frac{1}{\sqrt{n}}\bigg),$$

where $1/0=\infty$ and

$$\sigma^2:=\min_{1\leq i,j\leq n}\mathrm{Var}\Big(X_{ij}\mathbf{1}_{\{|X_{ij}|\leq a\}}\Big).$$

The proof of lemma A.1 follows mainly from [92, 113]. These works have already been used in the proof of the circular law, notably in [63]. As we shall see, the term $1/\sqrt{n}$ comes from the rate of convergence in the Berry-Esseen Theorem. Following [92], it could probably be improved by using finer results on the Littlewood-Offord problem [130]. Note however, that it is sufficient in the perspective of proving convergence in probability of spectral measures.

We emphasize that there is not any moments assumption on the entries in lemma A.1. However, (weak) moments assumptions may be used in order to obtain an upper bound on the quantity $\mathbb{P}(s_1(X+M)\geq s\sqrt{n})$. Also, the variance (of the truncated variables) σ may depend on n : this allows to deal with sparse matrix models (not considered here).

For the proof of the circular law and its heavy tailed analogue, lemma A.1 can be used typically with $t=\sigma^2/(s^2\sqrt{n})$ and $s=n^a$ large enough such that with high probability $s_1(X+M)\leq s\sqrt{n}$. In contrast with the Tao and Vu result, lemma A.1 cannot provide a summable bound usable with the first Borel-Cantelli lemma due to the presence of $1/\sqrt{n}$.

Let us give the idea behind the proof of lemma A.1. A geometric intuition says that the smallest singular value of a random matrix can be controlled by the minimum of the distances of each row to the span of the remaining rows. The distance of a vector to a subspace can be controlled with the scalar product of the vector with a unit norm vector belonging to the orthocomplement of the subspace. Also, when the entries of the matrix are independent, this boils down by conditioning to the control of a small ball probability involving a linear combination of independent random variables. The coefficients in this combination are the components of the orthogonal vector. The asymptotic behavior of this small ball probability depends in turn on the structure of these coefficients. When the coefficients are well spread, we expect an asymptotic Gaussian behavior thanks to the central limit theorem, more precisely its quantitative weighted version called the Berry-Esseen theorem. We will follow this scheme while keeping the geometric picture in mind.

The proof of lemma A.1 is divided into two parts which correspond to a subdivision of the unit sphere $\mathbb{S}^{n-1}$ of $\mathbb{C}^n$. Namely, for some real positive parameters $\delta,\rho>0$ that will be fixed later, we define the set of *sparse vectors*

$$\mathrm{Sparse}:=\{x\in\mathbb{C}^n:\mathrm{card}(\mathrm{supp}(x))\leq\delta n\}$$

and we split the unit sphere $\mathbb{S}^{n-1}$ into a set of *compressible vectors* and the complementary set of *incompressible vectors* as follows:

$$\mathrm{Comp}:=\{x\in\mathbb{S}^{n-1}:\mathrm{dist}(x,\mathrm{Sparse})\leq\rho\}\quad\text{and}\quad\mathrm{Incomp}:=\mathbb{S}^{n-1}\setminus\mathrm{Comp}.$$

We will use the variational formula, for $A\in\mathcal{M}_n(\mathbb{C})$,

$$s_n(A)=\min_{x\in\mathbb{S}^{n-1}}\|Ax\|_2=\min\bigg(\min_{x\in\mathrm{Comp}}\|Ax\|_2,\min_{x\in\mathrm{Incomp}}\|Ax\|_2\bigg).\tag{A.1}$$

Compressible vectors. Our treatment of compressible vectors differs significantly from [92, 113]. We start with a variation around lemma 4.12.

Lemma A.2 (Distance of a random vector to a small subspace)**.** *There exist $\varepsilon, c, \delta_0 > 0$ such that for all $n \gg 1$, all $1 \leq i \leq n$, any deterministic vector $v \in \mathbb{C}^n$ and any subspace H of $\mathbb{C}^n$ with $1 \leq \dim(H) \leq \delta_0 n$, we have, denoting $C := (X_{1i}, \dots, X_{ni}) + v$,*

$$\mathbb{P}\big(\mathrm{dist}(C, H) \leq \varepsilon\sigma\sqrt{n}\big) \leq \exp(-cn).$$

Proof. First, from Hoeffding's deviation inequality,

$$\mathbb{P}\left(\sum_{k=1}^{n} \mathbf{1}_{\{|X_{ki}| \leq a\}} \leq \frac{nb}{2}\right) \leq \exp\left(-\frac{nb^2}{2}\right).$$

It is thus sufficient to prove the result by conditioning on

$$E_m := \{|X_{1i}| \leq a, \dots, |X_{mi}| \leq a\} \quad \text{with} \quad m := \lceil nb/2 \rceil.$$

Let $\mathbb{E}_m[\,\cdot\,] := \mathbb{E}[\,\cdot\,|E_m; \mathcal{F}_m]$ denote the conditional expectation given E_m and the filtration $\mathcal{F}_m$ generated by $X_{m+1,i}, \dots, X_{n,i}$. Let W be the subspace spanned by H, v, and the vectors $u := (0, \dots, 0, X_{m+1,i}, \dots, X_{n,i})$ and

$$w := \big(\mathbb{E}\big[X_{1i} \;\big|\; |X_{1i}| \leq a\big], \dots, \mathbb{E}\big[X_{mi} \;\big|\; |X_{mi}| \leq a\big], 0, \dots, 0\big).$$

By construction $\dim(W) \leq \dim(H) + 3$ and W is $\mathcal{F}_m$-measurable. We note also that

$$\mathrm{dist}(C, H) \geq \mathrm{dist}(C, W) = \mathrm{dist}(Y, W),$$

where

$$Y := \big(X_{1i} - \mathbb{E}\big[X_{1i} \;\big|\; |X_{1i}| \leq a\big], \dots, X_{mi} - \mathbb{E}\big[X_{mi} \;\big|\; |X_{mi}| \leq a\big], 0, \dots, 0\big) = C - u - v - w.$$

By assumption, for $1 \leq k \leq m$,

$$\mathbb{E}_m Y_k = 0 \quad \text{and} \quad \mathbb{E}_m |Y_k|^2 \geq \sigma^2.$$

Let $D = \{z : |z| \leq a\}$. We define the function $f : D^m \to \mathbb{R}_+$ by

$$f(x) = \mathrm{dist}((x_1, \dots, x_m, 0, \dots, 0), W).$$

This function is convex and 1-Lipschitz. Hence, Talagrand's concentration inequality gives

$$\mathbb{P}_m(|\mathrm{dist}(Y, W) - M_m| \geq t) \leq 4\exp\left(-\frac{t^2}{16a^2}\right),$$

where M_m is the median of f under $\mathbb{P}_m$. In particular,

$$M_m \geq \sqrt{\mathbb{E}_m \mathrm{dist}^2(Y, W)} - ca.$$

Also, if P denotes the orthogonal projection on the orthogonal of W, we find

$$\begin{aligned}
\mathbb{E}_m \mathrm{dist}^2(Y, W) &= \sum_{k=1}^{m} \mathbb{E}_m |Y_k|^2 P_{kk} \\
&\geq \sigma^2 \left(\sum_{k=1}^{n} P_{kk} - \sum_{k=m+1}^{n} P_{kk}\right) \\
&\geq \sigma^2 (n - \dim(H) - 3 - (n - m)) \\
&\geq \sigma^2 \left(\frac{nb}{2} - \dim(H) - 3\right).
\end{aligned}$$

The latter, for n large enough, is lower bounded by $c\sigma^2 n$ if $\delta_0 = b/4$. □

Let $0 < \varepsilon < 1$ and $s \geq 1$ be as in lemma A.2. We set from now on

$$\rho = \frac{\varepsilon}{4s} \min(1, \sigma),$$

(in particular, $\rho \leq 1/4$). The parameter δ is still to be specified: at this stage, we simply assume that $\delta < \delta_0$. We note that if $A \in \mathcal{M}_n(\mathbb{C})$ and $y \in \mathbb{C}^n$ is such that $\mathrm{supp}(y) \subset \pi \subset \{1, \ldots, n\}$, then we have

$$\|Ay\|_2 \geq \|y\|_2 s_n(A_{|\pi}),$$

where $A_{|\pi}$ is the $n \times |\pi|$ matrix formed by the columns of A selected by π. We deduce that

$$\min_{x \in \mathrm{Comp}} \|Ax\|_2 \geq \frac{3}{4} \min_{\pi \subset \{1,\ldots,n\} : |\pi| = \lfloor \delta n \rfloor} s_n(A_{|\pi}) - \rho s_1(A). \tag{A.2}$$

However, by Pythagoras theorem, for any $x \in \mathbb{C}^{|\pi|}$,

$$\left\|A_{|\pi} x\right\|_2^2 = \left\| \sum_{i \in \pi} x_i C_i \right\|_2^2 = \sum_{i \in \pi} |x_i|^2 \mathrm{dist}^2(C_i, H_i) \geq \min_{i \in \pi} \mathrm{dist}^2(C_i, H_i) \sum_{i \in \pi} |x_i|^2$$

where C_i is the i-th column of A and $H_i := \mathrm{span}\{C_j : j \in \pi, j < i\}$. In particular,

$$s_n(A_{|\pi}) \geq \min_{i \in \pi} \mathrm{dist}(C_i, H_i),$$

Now, we apply this bound to $A = X + M$. Since H_i has dimension at most $n\delta$ and is independent of C_i, by lemma A.2, the event that,

$$\min_{i \in \pi} \mathrm{dist}(C_i, H_i) \geq \varepsilon \sigma \sqrt{n},$$

has probability at least $1 - n \exp(-cn)$ for $n \gg 1$. Hence

$$\mathbb{P}\big(s_n((X+M)_{|\pi}) \leq \varepsilon \sigma \sqrt{n}\big) \leq n \exp(-cn).$$

Therefore, using the union bound and our choice of ρ, we deduce from (A.2)

$$\mathbb{P}\left(\min_{x \in \mathrm{Comp}} \|(X+M)x\|_2 \leq \frac{\varepsilon\sigma}{2}\sqrt{n} \ ; \ s_1(X+M) \leq s\sqrt{n} \right) \leq \binom{n}{\lfloor \delta n \rfloor} n e^{-cn} = e^{n(H(\delta) - c + o(1))},$$

with $H(\delta) := -\delta \log \delta - (1-\delta)\log(1-\delta)$. Therefore, if δ is chosen small enough so that $H(\delta) < c$, we have proved that for some $c_1 > 0$,

$$\mathbb{P}\left(\min_{x \in \mathrm{Comp}} \|(X+M)x\|_2 \leq \frac{\varepsilon\sigma}{2}\sqrt{n} \ ; \ s_1(X+M) \leq s\sqrt{n} \right) \leq \exp(-c_1 n). \tag{A.3}$$

Incompressible vectors: invertibility via distance. We start our treatment of incompressible vectors with two key observations from [113].

Lemma A.3 (Incompressible vectors are spread)**.** *Let $x \in$ Incomp. There exists a subset $\pi \subset \{1, \ldots, n\}$ such that $|\pi| \geq \delta n/2$ and for all $i \in \pi$,*

$$\frac{\rho}{\sqrt{n}} \leq |x_i| \leq \sqrt{\frac{2}{\delta n}}.$$

Proof. For $\pi \subset \{1, \ldots, n\}$, we denote by P_π the orthogonal projection on $\mathrm{span}\{e_i ; i \in \pi\}$. Let $\pi_1 = \{k : |x_k| \leq \sqrt{2/(\delta n)}\}$ and $\pi_2 = \{k : |x_k| \geq \rho/\sqrt{n}\}$. Since $\|x\|_2^2 = 1$, we have

$$|\pi_1^c| \leq \frac{\delta n}{2}.$$

Note also that

$$\|x - P_{\pi_2} x\|_2 = \left\|P_{\pi_2^c} x\right\|_2 \leq \rho.$$

Hence, the definition of incompressible vectors implies that $|\pi_2| \geq \delta n$. We put $\pi = \pi_1 \cap \pi_2$. From what precedes,

$$|\pi| \geq n - |\pi_1^c| - |\pi_2^c| \geq n - \frac{\delta n}{2} - (n - \delta n) = \frac{\delta n}{2}.$$

□

Lemma A.4 (Invertibility via average distance)**.** *Let A be a random matrix taking its values in $\mathcal{M}_n(\mathbb{C})$, with columns $C_1, \ldots, C_n$, and for some arbitrary $1 \leq k \leq n$, let H_k be the span of all these columns except C_k. Then, for any $t \geq 0$,*

$$\mathbb{P}\bigg(\min_{x \in \mathrm{Incomp}} \|Ax\|_2 \leq \frac{t\rho}{\sqrt{n}}\bigg) \leq \frac{1}{\delta n} \sum_{k=1}^{n} \mathbb{P}(\mathrm{dist}(C_k, H_k) \leq t).$$

Proof. Let $x \in \mathbb{S}^{n-1}$, from $Ax = \sum_k C_k x_k$, we get

$$\|Ax\|_2 \geq \max_{1 \leq k \leq n} \mathrm{dist}(Ax, H_k) = \max_{1 \leq k \leq n} |x_k| \mathrm{dist}(C_k, H_k).$$

Now if $x \in \mathrm{Incomp}$ and π is as in lemma A.3, we get

$$\|Ax\|_2 \geq \frac{\rho}{\sqrt{n}} \max_{k \in \pi} \mathrm{dist}(C_k, H_k).$$

Then, the conclusion follows from the fact that for any reals $y_1, \ldots, y_n$ and $1 \leq m \leq n$,

$$\mathbf{1}_{\{\max_{1 \leq k \leq m} y_k \leq t\}} \leq \frac{1}{m} \sum_{k=1}^{m} \mathbf{1}_{\{y_k \leq t\}} \leq \frac{1}{m} \sum_{k=1}^{n} \mathbf{1}_{\{y_k \leq t\}}.$$

□

The strength of lemma A.4 lies in the fact that the control of $\|Ax\|_2$ over all incompressible vectors is done by an average of the distance between the columns of A.

Incompressible vectors: small ball probability. Now, we come back to our matrix $X + M$: let C be the k-th column of $X + M$ and H be the span of all columns but C. Our goal in this subsection is to establish the bound, for all $t \geq 0$,

$$\mathbb{P}\big(\mathrm{dist}(C, H) \leq \rho t\,;\, s_1(X + M) \leq s\sqrt{n}\big) \leq \frac{c\sqrt{\log \rho}}{\sigma}\bigg(t + \frac{1}{\sqrt{n}}\bigg). \tag{A.4}$$

To this end, we also consider a random vector ζ taking its values in $\mathbb{S}^{n-1} \cap H^{\perp}$, which is independent of C. Such a random vector ζ is not unique, we just pick one and we call it the *orthogonal vector* (to the subspace H). We have

$$\mathrm{dist}(C, H) \geq |\langle \zeta, C \rangle|. \tag{A.5}$$

Lemma A.5 (The random orthogonal vector is Incompressible)**.** *For our choice of ρ, δ and c_1 as in* (A.3)*, we have*

$$\mathbb{P}\big(\zeta \in \mathrm{Comp}\,; s_1(X + M) \leq s\sqrt{n}\big) \leq \exp(-c_1 n).$$

Proof. Let $A \in \mathcal{M}_{n-1,n}(\mathbb{C})$ be the matrix obtained from $(X + M)^*$ by removing the k-th row. Then, by construction : $A\zeta = 0$, $s_1((X + M)^*) = s_1(X + M)$, and

$$\min_{x \in \mathrm{Comp}} \|Ax\|_2 \geq \min_{x \in \mathrm{Comp}} \|(X + M)^* x\|_2.$$

The left hand side (and thus the right hand side) is zero if $\zeta \in \mathrm{Comp}$. In particular,

$$\mathbb{P}\big(\zeta \in \mathrm{Comp}\,; s_1(X + M) \leq s\sqrt{n}\big) \leq \mathbb{P}\bigg(\min_{x \in \mathrm{Comp}} \|(X + M)^* x\|_2 = 0\ ;\ s_1((X + M)^*) \leq s\sqrt{n}\bigg).$$

It remains to notice that obviously (A.3) holds with $(X + M)$ replaced by $(X + M)^*$. Indeed the statistical assumptions are the same on $X + M$ and $(X + M)^*$. □

We have reached now the final preparation step before the use of the Berry-Esseen theorem. This step consists in the reduction to a case where for a fixed set of coordinates, both the components of ζ and the random variables $X_{ik}+M_{ik}$ are well controlled. Namely, if $\zeta \in \mathrm{Incomp}$, let $\pi \subset \{1,\dots,n\}$ be as in lemma A.3 associated to vector ζ. Then conditioned on $\{\zeta \in \mathrm{Incomp}\}$, from Hoeffding's deviation inequality, the event that

$$\sum_{i\in\pi} \mathbf{1}_{\{|X_{ik}|\le a\}} \ge \frac{|\pi| b}{2} \ge \frac{\delta b n}{4},$$

has conditional probability at least $1-\exp(-|\pi|^2 b^2/2) \ge 1-\exp(-cn)$ (recall that ζ hence π are independent of C). In summary, by lemma A.5 and (A.5), in order to prove (A.4), it is sufficient to prove that for all $t \ge 0$,

$$\mathbb{P}_m(|\langle \zeta, C\rangle| \le \rho t) \le \frac{c}{\sigma}\sqrt{\frac{|\log \rho|}{n}}(t+1).$$

where $\mathbb{P}_m(\cdot) = \mathbb{P}(\cdot | E_m, \mathcal{F}_m)$ is the conditional probability given $\mathcal{F}_m$ the σ-algebra generated by all variables but $(X_{1k},\dots,X_{mk})$, $m = \lfloor \delta b n/4 \rfloor$, and the event

$$E_m = \left\{ \frac{\rho}{\sqrt{n}} \le |\zeta_i| \le \sqrt{\frac{2}{\delta n}}; 1 \le i \le m \right\} \bigcup \{|X_{ik}| \le a; 1 \le i \le m\}.$$

We may write

$$\langle \zeta, C\rangle = \sum_{i=1}^n \bar{\zeta}_i \langle C, e_i\rangle = \sum_{i=1}^m \bar{\zeta}_i X_{ik} + u,$$

where $u \in \mathcal{F}_m$ is independent of $(X_{1k},\dots,X_{mk})$. It follows that

$$\mathbb{P}_m(|\langle \zeta, C\rangle| \le \rho t) \le \sup_{z\in\mathbb{C}, \pi\subset\{1,\dots,m\}} \mathbb{P}_m\left(\left| \sum_{i\in\pi} \bar{\zeta}_i (X_{ik} - \mathbb{E}_m X_{ik}) - z \right| \le \rho t \right). \tag{A.6}$$

The idea, originated from [92], is now to use the rate of convergence given by the Berry-Esseen theorem to upper bound this last expression.

Lemma A.6 (Small ball probability via Berry-Esseen theorem)**.** *There exists $c>0$ such that if $Z_1,\dots,Z_n$ are independent centered complex random variables, then for all $t \ge 0$,*

$$\sup_{z\in\mathbb{C}} \mathbb{P}\left(\left| \sum_{i=1}^n Z_i - z \right| \le t \right) \le \frac{ct}{\sqrt{\sum_{i=1}^n \mathbb{E}(|Z_i|^2)}} + \frac{c\sum_{i=1}^n \mathbb{E}(|Z_i|^3)}{\left(\sum_{i=1}^n \mathbb{E}(|Z_i|^2)\right)^{3/2}}.$$

Proof. Let $\tau^2 = \sum_{i=1}^n \mathbb{E}|Z_i|^2$, then either $\sum_{i=1}^n \mathbb{E}(\mathfrak{Re} Z_i)^2$ or $\sum_{i=1}^n \mathbb{E}(\mathfrak{Im} Z_i)^2$ is larger or equal to $\tau^2/2$. Also

$$\mathbb{P}\left(\left| \sum_{i=1}^n Z_i - z \right| \le t \right) \le \mathbb{P}\left(\left| \sum_{i=1}^n \mathfrak{Re}(Z_i) - \mathfrak{Re}(z) \right| \le t \right)$$

and similarly with $\mathfrak{Im}$. Hence, up to loosing a factor 2, we can assume with loss of generality that the Z_i's are real random variables. Then, if G is a real centered Gaussian random variable with variance τ^2, Berry-Esseen theorem asserts that

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\sum_{i=1}^n Z_i \le t \right) - \mathbb{P}(G \le t) \right| \le c_0 \tau^{-3/2} \sum_{i=1}^n \mathbb{E}(|Z_i|^3).$$

In particular, for all $t \ge 0$ and $x \in \mathbb{R}$,

$$\mathbb{P}\left(\left| \sum_{i=1}^n Z_i - x \right| \le t \right) \le \mathbb{P}(|G-x| \le t) + 2c_0 \tau^{-3/2} \sum_{i=1}^n \mathbb{E}(|Z_i|^3).$$

We conclude by using the fact that G has a density upper bounded by $1/\sqrt{2\pi\tau^2}$. □

Define $L = \frac{1}{2}\log_2 \frac{2}{\delta\rho^2}$. Note that for our choice of ρ, δ, for some constant $c = c(a,b)$,

$$L \leq c|\log \rho|.$$

For $1 \leq j \leq L$, we define

$$\pi_j = \left\{ 1 \leq i \leq m : \frac{2^{j-1}\rho}{\sqrt{n}} \leq |\zeta_i| \leq \frac{2^j \rho}{\sqrt{n}} \right\}.$$

From the pigeonhole principle, there exists j such that $|\pi_j| \geq m/L$. We have

$$\sigma_j^2 = \sum_{i \in \pi_j} |\zeta_i|^2 \mathbb{E}_m(|X_{ik} - \mathbb{E}_m(X_{ik})|^2) \geq \frac{2^{2j-2}\rho^2\sigma^2|\pi_j|}{n}.$$

and,

$$\sum_{i \in \pi_j} |\zeta_i|^3 \mathbb{E}_m(|X_{ik} - \mathbb{E}_m(X_{ik})|^3) \leq \frac{2^j a\rho}{\sqrt{n}}\sigma_j^2.$$

We deduce by (A.6) and lemma A.6 that (by changing the value of c), for all $t \geq 0$,

$$\begin{aligned}
\mathbb{P}_m(|\langle \zeta, C\rangle| \leq \rho t) &\leq \frac{c\rho t}{\sigma_j} + \frac{c2^j a\rho}{\sigma_j\sqrt{n}} \\
&\leq \frac{ct\sqrt{n}}{\sigma\sqrt{|\pi_j|}} + \frac{c}{\sigma\sqrt{|\pi_j|}} \\
&\leq \frac{c\sqrt{|\log\rho|}}{\sigma}\left(t + \frac{1}{\sqrt{n}}\right).
\end{aligned}$$

The proof of (A.4) is complete.

Proof of lemma A.1. All ingredient have now been gathered. By lemma A.4 and (A.4) we find, for all $t \geq 0$,

$$\mathbb{P}\left(\min_{x \in \mathrm{Incomp}} \|(X+m)x\|_2 \leq \frac{\rho^2 t}{\sqrt{n}} \ ; \ s_1(X+M) \leq s\sqrt{n} \right) \leq \frac{c\sqrt{|\log\rho|}}{\sigma}\left(t + \frac{1}{\sqrt{n}}\right).$$

Using our choice of ρ, we obtain for some new constant $c > 0$,

$$\mathbb{P}\left(\min_{x \in \mathrm{Incomp}} \|(X+m)x\|_2 \leq \frac{t}{\sqrt{n}} \ ; \ s_1(X+M) \leq s\sqrt{n} \right) \leq \frac{c\sqrt{\log(cs/\sigma)}}{\sigma}\left(\frac{ts^2}{\sigma^2} + \frac{1}{\sqrt{n}}\right).$$

The desired result follows then by using (A.1) and (A.3). □

Acknowledgments

These notes are the expanded version of lecture notes written for the France-China summer school held in Changchun, China, July 2011. The authors are grateful to the organizers Z.-D. Bai, A. Guionnet, and J.-F. Yao for their invitation, to the local team for their wonderful hospitality, and to the participants for their feedback. It is also a pleasure to thank our collaborator Pietro Caputo for his useful remarks on the draft version of these notes, and Alexander Tikhomirov and Van Vu for our discussions on the circular law during the workshop in Oberwolfach organized by M. Ledoux, M. Rudelson, and G. Schechtman in May 2011. All numerics and graphics were done using the free software GNU-Octave and wxMaxima provided by the Debian GNU/Linux operating system. We are also grateful to the anonymous referee who pointed out a bug in the first version.

References

[1] R. Adamczak – "On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries", *Electronic Journal of Probability* **16** (2011), p. 1065–1095.

[2] G. Akemann and E. Kanzieper – "Integrable structure of Ginibre's ensemble of real random matrices and a Pfaffian integration theorem", *J. Stat. Phys.* **129** (2007), no. 5-6, p. 1159–1231.

[3] G. Akemann, M. J. Phillips and L. Shifrin – "Gap probabilities in non-Hermitian random matrix theory", *J. Math. Phys.* **50** (2009), no. 6, p. 063504, 32.

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

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

[6] 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, p. 1–72.

[7] G. W. Anderson, A. Guionnet and O. Zeitouni – *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.

[8] A. L. Andrew – "Eigenvalues and singular values of certain random matrices", *J. Comput. Appl. Math.* **30** (1990), no. 2, p. 165–171.

[9] J.-M. Azaïs and M. Wschebor – "On the roots of a random system of equations. The theorem on Shub and Smale and some extensions", *Found. Comput. Math.* **5** (2005), no. 2, p. 125–144.

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

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

[12] Z. D. Bai, J. W. Silverstein and Y. Q. Yin – "A note on the largest eigenvalue of a large-dimensional sample covariance matrix", *J. Multivariate Anal.* **26** (1988), no. 2, p. 166–168.

[13] Z. D. Bai and Y. Q. Yin – "Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang", *Probab. Theory Related Fields* **73** (1986), no. 4, p. 555–569.

[14] ———, "Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix", *Ann. Probab.* **21** (1993), no. 3, p. 1275–1294.

[15] Z. Bai and J. W. Silverstein – *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, New York, 2010.

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

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

[18] G. Ben Arous and O. Zeitouni – "Large deviations from the circular law", *ESAIM Probab. Statist.* **2** (1998), p. 123–134 (electronic).

[19] F. Benaych-Georges and F. Chapon – "Random right eigenvalues of Gaussian quaternionic matrices", preprint arXiv:1104.4455 [math.PR], 2011.

[20] M. Bender – "Edge scaling limits for a family of non-Hermitian random matrix ensembles", *Probab. Theory Related Fields* **147** (2010), no. 1-2, p. 241–271.

[21] P. Biane and F. Lehner – "Computation of some examples of Brown's spectral measure in free probability", *Colloq. Math.* **90** (2001), no. 2, p. 181–211.

[22] C. Bordenave – "On the spectrum of sum and product of non-Hermitian random matrices", *Electronic Communications in Probability* **16** (2011), p. 104–113.

[23] C. Bordenave, P. Caputo and D. Chafaï – "Circular Law Theorem for Random Markov Matrices", preprint arXiv:0808.1502 [math.PR] to appear in Probability Theory and Related Fields, 2011.

[24] ———, "The spectral measure of random generators", preprint in preparation, 2011.

[25] ———, "Spectrum of large random reversible Markov chains: Heavy-tailed weigths on the complete graph", *Annals of Probability* **39** (2011), no. 4, p. 1544–1590.

[26] ———, "Spectrum of non-Hermitian heavy tailed random matrices", preprint arXiv:1006.1713 [math.PR] to appear in Communications in Mathematical Physics, 2011.

[27] A. Borodin and C. D. Sinclair – "The Ginibre ensemble of real random matrices and its scaling limits", *Comm. Math. Phys.* **291** (2009), no. 1, p. 177–224.

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

[29] J. Bourgain, V. H. Vu and P. M. Wood – "On the singularity probability of discrete random matrices", *J. Funct. Anal.* **258** (2010), no. 2, p. 559–603.

[30] 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, p. 1–35.

[31] D. Chafaï – "Circular law for noncentral random matrices", *Journal of Theoretical Probability* **23** (2010), no. 4, p. 945–950.
[32] ———, "The Dirichlet Markov Ensemble", *J. of Multivariate Analysis* **101** (2010), p. 555–567.
[33] D. Chafaï, O. Guédon, G. Lecué and A. Pajor – "Interactions between compressed sensing, random matrices, and high dimensional geometry", forthcoming book in preparation available on the Internet, 2011.
[34] S. Chatterjee – "A generalization of the Lindeberg principle", *Ann. Probab.* **34** (2006), no. 6, p. 2061–2076.
[35] S. Chatterjee, P. Diaconis and A. Sly – "Properties of uniform doubly stochastic matrices", preprint arXiv:1010.6136 [math.PR], 2010.
[36] Z. Chen, J.-T. Zhang and F. Hu (eds.) – *Advances in statistics*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2008, Including the proceedings of the conference in honor of Professor Zhidong Bai on his 65th birthday, held at the National University of Singapore, Singapore, July 20, 2008.
[37] Z. Dong, T. Jiang and D. Li – "Circular Law and Arc Law for Truncation of Random Unitary Matrix", preprint, personal communication by Tiefeng Jiang, 2011.
[38] R. B. Dozier and J. W. Silverstein – "On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices", *J. Multivariate Anal.* **98** (2007), no. 4, p. 678–694.
[39] A. Edelman – "Eigenvalues and condition numbers of random matrices", *SIAM J. Matrix Anal. Appl.* **9** (1988), no. 4, p. 543–560.
[40] ———, "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, p. 203–232.
[41] A. Edelman, E. Kostlan and M. Shub – "How many eigenvalues of a random matrix are real?", *J. Amer. Math. Soc.* **7** (1994), no. 1, p. 247–267.
[42] A. Edelman and E. Kostlan – "How many zeros of a random polynomial are real?", *Bull. Amer. Math. Soc. (N.S.)* **32** (1995), no. 1, p. 1–37.
[43] ———, "Erratum: "How many zeros of a random polynomial are real?" [Bull. Amer. Math. Soc. (N.S.) **32** (1995), no. 1, 1–37; MR1290398 (95m:60082)]", *Bull. Amer. Math. Soc. (N.S.)* **33** (1996), no. 3, p. 325.
[44] I. Emiris, A. Galligo and E. Tsigaridas – "Random polynomials and expected complexity of bisection methods for real solving", preprint arXiv:1005.2001 [cs.SC], 2010.
[45] L. Erdős, B. Schlein and H.-T. Yau – "Wegner estimate and level repulsion for Wigner random matrices", *Int. Math. Res. Not. IMRN* (2010), no. 3, p. 436–479.
[46] J. Feinberg and A. Zee – "Non-Hermitian Random Matrix Theory: Method of Hermitian Reduction", *Nucl. Phys. B* (1997), no. 3, p. 579–608.
[47] P. J. Forrester – *Log-gases and random matrices*, London Mathematical Society Monographs Series, vol. 34, Princeton University Press, Princeton, NJ, 2010.
[48] P. J. Forrester and M. Krishnapur – "Derivation of an eigenvalue probability density function relating to the poincar disk", *Journal of Physics A: Mathematical and Theoretical* **42** (2009), no. 38, p. 385204.
[49] B. Fuglede and R. V. Kadison – "Determinant theory in finite factors", *Ann. of Math. (2)* **55** (1952), p. 520–530.
[50] A. Galligo – "Intriguing Patterns in the Roots of the Derivatives of some Random Polynomials", preprint oai:hal.inria.fr:inria-00552081, 2010.
[51] J. Ginibre – "Statistical ensembles of complex, quaternion, and real matrices", *J. Mathematical Phys.* **6** (1965), p. 440–449.
[52] V. L. Girko – "The circular law", *Teor. Veroyatnost. i Primenen.* **29** (1984), no. 4, p. 669–679.
[53] ———, *Theory of random determinants*, Mathematics and its Applications (Soviet Series), vol. 45, Kluwer Academic Publishers Group, Dordrecht, 1990, Translated from the Russian.
[54] ———, "The circular law: ten years later", *Random Oper. Stochastic Equations* **2** (1994), no. 3, p. 235–276.
[55] ———, *Theory of stochastic canonical equations. Vol. I*, Mathematics and its Applications, vol. 535, Kluwer Academic Publishers, Dordrecht, 2001.
[56] ———, "The strong circular law. Twenty years later. I", *Random Oper. Stochastic Equations* **12** (2004), no. 1, p. 49–104.
[57] ———, "The strong circular law. Twenty years later. II", *Random Oper. Stochastic Equations* **12** (2004), no. 3, p. 255–312.
[58] ———, "The circular law. Twenty years later. III", *Random Oper. Stochastic Equations* **13** (2005), no. 1, p. 53–109.

[59] V. L. Girko and A. I. Vladimirova – "L.I.F.E.: $\prod_{j=1}^m H_n^{(j)} \sim \widetilde{\text{L.I.F.E.}} \sim \{H_n^{(1)}\}^m$ and Halloween Law", *Random Oper. Stoch. Equ.* **18** (2010), no. 4, p. 327–353.
[60] A. Goldman – "The Palm measure and the Voronoi tessellation for the Ginibre process", *Ann. Appl. Probab.* **20** (2010), no. 1, p. 90–128.
[61] I. Y. Goldsheid and B. A. Khoruzhenko – "The Thouless formula for random non-Hermitian Jacobi matrices", *Israel J. Math.* **148** (2005), p. 331–346, Probability in mathematics.
[62] G. H. Golub and C. F. Van Loan – *Matrix computations*, third ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996.
[63] F. Götze and A. Tikhomirov – "The circular law for random matrices", *Ann. Probab.* **38** (2010), no. 4, p. 1444–1491.
[64] ———, "On the asymptotic spectrum of products of independent random matrices", arXiv:1012.2710 [math.PR], 2010.
[65] F. Götze and D. Zaporozhets – "On the Distribution of Complex Roots of Random Polynomials with Heavy-tailed Coefficients", arXiv:1104.5360 [math.PR], 2011.
[66] E. Gudowska-Nowak, A. Jarosz, M. Nowak and G. Pappe – "Towards non-Hermitian random lévy matrices", *Acta Physica Polonica B* **38** (2007), no. 13, p. 4089–4104.
[67] A. Guionnet, M. Krishnapur and O. Zeitouni – "The single ring theorem", preprint arXiv:0909.2214 [math.PR] to appear in the Annals of Mathematics, 2009.
[68] A. Guionnet and O. Zeitouni – "Support convergence in the single ring theorem", preprint arXiv:1012.2624 [math.PR], 2010.
[69] A. Guntuboyina and H. Leeb – "Concentration of the spectral measure of large Wishart matrices with dependent entries", *Electron. Commun. Probab.* **14** (2009), p. 334–342.
[70] U. Haagerup and F. Larsen – "Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras", *J. Funct. Anal.* **176** (2000), no. 2, p. 331–367.
[71] U. Haagerup and H. Schultz – "Brown measures of unbounded operators affiliated with a finite von Neumann algebra", *Math. Scand.* **100** (2007), no. 2, p. 209–263.
[72] J. M. Hammersley – "The zeros of a random polynomial", *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II* (Berkeley and Los Angeles), University of California Press, 1956, p. 89–111.
[73] D. E. Holz, H. Orland and A. Zee – "On the remarkable spectrum of a non-Hermitian random matrix model", *J. Phys. A* **36** (2003), no. 12, p. 3385–3400, Random matrix theory.
[74] A. Horn – "On the eigenvalues of a matrix with prescribed singular values", *Proc. Amer. Math. Soc.* **5** (1954), p. 4–7.
[75] R. A. Horn and C. R. Johnson – *Matrix analysis*, Cambridge University Press, Cambridge, 1990, Corrected reprint of the 1985 original.
[76] ———, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1994, Corrected reprint of the 1991 original.
[77] J. B. Hough, M. Krishnapur, Y. Peres and B. Virág – *Zeros of Gaussian analytic functions and determinantal point processes*, University Lecture Series, vol. 51, American Mathematical Society, Providence, RI, 2009.
[78] 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, p. 145–152.
[79] T. Jiang – "Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles", *Probab. Theory Related Fields* **144** (2009), no. 1-2, p. 221–246.
[80] K. Johansson – "From Gumbel to Tracy-Widom", *Probab. Theory Related Fields* **138** (2007), no. 1-2, p. 75–112.
[81] M. Kac – "A correction to "On the average number of real roots of a random algebraic equation."", *Bull. Amer. Math. Soc.* **49** (1943), p. 938.
[82] ———, "On the average number of real roots of a random algebraic equation", *Bull. Amer. Math. Soc.* **49** (1943), p. 314–320.
[83] ———, "On the average number of real roots of a random algebraic equation. II", *Proc. London Math. Soc. (2)* **50** (1949), p. 390–408.
[84] H. Kesten – "Symmetric random walks on groups", *Trans. Amer. Math. Soc.* **92** (1959), p. 336–354.
[85] A. Khorunzhy, B. Khoruzhenko, L. Pastur and M. Shcherbina – "The large-n limit in statistical mechanics and the spectral theory of disordered systems", *In: Phase transition and critical phenomena* **15** (1992), p. 74–239.
[86] E. Kostlan – "On the spectra of Gaussian matrices", *Linear Algebra Appl.* **162/164** (1992), p. 385–388, Directions in matrix theory (Auburn, AL, 1990).
[87] P. Lebœuf – "Random analytic chaotic eigenstates", *J. Statist. Phys.* **95** (1999), no. 3-4, p. 651–664.

[88] M. Ledoux – *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.
[89] ———, "Complex Hermite polynomials: from the semi-circular law to the circular law", *Commun. Stoch. Anal.* **2** (2008), no. 1, p. 27–32.
[90] R. LePage, M. Woodroofe and J. Zinn – "Convergence to a stable distribution via order statistics", *Ann. Probab.* **9** (1981), no. 4, p. 624–632.
[91] J. E. Littlewood and A. C. Offord – "On the number of real roots of a random algebraic equation. III", *Rec. Math. [Mat. Sbornik] N.S.* **12(54)** (1943), p. 277–286.
[92] A. E. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann – "Smallest singular value of random matrices and geometry of random polytopes", *Adv. Math.* **195** (2005), no. 2, p. 491–523.
[93] R. Lyons – "Identities and inequalities for tree entropy", *Combin. Probab. Comput.* **19** (2010), no. 2, p. 303–313.
[94] C. Male – "Free probability on traffics: the limiting distribution of heavy wigner and deterministic matrices", preprint, personal communication, 2011.
[95] 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, p. 148–188.
[96] B. D. McKay – "The expected eigenvalue distribution of a large regular graph", *Linear Algebra Appl.* **40** (1981), p. 203–216.
[97] M. Meckes – "Some results on random circulant matrices", *High Dimensional Probability V: The Luminy Volume* (2009), p. 213–223.
[98] M. L. Mehta – *Random matrices and the statistical theory of energy levels*, Academic Press, New York, 1967.
[99] ———, *Random matrices*, third ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004.
[100] S. O'Rourke and A. Soshnikov – "Products of Independent Non-Hermitian Random Matrices", preprint arXiv:1012.4497 [math.PR], 2010.
[101] G. Pan and W. Zhou – "Circular law, extreme singular values and potential theory", *J. Multivar. Anal.* **101** (2010), no. 3, p. 645–656.
[102] D. Panchenko and M. Talagrand – "On one property of Derrida-Ruelle cascades", *C. R. Math. Acad. Sci. Paris* **345** (2007), no. 11, p. 653–656.
[103] Y. Peres and B. Virág – "Zeros of the i.i.d. Gaussian power series: a conformally invariant determinantal process", *Acta Math.* **194** (2005), no. 1, p. 1–35.
[104] D. Petz and F. Hiai – "Logarithmic energy as an entropy functional", Advances in differential equations and mathematical physics (Atlanta, GA, 1997), Contemp. Math., vol. 217, Amer. Math. Soc., Providence, RI, 1998, p. 205–221.
[105] I. Popescu – "General tridiagonal random matrix models, limiting distributions and fluctuations", *Probab. Theory Related Fields* **144** (2009), no. 1-2, p. 179–220.
[106] 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.
[107] B. Rider – "A limit theorem at the edge of a non-Hermitian random matrix ensemble", *J. Phys. A* **36** (2003), no. 12, p. 3401–3409, Random matrix theory.
[108] ———, "Deviations from the circular law", *Probab. Theory Related Fields* **130** (2004), no. 3, p. 337–367.
[109] B. Rider and J. W. Silverstein – "Gaussian fluctuations for non-Hermitian random matrix ensembles", *Ann. Probab.* **34** (2006), no. 6, p. 2118–2143.
[110] B. Rider and B. Virág – "The noise in the circular law and the Gaussian free field", *Int. Math. Res. Not. IMRN* (2007), no. 2, p. Art. ID rnm006, 33.
[111] T. Rogers – "Universal sum and product rules for random matrices", *J. Math. Phys.* **51** (2010), no. 9, p. 093304, 15.
[112] T. Rogers and I. Castillo – "Cavity approach to the spectral density of non-Hermitian sparse matrices", *Phys. Rev. E* **79** (2009), no. 4, p. 012101.
[113] M. Rudelson and R. Vershynin – "The Littlewood-Offord problem and invertibility of random matrices", *Adv. Math.* **218** (2008), no. 2, p. 600–633.
[114] Ø. Ryan – "On the limit distributions of random matrices with independent or free entries", *Comm. Math. Phys.* **193** (1998), no. 3, p. 595–626.
[115] E. B. Saff and V. Totik – *Logarithmic potentials with external fields*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 316, Springer-Verlag, Berlin, 1997, Appendix B by Thomas Bloom.
[116] G. Schehr and S. N. Majumdar – "Real roots of random polynomials and zero crossing properties of diffusion equation", *J. Stat. Phys.* **132** (2008), no. 2, p. 235–273.

[117] D. I. Shparo and M. G. Shur – "On the distribution of roots of random polynomials", *Vestnik Moskov. Univ. Ser. I Mat. Meh.* **1962** (1962), no. 3, p. 40–43.
[118] M. Shub and S. Smale – "Complexity of Bezout's theorem. II. Volumes and probabilities", Computational algebraic geometry (Nice, 1992), Progr. Math., vol. 109, Birkhäuser Boston, Boston, MA, 1993, p. 267–285.
[119] J. W. Silverstein – "The spectral radii and norms of large-dimensional non-central random matrices", *Comm. Statist. Stochastic Models* **10** (1994), no. 3, p. 525–532.
[120] P. Śniady – "Random regularization of Brown spectral measure", *J. Funct. Anal.* **193** (2002), no. 2, p. 291–313.
[121] D. A. Spielman and S.-H. Teng – "Smoothed analysis of algorithms", *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)* (Beijing), Higher Ed. Press, 2002, p. 597–606.
[122] ———, "Smoothed analysis: motivation and discrete models", Algorithms and data structures, Lecture Notes in Comput. Sci., vol. 2748, Springer, Berlin, 2003, p. 256–270.
[123] T.-J. Stieltjes – "Recherches sur les fractions continues", *Ann. Fac. Sci. Toulouse Sci. Math. Sci. Phys.* **8** (1894), no. 4, p. J1–J122.
[124] M. Talagrand – "Concentration of measure and isoperimetric inequalities in product spaces", *Inst. Hautes Études Sci. Publ. Math.* (1995), no. 81, p. 73–205.
[125] T. Tao – "Outliers in the spectrum of iid matrices with bounded rank perturbations", preprint arXiv:1012.4818 [math.PR] to appear in Probability Theory and Related Fields, 2011.
[126] ———, "Topics in random matrix theory", forthcoming book in preraration available on the Internet, 2011.
[127] T. Tao and V. Vu – "Random matrices: A general approach for the least singular value problem", preprint `arXiv:0805.3167 [math.PR]`, 2008.
[128] ———, "Random matrices: the circular law", *Commun. Contemp. Math.* **10** (2008), no. 2, p. 261–307.
[129] ———, "From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices", *Bull. Amer. Math. Soc. (N.S.)* **46** (2009), no. 3, p. 377–396.
[130] ———, "Inverse Littlewood-Offord theorems and the condition number of random discrete matrices", *Ann. of Math. (2)* **169** (2009), no. 2, p. 595–632.
[131] ———, "Random matrices: the distribution of the smallest singular values", *Geom. Funct. Anal.* **20** (2010), no. 1, p. 260–297.
[132] ———, "Random matrices: universality of ESDs and the circular law", *Ann. Probab.* **38** (2010), no. 5, p. 2023–2065, With an appendix by Manjunath Krishnapur.
[133] L. N. Trefethen and M. Embree – *Spectra and pseudospectra*, Princeton University Press, Princeton, NJ, 2005, The behavior of nonnormal matrices and operators.
[134] D. V. Voiculescu – "Free entropy", *Bull. London Math. Soc.* **34** (2002), no. 3, p. 257–278.
[135] D. V. Voiculescu, K. J. Dykema and A. Nica – *Free random variables*, CRM Monograph Series, vol. 1, American Mathematical Society, Providence, RI, 1992, A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.
[136] H. Weyl – "Inequalities between the two kinds of eigenvalues of a linear transformation", *Proc. Nat. Acad. Sci. U. S. A.* **35** (1949), p. 408–411.
[137] P. M. Wood – "Universality and the circular law for sparse random matrices", preprint arXiv:1010.1726 [math.PR] to appear in the Annals of Applied Probability, 2011.
[138] X. Zhan – *Matrix inequalities*, Lecture Notes in Mathematics, vol. 1790, Springer-Verlag, Berlin, 2002.

E-mail address: `charles.bordenave(at)math.univ-toulouse.fr`
URL: `http://www.math.univ-toulouse.fr/~bordenave/`

Université de Toulouse, Institut de Mathématiques de Toulouse, UMR CNRS 5219, France

E-mail address: `djalil(at)chafai.net`
URL: `http://djalil.chafai.net/`

Université Paris-Est Marne-la-Vallée, UMR CNRS 8050, France