



**HAL**  
open science

# Circular Law Theorem for Random Markov Matrices

Djalil Chafai

► **To cite this version:**

| Djalil Chafai. Circular Law Theorem for Random Markov Matrices. 2008. hal-00310528v4

**HAL Id: hal-00310528**

**<https://hal.science/hal-00310528v4>**

Preprint submitted on 11 Aug 2008 (v4), last revised 8 Jun 2010 (v5)

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

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

# Circular Law Theorem for Random Markov Matrices

Djalil CHAFAÏ

August 11, 2008

## Abstract

Let  $(X_{i,j})$  be an infinite array of i.i.d. non negative real random variables with unit mean, finite positive variance  $\sigma^2$ , and finite fourth moment. Let  $M$  be the  $n \times n$  random Markov matrix with i.i.d. rows defined by  $M_{i,j} = X_{i,j}/(X_{i,1} + \dots + X_{i,n})$ . It belongs to the Dirichlet Markov Ensemble when  $X_{1,1}$  follows an exponential law. We show that with probability one, the empirical spectral distribution  $\frac{1}{n}(\delta_{\lambda_1} + \dots + \delta_{\lambda_n})$  of  $\sqrt{n}M$  converges weakly as  $n \rightarrow \infty$  to the uniform law on the disc  $\{z \in \mathbb{C}; |z| \leq \sigma\}$  and moreover, the spectral gap of  $M$  is of order  $n^{-1/2}$  for large enough  $n$ . **There is for now a gap in the proof of Theorem 1.3 page 16.**

**AMS 2000 Mathematical Subject Classification:** 60F15; 15A52; 62H99.

**Keywords:** Random matrices; Spectral gap; Singular values; Eigenvalues; Circular law; Rank-one deformations; Non-normal operators; Logarithmic potential; Markov matrices.

## 1 Introduction

Random Matrix Theory takes its roots in the works of Wishart [1] in Statistics and of Wigner [55] and Dyson [15] in Nuclear Physics. The spectral analysis of large random matrices is still the source of many works and nice open problems. The *eigenvalues* of a square matrix with complex entries are the roots in  $\mathbb{C}$  of its characteristic polynomial. For any  $n \times n$  complex matrix  $A$ , we label its eigenvalues  $\lambda_1(A), \dots, \lambda_n(A)$  so that

$$|\lambda_1(A)| \geq \dots \geq |\lambda_n(A)| \geq 0.$$

The *empirical spectral distribution*  $\mu_A$  of  $A$  is the finite discrete law on  $\mathbb{C}$  defined by

$$\mu_A = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(A)}.$$

The *spectral radius* of  $A$  is  $|\lambda_1(A)|$ . We also denote by

$$s_1(A) \geq \dots \geq s_n(A) \geq 0$$

the *singular values* of  $A$  defined for every  $1 \leq k \leq n$  by

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

where  $A^* = \bar{A}^\top$  is the conjugate transpose of  $A$ . The *operator norm* of  $A$  is

$$\|A\|_{2 \rightarrow 2} = \max_{\|x\|_2=1} \|Ax\|_2 = s_1(A).$$

The matrix  $A$  is singular if and only if  $s_n(A) = 0$  and if not then

$$s_n(A) = s_1(A^{-1})^{-1} = \|A^{-1}\|_{2 \rightarrow 2}^{-1}.$$

If  $A$  is normal then  $s_i(A) = |\lambda_i(A)|$  for every  $1 \leq i \leq n$ , and in general, we know that

$$s_n(A) \leq |\lambda_n(A)| \quad \text{and} \quad |\lambda_1(A)| \leq s_1(A).$$

Let  $(X_{i,j})_{i,j \geq 1}$  be an infinite array of i.i.d. complex random variables with finite positive variance  $0 < \sigma^2 < \infty$ . If  $X = (X_{i,j})_{1 \leq i,j \leq n}$  then the strong Quartercircular Law theorem (Universal version of the Marchenko-Pastur theorem, see [32], [52], [56]) states that

$$\mathbb{P} \left( \mu_{\sqrt{\frac{1}{n} X X^*}} \xrightarrow[n \rightarrow \infty]{w} \mathcal{Q}_\sigma \right) = 1 \tag{1}$$

where “ $\xrightarrow{w}$ ” denotes weak convergence of laws (with respect to continuous bounded functions) and where  $\mathcal{Q}_\sigma$  is the Quartercircular Law on  $\mathbb{R}$  with Lebesgue density

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

This law has mean  $(3\pi)^{-1}8\sigma$  and variance  $\sigma^2$ . If  $Z \sim \mathcal{Q}_\sigma$  then  $Z^2$  follows a Marchenko-Pastur law [32] on  $[0, 4\sigma^2]$  without atom. Additionally, it is shown in [7, 5, 4] that

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} s_1(\frac{1}{\sqrt{n}} X) = 2\sigma \right) = 1 \quad \text{iff} \quad \mathbb{E}(X_{1,1}) = 0 \quad \text{and} \quad \mathbb{E}(|X_{1,1}|^4) < \infty. \tag{2}$$

The famous strong Circular Law theorem [47] states that

$$\mathbb{P} \left( \mu_{\frac{1}{\sqrt{n}} X} \xrightarrow[n \rightarrow \infty]{w} \mathcal{U}_\sigma \right) = 1 \tag{3}$$

where  $\mathcal{U}_\sigma$  is the uniform law on the centered disc  $\{z \in \mathbb{C}; |z| \leq \sigma\}$  of radius  $\sigma$ . This statement, with the sole finite positive variance assumption, appears as the culminating point of a long series of partial results [19], [33], [20], [21], [16], [3], [4], [35], [25], [46], [47] concerning the convergence to the Circular Law of the empirical spectral distribution of random matrices with i.i.d. entries. By (3), we have for every fixed integer  $k \geq 1$ ,

$$\mathbb{P} \left( \liminf_{n \rightarrow \infty} |\lambda_k(\frac{1}{\sqrt{n}} X)| \geq \sigma \right) = 1.$$

We get from [6], [35] and (2) that if additionally  $\mathbb{E}(X_{1,1}) = 0$  and  $\mathbb{E}(|X_{1,1}|^4) < \infty$  then

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} |\lambda_1(\frac{1}{\sqrt{n}} X)| = \sigma \right) = 1 \quad \text{and} \quad \mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{s_1(X)}{|\lambda_1(X)|} = 2 \right) = 1. \tag{4}$$

The behavior of the ratio norm/radius suggests in a sense that  $X$  is not “asymptotically normal”. Following [42] and [13], if  $\mathbb{E}(X_{1,1}) \neq 0$  while  $\mathbb{E}(|X_{1,1}|^4) < \infty$  then almost surely,  $|\lambda_1(n^{-1/2}X)| \rightarrow +\infty$  at speed  $\sqrt{n}$  while  $|\lambda_2(n^{-1/2}X)|$  remains bounded.

The strong Circular Law theorems derived in [46, 47] and [13] are partly (but crucially) based on the Tao & Vu polynomial bound on the least singular value of random matrices with i.i.d. entries [46, 45] (see also [40, 39] and references therein). They showed for instance that for every  $a > 0$  there exists  $b > 0$  such that

$$\mathbb{P}(s_n(X) \leq n^{-b}) = O(n^{-a}). \quad (5)$$

## 1.1 Random Markov Matrices and main results

The present work concerns some kind of random Markov matrices with i.i.d. rows. The entries of these random matrices are dependent. Namely, let  $\mathcal{L}$  be a law on  $[0, \infty)$  with **mean  $m$ , variance  $0 < \sigma^2 < \infty$ , and finite fourth moment**, and set

$$\kappa = \frac{\sigma}{m}.$$

In the whole sequel, and unless otherwise stated,  $(X_{i,j})_{i,j \geq 1}$  is an infinite array of real non negative i.i.d. random variables of law  $\mathcal{L}$ . Note that  $X_{1,1} \geq 0$  with

$$m = \mathbb{E}(X_{1,1}) > 0 \quad \text{and} \quad 0 < \sigma^2 = \mathbb{E}(X_{1,1}^2) - m^2 < \infty \quad \text{and} \quad \mathbb{E}(X_{1,1}^4) < \infty.$$

We define the  $n \times n$  real matrix  $M$  by

$$M_{i,j} = \frac{X_{i,j}}{X_{i,1} + \cdots + X_{i,n}}. \quad (6)$$

The matrix  $M$  is Markov since its entries belong to  $[0, 1]$  and each row sum up to 1. We have  $M = DX$  where  $X = (X_{i,j})_{1 \leq i,j \leq n}$  and  $D$  is the  $n \times n$  diagonal matrix defined by

$$D_{i,i} = \frac{1}{X_{i,1} + \cdots + X_{i,n}}. \quad (7)$$

The matrix  $M$  has equally distributed dependent entries. However, the rows of  $M$  are i.i.d. and follow an exchangeable law on  $\mathbb{R}^n$  supported by the non negative portion of the  $\ell_n^1$  unit sphere. The law of the entries of  $M$  is invariant if  $\mathcal{L}$  is dilated and one can for instance assume that  $m = 1$ . We have  $\sqrt{n}M = nDn^{-1/2}X$  and by the strong Law of Large Numbers, the diagonal matrix  $nD$  is close to the identity matrix  $I$  for large  $n$ , and thus  $\sqrt{n}M$  is close to the matrix  $n^{-1/2}X$  which has i.i.d. entries and satisfies to (1) and (3). Unfortunately, the spectrum of non normal matrices is very sensitive to perturbations [49], and some rigorous work is needed.

Our first result below is the analog of the strong Universal Marchenko-Pastur theorem (1). It generalizes the result of the same kind obtained in [14] in the case where  $\mathcal{L}$  is an exponential law. Note that since  $M$  has real entries, we have  $M^* = M^\top$ .

**Theorem 1.1 (Strong Quartercircular Law theorem).** *We have*

$$\mathbb{P}\left(\mu_{\sqrt{n}MM^\top} \xrightarrow[n \rightarrow \infty]{w} \mathcal{Q}_\kappa\right) = 1$$

and moreover

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} s_2(\sqrt{n}M) = 2\kappa\right) = 1 \quad \text{and} \quad \mathbb{P}\left(\lim_{n \rightarrow \infty} s_1(M) = 1\right) = 1.$$

Our second result below provides a strong spectrum localization for  $\sqrt{n}M$ .

**Theorem 1.2 (Strong Spectrum Localization).** *We have*

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} |\lambda_2(\sqrt{n}M)| \leq 2\kappa\right) = 1 \quad \text{and} \quad \lambda_1(M) = 1.$$

Theorem 1.2 shows that almost surely, the spectrum of  $\sqrt{n}M$  contains one copy of  $\sqrt{n}$  while the rest is included in  $\{z \in \mathbb{C}; |z| \leq 2\kappa + \varepsilon\}$  for every  $\varepsilon$  and large enough  $n$ . Our third result states an almost sure weak convergence of  $\mu_{\sqrt{n}M}$  to the Circular Law of radius  $\kappa$ . This result is the analogue of (3) for our random Markov matrices.

**Theorem 1.3 (Strong Circular Law theorem).** *We have*

$$\mathbb{P}\left(\mu_{\sqrt{n}M} \xrightarrow[n \rightarrow \infty]{w} \mathcal{U}_\kappa\right) = 1.$$

If  $\mathcal{L}$  does not have atoms then  $\mathbb{P}(s_n(M) > 0) = \mathbb{P}(\det(M) = \det(D)\det(X) \neq 0) = 1$ . However, if  $\mathcal{L}$  has an atom of weight  $q > 0$  then  $\mathbb{P}(s_n(M) = 0) \geq q^{n^2} > 0$ . It is quite natural to ask about a bound similar to (5) for  $M$ . This random matrix does not have i.i.d. entries, but concentration of measure arguments related to the law of large numbers may help. This simple intuition is confirmed by the following theorem. Its proof was suggested to the author by Manjunath Krishnapur. Note that only a finite variance is needed here. We do not use theorem 1.4 in our proof of theorem 1.3.

**Theorem 1.4 (Polynomial bound on least singular value).** *Let us drop the finite fourth moment assumption for  $\mathcal{L}$ . For every  $a > 0$ , there exists  $b' > 0$  such that*

$$\mathbb{P}(s_n(M) \leq n^{-b'}) = O(n^{-a}).$$

As an immediate consequence, by taking  $a > 1$  and by using the first Borel-Cantelli lemma, we obtain that almost surely, for large enough  $n$ ,

$$|\lambda_n(M)| \geq s_n(M) \geq n^{-b'}.$$

## 1.2 Discussion and open problems

When  $\mathcal{L}$  is an exponential law, theorem 1.3 appears as a strong Circular Law theorem for the Dirichlet Markov Ensemble [14]. For an exponential law,  $m = \sigma$  and  $\kappa = 1$ . Rigorously, the definitions (7) and (6) of  $D$  and  $M$  that we used make sense on the event

$$\mathcal{D}_n = \{D_{1,1} \cdots D_{n,n} > 0\}.$$

Since  $\sigma > 0$  we have  $q = \mathcal{L}(\{0\}) < 1$  and hence

$$\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{D}_n^c) = \sum_{n=1}^{\infty} (1 - (1 - q^n)^n) \leq \sum_{n=1}^{\infty} nq^n < \infty.$$

Therefore, by the first Borel-Cantelli lemma, almost surely, for large enough  $n$ , both  $D$  and  $M$  are well defined. From a Markovian point of view,  $M$  can be seen as the Markov kernel associated to the oriented complete graph with  $n$  vertices with one loop per vertex for which each edge  $(i, j)$  has weight  $X_{i,j}$ . The adjacency matrix of this graph is full of ones. The skeleton of the Markov kernel  $M$  is an oriented Erdős-Rényi random graph where each edge exists independently of the others with probability  $p = 1 - q$ . If  $p = 1$  then almost surely, the skeleton of  $M$  is the complete graph,  $M$  is irreducible, and by theorem 1.2  $M$  is aperiodic for large enough  $n$  since 1 is the sole eigenvalue of  $M$  of module 1. The non oriented version of this graphical construction gives rise to random reversible Markov kernels for which a Semicircular theorem is available [10].

Since  $M$  is Markov, we have from (18) that for every integer  $r \geq 0$ ,

$$\int_{\mathbb{C}} z^r d\mu_M(x, y) = \frac{1}{n} \sum_{i=1}^n \lambda_i^r(M) = \frac{1}{n} \sum_{i=1}^n p_M(r, i)$$

where  $z = x + \sqrt{-1}y$  and

$$p_M(r, i) = \sum_{\substack{1 \leq i_1, \dots, i_r \leq n \\ i_1 = i_r}} M_{i_1, i_2} \cdots M_{i_{r-1}, i_1}$$

is simply, conditional on  $M$ , the probability of a loop of length  $r$  rooted at  $i$  for a Markov chain with transition kernel  $M$ . This provides a probabilistic interpretation of the moments of the empirical spectral distribution  $\mu_M$  of  $M$ . The random Markov matrix  $M$  is a *random environment* in the language of Probability Theory. Note that since  $M$  has real entries, its spectrum (which is complex) is symmetric with respect to the real axis. By combining theorem 1.2 with theorem 1.3 and the identity (19), we get that for every fixed integer  $r \geq 0$ , almost surely,

$$\lim_{n \rightarrow \infty} n^{\frac{r}{2}} \left( \frac{1}{n} \sum_{i=1}^n p_M(r, i) - \frac{1}{n} \right) = 0.$$

The finite fourth moment assumption on  $\mathcal{L}$  is technical and allows a reasonably simple proof of theorem 1.3. One may try to adapt the recent work [47]. It is also likely that the localization result of theorem 1.2 is not optimal. By theorem 1.2 and 1.3 we get

$$\mathbb{P} \left( \kappa \leq \liminf_{n \rightarrow \infty} |\lambda_2(\sqrt{n}M)| \leq \limsup_{n \rightarrow \infty} |\lambda_2(\sqrt{n}M)| \leq 2\kappa \right) = 1.$$

Also, almost surely, the “spectral gap” of  $M$  is of order  $n^{-1/2}$  for large  $n$  (compare with the results of [22, 23]). Note that in contrast with (4), we have from theorems 1.1-1.2

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{s_1(M)}{|\lambda_1(M)|} = 1 \right) = 1.$$

Numerical simulations (see figure 1) suggest that

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} |\lambda_2(\sqrt{n}M)| = \kappa \right) = 1 \quad \text{and thus} \quad \mathbb{P} \left( \lim_{n \rightarrow \infty} \frac{s_2(M)}{|\lambda_2(M)|} = 2 \right) = 1.$$

Unfortunately, our method of proof of theorem 1.2 is too perturbative to extract this result. The almost sure convergence of the sub-dominant eigenvalue to the critical radius is not completely elucidated in the case of random matrices with i.i.d. entries (except certain Ginibre Ensembles for which their Gaussian nature provides the law of the spectrum explicitly). We refer to [30], [37], [18], [31], [6], [4, Chapter 10] for the spectral radius of random matrices with i.i.d. centered entries, and to [42] and [13] for the non central case. Another very interesting open problem is the fluctuation of the global and extremal spectrum for which one can find some partial answers in [43], [36], [42], [38].

Recall that the rows of  $M$  are i.i.d. and follow an exchangeable law  $\eta_n$  on the simplex

$$\{(x_1, \dots, x_n) \in [0, 1]^n; |x|_1 = x_1 + \dots + x_n = 1\}.$$

By “exchangeable” we mean that if  $Z \sim \eta_n$  then for every permutation  $\pi$  of  $\{1, \dots, n\}$  the random vector  $(Z_{\pi_1}, \dots, Z_{\pi_n})$  follows also the law  $\eta_n$ . This gives

$$0 = \text{Var}(1) = \text{Var}(Z_1 + \dots + Z_n) = n\text{Var}(Z_1) + n(n-1)\text{Cov}(Z_1, Z_2)$$

and therefore

$$\text{Cov}(Z_1, Z_2) = -\frac{1}{n-1}\text{Var}(Z_1) \leq 0.$$

We believe that an averaged form of theorems 1.1-1.2-1.3-1.4 remain valid at least if  $M$  is a real  $n \times n$  random matrix with i.i.d. rows such that for every  $1 \leq i, j \neq j' \leq n$

$$\mathbb{E}(M_{i,j}) = \frac{1}{n} \quad \text{and} \quad 0 < \text{Var}(M_{i,j}) = O(n^{-2}) \quad \text{and} \quad |\text{Cov}(M_{i,j}, M_{i,j'})| = O(n^{-3}).$$

These rates in  $n$  correspond to the Dirichlet Markov Ensemble for which  $\mathcal{L}$  is exponential and  $\eta_n$  is Dirichlet  $\mathcal{D}_n(1, \dots, 1)$ . Another interesting problem is the spectral analysis of  $M$  or even of  $X$  when the law  $\mathcal{L}$  has heavy tails (e.g. the law of  $W^{-\beta}$  with  $2\beta > 1$  and

$W$  uniform on  $[0, 1]$ ), or the case where  $X$  has independent entries, equally distributed on the diagonal and outside the diagonal, and with heavy tails on the diagonal.

The rest of the present article is divided as follows. The proofs of theorems 1.1-1.2-1.3-1.4 are given in the next sections (one section for each theorem). We give two proofs of theorem 1.4. In the Appendix, we gather for convenience some tools from Potential Theory used in our proof of theorem 1.3. The article ends with numerical simulations (figures 1 and 2) performed with GNU Octave 3.0.1 on the Debian GNU/Linux personal desktop computer of the author. The spectra were computed with the Octave function `eig` which is based on LAPACK numerical algorithms for the Schur decomposition (Hessenberg decomposition, Householder reflections, QR iteration, see [24]).

## 2 Proof of theorem 1.1

For a complex  $n \times n$  matrix  $A$ , the second moment of  $\frac{1}{n} \sum_{i=1}^n \delta_{s_i(A)}$  is given by

$$\frac{1}{n} (s_1(A)^2 + \dots + s_n(A)^2) = \frac{1}{n} \sum_{i,j=1}^n |A_{i,j}|^2 = \frac{1}{n} \|A\|_{2 \rightarrow 2}^2.$$

In particular, the second moment  $\rho_n = \frac{1}{n} \sum_{i=1}^n s_i(\sqrt{n}M)^2$  of  $\frac{1}{n} \sum_{i=1}^n \delta_{s_i(\sqrt{n}M)}$  satisfies

$$\rho_n = \sum_{i=1}^n \frac{X_{i,1}^2 + \dots + X_{i,n}^2}{(X_{i,1} + \dots + X_{i,n})^2}.$$

Since  $\mathcal{L}$  has finite second moment, the strong Uniform Law of Large Numbers [7, Lemma 2] together with the standard strong Law of Large Numbers give that almost surely,

$$\rho_n \leq \frac{1}{n^2(m + o(1))^2} \sum_{i,j=1}^n X_{i,j}^2 = (1 + \kappa^2) + o(1) = O(1). \quad (8)$$

It follows by Markov's inequality that almost surely, the sequence  $(\frac{1}{n} \sum_{i=1}^n \delta_{s_i(\sqrt{n}M)})$  is tight. Theorem 1.1 goes beyond tightness and provides weak convergence.

*Proof of theorem 1.1.* One may assume that  $m = 1$  by a simple scaling. We follow the method used by Aubrun in [2]. The Courant-Fischer variational formulas for the singular values imply (see [29, Theorem 3.3.16 page 178]) that for every  $n \times n$  complex matrix  $A$ , every complex diagonal matrix  $D$ , and every  $1 \leq i \leq n$ ,

$$s_n(D)s_i(A) \leq s_i(DA) \leq s_1(D)s_i(A) \quad (9)$$

Since  $\sqrt{n}M = nDn^{-1/2}X$ , we get for every  $1 \leq i \leq n$

$$s_n(nD)s_i(n^{-1/2}X) \leq s_i(\sqrt{n}M) \leq s_1(nD)s_i(n^{-1/2}X). \quad (10)$$



Since  $\mathcal{L}$  has finite second moment, the Uniform Law of Large Numbers taken from [7, Lemma 2] gives that almost surely,

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} |nD_{i,i} - 1| = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} |n^{-1}D_{i,i}^{-1} - 1| = 0. \quad (11)$$

In particular, this gives that almost surely,

$$s_1(nD) = \max_{1 \leq i \leq n} |nD_{i,i}| = 1 + o(1) \quad \text{and} \quad s_n(nD) = \min_{1 \leq i \leq n} |nD_{i,i}| = 1 + o(1). \quad (12)$$

Let us define the column  $n$ -vector  $u$  and the  $n \times n$  matrix  $S$  by

$$u = (1, \dots, 1)^\top \quad \text{and} \quad S = u^\top u.$$

The random matrix  $n^{-1/2}(X - S)$  has i.i.d. centered entries with variance  $\sigma^2$  and finite fourth moment, and consequently, by the main result of [7], we have, almost surely,

$$s_1(n^{-1/2}(X - S)) = 2\sigma + o(1).$$

Now, since  $n^{-1/2}S$  has rank one we have, by a result of Thompson [48],

$$s_2(n^{-1/2}X) \leq s_1(n^{-1/2}(X - S))$$

and therefore, almost surely,

$$s_2(n^{-1/2}X) \leq 2\sigma + o(1). \quad (13)$$

By combining (10) with (12) and (13) we get that almost surely, for every  $1 \leq i \leq n$ ,

$$\max_{2 \leq i \leq n} |s_i(\sqrt{n}M) - s_i(n^{-1/2}X)| = o(1).$$

In particular, this gives from (13) that almost surely,

$$s_2(\sqrt{n}M) \leq 2\sigma + o(1). \quad (14)$$

Another consequence, by lemma 4.1, is that almost surely, the empirical distributions

$$\frac{1}{n} \sum_{i=2}^n \delta_{s_i(\sqrt{n}M)} \quad \text{and} \quad \frac{1}{n} \sum_{i=2}^n \delta_{s_i(n^{-1/2}X)}$$

have the same weak limit. Now, the contribution of  $s_1$  in

$$\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{s_i(n^{-1/2}X)}$$

is negligible (its weight is  $1/n$ ). Moreover, since  $X$  has i.i.d. entries with variance  $\kappa = \sigma$ , almost surely, the sequence  $(\nu_n)$  converges to  $\mathcal{Q}_\sigma$  by (1). This shows that almost surely,

$$\frac{1}{n} \sum_{i=1}^n \delta_{s_i(\sqrt{n}M)}$$

converges weakly to  $\mathcal{Q}_\sigma$ , which is the first result of theorem 1.1. Since  $\mathcal{Q}_\sigma$  is supported by  $[0, 2\sigma]$ , we get by using (14) that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} s_2(\sqrt{n}M) = 2\sigma\right) = 1$$

which is nothing else but the second result of theorem 1.1. It remains to show now that  $\mathbb{P}(\lim_{n \rightarrow \infty} s_1(M) = 1) = 1$ . Almost surely, we have

$$\begin{aligned} s_1(\sqrt{n}M) &\leq s_1(nD)s_1(n^{-1/2}X) \\ &\leq s_1(nD)(s_1(n^{-1/2}(X - S) + n^{-1/2}S)) \\ &\leq s_1(nD)(s_1(n^{-1/2}(X - S)) + s_1(n^{-1/2}S)) \\ &= (1 + o(1))(2\sigma + o(1) + n^{1/2}) \end{aligned}$$

which gives

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} s_1(M) \leq 1\right) = 1. \quad (15)$$

Let us show now that  $\mathbb{P}(\liminf_{n \rightarrow \infty} s_1(M) \geq 1) = 1$ . It is known [29, Theorem 3.3.16 page 178] that for any  $n \times n$  complex matrices  $A$  and  $B$  and every  $1 \leq i \leq n$ , we have

$$|s_i(A) - s_i(B)| \leq s_1(A - B). \quad (16)$$

By using this bound, we get, almost surely,

$$|s_1(X) - s_1(S)| \leq s_1(X - S) = (2\sigma + o(1))\sqrt{n}.$$

Since  $s_1(S) = n$  we obtain that  $s_1(X) \geq n - (2\sigma + o(1))\sqrt{n}$  almost surely. By combining this result with (10) and (12) we get that almost surely,

$$s_1(M) \geq s_1(nD)\frac{1}{n}s_1(X) \geq (1 + o(1))(1 - o(1)) = 1 + o(1)$$

and therefore  $\mathbb{P}(\liminf_{n \rightarrow \infty} s_1(M) \geq 1) = 1$ , which achieves the proof.  $\square$

### 3 Proof of theorem 1.2

For every  $n \times n$  complex matrix  $A$  and for every  $1 \leq i \leq n$ ,

$$|\lambda_1(A)|^2 + \cdots + |\lambda_i(A)|^2 \leq s_1(A)^2 + \cdots + s_i(A)^2,$$

see [29, Theorem 3.3.13 pages 175–176]. In particular, the second moment of

$$\frac{1}{n} \sum_{i=1}^n \delta_{|\lambda_i(A)|}$$

is majorized by the second moment  $\rho_n$  of

$$\frac{1}{n} \sum_{i=1}^n \delta_{s_i(A)}.$$

This shows via Markov's inequality and (8) that almost surely,  $(\mu_{\sqrt{n}M})$  is tight. Theorem 1.2 goes beyond tightness and provides strong support localization.

*Proof of theorem 1.2.* Since  $M$  is Markov, it is known that  $\lambda_1(M) = 1$ . Let us briefly recall the proof for the sake of completeness. If  $u = (1, \dots, 1)^\top$  then  $Mu = u$  and thus 1 belongs to the spectrum of  $M$ . Next, let  $\lambda \in \mathbb{C}$  be an eigenvalue of  $M$  and let  $x \in \mathbb{C}^n \setminus \{0\}$  be such that  $Mx = \lambda x$ . There exists  $i \in \{1, \dots, n\}$  such that  $|x_i| = \max\{|x_1|, \dots, |x_n|\}$ . Since  $|x_i| \neq 0$  and

$$|\lambda| |x_i| \leq \sum_{j=1}^n M_{i,j} |x_j| \leq |x_i| \sum_{j=1}^n M_{i,j} = |x_i|,$$

we get  $|\lambda| \leq 1$ . This shows that  $\lambda_1(\sqrt{n}M) = \sqrt{n}$ . It remains to establish that

$$\mathbb{P}(\limsup_{n \rightarrow \infty} |\lambda_2(\sqrt{n}M)| \leq 2\kappa) = 1.$$

For every square  $n \times n$  complex matrix  $A$  and every  $1 \leq i \leq n$ ,

$$|\lambda_1(A)| \cdots |\lambda_i(A)| \leq s_1(A) \cdots s_i(A),$$

see the paper [54] by Hermann Weyl. In particular, we have the useful bound

$$|\lambda_2(A)| \leq \frac{s_1(A)s_2(A)}{|\lambda_1(A)|}.$$

With  $A = \sqrt{n}M$  and by using theorem 1.1 and  $\lambda_1(M) = 1$ , we obtain, almost surely,

$$\limsup_{n \rightarrow \infty} |\lambda_2(\sqrt{n}M)| \leq \limsup_{n \rightarrow \infty} s_2(\sqrt{n}M) \limsup_{n \rightarrow \infty} s_1(M) = 2\kappa.$$

□

We used  $s_2(\sqrt{n}M)$  to bound  $\lambda_2(\sqrt{n}M)$  and this explains the presence of a factor 2 in the result. It is tempting to try to remove this factor 2 by using a *Wedderburn rank one reduction* [53, p. 69]. Namely by the Schur unitary triangularization theorem [28, Theorem 2.3.1 page 79], for every  $n \times n$  complex matrix  $A$ , there exists a unitary matrix  $U$  and an upper triangular matrix  $T$  with diagonal  $\lambda_1(A), \dots, \lambda_n(A)$  such that

$$A = UTU^*. \tag{17}$$

Moreover, if  $Av = \lambda_1 v$  for some column vector  $v$  in the unit sphere of  $\mathbb{C}^n$  then one can choose  $U$  such that the first column of  $U$  is  $v$ . It follows that

$$A - \lambda_1(A)v^*v = U(T - \text{diag}(0, \lambda_2(A), \dots, \lambda_n(A)))U^*.$$

As a consequence, the matrix  $A - \lambda_1(A)v^*v$  has spectrum

$$\{0, \lambda_2(A), \dots, \lambda_n(A)\}.$$

This allows to control  $|\lambda_2(A)|$  by the spectral radius of a rank one perturbation since

$$|\lambda_2(A)| = |\lambda_1(A - \lambda_1 v^* v)|.$$

If  $A = M$  is Markov then  $\lambda_1(M) = 1$ , and with  $v = n^{-1/2}(1, \dots, 1)$  we get

$$|\lambda_2(M)| = |\lambda_1(M - n^{-1}S)|$$

where  $S = (1, \dots, 1)(1, \dots, 1)^\top$  is the  $n \times n$  matrix full of ones (it is symmetric, with rank 1 and norm  $n$ ). If  $M = DX$  is our random Markov matrix, we have, with  $m = 1$ , by arguing as in the proof of theorem 1.1,

$$\sqrt{n}M - n^{-1/2}S = (nD - I)n^{-1/2}X + n^{-1/2}(X - S).$$

It turns out that this approach via a rank one perturbation is successful for random reversible Markov kernels [10]. We ignore if it helps in our case. Note that since  $\mathcal{L}$  has finite fourth moment, the spectral radius of  $n^{-1/2}(X - S)$  tends to  $\kappa$  almost surely.

## 4 Proof of theorem 1.3

If  $\mathcal{L}$  is diffuse (no atoms) then  $MM^* \neq M^*M$  almost surely which means that  $M$  is not normal with probability one. The spectrum of non normal matrices is a delicate object very sensitive to perturbations, even of finite rank. Non normal matrices include non diagonalizable matrices such as nilpotent matrices. Let us recall a striking example taken from [44] (see also [4, p. 292]). Consider the  $n \times n$  matrices  $A$  and  $B$  defined by

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \varepsilon_n & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where  $\varepsilon_n = n^{-\alpha}$  with  $\alpha > 0$ . We have  $\text{rank}(A - B) = 1$  and  $\|A - B\|_{2 \rightarrow 2} = \varepsilon_n$ . However, the empirical spectral distribution of  $A$  is  $\delta_0$  while the empirical spectral distribution of  $B$  is supported on the centered circle of radius  $\varepsilon_n^{1/n}$  of the complex plane. Moreover,  $\varepsilon_n^{1/n} \rightarrow 1$  and the limiting spectral distribution of  $B$  is the uniform distribution on the unit circle of the complex plane! In numerical analysis, the sensitivity of the spectrum is captured by the notion of *pseudo-spectrum* [49]. For any any  $\varepsilon > 0$  and any fixed matrix norm  $\|\cdot\|$ , the  $\varepsilon$ -pseudo-spectrum  $\Lambda_\varepsilon(A)$  of  $A$  is given by

$$\Lambda_\varepsilon(A) = \bigcup_{\|A - B\| \leq \varepsilon} \{\lambda_1(B), \dots, \lambda_n(B)\} = \{z \in \mathbb{C} : s_n(A - zI) \leq \varepsilon\}.$$

If  $A$  is a normal matrix, its pseudo-spectrum for the operator norm  $\|\cdot\|_{2 \rightarrow 2}$  is just the Euclidean  $\varepsilon$ -neighborhood of its spectrum. The pseudo-spectrum can be larger for non-normal matrices, due to the possible big difference between singular values and eigenvalues. The continuity of the spectrum with respect to the matrix entries implies that  $\Lambda_\varepsilon(A) \rightarrow \Lambda(A)$  as  $\varepsilon \rightarrow 0$ , for any fixed  $n \times n$  matrix  $A$ . However, this says nothing on the behavior of  $\Lambda_{\varepsilon_n}(A)$  as  $n \rightarrow \infty$  where  $A$  is the  $n \times n$  truncation of an infinite array.

Let us recall some basic useful and enlightening facts before we enter the proof of theorem 1.3. If  $A$  is a complex  $n \times n$  random matrix then for every integer  $r \geq 0$ ,

$$n \int_{\mathbb{C}} z^r d\mu_A(x, y) = \sum_{i=1}^n \lambda_i^r(A) = \text{Tr}(A^r) = \sum_{\substack{1 \leq i_1, \dots, i_r \leq n \\ i_1 = i_r}} A_{i_1, i_2} \cdots A_{i_{r-1}, i_r} \quad (18)$$

where  $z = x + \sqrt{-1}y$ . If  $A$  is random then (18) allows to express the  $r^{\text{th}}$  moment of the mean empirical spectral distribution  $\mathbb{E}\mu_A$  of  $A$  in terms of a sum of moments of the entries of  $A$  leading to combinatorics of paths. Suppose for instance that we want to show that  $\mathbb{E}\mu_A$  converges weakly to some compactly supported law  $\mu$  on the real line. By the Weierstrass theorem,  $\mu$  is fully characterized by its sequence of moments. Moreover, if  $\mu_A$  is also supported on the real line then the convergence of its moments to the moments of  $\mu$  implies weak convergence, and actually convergence in any Wasserstein transportation distance [50]. This method of moments is a very efficient tool for the spectral analysis of large dimensional Hermitian random matrices because their empirical spectral distribution is supported by the real line. It allows to establish the Universal Wigner Semicircular Law theorem and the Universal Marchenko-Pastur Quartercircular Law theorem via some paths combinatorics. The case where the entries of the matrix have unbounded moments is typically addressed by using truncation techniques, and the strong convergence without the expectation (i.e. almost sure weak convergence) can be obtained via the control of the speed of convergence of moments (see for instance [27] and [4] for more details). 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  $\mathcal{U}_\kappa$  on the centered disc  $\{z \in \mathbb{C}; |z| \leq \kappa\}$  of radius  $\kappa$  then for every  $r \geq 0$ ,

$$\mathbb{E}(Z^r) = 0 \quad (19)$$

and thus  $\mathcal{U}_\kappa$  is not characterized by its moments. Any rotational invariant law on  $\mathbb{C}$  with light tails shares with  $\mathcal{U}_\kappa$  the same sequence of null moments. One can try to circumvent the problem by using “mixed moments” which uniquely determine  $\mu$  by the Stone-Weierstrass theorem. Namely, if  $A = UTU^*$  is the Schur unitary triangularization of  $A$  as in (17) then for every integers  $r, r' \geq 0$  and with  $z = x + \sqrt{-1}y$ ,

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

Equality holds true when  $\overline{T} = T^*$ , i.e. when  $A$  is normal. This explains why the method of moments loses its strength for non normal matrices. Another tool commonly used in Random Matrix Theory is the Cauchy-Hilbert-Stieltjes transform  $S_\mu : \mathbb{C}^+ \rightarrow \mathbb{C}$  of a law  $\mu$  on  $\mathbb{R}$ , defined for every  $z \in \mathbb{C}^+ = \{z \in \mathbb{C} : \Im(z) > 0\}$  by

$$S_\mu(z) = \int_{\mathbb{C}} \frac{1}{z' - z} d\mu(z').$$

The transform  $S_\mu$  determines uniquely  $\mu$  and moreover, the pointwise convergence on  $\mathbb{C}^+$  of this transform along a sequence of laws is equivalent to the weak convergence of the sequence. When  $\mu$  is supported in a compact subset of  $\mathbb{R}$ , then  $S_\mu$  encodes, at a neighborhood of infinity, the moments of  $\mu$  via the expansion (for any  $r \geq 0$ )

$$S_\mu(z) = \sum_{k=0}^r \frac{m_k}{z^{k+1}} + o\left(z^{-(r+1)}\right) \quad \text{where} \quad m_k = \int_{\mathbb{R}} x^k d\mu(x).$$

See [27] and [4] for an account on the properties and usage of  $S_\mu$ . For a complex  $n \times n$  matrix  $A$ , the quantity  $S_{\mu_A}(z)$  makes sense at any point  $z \notin \{\lambda_1(A), \dots, \lambda_n(A)\}$  and is linked to the *resolvent*  $(A - zI)^{-1}$  of  $A$  since

$$S_{\mu_A}(z) = \frac{1}{n} \sum_{k=1}^n \frac{1}{\lambda_k(A) - z} = \frac{1}{n} \text{Tr}((A - zI)^{-1}). \quad (20)$$

In particular,  $|S_{\mu_A}(z)| \leq |\Im(z)|^{-1}$  for every  $z \in \mathbb{C}^+$  when  $A$  is Hermitian. Unfortunately,  $S_\mu$  is not necessarily bounded when  $\mu$  is a law on  $\mathbb{C}$  and this makes problematic its usage for the characterization of weak convergence for the empirical spectral distribution of non normal matrices. However, since  $S_{\mu_A}$  is analytic on  $\mathbb{C}$  except on a finite set of poles  $\{\lambda_1(A), \dots, \lambda_n(A)\}$ , the function  $\Re(S_{\mu_A})$  determines the spectrum of  $A$ .

We have seen that the method of moments and the Cauchy-Hilbert-Stieltjes transform, which are so successful for the spectral analysis of large dimensional Hermitian random matrices, are not efficient for non normal random matrices with complex spectrum. This explains in part why Circular Law theorems are more difficult to establish than Semicircular Law theorems (Wigner) or Quartercircular Law theorems (Marchenko-Pastur). A successful hermitization approach due to Girko [20] (see also the work of Bai [3]) consists in replacing the study of the empirical spectral distribution of a non normal square random matrix  $A$  by the study of the empirical spectral distribution of the family of Hermitian matrices  $(H_z)_{z \in \mathbb{C}}$  where

$$H_z = \sqrt{(A - zI)(A - zI)^*}.$$

The eigenvalues of  $H_z$  are the singular values of  $A - zI$ . The technique that we use to establish theorem 1.3 is based on this hermitization and on logarithmic potentials, as in [35], [25], [13], [46], [47]. For convenience, we gather in the Appendix some tools used in our proof of theorem 1.3. The *logarithmic potential*  $U_{\mu_A} : \mathbb{C} \rightarrow (-\infty, +\infty]$  of the

empirical spectral distribution  $\mu_A$  of  $A$  is given for every  $z \in \mathbb{C}$  by

$$\begin{aligned}
U_{\mu_A}(z) &= -\frac{1}{n} \sum_{i=1}^n \log |\lambda_i(A) - z| \\
&= -\frac{1}{n} \log |\det(A - zI)| \\
&= -\frac{1}{n} \log \det(H_z) \\
&= -\frac{1}{n} \sum_{i=1}^n \log s_i(A - zI) \\
&= U_{\mu_{H_z}}(0).
\end{aligned}$$

Roughly speaking, the logarithmic potential characterizes the law and for a sequence of laws supported in a common compact set, the convergence of their logarithmic potentials for every  $z$  is equivalent to weak convergence (see the Appendix). The logarithmic potential is related to the Cauchy-Hilbert-Stieltjes transform via the identity

$$\partial_{\Re(z)} U_{\mu_A}(z) = \Re(S_{\mu_A}(z)).$$

The main problem when using logarithmic potentials for the convergence of empirical spectral distributions is their singularity at 0 and  $\infty$  in the right hand side of the identity

$$U_{\mu_A}(z) = - \int_0^\infty \log(t) d\mu_{H_z}(t).$$

The singularity at  $\infty$  is not so serious and can be circumvented by using upper bounds on the norm  $s_1(A - zI)$  of  $A - zI$  via the moments of the entries of  $A$ . The singularity at point 0 is more serious, and leads to the control of the least singular value  $s_n(A - zI)$  of  $A - zI$ . In the works [35], [25], [46, 47], [13], the regularization of the logarithmic potential at point 0 is obtained by using lower bound on  $s_n(A - zI)$  similar to (5). Our proof of theorem 1.3 is perturbative, and does not involve these delicate singular aspects.

Free Probability Theory (tracial von Neumann operators algebras) provides a natural algebraic language for large (infinite!) dimensional random matrices [51]. The notions of *operator*, *trace*, and *freeness* are the analogue of the notions of *matrix*, *expectation*, and *independence* respectively. In this theory, the analogue of the empirical spectral distribution of a non normal operator is the so called *Brown spectral measure* [11] while the analogue of its logarithmic potential is up to a sign the *Fuglede-Kadison determinant* [17] (via the identity  $\text{tr} \log(aa^*) = \log(\det(aa^*))$ ). Roughly speaking, the Circular Law is the Brown spectral measure of Voiculescu's Circular operators [27]. If  $W_1$  and  $W_2$  are two free Semicircular operators then the sum  $W_1 + \sqrt{-1}W_2$  is Circular. If  $U$  and  $R$  are two free operators such that  $U$  is Haar unitary and  $H$  is Hermitian Quartercircular then the product  $UH$  is Circular (think about the polar factorization of matrices). This last construction is related to the notion of  $R$ -diagonal operators/pairs considered by Nica & Speicher [34] and Haagerup & Larsen [26, Example 5.1 page 358]. In [44], Śniady shows

that the convergence of “\*-moments” (i.e. trace of words with letters  $A$  and  $A^*$ ) implies some sort of weak Circular Law theorem by using least singular value regularization and the Fuglede-Kadison determinant. Several examples of Brown spectral measures are given by Biane and Leherer in [8].

*Proof of theorem 1.3.* One can assume that  $m = 1$  by a simple scaling. By theorem 1.2, the sequence  $(\mu_{\sqrt{n}M})$  is almost surely tight, and therefore it suffices to show that almost surely, the law  $\mathcal{U}_\sigma$  is the unique adherence value of  $(\mu_{\sqrt{n}M})$ . For notational convenience, we will show that almost surely, if  $(\mu_{\sqrt{n}M})$  converges to a compactly supported law  $\mu$  then  $\mu = \mathcal{U}_\sigma$ . Let  $(\Omega, \mathbb{B}, \mathbb{P})$  be the probability space associated to the array  $(X_{i,j})_{i,j \geq 1}$ . By (11), we have, on an event  $\mathcal{A} \in \mathbb{B}$  of probability one,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\det(nD)| = 0. \quad (21)$$

Let us fix  $\omega \in \mathcal{A}$  and let us assume that  $(\mu_{\sqrt{n}M})$  converges to a compactly supported law  $\mu$  (which may depend on  $\omega$  a priori). We have to show that  $\mu = \mathcal{U}_\sigma$ . By theorems 1.2 and A.2, there exists  $C \subset \mathbb{C}$  with zero capacity such that for every  $z \notin C$ ,

$$\liminf_{n \rightarrow \infty} U_{\mu_{\sqrt{n}M}}(z) = U_\mu(z). \quad (22)$$

We have  $U_{\mu_{\sqrt{n}M}}(z) < \infty$  iff  $z \notin \{\lambda_1(\sqrt{n}M), \dots, \lambda_n(\sqrt{n}M)\}$  and if we write

$$\sqrt{n}M - zI = nDY \quad \text{with} \quad Y = n^{-1/2}X - zn^{-1}D^{-1}$$

then thanks to (21), for large enough  $n$  and every  $z \notin \{\lambda_1(\sqrt{n}M), \dots, \lambda_n(\sqrt{n}M)\}$

$$\begin{aligned} U_{\mu_{\sqrt{n}M}}(z) &= -\frac{1}{n} \log |\det(\sqrt{n}M - zI)| \\ &= -\frac{1}{n} \log |\det(nD)| + U_{\mu_Y}(z). \end{aligned}$$

Therefore, thanks to (21) and (22), for every  $z \notin C$  such that  $U_\mu(z) < \infty$ ,

$$U_\mu(z) = \liminf_{n \rightarrow \infty} U_{\mu_Y}(z). \quad (23)$$

The strong Uniform Law of Large Numbers (11) suggests that  $Y$  is close to  $n^{-1/2}X - zI$  when  $n$  is large. Let us make precise this intuitive fact. Since  $X_{1,1} - 1$  is centered with finite positive variance and finite fourth moment, we get from [35], for every  $z \in \mathbb{C}$ ,

$$\nu_{n,z} = \frac{1}{n} \sum_{i=1}^n \delta_{s_i(n^{-1/2}(X-S)-zI)} \xrightarrow[n \rightarrow \infty]{w} \nu_z$$

where  $S = \mathbb{E}(X) = (1, \dots, 1)(1, \dots, 1)^\top$  and where  $\nu_z$  is a law on  $\mathbb{C}$  such that

$$U_{\nu_z}(0) = U_{\mathcal{U}_\sigma}(z). \quad (24)$$



From [48] (see also [4, Lemma 2.6 p. 621]), for every  $n \times n$  complex matrices  $A$  and  $B$ ,

$$\|F_{\sqrt{AA^*}} - F_{\sqrt{BB^*}}\|_\infty \leq \frac{1}{n} \text{rank}(A - B)$$

where  $F_{\sqrt{AA^*}}$  and  $F_{\sqrt{BB^*}}$  are the cumulative distribution functions of the empirical spectral distributions of  $\sqrt{AA^*}$  and  $\sqrt{BB^*}$ . Since  $\text{rank}(S) = 1$ , we get, for every  $z \in \mathbb{C}$ ,

$$\bar{\nu}_{n,z} = \frac{1}{n} \sum_{i=1}^n \delta_{s_i(n^{-1/2}X - zI)} \xrightarrow[n \rightarrow \infty]{w} \nu_z.$$

By using (16) together with (11) and lemma 4.1 we deduce that for every  $z \in \mathbb{C}$ ,

$$\tilde{\nu}_{n,z} = \frac{1}{n} \sum_{i=1}^n \delta_{s_i(Y)} \xrightarrow[n \rightarrow \infty]{w} \nu_z.$$

Moreover, we have  $s_1(n^{-1/2}X) \geq |\lambda_1(n^{-1/2}X)|$  which blows up to  $+\infty$  at speed  $\sqrt{n}$  while  $s_2(n^{-1/2}X) = O(1)$  (see [42] or [13] and (13)). As a consequence, by using (16) and (11), we see that  $s_1(Y) \rightarrow +\infty$  at speed  $\sqrt{n}$  while  $s_2(Y)$  remains bounded. Also, the assumptions of theorem A.2 are satisfied by  $\tilde{\nu}_{n,z} \rightarrow \nu_z$  for every  $z \in \mathbb{C}$ . Now, for every  $z \in \mathbb{C}$ , theorem A.2 applied at point 0 gives,

$$\liminf_{n \rightarrow \infty} U_{\tilde{\nu}_{n,z}}(0) \geq U_{\nu_z}(0).$$

By using the identities  $U_{\tilde{\nu}_{n,z}}(0) = U_{\mu_Y}(0)$  and (24) we thus have

$$\liminf_{n \rightarrow \infty} U_{\mu_Y}(0) \geq U_{\mathcal{U}_\sigma}(z).$$

From (23), we obtain that  $U_\mu(z) \geq U_{\mathcal{U}_\sigma}(z)$  for every  $z \notin C$  such that  $U_\mu(z) < \infty$ , and this holds true obviously for every  $z \notin C$ . **The gap is here: we have only shown that  $U_\mu(z) \geq U_{\mathcal{U}_\sigma}(z)$  for every  $z \in C$ , and thus for almost every  $z \in \mathbb{C}$ . It remains to show that  $U_\mu(z) \leq U_{\mathcal{U}_\sigma}(z)$  for almost every  $z \in \mathbb{C}$ . If so, we will then get  $\mu = \mathcal{U}_\sigma$  by theorem A.5.**  $\square$

**Lemma 4.1 (Weak convergence of atomic laws under uniform perturbation).**

*For every couple of triangular arrays of complex numbers*

$$(\lambda_{n,i})_{1 \leq i \leq n < \infty} \quad \text{and} \quad (\tilde{\lambda}_{n,i})_{1 \leq i \leq n < \infty}$$

*and every law  $\mu$  on  $\mathbb{C}$ ,*

$$\text{if } \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_{n,i}} \xrightarrow[n \rightarrow \infty]{w} \mu \quad \text{and} \quad \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} |\lambda_{n,i} - \tilde{\lambda}_{n,i}| = 0 \quad \text{then} \quad \frac{1}{n} \sum_{i=1}^n \delta_{\tilde{\lambda}_{n,i}} \xrightarrow[n \rightarrow \infty]{w} \mu.$$

*Proof.* Let us fix  $\varepsilon > 0$ . Since  $\mu$  is a law, there exists a compact subset  $K \subset \mathbb{C}$  such that  $\mu(\mathbb{C} \setminus K) \leq \varepsilon$ , and a disc  $D = \{z \in \mathbb{C}; |z| \leq r\}$  such that  $K \subset D$  and

$$\kappa = \text{dist}(K, \mathbb{C} \setminus D) > 0.$$

Let  $f : \mathbb{C} \rightarrow \mathbb{R}$  be a bounded continuous function. By Heine's theorem, there exists  $\eta > 0$  such that  $|f(z) - f(z')| \leq \varepsilon$  for every  $z, z' \in D$  with  $|z - z'| \leq \eta$ . On the other hand, there exists an integer  $n_\eta$  such that if  $n \geq n_\eta$  then

$$\max_{1 \leq i \leq n} |\lambda_{n,i} - \tilde{\lambda}_{n,i}| \leq \max(\eta, \kappa).$$

Consequently, if  $\lambda_{n,i} \in K$  and  $n \geq n_\eta$  then  $\lambda_{n,i}$  and  $\tilde{\lambda}_{n,i}$  belong to  $D$  and therefore

$$|f(\lambda_{n,i}) - f(\tilde{\lambda}_{n,i})| \leq \varepsilon.$$

Thus, for  $n \geq n_\eta$  and by denoting  $I_n = \{1 \leq i \leq n; \lambda_{n,i} \notin D\}$ , we have

$$\frac{1}{n} \sum_{i=1}^n |f(\lambda_{n,i}) - f(\tilde{\lambda}_{n,i})| \leq 2\|f\|_\infty \frac{1}{n} \text{card}(I_n) + \varepsilon.$$

It remains to control  $\text{card}(I_n)$  by using the tightness of  $\mu$ . Namely, by considering a continuous function  $g_D : \mathbb{C} \rightarrow [0, 1]$  such that  $g_D \equiv 0$  on  $K$  and  $g_D \equiv 1$  on  $\mathbb{C} \setminus D$ , we get

$$\frac{1}{n} \text{card}(I_n) \leq \frac{1}{n} \sum_{i=1}^n g_D(\lambda_{n,i}) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{C}} g_D d\mu(z) \leq \mu(\mathbb{C} \setminus K) \leq \varepsilon.$$

□

## 5 Proofs of theorem 1.4

We give two similar but different proofs of theorem 1.4. The first one, based on lemma 5.1, was suggested to the author by Manjunath Krishnapur. The second one is based on (9). Both proofs are consequences of (5).

*First proof of theorem 1.4.* Recall that  $M = DX$  and set  $\tilde{X} = M$  for convenience. For any real  $\delta > 0$  and integer  $n > 0$ , let us define the event  $\mathcal{A}_{n,\delta}$  by

$$\mathcal{A}_{n,\delta} = \bigcap_{i=1}^n \{|X_{i,1} + \cdots + X_{i,n} - nm| \leq n\delta\} = \left\{ \max_{1 \leq i \leq n} \left| \frac{X_{i,1} + \cdots + X_{i,n}}{n} - m \right| \leq \delta \right\}.$$

Let  $R_1, \dots, R_n$  and  $\tilde{R}_1, \dots, \tilde{R}_n$  be the rows of  $X$  and  $\tilde{X}$  respectively. By using lemma 5.1 for the rows of  $X$  and  $\tilde{X}$ , we get on the event  $\mathcal{A}_{n,\delta}$  and for every  $t \geq 0$

$$\begin{aligned} \{s_n(\tilde{X}) \leq t\} &\leq \cup_{1 \leq i \leq n} \{\text{dist}(\tilde{R}_i, \tilde{R}_{-i}) \leq n^{1/2}t\} \\ &\leq \cup_{1 \leq i \leq n} \{\text{dist}(R_i, R_{-i}) \leq n^{3/2}t(m + \delta)\} \\ &\leq \{s_n(X) \leq n^{3/2}t(m + \delta)\}. \end{aligned}$$

Now the Bienaymé-Chebyshev inequality gives

$$\mathbb{P}(\mathcal{A}_{n,\delta}^c) \leq 1 - (1 - \sigma^2 \delta^{-2})^n.$$

For some prescribed  $a > 0$  one may select  $c > 0$  and  $\delta = \delta_n = n^c$  such that

$$\mathbb{P}(\mathcal{A}_{n,\delta_n}^c) = O(n^{-a}) \tag{25}$$

for large enough  $n$ . On the other hand, for any prescribed  $b > 0$  one may select

$$t = t_n = n^{-b-3/2}(m + \delta_n)^{-1}$$

which gives for some  $\varepsilon > 0$ , on the event  $\mathcal{A}_{n,\delta_n}$  and for large enough  $n$ ,

$$\tilde{\mathcal{S}}_n := \{s_n(\tilde{X}) \leq n^{-b-3/2-c-\varepsilon}\} \subset \{s_n(\tilde{X}) \leq t_n\} \subset \{s_n(X) \leq n^{-b}\} =: \mathcal{S}_n.$$

Now by selecting  $b$  with (5) we finally obtain

$$\mathbb{P}(\tilde{\mathcal{S}}_n) = \mathbb{P}(\mathcal{A}_{n,\delta_n} \cap \tilde{\mathcal{S}}_n) + \mathbb{P}(\mathcal{A}_{n,\delta_n}^c \cap \tilde{\mathcal{S}}_n) \leq \mathbb{P}(\mathcal{S}_n) + \mathbb{P}(\mathcal{A}_{n,\delta_n}^c) = O(n^{-a})$$

which is exactly the desired result with  $b' = b + 3/2 + c + \varepsilon$ . □

*Alternative proof of theorem 1.4.* Let  $\tilde{X}$  and  $\mathcal{A}_{\delta,n}$  as above. From (9), we have

$$s_n(D)s_n(X) \leq s_n(\tilde{X}).$$

Note that

$$s_n(D)^{-1} = \max_{1 \leq i \leq n} |X_{i,1} + \cdots + X_{i,n}|.$$

As before, for some  $\delta = \delta_n = n^c$  we have (25), and on the event  $\mathcal{A}_{n,\delta_n}$  we have

$$\{s_n(\tilde{X}) \leq t\} \subset \{s_n(X)s_n(D) \leq t\} \subset \{s_n(X) \leq t(nm + n^{1+c})\}.$$

For any prescribed  $b > 0$  one may select  $b' > 0$  and  $t = t_n = t^{-b'}$  such that

$$t_n(nm + n^{1+c}) \leq n^{-b}$$

for large enough  $n$ . In particular, on the event  $\mathcal{A}_{n,\delta_n}$  and for large enough  $n$ ,

$$\tilde{\mathcal{S}}_n := \{s_n(\tilde{X}) \leq n^{-b'}\} \subset \{s_n(X) \leq n^{-b}\} =: \mathcal{S}_n.$$

Now, if  $b$  is as in (5), one can write

$$\mathbb{P}(\tilde{\mathcal{S}}_n) = \mathbb{P}(\tilde{\mathcal{S}}_n \cap \mathcal{A}_{n,\delta_n}) + \mathbb{P}(\tilde{\mathcal{S}}_n \cap \mathcal{A}_{n,\delta_n}^c) \leq \mathbb{P}(\mathcal{S}_n) + \mathbb{P}(\mathcal{A}_{n,\delta_n}^c) = O(n^{-a}).$$

□

**Lemma 5.1 (Least singular value under diagonal perturbation).** For a square  $n \times n$  real matrix  $A$  with columns  $C_1, \dots, C_n$  and rows  $R_1, \dots, R_n$ , we define for every  $1 \leq i \leq n$

$$C_{-i} = \text{span}(\{C_j; j \neq i\}) \quad \text{and} \quad R_{-i} = \text{span}(\{R_j; j \neq i\}).$$

Then for every  $t \geq 0$ ,

$$\cup_{1 \leq i \leq n} \{\text{dist}(C_i, C_{-i}) \leq t\} \subset \{s_n(A) \leq t\} \subset \cup_{1 \leq i \leq n} \{\text{dist}(C_i, C_{-i}) \leq n^{1/2}t\}$$

and

$$\cup_{1 \leq i \leq n} \{\text{dist}(R_i, R_{-i}) \leq t\} \subset \{s_n(A) \leq t\} \subset \cup_{1 \leq i \leq n} \{\text{dist}(R_i, R_{-i}) \leq n^{1/2}t\}.$$

*Proof.* Since  $A$  is a square real matrix, we have  $s_n(A) = s_n(A^*)$  and it suffices for instance to prove the statement for the columns of  $A$ . For every  $x \in \mathbb{R}^n$  and  $1 \leq i \leq n$ , by using the triangle inequality and the identity  $Ax = x_1C_1 + \dots + x_nC_n$  we get

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

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

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

Conversely, for every  $1 \leq i \leq n$ , there exists  $y \in \mathbb{R}^n$  with  $y_i = 1$  such that

$$\text{dist}(C_i, C_{-i}) = \|C_1 + y_2C_2 + \dots + y_nC_n\| = \|Ay\| \geq \|y\| \min_{\|x\|=1} \|Ax\| \geq s_n(A)$$

where we used the fact that  $\|y\|^2 = 1 + y_2^2 + \dots + y_n^2 \geq 1$ . □

Note that for a square  $n \times n$  complex matrix  $A$  with rows  $R_1, \dots, R_n$ , we also have

$$|\det(A)| = \prod_{i=1}^n |\lambda_i(A)| = \prod_{i=1}^n s_i(A) = \prod_{i=1}^n d_i(A)$$

where  $d_i = \text{dist}(R_i, \text{span}(R_1, \dots, R_{i-1}))$ . This identity is used for instance in [47].

## A Logarithmic Potential Tools

We give here for convenience some logarithmic potentials tools. The *logarithmic potential*  $U_\mu$  of a compactly supported law  $\mu$  on  $\mathbb{C}$  is the function  $U_\mu : \mathbb{C} \rightarrow (-\infty, +\infty]$  defined by

$$U_\mu(z) = - \int_{\mathbb{C}} \log |z - z'| d\mu(z')$$

for every  $z \in \mathbb{C}$ . The *logarithmic energy*  $\mathcal{E}(\mu) \in (-\infty, +\infty]$  of  $\mu$  is defined by

$$\mathcal{E}(\mu) = \int_{\mathbb{C}} U_\mu(z) d\mu(z) = - \int_{\mathbb{C}} \int_{\mathbb{C}} \log |z - z'| d\mu(z) d\mu(z').$$

In Free Probability Theory [51], the quantity  $-\mathcal{E}(\mu)$  is known as Voiculescu's *free entropy*, see for instance [27, Section 5.3]. The *energy*  $\mathcal{E}(K)$  of a compact subset  $K$  of  $\mathbb{C}$  is

$$\mathcal{E}(K) = \inf\{\mathcal{E}(\mu); \mu \text{ is a law supported in } K\}.$$

The *logarithmic capacity* of a compact subset  $K$  of  $\mathbb{C}$  is defined by

$$\text{Cap}(K) = e^{-\mathcal{E}(K)}.$$

The *capacity* of a Borel subset  $B$  of  $\mathbb{C}$  is defined by

$$\text{Cap}(B) = \sup\{\text{Cap}(K); K \text{ is a compact subset of } \mathbb{C} \text{ included in } B\}.$$

We have  $\text{Cap}(B) > 0$  if and only if  $B$  carries a law of finite logarithmic energy. If  $\text{Cap}(B) = 0$  then  $B$  has zero Lebesgue measure but the converse is false (e.g. compact sets with positive Hausdorff dimension [12]). A property is said to hold *quasi-everywhere* if the set where it does not hold has zero capacity. Also, "quasi-everywhere" implies "almost everywhere".

**Theorem A.1.** *Let  $(\nu_n)$  be a sequence of laws on  $\mathbb{C}$  with support in a common compact subset of  $\mathbb{C}$ ,  $(a_n)$  be a sequence of complex numbers such that  $|a_n| \rightarrow +\infty$ , and  $(p_n)$  be a sequence of real numbers in  $[0, 1]$  such that  $p_n \rightarrow 0$ . Define  $\mu_n = (1 - p_n)\nu_n + p_n\delta_{a_n}$ . If  $p_n \log |a_n| \rightarrow 0$  and  $\liminf_{n \rightarrow \infty} U_{\mu_n} = U_\mu$  quasi-everywhere on  $\mathbb{C}$  for some compactly supported law  $\mu$  on  $\mathbb{C}$ , then  $\mu_n \rightarrow \mu$  and  $\nu_n \rightarrow \mu$  weakly as  $n \rightarrow \infty$ .*

*Proof.* For every fixed  $z \in \mathbb{C}$  and large enough  $n$ ,

$$U_{\mu_n}(z) = -p_n \log |z - a_n| + (1 - p_n)U_{\nu_n}(z).$$

By hypothesis, we have for every fixed  $z \in \mathbb{C}$  and large enough  $n$ ,

$$p_n \log |z - a_n| = p_n \log |a_n| + p_n \log \left| \frac{z}{a_n} - 1 \right| \rightarrow 0. \quad (26)$$

Consequently,  $\liminf_{n \rightarrow \infty} U_{\nu_n} = \liminf_{n \rightarrow \infty} U_{\mu_n} = U_\mu$  quasi-everywhere on  $\mathbb{C}$ . Since the sequence  $(\nu_n)$  is supported in a common compact subset of  $\mathbb{C}$ , it is tight and thus weakly relatively compact (a standard result due to Prohorov states that the set of laws on a compact space equipped with the weak topology is metrizable and compact [9]). Let  $\mu_{n_k} \rightarrow \eta$  be a weakly converging subsequence, to some law  $\eta$  (necessarily supported in the same compact subset). By the Lower Envelope theorem A.4,  $\liminf_{n \rightarrow \infty} U_{\nu_n} = U_\eta$  quasi-everywhere in  $\mathbb{C}$ . In particular,  $U_\mu = U_\eta$  quasi-everywhere in  $\mathbb{C}$ , and thus almost everywhere for the Lebesgue measure. Now, by the Unicity theorem A.5, we get  $\eta = \mu$ . Therefore,  $\mu$  is the unique weak adherence value of  $(\nu_n)$ . It follows then that  $\nu_n \rightarrow \mu$  weakly. Since  $p_n \rightarrow 0$ , we get also that  $\mu_n = (1 - p_n)\nu_n + p_n\delta_{a_n} \rightarrow \mu$  weakly.  $\square$

**Theorem A.2.** *Let  $(\nu_n)$  be a sequence of laws on  $\mathbb{C}$  with support in a common compact subset of  $\mathbb{C}$ ,  $(a_n)$  be a sequence of complex numbers such that  $|a_n| \rightarrow +\infty$ , and  $(p_n)$  be a sequence of real numbers in  $[0, 1]$  such that  $p_n \rightarrow 0$ . If  $(\mu_n)$  (equivalently  $(\nu_n)$ ) converges weakly to some law  $\mu$  then  $U_\mu(z) \leq \liminf_{n \rightarrow \infty} U_{\mu_n}(z)$  for every  $z \in \mathbb{C}$ , and equality is achieved for quasi-every  $z$ .*

*Proof.* The desired results follow from the Principle of Descent (theorem A.3) and the Lower Envelope theorem (theorem A.4) applied for  $\nu_n \xrightarrow{w} \mu$  together with (26).  $\square$

Theorems A.3-A.4-A.5 below are extracted from the book [41] by Saff & Totik.

**Theorem A.3 (Principle of Descent).** *If  $(\mu_n)$  is a sequence of laws all having support in a fixed compact set of  $\mathbb{C}$  and converging weakly to some law  $\mu$  then for every  $z \in \mathbb{C}$ ,*

$$U_\mu(z) \leq \liminf_{n \rightarrow \infty} U_{\mu_n}(z). \quad (27)$$

**Theorem A.4 (Lower Envelope).** *Equality is achieved in (27) for quasi-every  $z$ .*

**Theorem A.5 (Unicity).** *If  $\mu$  and  $\nu$  are two compactly supported laws on  $\mathbb{C}$  and if their logarithmic potential  $U_\mu$  and  $U_\nu$  coincide almost everywhere with respect to the Lebesgue measure on  $\mathbb{R}^2$ , then  $\mu = \nu$ .*

Note that the uniform law  $\mathcal{U}_1$  on the disc  $\{z \in \mathbb{C}; |z| \leq 1\}$  has finite logarithmic energy and logarithmic potential given for every  $z \in \mathbb{C}$  by (see [35] for instance)

$$U(z) = \begin{cases} -\log |z| & \text{if } |z| > 1 \\ \frac{1}{2}(1 - |z|^2) & \text{if } |z| \leq 1. \end{cases}$$

## References

- [1] T. W. Anderson, *An introduction to multivariate statistical analysis*, third ed., Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2003. 1
- [2] G. Aubrun, *Random points in the unit ball of  $l_p^n$* , Positivity **10** (2006), no. 4, 755–759. 7
- [3] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), no. 1, 494–529. 2, 13
- [4] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Mathematics Monograph Series 2, Science Press, Beijing, 2006. 2, 6, 11, 12, 13, 16
- [5] 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, 166–168. 2
- [6] 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, 555–569. 2, 6
- [7] ———, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Ann. Probab. **21** (1993), no. 3, 1275–1294. 2, 7, 8
- [8] Ph. Biane and F. Lehner, *Computation of some examples of Brown’s spectral measure in free probability*, Colloq. Math. **90** (2001), no. 2, 181–211. 15
- [9] P. Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication. 20

- [10] Ch. Bordenave, P. Caputo, and Chafaï, *Spectral Analysis of Large Reversible Markov Kernels*, work in progress, 2008. 5, 11
- [11] L. G. Brown, *Lidskiĭ's theorem in the type II case*, Geometric methods in operator algebras (Kyoto, 1983), Pitman Res. Notes Math. Ser., vol. 123, Longman Sci. Tech., Harlow, 1986, pp. 1–35. 14
- [12] L. Carleson, *Selected problems on exceptional sets*, Van Nostrand Mathematical Studies, No. 13, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1967. 20
- [13] D. Chafaï, *Circular law for non-central random matrices*, preprint [arXiv:0709.0036](https://arxiv.org/abs/0709.0036) [math.PR], 2007. 3, 6, 13, 14, 16
- [14] ———, *The Dirichlet Markov Ensemble*, preprint [arXiv:0709.4678](https://arxiv.org/abs/0709.4678) [math.PR], 2007. 3, 5
- [15] F. J. Dyson, *Selected papers of Freeman Dyson with commentary*, Collected Works, vol. 5, American Mathematical Society, Providence, RI, 1996, With a foreword by Elliott H. Lieb. 1
- [16] A. Edelman, *The probability that a random real Gaussian matrix has  $k$  real eigenvalues, related distributions, and the circular law*, J. Multivariate Anal. **60** (1997), no. 2, 203–232. 2
- [17] B. Fuglede and R. V. Kadison, *Determinant theory in finite factors*, Ann. of Math. (2) **55** (1952), 520–530. 14
- [18] S. Geman, *The spectral radius of large random matrices*, Ann. Probab. **14** (1986), no. 4, 1318–1328. 6
- [19] J. Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, J. Mathematical Phys. **6** (1965), 440–449. 2
- [20] V. L. Girko, *The circular law*, Teor. Veroyatnost. i Primenen. **29** (1984), no. 4, 669–679. 2, 13
- [21] ———, *The circular law. Twenty years later. III*, Random Oper. Stochastic Equations **13** (2005), no. 1, 53–109. 2
- [22] G. Goldberg and M. Neumann, *Distribution of subdominant eigenvalues of matrices with random rows*, SIAM J. Matrix Anal. Appl. **24** (2003), no. 3, 747–761 (electronic). 6
- [23] G. Goldberg, P. Okunev, M. Neumann, and H. Schneider, *Distribution of subdominant eigenvalues of random matrices*, Methodol. Comput. Appl. Probab. **2** (2000), no. 2, 137–151. 6
- [24] G. H. Golub and Ch. F. Van Loan, *Matrix computations*, third ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996. 7
- [25] F. Götze and A. Tikhomirov, *The Circular Law for Random Matrices*, preprint available at [arXiv:math/07093995](https://arxiv.org/abs/math/07093995) [math.PR], 2007. 2, 13, 14
- [26] 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, 331–367. 14
- [27] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy*, Mathematical Surveys and Monographs, vol. 77, American Mathematical Society, Providence, RI, 2000. 12, 13, 14, 20

- [28] R. A. Horn and Ch. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1990, Corrected reprint of the 1985 original. 10
- [29] ———, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1994, Corrected reprint of the 1991 original. 7, 9
- [30] C.-R. Hwang, *A brief survey on the spectral radius and the spectral distribution of large random matrices with i.i.d. entries*, Random matrices and their applications (Brunswick, Maine, 1984), Contemp. Math., vol. 50, Amer. Math. Soc., Providence, RI, 1986, pp. 145–152. 6
- [31] E. Kostlan, *On the spectra of Gaussian matrices*, Linear Algebra Appl. **162/164** (1992), 385–388, Directions in matrix theory (Auburn, AL, 1990). 6
- [32] V.A. Marchenko and L.A. Pastur, *The distribution of eigenvalues in certain sets of random matrices*, Mat. Sb. **72** (1967), 507–536. 2
- [33] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Academic Press, New York, 1967. 2
- [34] A. Nica and R. Speicher, *R-diagonal pairs—a common approach to Haar unitaries and circular elements*, Free probability theory (Waterloo, ON, 1995), Fields Inst. Commun., vol. 12, Amer. Math. Soc., Providence, RI, 1997, pp. 149–188. 14
- [35] G. Pan and W. Zhou, *Circular law, extreme singular values and potential theory*, arXiv:0705.3773 [math.PR], 2007. 2, 13, 14, 15, 21
- [36] B. Rider, *A limit theorem at the edge of a non-Hermitian random matrix ensemble*, J. Phys. A **36** (2003), no. 12, 3401–3409, Random matrix theory. 6
- [37] ———, *Order statistics and Ginibre’s ensembles*, J. Statist. Phys. **114** (2004), no. 3-4, 1139–1148. 6
- [38] B. Rider and B. Virág, *The noise in the circular law and the Gaussian free field*, Int. Math. Res. Not. IMRN (2007), no. 2, Art. ID rnm006, 33. 6
- [39] M. Rudelson and R. Vershynin, *The least singular value of a random square matrix is  $o(n^{-1/2})$* , preprint arXiv:0805.3407 [math.PR], 2008. 3
- [40] ———, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633. 3
- [41] 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. 21
- [42] J. W. Silverstein, *The spectral radii and norms of large-dimensional non-central random matrices*, Comm. Statist. Stochastic Models **10** (1994), no. 3, 525–532. 3, 6, 16
- [43] J.W. Silverstein and B. Rider, *Gaussian fluctuations for non-Hermitian random matrix ensembles*, Ann. Probab. (2006), no. 34, 2118–2143. 6
- [44] P. Śniady, *Random regularization of Brown spectral measure*, J. Funct. Anal. **193** (2002), no. 2, 291–313. 11, 14
- [45] T. Tao and V. Vu, *Random matrices: A general approach for the least singular value problem*, preprint arXiv:0805.3167 [math.PR], 2008. 3



- [46] ———, *Random matrices: the circular law*, Commun. Contemp. Math. **10** (2008), no. 2, 261–307. 2, 3, 13, 14
- [47] ———, *Random matrices: Universality of ESDs and the circular law*, preprint [arXiv:0807.4898 \[math.PR\]](https://arxiv.org/abs/0807.4898), 2008. 2, 3, 6, 13, 14, 19, 25
- [48] R. C. Thompson, *The behavior of eigenvalues and singular values under perturbations of restricted rank*, Linear Algebra and Appl. **13** (1976), no. 1/2, 69–78, Collection of articles dedicated to Olga Taussky Todd. 8, 16
- [49] L. N. Trefethen and M. Embree, *Spectra and pseudospectra*, Princeton University Press, Princeton, NJ, 2005, The behavior of nonnormal matrices and operators. 3, 11
- [50] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. 12
- [51] 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. 14, 20
- [52] K. W. Wachter, *The strong limits of random matrix spectra for sample matrices of independent elements*, Ann. Probability **6** (1978), no. 1, 1–18. 2
- [53] J. H. M. Wedderburn, *Lectures on matrices*, Dover Publications Inc., New York, 1964. 10
- [54] H. Weyl, *Inequalities between the two kinds of eigenvalues of a linear transformation*, Proc. Nat. Acad. Sci. U. S. A. **35** (1949), 408–411. 10
- [55] E. P. Wigner, *The collected works of Eugene Paul Wigner. Part A. The scientific papers. Vol. II. Nuclear physics*, Springer-Verlag, Berlin, 1996, Annotated by Herman Feshbach, Edited and with a preface by Arthur S. Wightman and Jagdish Mehra. 1
- [56] Y. Q. Yin, *Limiting spectral distribution for a class of random matrices*, J. Multivariate Anal. **20** (1986), no. 1, 50–68. 2

**Acknowledgements.** The first proof of theorem 1.4 was kindly suggested to the author by Manjunath Krishnapur. The author would also thank Terence Tao for pointing out a gap in the proof of Theorem 1.3 page 16.

Djalil CHAFAÏ <mailto:chafai@math.univ-toulouse.fr>  
 UMR181 ENVT INRA & INSTITUT DE MATHÉMATIQUES DE TOULOUSE UMR5219 CNRS  
 UNIVERSITÉ DE TOULOUSE, FRANCE.

Compiled August 11, 2008

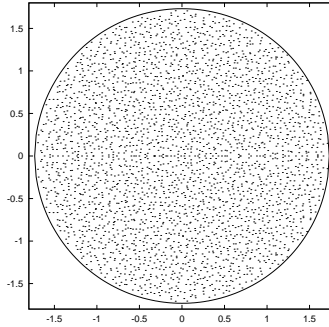


Figure 1: Plot of the circle of radius  $\kappa$  and the set of points  $\{\lambda_2(\sqrt{n}M), \dots, \lambda_n(\sqrt{n}M)\}$  for a single realization of  $\sqrt{n}M$  with  $n = 2000$  and  $X_{1,1} = W^{-1/\alpha} - 1$  where  $W$  is uniformly distributed on  $[0, 1]$  and  $\alpha = 3.001$ . Here  $M = DX$  as in (6-7) and the law  $\mathcal{L}$  of the i.i.d. entries of  $X$  has Lebesgue density  $t \mapsto \alpha(1+t)^{-(1+\alpha)}$  on  $[0, \infty)$ . It has finite positive variance with  $\kappa \approx 1.73$ , finite third moment, but infinite fourth moment. This simulation illustrates theorems 1.2 and 1.3. It suggests that the moment assumption can be weakened and also that  $2\kappa$  might be replaced by  $\kappa$  in theorem 1.2.

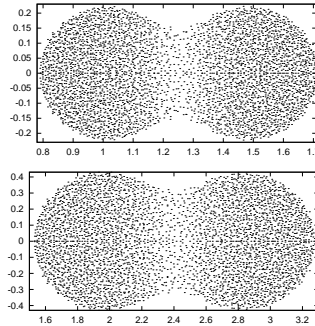


Figure 2: We consider here a random matrix shifted by a deterministic matrix as in [47]. The upper plot is the set of points  $\{\lambda_2(n^{-1/2}X + S_n), \dots, \lambda_n(n^{-1/2}X + S_n)\}$  with  $n = 4000$  where  $X_{1,1}$  is uniformly distributed on  $[0, 1]$  and  $S_n = \text{diag}(1, \dots, 1, 1.5, \dots, 1.5)$  where 1 appears  $\lfloor n/2 \rfloor$  times. The limiting spectral distribution of the deterministic shift matrix  $S_n$  is the Bernoulli law  $\frac{1}{2}(\delta_1 + \delta_{1.5})$ . The lower picture is the plot of the set of points  $\{\lambda_2(\sqrt{n}M), \dots, \lambda_n(\sqrt{n}M)\}$  where the Markov matrix  $M$  is built from  $n^{-1/2}X + S_n$  by dividing each row by its  $\ell^1$ -norm. This simulation suggests to explore the universality approach of [47] for our Markov matrices with i.i.d. rows.