Probab. Theory Relat. Fields (2012) 152:751–779
DOI 10.1007/s00440-010-0336-1

Circular law theorem for random Markov matrices

Charles Bordenave · Pietro Caputo · Djalil Chafaï

Received: 12 May 2010 / Revised: 3 December 2010 / Published online: 12 January 2011
© Springer-Verlag 2010

Abstract Let $(X_{jk})_{jk\geq 1}$ be i.i.d. nonnegative random variables with bounded density, mean m, and finite positive variance σ^2. Let M be the $n \times n$ random Markov matrix with i.i.d. rows defined by $M_{jk} = X_{jk}/(X_{j1}+\cdots+X_{jn})$. In particular, when X_{11} follows an exponential law, the random matrix M belongs to the Dirichlet Markov Ensemble of random stochastic matrices. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\sqrt{n}M$ i.e. the roots in $\mathbb{C}$ of its characteristic polynomial. Our main result states that with probability one, the counting probability measure $\frac{1}{n}\delta_{\lambda_1} + \cdots + \frac{1}{n}\delta_{\lambda_n}$ converges weakly as $n \to \infty$ to the uniform law on the disk $\{z \in \mathbb{C} : |z| \leq m^{-1}\sigma\}$. The bounded density assumption is purely technical and comes from the way we control the operator norm of the resolvent.

Keywords Random matrices · Eigenvalues · Spectrum · Stochastic matrices · Markov chains

Mathematics Subject Classification (2000) 15A52

Ch. Bordenave
Institut de Mathématiques de Toulouse, UMR 5219 CNRS, Université de Toulouse,
118 route de Narbonne, 31062 Toulouse, France
e-mail: charles.bordenave@math.univ-toulouse.fr
URL: http://www.math.univ-toulouse.fr/~bordenave/

P. Caputo
Dipartimento di Matematica, Università Roma Tre, Largo San Murialdo 1, 00146 Rome, Italy
e-mail: caputo@mat.uniroma3.it
URL: http://www.mat.uniroma3.it/users/caputo/

D. Chafaï (✉)
UMR 8050 CNRS, Université Paris-Est Marne-la-Vallée, 5 boulevard Descartes,
77454 Champs-sur-Marne, France
e-mail: djalil@chafai.net
URL: http://djalil.chafai.net/

1 Introduction

The *eigenvalues* of an $n \times n$ complex matrix A are the roots in $\mathbb{C}$ of its characteristic polynomial. We label them $\lambda_1(A), \ldots, \lambda_n(A)$ so that $|\lambda_1(A)| \geq \cdots \geq |\lambda_n(A)|$ with growing phases. The *spectral radius* is $|\lambda_1(A)|$. We also denote by $s_1(A) \geq \cdots \geq s_n(A)$ the *singular values* of A, defined for all $1 \leq k \leq n$ by $s_k(A) := \lambda_k(\sqrt{AA^*})$ where $A^* = \bar{A}^\top$ is the conjugate-transpose. The matrix A maps the unit sphere to an ellipsoid, the half-lengths of its principal axes being the singular values of A. The *operator 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 matrix A is singular iff $s_n(A) = 0$, and if not then $s_n(A) = s_1(A^{-1})^{-1} = \|A^{-1}\|_{2\to 2}^{-1}$. If A is normal (i.e. $A^*A = A^*A$) then $s_i(A) = |\lambda_i(A)|$ for every $1 \leq i \leq n$. Beyond normal matrices, the relationships between the eigenvalues and the singular values are captured by the Weyl inequalities (see Lemma B.6). Let us define the discrete probability measures

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

From now on, we denote "$\xrightarrow{\mathscr{C}_b}$" the weak convergence of probability measures with respect to bounded continuous functions. We use the abbreviations *a.s.*, *a.a.*, and *a.e.* for *almost surely*, *Lebesgue almost all*, and *Lebesgue almost everywhere* respectively. The notation $n \gg 1$ means *large enough* n. Let $(X_{i,j})_{i,j\geq 1}$ be an infinite table of i.i.d. complex random variables with finite positive variance $0 < \sigma^2 < \infty$. If one defines the square $n \times n$ complex random matrix $X := (X_{i,j})_{1\leq i,j\leq n}$ then the quartercircular law theorem (universal square version of the Marchenko-Pastur theorem, see [31,45,47]) states that a.s.

$$\nu_{\frac{1}{\sqrt{n}}X} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mathcal{Q}_\sigma \tag{1.1}$$

where $\mathcal{Q}_\sigma$ is the quartercircular law on the real interval $[0, 2\sigma]$ with Lebesgue density

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

Additionally, it is shown in [5,6,8] that

$$\lim_{n\to\infty} s_1\left(\tfrac{1}{\sqrt{n}}X\right) = 2\sigma \text{ a.s. iff } \mathbb{E}(X_{1,1}) = 0 \quad \text{and} \quad \mathbb{E}(|X_{1,1}|^4) < \infty. \tag{1.2}$$

Concerning the eigenvalues, the famous Girko circular law theorem states that a.s.

$$\mu_{\frac{1}{\sqrt{n}}X} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mathcal{U}_\sigma \tag{1.3}$$

where $\mathcal{U}_\sigma$ is the uniform law on the disc $\{z \in \mathbb{C} : |z| \leq \sigma\}$, known as the circular law. This statement was established through a long sequence of partial results [4,5,19,21,23,24,26,29,32,33,40,41], the general case (1.3) being finally obtained by Tao and Vu [41]. From (1.3) we have a.s. $\underline{\lim}_{n\to\infty} |\lambda_k(n^{-1/2}X)| \geq \sigma$ for any fixed $k \geq 1$, and we get from [7,33] and (1.2) that if additionally $\mathbb{E}(X_{1,1}) = 0$ and $\mathbb{E}(|X_{1,1}|^4) < \infty$ then a.s.

$$\lim_{n\to\infty} |\lambda_1(\tfrac{1}{\sqrt{n}}X)| = \sigma \quad \text{and} \quad \lim_{n\to\infty} \frac{s_1(X)}{|\lambda_1(X)|} = 2. \tag{1.4}$$

The behavior of the ratio operator-norm/spectral-radius suggests that X is far from being an asymptotically normal matrix. Following [16,37], if $\mathbb{E}(X_{1,1}) \neq 0$ while $\mathbb{E}(|X_{1,1}|^4) < \infty$ then a.s. $|\lambda_1(n^{-1/2}X)| \to +\infty$ at speed $\sqrt{n}$ while $|\lambda_2(n^{-1/2}X)|$ remains bounded.

The proof of (1.3) is partly but crucially based on a polynomial lower bound on the smallest singular value proved in [40]: for every $a, d > 0$, there exists $b > 0$ such that for any deterministic complex $n \times n$ matrix A with $s_1(A) \leq n^d$ we have

$$\mathbb{P}(s_n(X + A) \leq n^{-b}) \leq n^{-a}. \tag{1.5}$$

In particular, by the first Borel–Cantelli lemma, there exists $b > 0$ which may depend on d such that a.s. $X + A$ is invertible with $s_n(X + A) \geq n^{-b}$ for $n \gg 1$.

1.1 Random Markov matrices and main results

From now on and unless otherwise stated $(X_{i,j})_{i,j\geq 1}$ is an infinite array of nonnegative real random variables with mean $m := \mathbb{E}(X_{1,1}) > 0$ and finite positive variance $\sigma^2 := \mathbb{E}(X_{1,1}^2) - m^2$. Let us define the event

$$\mathcal{D}_n := \{\rho_{n,1} \cdots \rho_{n,n} > 0\} \quad \text{where } \rho_{n,i} := X_{i,1} + \cdots + X_{i,n}.$$

Since $\sigma > 0$ we get $q := \mathbb{P}(X_{1,1} = 0) < 1$ and thus

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

By the first Borel–Cantelli lemma, a.s. for $n \gg 1$, one can define the $n \times n$ matrix M by

$$M_{i,j} := \frac{X_{i,j}}{\rho_{n,i}}.$$

The matrix M is Markov since its entries belong to $[0, 1]$ and each row sums 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

$$D_{i,i} := \frac{1}{\rho_{n,i}}.$$

We may define M and D for all $n \geq 1$ by setting, when $\rho_{n,i} = 0$, $M_{i,j} = \delta_{i,j}$ for all $1 \leq j \leq n$ and $D_{i,i} = 1$. 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 simplex

$$\Lambda_n := \{(p_1, \ldots, p_n) \in [0, 1]^n : p_1 + \cdots + p_n = 1\}.$$

From now on, we set $m = 1$. This is actually no loss of generality since the law of the random matrix M is invariant under the linear scaling $t \to t\, X_{i,j}$ for any $t > 0$. Since $\sigma < \infty$, the uniform law of large numbers of Bai and Yin [8, lem. 2] states that a.s.

$$\max_{1\leq i\leq n} \left|\rho_{n,i} - n\right| = o(n). \tag{1.6}$$

This suggests that $\sqrt{n}M$ is approximately equal to $n^{-1/2}X$ for $n \gg 1$. One can then expect that (1.1) and (1.3) hold for $\sqrt{n}M$. Our work shows that this heuristics is valid. There is however a complexity gap between (1.1) and (1.3), due to the fact that for nonnormal operators such as M, the eigenvalues are less stable than the singular values under perturbations, see e.g. the book [43]. Our first result below constitutes the analog of the universal Marchenko–Pastur theorem (1.1) for $\sqrt{n}M$, and generalizes the result of the same kind obtained in [17] in the case where $X_{1,1}$ follows an exponential law.

Theorem 1.1 (Quartercircular law theorem) *We have a.s.*

$$\nu_{\sqrt{n}M} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mathcal{Q}_\sigma.$$

Our second result provides some estimates on the largest singular values and eigenvalues.

Theorem 1.2 (Extremes) *We have $\lambda_1(M) = 1$. Moreover, if $\mathbb{E}(|X_{1,1}|^4) < \infty$ then a.s.*

$$\lim_{n\to\infty} s_1(M) = 1 \quad \textit{and} \quad \lim_{n\to\infty} s_2(\sqrt{n}M) = 2\sigma \quad \textit{while} \quad \overline{\lim_{n\to\infty}}\, |\lambda_2(\sqrt{n}M)| \leq 2\sigma.$$

Our third result below is the analogue of (1.3) for our random Markov matrices. When $X_{1,1}$ follows the exponential distribution of unit mean, Theorem 1.3 is exactly the circular law theorem for the Dirichlet Markov Ensemble conjectured in [15, 17]. Note that we provide, probably for the first time, an almost sure circular law theorem for a matrix model with dependent entries under a finite positive variance assumption.

Theorem 1.3 (Circular law theorem) *If $X_{1,1}$ has a bounded density then a.s.*

$$\mu_{\sqrt{n}M} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mathcal{U}_\sigma.$$

The proof of Theorem 1.3 is crucially based on the following estimate on the norm of the resolvent of $\sqrt{n}M$. It is the analogue of (1.5) for our random Markov matrices.

Theorem 1.4 (Smallest singular value) *If $X_{1,1}$ has a bounded density then for every $a, C > 0$ there exists $b > 0$ such that for any $z \in \mathbb{C}$ with $|z| \le C$, for $n \gg 1$,*

$$\mathbb{P}(s_n(\sqrt{n}M - zI) \le n^{-b}) \le n^{-a}.$$

In particular, for some $b > 0$ which may depend on C, a.s. for $n \gg 1$, the matrix $\sqrt{n}M - zI$ is invertible with $s_n(\sqrt{n}M - zI) \ge n^{-b}$.

The proofs of Theorems 1.1–1.4 are given in Sects. 2–5 respectively. These proofs make heavy use of lemmas given in the Appendices A–C.

The matrix M is the Markov kernel associated to the weighted oriented complete graph with n vertices with one loop per vertex, for which each edge $i \to j$ has weight $X_{i,j}$. The skeleton of this kernel is an oriented Erdős–Rényi random graph where each edge exists independently of the others with probability $1-q$. If $q = 0$ then M has a complete skeleton, is aperiodic, and 1 is the sole eigenvalue of unit module [36]. The nonoriented version of this graphical construction gives rise to random reversible kernels for which a semicircular theorem is available [12]. The bounded density assumption forces $q = 0$.

Since M is Markov, we have that for every integer $r \ge 0$,

$$\int_{\mathbb{C}} z^r \, \mu_M(dz) = \frac{1}{n}\sum_{i=1}^n \lambda_i^r(M) = \frac{1}{n}\sum_{i=1}^n p_M(r,i) \tag{1.7}$$

where

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

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 μ_M of M. The random Markov matrix M is a *random environment*. By combining Theorem 1.2 with Theorem 1.3 and the identity (1.7), we get that for every fixed $r \ge 0$, a.s.

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

1.2 Discussion and open questions

The bounded density assumption in Theorem 1.3 is only due to the usage of Theorem 1.4 in the proof. We believe that Theorem 1.4 (and thus 1.3) is valid without this assumption, but this is outside the scope of the present work, see Remark 5.1 and Fig. 1. Our proof of Theorem 1.3 is inspired from the Tao and Vu proof of (1.3) based on Girko Hermitization, and allows actually to go beyond the circular law, see Remarks 4.1 and 4.2. Concerning the extremes, by Theorems 1.2 and 1.3, a.s.

$$\sigma \leq \varliminf_{n\to\infty} |\lambda_2(\sqrt{n}M)| \leq \varlimsup_{n\to\infty} |\lambda_2(\sqrt{n}M)| \leq 2\sigma.$$

Also, a.s. the "spectral gap" of M is a.s. of order $1 - O(n^{-1/2})$ (compare with the results of [25]). Note that in contrast with (1.4), we have from Theorems 1.1 and 1.2, a.s.

$$\lim_{n\to\infty} \frac{s_1(M)}{|\lambda_1(M)|} = 1.$$

Numerical simulations suggest that if $\mathbb{E}(|X_{1,1}|^4) < \infty$ then a.s.

$$\lim_{n\to\infty} |\lambda_2(\sqrt{n}M)| = \sigma \quad \text{and thus} \quad \lim_{n\to\infty} \frac{s_2(M)}{|\lambda_2(M)|} = 2.$$

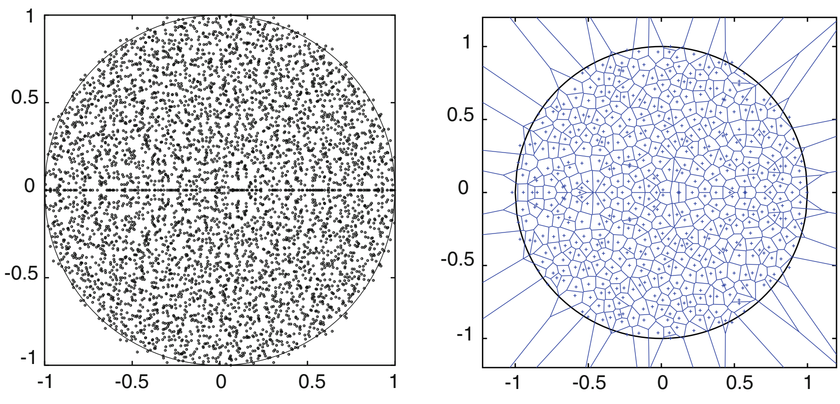

Fig. 1 Here we have fixed $n = 250$, and $X_{1,1}$ follows the Bernoulli law $\frac{1}{2}(\delta_0 + \delta_1)$. In both graphics, the solid circle has radius $m^{-1}\sigma = 1$. The left hand side graphic is the superposition of the plot of $\lambda_2, \ldots, \lambda_n$ for 10 i.i.d. simulations of $\sqrt{n}M$, made with the GNU Octave free software. The right hand side graphic is the Voronoï tessellation of $\lambda_2, \ldots, \lambda_n$ for a single simulation of $\sqrt{n}M$. Since $\sqrt{n}M$ has real entries, its spectrum is symmetric with respect to the real axis. On the left hand side graphic, it seems that the spectrum is slightly more concentrated on the real axis. This phenomenon, which disappears as $n \to \infty$, was already described for random matrices with i.i.d. real Gaussian entries by Edelman [19], see also the work of Akemann and Kanzieper [2]. Our simulations suggest that Theorem 1.3 remains valid beyond the bounded density assumption

Unfortunately, our proof of Theorem 1.2 is too perturbative to extract this result. Following [17, fig. 2], one can also ask if the phase $\mathrm{Phase}(\lambda_2(M)) = \lambda_2(M)/|\lambda_2(M)|$ converges in distribution to the uniform law on $[0, 2\pi]$ as $n \to \infty$. Another interesting problem concerns the behavior of $\max_{2\le k\le n} \mathfrak{Re}(\lambda_k(\sqrt{n}M))$ and the fluctuation as $n \to \infty$ of the extremal singular values and eigenvalues of $\sqrt{n}M$, in particular the fluctuation of $\lambda_2(\sqrt{n}M)$.

Classical results on the connectivity of Erdős–Rényi random graphs [10,20] imply that a.s. for $n \gg 1$ the Markov matrix M is irreducible. Hence, a.s. for $n \gg 1$, the Markov matrix M admits a unique invariant probability measure κ. If κ is seen as a row vector in Λ_n then we have $\kappa M = \kappa$. Let $\Upsilon := n^{-1}(\delta_1 + \cdots + \delta_n)$ be the uniform law on $\{1, \ldots, n\}$, which can be viewed as the vector $n^{-1}(1, \ldots, 1) \in \Lambda_n$. By denoting $\|\cdot\|_{\mathrm{TV}}$ the total variation (or ℓ^1) distance on Λ_n, one can ask if a.s.

$$\lim_{n\to\infty} \|\kappa - \Upsilon\|_{\mathrm{TV}} = 0.$$

Recall that the rows of M are i.i.d. and follow an exchangeable law η_n on the simplex Λ_n. By "exchangeable" we mean that if $Z \sim \eta_n$ then for every permutation π of $\{1, \ldots, n\}$ the random vector $(Z_{\pi(1)}, \ldots, Z_{\pi(n)})$ follows also the law η_n. This gives

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

and therefore $\mathrm{Cov}(Z_1, Z_2) = -(n-1)^{-1}\mathrm{Var}(Z_1) \le 0$. One can ask if the results of Theorems 1.1–1.4 remain essentially valid at least if M is a real $n \times n$ random matrix with i.i.d. rows such that for every $1 \le i, j \ne j' \le n$,

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

These rates in n correspond to the Dirichlet Markov Ensemble for which $X_{1,1}$ follows an exponential law and where η_n is the Dirichlet law $\mathcal{D}_n(1, \ldots, 1)$ on the simplex Λ_n. Another interesting problem is the spectral analysis of M when the law of $X_{1,1}$ has heavy tails, e.g. $X_{1,1} = V^{-\beta}$ with $2\beta > 1$ and where V is a uniform random variable on $[0, 1]$, see for instance [11] for the reversible case. A logical first step consists in the derivation of a heavy tailed version of (1.3) for X. This program is addressed in a separate paper [13]. In the same spirit, one may ask about the behavior of $\mu_{X\circ A}$ where "$\circ$" denotes the Schur–Hadamard entrywise product and where A is some prescribed profile matrix.

2 Proof of Theorem 1.1

Let us start by an elementary observation. The second moment ς_n of $\nu_{\sqrt{n}M}$ is given by

$$\varsigma_n := \int t^2\, \nu_{\sqrt{n}M}(dt) = \frac{1}{n}\sum_{i=1}^{n} s_i(\sqrt{n}M)^2 = \mathrm{Tr}(MM^*) = \sum_{i=1}^{n} \frac{X_{i,1}^2 + \cdots + X_{i,n}^2}{(X_{i,1} + \cdots + X_{i,n})^2}.$$

By using (1.6) together with the standard law of large numbers, we get that a.s.

$$\varsigma_n \leq \frac{1}{n^2(1+o(1))^2} \sum_{i,j=1}^{n} X_{i,j}^2 = (1+\sigma^2) + o(1) = O(1). \tag{2.1}$$

It follows by the Markov inequality that a.s. the sequence $(\nu_{\sqrt{n}M})_{n\geq 1}$ is tight. However, we will not rely on tightness and the Prohorov theorem in order to establish the convergence of $(\nu_{\sqrt{n}M})_{n\geq 1}$. We will use instead a perturbative argument based on the special structure of M and on (1.6). Namely, since $\sqrt{n}M = nDn^{-1/2}X$, we get from (B.4), for all $1 \leq i \leq n$,

$$s_n(nD)s_i(n^{-1/2}X) \leq s_i(\sqrt{n}M) \leq s_1(nD)s_i(n^{-1/2}X). \tag{2.2}$$

Additionally, we get from (1.6) that a.s.

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

This gives that a.s.

$$s_1(nD) = \max_{1\leq i\leq n} |nD_{i,i}| = 1 + o(1) \quad \text{and} \quad s_n(nD) = \min_{1\leq i\leq n} |nD_{i,i}| = 1 + o(1). \tag{2.3}$$

From (2.2), (2.3), and (1.5), we get that a.s. for $n \gg 1$,

$$s_n(\sqrt{n}M) > 0 \quad \text{and} \quad s_n(n^{-1/2}X) > 0 \tag{2.4}$$

and from (2.2) and (2.3) again we obtain that a.s.

$$\max_{1\leq i\leq n} \left| \log(s_i(\sqrt{n}M)) - \log(s_i(n^{-1/2}X)) \right| = o(1). \tag{2.5}$$

Now, from (1.1), and by denoting $\mathcal{L}_\sigma$ the image probability measure of $\mathcal{Q}_\sigma$ by $\log(\cdot)$, a.s.

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\log(s_i(n^{-1/2}X))} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mathcal{L}_\sigma.$$

Next, using (2.5) with Lemma C.1 provides that a.s.

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\log(s_i(\sqrt{n}M))} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mathcal{L}_\sigma.$$

This implies by the change of variable $t \mapsto e^t$ that a.s.

$$\nu_{\sqrt{n}M} = \frac{1}{n}\sum_{i=1}^{n} \delta_{s_i(\sqrt{n}M)} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mathcal{Q}_\sigma$$

which is the desired result.

Remark 2.1 (Alternative arguments) From (2.2) and (2.3), a.s. for $n \gg 1$, $s_k(M) = 0$ iff $s_k(X) = 0$, for all $1 \le k \le n$, and thus $\nu_{\sqrt{n}M}(\{0\}) = \nu_{n^{-1/2}X}(\{0\})$, and hence the reasoning can avoid the usage of (2.4). Note also that (2.4) is automatically satisfied when $X_{1,1}$ is absolutely continuous since the set of singular matrices has zero Lebesgue measure. On the other hand, it is also worthwhile to mention that if $\mathbb{E}(|X_{1,1}|^4) < \infty$ then one can obtain the desired result without using (2.4), by using Lemma C.1 together with (3.2), and this reasoning was already used by Aubrun for a slightly different model [3].

3 Proof of Theorem 1.2

Let us define the $n \times n$ deterministic matrix $S := \mathbb{E}(X) = (1, \dots, 1)^\top (1, \dots, 1)$. The random matrix $n^{-1/2}(X - S)$ has i.i.d. centered entries with finite positive variance σ^2 and finite fourth moment, and consequently, by (1.2), a.s.

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

Now, since $\operatorname{rank}(n^{-1/2}S) = 1$ we have, by Lemma B.5,

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

and therefore, a.s.

$$s_2(n^{-1/2}X) \le 2\sigma + o(1). \tag{3.1}$$

By combining (2.2) with (2.3) and (3.1) we get that a.s.

$$\max_{2\le i\le n} |s_i(\sqrt{n}M) - s_i(n^{-1/2}X)| = o(1). \tag{3.2}$$

In particular, this gives from (3.1) that a.s.

$$s_2(\sqrt{n}M) \le 2\sigma + o(1). \tag{3.3}$$

From Theorem 1.1, since $\mathcal{Q}_\sigma$ is supported by $[0, 2\sigma]$, we get by using (3.3) that a.s.

$$\lim_{n\to\infty} s_2(\sqrt{n}M) = 2\sigma. \tag{3.4}$$

Next, since M is a Markov matrix, it is easy to see and well known (see [36]) that

$$\lambda_1(M) = 1. \tag{3.5}$$

Let us show now that a.s.

$$\lim_{n\to\infty} s_1(M) = 1. \tag{3.6}$$

Let S be as in the proof of Theorem 1.1. From (B.1) and (1.2), we get a.s.

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

which gives $\overline{\lim}_{n\to\infty} s_1(M) \le 1$ a.s. On the other hand, from (B.6) and (3.5) we get $s_1(M) \ge |\lambda_1(M)| = 1$, which gives (3.6). It remains to establish that a.s.

$$\varlimsup_{n\to\infty} |\lambda_2(\sqrt{n}M)| \le 2\sigma.$$

Indeed, from (B.6) we get for every non null $n \times n$ complex matrix A,

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

With $A = \sqrt{n}M$ and by using (3.4–3.6), we obtain that a.s.

$$\varlimsup_{n\to\infty} |\lambda_2(\sqrt{n}M)| \le \varlimsup_{n\to\infty} s_2(\sqrt{n}M)\ \varlimsup_{n\to\infty} s_1(M) = 2\sigma.$$

4 Proof of Theorem 1.3

Let us start by observing that from the Weyl inequality (B.9) and (2.1), a.s.

$$\int_{\mathbb{C}} |z|^2\, \mu_{\sqrt{n}M}(dt) \le \int_0^\infty t^2\, \nu_{\sqrt{n}M}(dt) = O(1).$$

This shows via the Markov inequality that a.s. the sequence $(\mu_{\sqrt{n}M})_{n\ge1}$ is tight. However, we will not rely directly on this tightness and the Prohorov theorem in order to establish the convergence of $(\mu_{\sqrt{n}M})_{n\ge1}$. We will use instead the Girko Hermitization of Lemma A.2. We know, from the work of Dozier and Silverstein [18], that for all $z \in \mathbb{C}$, there exists a probability measure ν_z on $[0,\infty)$ such that a.s. $(\nu_{n^{-1/2}X-zI})_{n\ge1}$

converges weakly to ν_z. Moreover, following e.g. Pan and Zhou [33, lem. 3], one can check that for all $z \in \mathbb{C}$,

$$U_{\mathcal{U}_\sigma}(z) = -\int_0^\infty \log(t)\,\nu_z(dt)$$

where $U_{\mathcal{U}_\sigma}$ is the logarithmic potential of $U_{\mathcal{U}_\sigma}$ as in (A.2). To prove that a.s. $(\mu_{\sqrt{n}M})_{n\geq 1}$ tends weakly to $\mathcal{U}_\sigma$, we start from the decomposition

$$\sqrt{n}M - zI = nDW \quad \text{where } W := n^{-1/2}X - zn^{-1}D^{-1}.$$

By using (B.2), (1.6), and Lemma C.1, we obtain that for a.a. $z \in \mathbb{C}$, a.s.

$$\nu_W \xrightarrow[n\to\infty]{\mathscr{C}_b} \nu_z.$$

Now, arguing as in the proof of Theorem 1.1, it follows that for all $z \in \mathbb{C}$, a.s.

$$\nu_{\sqrt{n}M-zI} \xrightarrow[n\to\infty]{\mathscr{C}_b} \nu_z.$$

Suppose for the moment that for a.a. $z \in \mathbb{C}$, a.s. the function $\log(\cdot)$ is uniformly integrable for $(\nu_{\sqrt{n}M-zI})_{n\geq 1}$. Let $\mathcal{P}(\mathbb{C})$ be as in Appendix A. Lemma A.2 implies that there exists $\mu \in \mathcal{P}(\mathbb{C})$ such that a.s.

$$\mu_{\sqrt{n}M} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mu \quad \text{and} \quad U_\mu = U_{\mathcal{U}_\sigma} \text{ a.e.}$$

where U_μ is the logarithmic potential of μ as defined in Sect. A. Now by Lemma A.1, we obtain $\mu = \mathcal{U}_\sigma$, which is the desired result. It thus remains to show that for a.a. $z \in \mathbb{C}$, a.s. the function $\log(\cdot)$ is uniformly integrable for $(\nu_{\sqrt{n}M-zI})_{n\geq 1}$. For every $z \in \mathbb{C}$, a.s. by the Cauchy–Schwarz inequality, for all $t \geq 1$, and $n \gg 1$,

$$\left(\int_t^\infty \log(s)\,\nu_{\sqrt{n}M-zI}(ds)\right)^2 \leq \nu_{\sqrt{n}M-zI}([t,\infty)) \int_0^\infty s^2\,\nu_{\sqrt{n}M-zI}(ds).$$

Now the Markov inequality and (2.1) give that for all $z \in \mathbb{C}$, a.s. for all $t \geq 1$

$$\int_t^\infty \log(s)\,\nu_{\sqrt{n}M-zI}(ds) \leq \frac{O(1)}{t^2}$$

where the $O(1)$ is uniform in t. Consequently, for all $z \in \mathbb{C}$, a.s.

$$\lim_{t\to\infty} \varlimsup_{n\to\infty} \int_t^\infty \log(s)\, \nu_{\sqrt{n}M-zI}(ds) = 0.$$

This means that for all $z \in \mathbb{C}$, a.s. the function $\mathbb{1}_{[1,\infty)} \log(\cdot)$ is uniformly integrable for $(\nu_{\sqrt{n}M-zI})_{n\geq 1}$. It remains to show that for all $z \in \mathbb{C}$, a.s. the function $\mathbb{1}_{(0,1)} \log(\cdot)$ is uniformly integrable for $(\nu_{\sqrt{n}M-zI})_{n\geq 1}$. This is equivalent to show that for all $z \in \mathbb{C}$, a.s.

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} \int_0^\delta -\log(s)\, \nu_{\sqrt{n}M-zI}(ds) = 0.$$

For convenience, we fix $z \in \mathbb{C}$ and set $s_i := s_i(\sqrt{n}M - zI)$ for all $1 \leq i \leq n$. Now we write

$$\begin{aligned}
-\int_0^\delta \log(t)\, \nu_{\sqrt{n}M-zI}(dt) &= \frac{1}{n}\sum_{i=0}^{\lfloor 2n^{0.99}\rfloor} \mathbb{1}_{(0,\delta)}(s_{n-i}) \log(s_{n-i}^{-1}) \\
&\quad + \frac{1}{n}\sum_{i=\lfloor 2n^{0.99}\rfloor+1}^{n-1} \mathbb{1}_{(0,\delta)}(s_{n-i}) \log(s_{n-i}^{-1}) \\
&\leq \frac{\log(s_n^{-1})}{n}\sum_{i=0}^{\lfloor 2n^{0.99}\rfloor} \mathbb{1}_{(0,\delta)}(s_{n-i}) \\
&\quad + \frac{1}{n}\sum_{i=\lfloor 2n^{0.99}\rfloor+1}^{n-1} \mathbb{1}_{(0,\delta)}(s_{n-i}) \log(s_{n-i}^{-1}).
\end{aligned}$$

From Theorem 1.4 (here we need the bounded density assumption) we get that a.s.

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} \frac{\log(s_n^{-1})}{n}\sum_{i=0}^{\lfloor 2n^{0.99}\rfloor} \mathbb{1}_{(0,\delta)}(s_{n-i}) = 0$$

and it thus remains to show that a.s.

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} \frac{1}{n}\sum_{i=\lfloor 2n^{0.99}\rfloor+1}^{n-1} \mathbb{1}_{(0,\delta)}(s_{n-i}) \log\left(s_{n-i}^{-1}\right) = 0.$$

This boils down to show that there exists $c_0 > 0$ such a.s. for $n \gg 1$ and $2n^{0.99} \leq i \leq n-1$,

$$s_{n-i} \geq c_0 \frac{i}{n}. \tag{4.1}$$

To prove it, we adapt an argument due to Tao and Vu [41]. We fix $2n^{0.99} \leq i \leq n-1$ and we consider the matrix M' formed by the first $n - \lceil i/2 \rceil$ rows of

$$\sqrt{n}(\sqrt{n}M - zI) = nDX - \sqrt{n}zI.$$

By the Cauchy interlacing Lemma B.4, we get

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

where $s'_j := s_j(M')$ for all $1 \leq j \leq n - \lceil i/2 \rceil$ are the singular values of the rectangular matrix M' in nonincreasing order. Next, by the Tao and Vu negative moment Lemma B.3,

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

where dist_j is the distance from the jth row of M' to H_j, the subspace spanned by the other rows of M'. In particular, we have

$$\frac{i}{2} s_{n-i}^{-2} \leq n \sum_{j=1}^{n-\lceil i/2 \rceil} \mathrm{dist}_j^{-2}. \tag{4.2}$$

Let R_j be the jth row of X. Since the jth row of M is $D_{j,j} R_j$, we deduce that

$$\mathrm{dist}_j = \mathrm{dist}(nD_{j,j}R_j - z\sqrt{n}e_j, H_j) \geq nD_{j,j}\mathrm{dist}(R_j, \mathrm{span}(H_j, e_j))$$

where $e_1, \ldots, e_n$ is the canonical basis of $\mathbb{R}^n$. Since $\mathrm{span}(H_j, e_j)$ is independent of R_j and

$$\dim(\mathrm{span}(H_j, e_j)) \leq n - \frac{i}{2} \leq n - n^{0.99},$$

Lemma C.2 gives

$$\sum_{n \gg 1} \mathbb{P}\left(\bigcup_{i=2n^{0.99}}^{n-1} \bigcup_{j=1}^{n-\lceil i/2 \rceil} \left\{ \mathrm{dist}(R_j, \mathrm{span}(H_j, e_j)) \leq \frac{\sigma\sqrt{i}}{2\sqrt{2}} \right\} \right) < \infty$$

(note that the exponential bound in Lemma C.2 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 $2n^{0.99} \leq i \leq n-1$, and all $1 \leq j \leq n - \lceil i/2 \rceil$,

$$\mathrm{dist}_j \geq nD_{j,j} \frac{\sigma\sqrt{i}}{2\sqrt{2}} = \sqrt{i} \frac{\sigma}{2\sqrt{2}} \frac{n}{\rho_{n,j}}.$$

Now, the uniform law of large numbers (1.6) gives that a.s.

$$\lim_{n\to\infty}\max_{1\le j\le n}\left|\frac{\rho_{n,j}}{n}-1\right|=0.$$

We deduce that a.s. for $n\gg 1$, all $2n^{0.99}\le i\le n-1$, and all $1\le j\le n-\lceil i/2\rceil$,

$$\mathrm{dist}_j\ge\sqrt{i}\frac{\sigma}{4}$$

Finally, from (4.2) we get

$$s_{n-i}^2\ge\frac{i^2}{n^2}\frac{\sigma^2}{32},$$

and (4.1) holds with $c_0:=\sigma/(4\sqrt{2})$.

Remark 4.1 (Proof of the circular law (1.3) and beyond) The same strategy allows a relatively short proof of (1.3). Indeed, the a.s. weak convergence of $(\mu_{n^{-1/2}X})_{n\ge1}$ to $\mathcal{U}_\sigma$ follows from the Girko Hermitization Lemma A.2 and the uniform integrability of $\log(\cdot)$ for $(\nu_{n^{-1/2}X-zI})_{n\ge1}$ as above using (1.5). This direct strategy does not rely on the *replacement principle* of Tao and Vu. The replacement principle allows a statement which is more general than (1.3) involving two sequences of random matrices (the main result of [41] on universality). Our strategy allows to go beyond the circular law (1.3), by letting $\mathbb{E}(X_{i,j})$ possibly depend on i, j, n, provided that (1.5) and the result of Dozier and Silverstein [18] hold. Set $A:=(\mathbb{E}(X_{i,j}))_{1\le i,j\le n}$. If $\mathrm{Tr}(AA^*)$ is large enough for $n\gg 1$, the limit is no longer the circular law, and can be interpreted by using free probability theory [38].

Remark 4.2 (Beyond the circular law) It is likely that the Tao and Vu replacement principle [41] allows a universal statement for our random Markov matrices of the form $M=DX$, beyond the circular law, by letting $\mathbb{E}(X_{i,j})$ possibly depend on i, j, n. This is however beyond the scope of the present work.

5 Proof of Theorem 1.4

Note that when $z=0$, one can get some b immediately from (B.3, 1.5, 1.6). Thus, our problem is actually to deal with $z\ne 0$. Fix $a, C>0$ and $z\in\mathbb{C}$ with $|z|\le C$. We have

$$\sqrt{n}M-zI=\sqrt{n}DY\quad\text{where } Y:=X-n^{-1/2}zD^{-1}.$$

For an arbitrary $\delta_n>0$, let us define the event

$$\mathcal{A}_n:=\bigcap_{i=1}^{n}\left\{\left|\frac{\rho_{n,i}}{n}-1\right|\le\delta_n\right\}.$$

By using the union bound and the Chebyshev inequality, we get $\mathbb{P}(\mathcal{A}_n^c) \le \sigma^2\delta_n^{-2}$. Now with $c > a/2$ and $\delta_n = n^c$ we obtain $\mathbb{P}(\mathcal{A}_n^c) \le n^{-a}$ for $n \gg 1$. Since we have

$$s_n(D)^{-1} = \max_{1\le i\le n} |\rho_{n,i}|,$$

we get by (B.3), on the event $\mathcal{A}_n$, for $n \gg 1$,

$$\{s_n(\sqrt{n}M - zI) \le t_n\} \subset \{\sqrt{n}s_n(D)s_n(Y) \le t_n\} \subset \{s_n(Y) \le \sqrt{n}t_n(1+n^c)\}$$

for every $t_n > 0$. Now, for every $b' > 0$, one may select $b > 0$ and set $t_n = n^{-b}$ such that $\sqrt{n}t_n(1+n^c) \le n^{-b'}$ for $n \gg 1$. Thus, on the event $\mathcal{A}_n$, for $n \gg 1$,

$$\mathcal{M}_n := \{s_n(\sqrt{n}M - zI) \le n^{-b}\} \subset \{s_n(Y) \le n^{-b'}\} =: \mathcal{Y}_n.$$

Consequently, for every $b' > 0$ there exists $b > 0$ such that for $n \gg 1$,

$$\mathbb{P}(\mathcal{M}_n) = \mathbb{P}(\mathcal{M}_n \cap \mathcal{A}_n) + \mathbb{P}(\mathcal{M}_n \cap \mathcal{A}_n^c) \le \mathbb{P}(\mathcal{Y}_n) + \mathbb{P}(\mathcal{A}_n^c) \le \mathbb{P}(\mathcal{Y}_n) + n^{-a}.$$

The desired result follows if we show that for some $b' > 0$ depending on a, C, for $n \gg 1$,

$$\mathbb{P}(\mathcal{Y}_n) = \mathbb{P}(s_n(Y) \le n^{-b'}) \le n^{-a}. \tag{5.1}$$

Let us prove (5.1). At this point, it is very important to realize that (5.1) cannot follow form a perturbative argument based on (1.5,B.2,1.6) since the operator norm of the perturbation is much larger that the least singular value of the perturbed matrix. We thus need a more refined argument. We have $Y = X - wD^{-1}$ with $w := n^{-1/2}z$. Let $A_w = A_{n^{-1/2}z}$ be as in Lemma C.3. For every $1 \le k \le n$, let P_k be the $n \times n$ permutation matrix for the transposition $(1, k)$. Note that $P_1 = I$. For every column vector e_i of the canonical basis of $\mathbb{R}^n$,

$$(P_k A_w P_k)e_i = \begin{cases} e_i & \text{if } i \neq k, \\ e_k - w(e_1 + \cdots + e_n) & \text{if } i = k. \end{cases}$$

Now, if $R_1, \ldots, R_n$ and $R'_1, \ldots, R'_n$ are the rows of the matrices X and Y then

$$Y = \begin{pmatrix} R'_1 \\ \vdots \\ R'_n \end{pmatrix} = \begin{pmatrix} R_1 P_1 A_w P_1 \\ \vdots \\ R_n P_n A_w P_n \end{pmatrix}.$$

Define the vector space $R'_{-i} := \operatorname{span}\{R'_j : j \neq i\}$ for every $1 \le i \le n$. From Lemma B.2,

$$\min_{1\le i\le n} \operatorname{dist}\left(R'_i, R'_{-i}\right) \le \sqrt{n}\, s_n(Y).$$

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

$$\mathbb{P}(\sqrt{n}\, s_n(Y) \leq u) \leq n \max_{1 \leq i \leq n} \mathbb{P}\left(\operatorname{dist}\left(R'_i, R'_{-i}\right) \leq u\right).$$

The law of $\operatorname{dist}(R'_i, R'_{-i})$ does not depend on i. We take $i = 1$. Let V' be a unit normal vector to R'_{-1}. Such a vector is not unique, but we just pick one, and this defines a random variable on the unit sphere $\mathbb{S}^{n-1} := \{x \in \mathbb{C}^n : \|x\|_2 = 1\}$. Since $V' \in R'^{\perp}_{-1}$ and $\left\|V'\right\|_2 = 1$,

$$|R'_1 \cdot V'| \leq \operatorname{dist}\left(R'_1, R'_{-1}\right).$$

Let ν be the distribution of V' on $\mathbb{S}^{n-1}$. Since V' and R'_1 are independent, for any $u \geq 0$,

$$\mathbb{P}\left(\operatorname{dist}\left(R'_1, R'_{-1}\right) \leq u\right) \leq \mathbb{P}(|R'_1 \cdot V'| \leq u) = \int_{\mathbb{S}^{n-1}} \mathbb{P}(|R'_1 \cdot v'| \leq u)\, d\nu(v').$$

Let us fix $v' \in \mathbb{S}^{n-1}$. If A_w, $P_1 = I$, R_1 are as above then

$$R'_1 \cdot v' = R_1 \cdot v \quad \text{where } v := P_1 A_w P_1 v' = A_w v'.$$

Now, since $v' \in \mathbb{S}^{n-1}$, Lemma C.3 provides a constant $K > 0$ such that for $n \gg 1$,

$$\|v\|_2 = \left\|A_w v'\right\|_2 \geq \min_{x \in \mathbb{S}^{n-1}} \|A_w x\|_2 = s_n(A_w) \geq K^{-1}.$$

But $\|v\|_2 \geq K^{-1}$ implies $|v_j|^{-1} \leq K\sqrt{n}$ for some $j \in \{1, \ldots, n\}$, and therefore

$$|\mathfrak{Re}(v_j)|^{-1} \leq K\sqrt{2n} \quad \text{or} \quad |\mathfrak{Im}(v_j)|^{-1} \leq K\sqrt{2n}.$$

Suppose for instance that we have $|\mathfrak{Re}(v_j)|^{-1} \leq K\sqrt{2n}$. We first observe that

$$\mathbb{P}\left(|R'_1 \cdot v'| \leq u\right) = \mathbb{P}(|R_1 \cdot v| \leq u) \leq \mathbb{P}(|\mathfrak{Re}(R_1 \cdot v)| \leq u).$$

The real random variable $\mathfrak{Re}(R_1 \cdot v)$ is a sum of independent real random variables and one of them is $X_{1,j}\mathfrak{Re}(v_j)$, which is absolutely continuous with a density bounded above by $BK\sqrt{2n}$ where B is the bound on the density of $X_{1,1}$. Consequently, by a basic property of convolutions of probability measures, the real random variable $\mathfrak{Re}(R_1 \cdot v)$ is also absolutely continuous with a density φ bounded above by $BK\sqrt{2n}$, and therefore,

$$\mathbb{P}(|\mathfrak{Re}(R_1 \cdot v)| \leq u) = \int_{[-u,u]} \varphi(s)\, ds \leq BK\sqrt{2n}2u.$$

To summarize, for $n \gg 1$ and every $u \ge 0$,

$$\mathbb{P}(\sqrt{n}s_n(Y) \le u) \le BK(2n)^{3/2}u.$$

Lemma C.3 shows that the constant K may be chosen depending on C and not on z, and (5.1) holds with $b' = d + 1/2$ by taking $u = n^{-d}$ such that $BK(2n)^{3/2}n^{-d} \le n^{-a}$ for $n \gg 1$.

Remark 5.1 (Assumptions) Our proof of Theorem 1.4 still works if the entries of X are just independent and not necessarily i.i.d. provided that the densities are uniformly bounded and that (1.6) holds. The bounded density assumption allows to bound the small ball probability $\mathbb{P}(|R_1 \cdot v| \le u)$ uniformly over v. If this assumption does not hold, then the small ball probability may depend on the additive structure of v, but the final result is probably still valid. A possible route, technical and uncertain, is to adapt the Tao and Vu proof of (1.5). On the opposite side, if $X_{1,1}$ has a log-concave density (e.g. exponential) then a finer bound might follow from a noncentered version of the results of Adamczak et al. [1]. Alternatively, if $X_{1,1}$ has sub-Gaussian or sub-exponential moments then one may also try to adapt the proof of Rudelson and Vershynin [34] to the noncentered settings.

Remark 5.2 (Away from the limiting support) The derivation of an a.s. lower bound on $s_n(\sqrt{n}M - zI)$ is an easy task when $|z| > 2\sigma$ and $\mathbb{E}(|X_{1,1}|^4) < \infty$, without assuming that $X_{1,1}$ has a bounded density. Let us show for instance that for every $z \in \mathbb{C}$, a.s.

$$s_n(\sqrt{n}M - zI) \ge |z| - 2\sigma + o(1). \tag{5.2}$$

This lower bound is meaningful only when $|z| > 2\sigma$. For proving (5.2), we adopt a perturbative approach. Let us fix $z \in \mathbb{C}$. By (B.3) and (1.6) we get that a.s.

$$s_n(\sqrt{n}M - zI) \ge n^{-1}(1 + o(1))\, s_n(\sqrt{n}X - zD^{-1}). \tag{5.3}$$

Now we write, with $S = \mathbb{E}X = (1, \dots, 1)(1, \dots, 1)^\top$,

$$\sqrt{n}X - zD^{-1} = \sqrt{n}S - znI + W \quad \text{where } W := \sqrt{n}(X - S) + nzI - zD^{-1}.$$

We observe that (B.2) gives

$$s_n(\sqrt{n}X - zD^{-1}) \ge s_n(\sqrt{n}S - znI) - s_1(W).$$

For the symmetric complex matrix $\sqrt{n}S - znI$ we have for any $z \in \mathbb{C}$ and $n \gg 1$,

$$s_n(\sqrt{n}S - znI) = n \min(|z|, |\sqrt{n} - z|) = n|z|.$$

On the other hand, since $\mathbb{E}(|X_{1,1}|^4) < \infty$, by (1.2) and (1.6), a.s. for every $z \in \mathbb{C}$,

$$s_1(W) \le s_1(\sqrt{n}(X - S)) + s_1(nzI - zD^{-1}) = n(2\sigma + o(1)) + |z|no(1).$$

Putting all together, we have shown that a.s. for any $z \in \mathbb{C}$,

$$s_n(\sqrt{n}X - zD^{-1}) \geq n|z|(1 - o(1)) - n(2\sigma + o(1)).$$

Combined with (5.3), this gives finally (5.2).

Remark 5.3 (Invertibility) 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 $(\Omega, \mathcal{A}, \mathbb{P})$. For every $\omega \in \Omega$, the set $\cup_{n\geq 1}\{\lambda_1(A_n(\omega)), \dots, \lambda_n(A_n(\omega))\}$ 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 $s_n(A_n(\omega) - zI) > 0$ for all **all** $n \geq 1$. Note that (1.5) and Theorem 1.4 imply respectively that for $A_n = X$ or $A_n = M$, this holds for **all** $z \in \mathbb{C}$, **a.s.** on ω, and $n \gg 1$.

Acknowledgments DC is grateful to O. Guédon, M. Krishnapur, and T. Tao for helpful discussions. CB and DC are grateful to Dipartimento di Matematica Università Roma Tre for kind hospitality.

Appendix A: 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. For every $\mu \in \mathcal{P}(\mathbb{C})$, the *logarithmic potential* U_μ of μ on $\mathbb{C}$ is the function $U_\mu : \mathbb{C} \to (-\infty, +\infty]$ defined for every $z \in \mathbb{C}$ by

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

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

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

see e.g. [35]. Let $\mathcal{D}'(\mathbb{C})$ be the set of Schwartz–Sobolev distributions (generalized functions). Since $\log|\cdot|$ is Lebesgue locally integrable on $\mathbb{C}$, one can check by using the Fubini theorem that U_μ is Lebesgue locally integrable on $\mathbb{C}$. In particular, $U_\mu < \infty$ a.e. and $U_\mu \in \mathcal{D}'(\mathbb{C})$. Since $\log|\cdot|$ is the fundamental solution of the Laplace equation in $\mathbb{C}$, we have, in $\mathcal{D}'(\mathbb{C})$,

$$\Delta U_\mu = -2\pi\mu. \tag{A.3}$$

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

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

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

Proof Since $U_\mu = U_\nu$ in $\mathcal{D}'(\mathbb{C})$, we get $\Delta U_\mu = \Delta U_\nu$ in $\mathcal{D}'(\mathbb{C})$. Now (A.3) gives $\mu = \nu$ in $\mathcal{D}'(\mathbb{C})$, and thus $\mu = \nu$ as measures since μ and ν are Radon measures. □

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

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

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

$$U_{\mu_A}(z) = -\frac{1}{n}\log\det(\sqrt{(A - zI)(A - zI)^*}) = -\int_0^\infty \log(t)\, \nu_{A-zI}(dt). \quad \text{(A.4)}$$

The identity (A.4) 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 in contrary to the eigenvalues, one can control the singular values with the entries of the matrix. The price payed here is the introduction of the auxiliary variable z and the uniform integrability. We recall that on a Borel measurable space $(E, \mathcal{E})$, we say that a Borel function $f : E \to \mathbb{R}$ is *uniformly integrable* for a sequence of probability measures $(\eta_n)_{n\geq 1}$ on E when

$$\lim_{t\to\infty} \overline{\lim_{n\to\infty}} \int_{\{|f|>t\}} |f|\, d\eta_n = 0. \quad \text{(A.5)}$$

We will use this property as follows: if $(\eta_n)_{n\geq 1}$ converges weakly to η and f is continuous and uniformly integrable for $(\eta_n)_{n\geq 1}$ then f is η-integrable and $\lim_{n\to\infty} \int f\, d\eta_n = \int f\, \eta$. The idea of using Hermitization goes back at least to Girko [22]. However, the proofs of Lemmas A.2 and A.3 below are inspired from the approach of Tao and Vu [41].

Lemma A.2 (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 a.s.*

(i) *$(\nu_{A_n - zI})_{n\geq 1}$ converges weakly to ν_z as $n \to \infty$*
(ii) *$\log(\cdot)$ is uniformly integrable for $\left(\nu_{A_n - zI}\right)_{n\geq 1}$*

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

(j) *a.s. $(\mu_{A_n})_{n\geq 1}$ converges weakly to μ as $n \to \infty$*
(jj) *for a.a. $z \in \mathbb{C}$,*

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

Moreover, if $(A_n)_{n\geq 1}$ is deterministic, then the statements hold without the “a.s.”

Proof Let z and ω be such that (i-ii) hold. For every $1 \le k \le n$, 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 (B.6) and to the assumptions (i, ii), one can use Lemma A.3 below, which gives that $(\mu_{A_n})_{n\ge1}$ is tight, that $\log|z - \cdot|$ is uniformly integrable for $(\mu_{A_n})_{n\ge1}$, and that

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

Consequently, a.s. $\mu \in \mathcal{P}(\mathbb{C})$ and $U_\mu = U$ a.e. for every adherence value μ of $(\mu_{A_n})_{n\ge1}$. Now, since U does not depend on μ, by Lemma A.1, a.s. $\left(\mu_{A_n}\right)_{n\ge1}$ has a unique adherence value μ, and since $(\mu_n)_{n\ge1}$ is tight, $(\mu_{A_n})_{n\ge1}$ converges weakly to μ by the Prohorov theorem. Finally, by (A.3), μ is deterministic since U is deterministic, and (j, jj) hold. □

The following lemma is in a way the skeleton of the Girko Hermitization of Lemma A.2. 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.

Lemma A.3 (Logarithmic majorization and uniform integrability) *Let* $(a_{n,k})_{1\le k\le n}$ *and* $(b_{n,k})_{1\le k\le n}$ *be two triangular arrays in* $[0,\infty)$. *Define the discrete probability measures*

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

If the following properties hold

- (i) $a_{n,1} \ge \cdots \ge a_{n,n}$ *and* $b_{n,1} \ge \cdots \ge b_{n,n}$ *for* $n \gg 1$,
- (ii) $\prod_{i=1}^k a_{n,i} \le \prod_{i=1}^k b_{n,i}$ *for every* $1 \le k \le n$ *for* $n \gg 1$,
- (iii) $\prod_{i=k}^n b_{n,i} \le \prod_{i=k}^n a_{n,i}$ *for every* $1 \le k \le n$ *for* $n \gg 1$,
- (iv) $(\nu_n)_{n\ge1}$ *converges weakly to some probability measure* ν *as* $n \to \infty$,
- (v) $\log(\cdot)$ *is uniformly integrable for* $(\nu_n)_{n\ge1}$,

then

- (j) $(\mu_n)_{n\ge1}$ *is tight,*
- (jj) $\log(\cdot)$ *is uniformly integrable for* $(\mu_n)_{n\ge1}$,
- (jjj) *we have, as* $n \to \infty$,

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

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

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

Proof Proof of (jjj). From the logarithmic majorizations (ii, iii) we get, for $n \gg 1$,

$$\prod_{k=1}^n a_{n,k} = \prod_{k=1}^n b_{n,k},$$

and (v) gives $b_{n,k} > 0$ and $a_{n,k} > 0$ for every $1 \leq k \leq n$ and $n \gg 1$. Now, (iv-v) give

$$\begin{aligned}\int_0^\infty \log(t)\,\mu_n(dt) &= \frac{1}{n}\log\prod_{k=1}^n a_{n,k}\\ &= \frac{1}{n}\log\prod_{k=1}^n b_{n,k}\\ &= \int_0^\infty \log(t)\,\nu_n(dt) \to \int_0^\infty \log(t)\,\nu(dt).\end{aligned}$$

Proof of (j). From (ii) and (v) we get

$$\sup_{1\leq k\leq n}\sum_{i=1}^k \log(a_{n,i}) \leq \sup_{1\leq k\leq n}\sum_{i=1}^k \log(b_{n,i}) \quad \text{and} \quad C := \sup_{n\geq N}\int_0^\infty |\log(s)|\,\nu_n(ds) < \infty$$

(for large enough N) respectively. Now the tightness of $(\mu_n)_{n\geq 1}$ follows from

$$\int_1^\infty \log(s)\,\mu_n(ds) \leq \int_1^\infty \log(s)\,\nu_n(ds) \leq C. \tag{A.6}$$

Proof of (jj). We start with the uniform integrability in the neighborhood of infinity. Let us show that for $n \gg 1$, for any $\varepsilon > 0$ there exists $t \geq 1$ such that

$$\int_t^\infty \log(s)\,\mu_n(ds) < \varepsilon. \tag{A.7}$$

If $\nu((1,\infty)) = 0$ then (iv) implies

$$\int_1^\infty \log(t)\,\nu_n(dt) < \varepsilon$$

for $n \gg 1$ and (A.7) follows then from (A.6). If otherwise $\nu((1,\infty)) > 0$ then

$$c := \int_1^\infty \log(t)\,\nu(dt) > 0$$

and one can assume that $\varepsilon < c$. Let us show that there exists a sequence of integers $(k_n)_{n\geq 1}$ such that $\lim_{n\to\infty} k_n/n \to \sigma > 0$ and for $n \gg 1$,

$$\sup_{1\leq k\leq k_n} \frac{1}{n}\sum_{i=1}^{k_n} \log(b_{n,i}) < \varepsilon. \tag{A.8}$$

For $0 < \varepsilon/2 < c$, let t be the infimum over all $s > 1$ such that

$$\int_s^\infty \log(u)\,\nu(du) < \frac{1}{2}\varepsilon.$$

There exists $s \geq t$ such that $\nu(\{s\}) = 0$, and from (v) we get

$$\lim_{n\to\infty} \nu_n((s,\infty)) = \nu((s,\infty)) \geq 0 \quad \text{and}$$
$$\lim_{n\to\infty} \int_s^\infty \log(u)\,\nu_n(du) = \int_s^\infty \log(u)\,\nu(du) \leq \frac{1}{2}\varepsilon.$$

If $\nu((s,\infty)) > 0$ then (A.8) holds with $\sigma := \nu((s,\infty))$ and $k_n := \lfloor n\nu_n((s,\infty))\rfloor$. Otherwise, $\nu((s,\infty)) = 0$, and if $k'_n := \lfloor n\nu_n((s,\infty))\rfloor$ then $\lim_{n\to\infty} k'_n/n = 0$, while for any $\delta > 0$,

$$\frac{1}{n}\sum_{i=1}^{k'_n+\lfloor n\delta\rfloor} \log(b_{n,i}) \leq \frac{\varepsilon}{2} + \delta\log(s).$$

Taking $k_n := k'_n + \lfloor n\delta\rfloor$ with δ small enough, we deduce that (A.8) holds. We have thus shown that (A.8) holds in all case. Now, from (ii) and (A.8) we get for every $1 \leq k \leq k_n$,

$$\frac{1}{n}\sum_{i=1}^{k} \log(a_{n,i}) < \varepsilon.$$

In particular, by using (i), we get $\log(a_{n,k_n}) \le \varepsilon n/k_n$ and

$$\int_{e^{\varepsilon n/k_n}}^{\infty} \log(u)\mu_n(du) < \varepsilon.$$

Since $\underline{\lim}_{n\to\infty} k_n/n = \delta > 0$, we deduce that (A.7) holds with $t := e^{\varepsilon\delta}$. Now, by following the same reasoning, with (ii) replaced by (iii), we obtain that for all $\varepsilon > 0$, there exists $0 < t < 1$ such that for $n \gg 1$,

$$-\int_0^t \log(s)\mu_n(ds) < \varepsilon,$$

which is the counterpart of (A.7) needed for (jj). □

Remark A.4 (Other fundamental aspects of the logarithmic potential) The logarithmic potential is related to the Cauchy-Stieltjes transform of μ via

$$S_\mu(z) := \int_{\mathbb{C}} \frac{1}{z'-z}\,\mu(dz') = (\partial_x - i\partial_y)U_\mu(z) \quad \text{and thus} \quad (\partial_x + i\partial_y)S_\mu = -2\pi\mu$$

in $\mathcal{D}'(\mathbb{C})$. The term "logarithmic potential" comes from the fact that U_μ is the electrostatic potential of μ viewed as a distribution of charges in $\mathbb{C} \equiv \mathbb{R}^2$ [35]. The logarithmic energy

$$\mathcal{E}(\mu) := \int_{\mathbb{C}} U_\mu(z)\,\mu(dz) = -\int_{\mathbb{C}}\int_{\mathbb{C}} \log\left|z-z'\right|\mu(dz)\mu(dz')$$

is up to a sign the Voiculescu free entropy of μ in free probability theory [44]. The circular law $\mathcal{U}_\sigma$ minimizes $\mu \mapsto \mathcal{E}(\mu)$ under a second moment constraint [35]. In the spirit of (A.4) and beyond matrices, the Brown [14] spectral measure of a nonnormal bounded operator a is $\mu_a := (-4\pi)^{-1}\Delta \int_0^\infty \log(t)\,\nu_{a-zI}(dt)$ where ν_{a-zI} is the spectral distribution of the self-adjoint operator $(a-zI)(a-zI)^*$. Due to the logarithm, the Brown spectral measure μ_a depends discontinuously on the $*$-moments of a [9,38]. For random matrices, this problem is circumvented in the Girko Hermitization by requiring a uniform integrability, which turns out to be a.s. satisfied for random matrices such as $n^{-1/2}X$ or $\sqrt{n}M$.

Appendix B: General spectral estimates

We gather in this section useful lemmas on deterministic matrices. We provide mainly references for the most classical results, and sometimes proofs for the less classical ones.

Lemma B.1 (Basic inequalities [28]) *If A and B are $n \times n$ complex matrices then*

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

and

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

and

$$s_n(AB) \ge s_n(A)s_n(B). \tag{B.3}$$

Moreover, if $A = D$ is diagonal, then for every $1 \le i \le n$

$$s_n(D)s_i(B) \le s_i(DB) \le s_1(D)s_i(B). \tag{B.4}$$

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

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

The argument behind Lemma B.2 is buried in [34]. We give a proof below for convenience.

Proof of Lemma B.2 Since $A, A^\top$ have same singular values, one can consider the columns $C_1, \dots, C_n$ of A instead of the rows. For every column vector $x \in \mathbb{C}^n$ and $1 \le i \le n$, the triangle inequality and the identity $Ax = x_1C_1 + \cdots + x_nC_n$ give

$$\begin{aligned}\|Ax\|_2 \ge \operatorname{dist}(Ax, C_{-i}) &= \min_{y \in C_{-i}} \|Ax - y\|_2 \\ &= \min_{y \in C_{-i}} \|x_iC_i - y\|_2 = |x_i| \operatorname{dist}(C_i, C_{-i}).\end{aligned}$$

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

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

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

$$\operatorname{dist}(C_i, C_{-i}) = \|y_1C_1 + \cdots + y_nC_n\|_2 = \|Ay\|_2 \ge \|y\|_2 \min_{\|x\|_2 = 1} \|Ax\|_2 \ge s_n(A)$$

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

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

Lemma B.3 (Tao–Vu negative second moment [41, lem. A4]) *If A is a full rank $n' \times n$ complex matrix ($n' \leq n$) with rows $R_1, \dots, R_{n'}$, and $R_{-i} := \mathrm{span}\{R_j : j \neq i\}$, then*

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

Lemma B.4 (Cauchy interlacing by rows deletion [28]) *Let A be an $n \times n$ complex matrix. If B is $n' \times n$, obtained from A by deleting $n - n'$ rows, then for every $1 \leq i \leq n'$,*

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

Lemma B.4 gives $[s_{n'}(B), s_1(B)] \subset [s_n(A), s_1(A)]$, i.e. row deletions produce a compression of the singular values interval. Another way to express this phenomenon consists in saying that if we add a row to B then the largest singular value increases while the smallest is diminished. Closely related, the following result on finite rank additive perturbations. If A is an $n \times n$ complex matrix, let us set $s_i(A) := +\infty$ if $i < 1$ and $s_i(A) := 0$ if $i > n$.

Lemma B.5 (Thompson–Lidskii interlacing for finite rank perturbations [42]) *For any $n \times n$ complex matrices A and B with $\mathrm{rank}(A - B) \leq k$, we have, for any $i \in \{1, \dots, n\}$,*

$$s_{i-k}(A) \geq s_i(B) \geq s_{i+k}(A). \tag{B.5}$$

Even if Lemma B.5 gives nothing on the extremal singular values $s_i(B)$ where $i \leq k$ or $n - i < k$, it provides however the useful "bulk" inequality $\|F_A - F_B\|_\infty \leq \mathrm{rank}(A - B)/n$ where F_A and F_B are the cumulative distribution functions of ν_A and ν_B respectively.

Lemma B.6 (Weyl inequalities [46]) *For every $n \times n$ complex matrix A, we have*

$$\prod_{i=1}^{k} |\lambda_i(A)| \leq \prod_{i=1}^{k} s_i(A) \quad \textit{and} \quad \prod_{i=n-k+1}^{n} s_i(A) \leq \prod_{i=n-k+1}^{n} |\lambda_i(A)| \tag{B.6}$$

for all $1 \leq k \leq n$, with equality for $k = n$. In particular, by viewing $|\det(A)|$ as a volume,

$$|\det(A)| = \prod_{k=1}^{n} |\lambda_k(A)| = \prod_{k=1}^{n} s_k(A) = \prod_{k=1}^{n} \mathrm{dist}(R_k, \mathrm{span}\{R_1, \dots, R_{k-1}\}) \tag{B.7}$$

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

$$\sum_{i=1}^{k} \varphi(|\lambda_i(A)|^2) \leq \sum_{i=1}^{k} \varphi(s_i(A)^2) \tag{B.8}$$

for every $1 \le k \le n$. *In particular, with* $\varphi(t) = t$ *for every* $t > 0$ *and* $k = n$, *we obtain*

$$\sum_{k=1}^{n} |\lambda_k(A)|^2 \le \sum_{k=1}^{n} s_k(A)^2 = \mathrm{Tr}(AA^*) = \sum_{i,j=1}^{n} |A_{i,j}|^2. \tag{B.9}$$

It is worthwhile to mention that (B.5) and (B.6) are optimal in the sense that every sequences of numbers satisfying these inequalities are associated to matrices, see [27,42].

Appendix C: Additional lemmas

Lemma C.1 below is used in the proof of Theorem 1.1. We omit its proof since it follows for instance quite easily from the Paul Lévy criterion on characteristic functions.

Lemma C.1 (Convergence under uniform perturbation) *Let* $(a_{n,k})_{1\le k\le n}$ *and* $(b_{n,k})_{1\le k\le n}$ *be triangular arrays of complex numbers. Let* μ *be a probability measure on* $\mathbb{C}$.

$$\text{If}\quad \frac{1}{n}\sum_{k=1}^{n}\delta_{a_{n,k}} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mu \quad\text{and}\quad \lim_{n\to\infty}\max_{1\le k\le n}|a_{n,k}-b_{n,k}|=0 \quad\text{then}\quad \frac{1}{n}\sum_{k=1}^{n}\delta_{b_{n,k}} \xrightarrow[n\to\infty]{\mathscr{C}_b} \mu.$$

Lemma C.2 below is used for the rows of random matrices in the proof of Theorem 1.3.

Lemma C.2 (Tao–Vu distance Lemma [41, prop. 5.1]) *Let* $(X_i)_{i\ge1}$ *be i.i.d. random variables on* $\mathbb{C}$ *with finite positive variance* $\sigma^2 := \mathbb{E}(|X_1 - \mathbb{E}X_1|^2)$. *For* $n \gg 1$ *and every deterministic subspace* H *of* $\mathbb{C}^n$ *with* $1 \le \dim(H) \le n - n^{0.99}$, *setting* $R := (X_1, \ldots, X_n)$,

$$\mathbb{P}\Big(\mathrm{dist}(R, H) \le \frac{\sigma}{2}\sqrt{n - \dim(H)}\Big) \le \exp(-n^{0.01}).$$

The proof of Lemma C.2 is based on a concentration inequality for convex Lipschitz functions and product measures due to Talagrand [39], see also [30, cor. 4.9]. The power 0.01 is used here to fix ideas and is obviously not optimal. This is more than enough for our purposes (proof of Theorem 1.3). A careful reading of the proof of Theorem 1.3 shows that a polynomial bound on the probability with a large enough power on n suffices.

We end up this section by a lemma used in the proof of Theorem 1.4.

Lemma C.3 (A special matrix) *For every* $w \in \mathbb{C}$, *let us define the* $n \times n$ *complex matrix*

$$A_w = I - w\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}.$$

Then for every $z \in \mathbb{C}$ *we have* $s_2\left(A_{\frac{z}{\sqrt{n}}}\right) = \cdots = s_{n-1}\left(A_{\frac{z}{\sqrt{n}}}\right) = 1$ *for* $n \gg 1$ *while*

$$\lim_{n\to\infty} s_n\left(A_{\frac{z}{\sqrt{n}}}\right) = \lim_{n\to\infty} s_1\left(A_{\frac{z}{\sqrt{n}}}\right)^{-1} = \frac{\sqrt{2}}{\sqrt{2+|z|^2+|z|\sqrt{4+|z|^2}}}$$

and the convergence is uniform on every compact subset of $\mathbb{C}$.

Proof Note that $A_0 = I$ and $A_w A_{w'} = A_{ww'-(w+w')}$ for every $w, w' \in \mathbb{C}$. Moreover, A_w is invertible if and only if $w \neq 1$ and in that case $(A_w)^{-1} = A_{w/(w-1)}$. It is a special case of the Sherman–Morrison formula for the inverse of rank one perturbations. It is immediate to check that $s_1(A_w - I) = \|A_w - I\|_{2\to 2} = \sqrt{n}|w|$ for every $w \in \mathbb{C}$. An elementary explicit computation reveals that the symmetric matrix $A_w A_w^* - I$ has rank at most 2, and thus A_w has at least $n-2$ singular values equal to 1 and in particular $s_n(A_w) \leq 1 \leq s_1(A_w)$. From now, let us fix $z \in \mathbb{C}$ and set $w = n^{-1/2}z$ and $A = A_w$ for convenience. The matrix A is nonsingular for $n \gg 1$ since $w \to 0$ as $n \to \infty$. Also, we have $s_n(A) > 0$ for $n \gg 1$. Since A is lower triangular with eigenvalues $1-w, 1, \ldots, 1$, by (B.7),

$$|1-w| = \prod_{i=1}^{n} |\lambda_i(A)| = |\det(A)| = \prod_{i=1}^{n} s_i(A) = u_- u_+ \tag{C.1}$$

where $u_- \leq u_+$ are two singular values of A. We have also

$$\begin{aligned} u_-^2 + u_+^2 + (n-2) &= s_1(A)^2 + \cdots + s_n(A)^2 = \operatorname{Tr}(AA^*) \\ &= |1-w|^2 + (n-1)(1+|w|^2) \end{aligned}$$

which gives $u_-^2 + u_+^2 = 1 + |1-w|^2 + (n-1)|w|^2$. Combined with (C.1), we get that $u_\pm^2$ are the solution of $X^2 - (1+(n-1)|w|^2 + |1-w|^2)X + |1-w|^2 = 0$. This gives

$$2u_\pm^2 = 2 + |z|^2 + O(n^{-1/2}) \pm |z|\sqrt{4+|z|^2+O(n^{-1/2})}$$

and the $O(n^{-1/2})$ is uniform in z on every compact. From this formula we get that $u_- \leq 1$ and $u_+ \geq 1$ for $n \gg 1$, and thus $u_- = s_n(A)$ and $u_+ = s_1(A)$. □

The result of Lemma C.3 is more than enough for our purposes. More precisely, a careful reading of the proof of Theorem 1.4 shows that a polynomial (in n) lower bound on $s_n(A_{n^{-1/2}z})$ for $n \gg 1$, uniformly on compact sets on z, is actually enough.

References

1. Adamczak, R., Guédon, O., Litvak, A., Pajor, A., Tomczak-Jaegermann, N.: Smallest singular value of random matrices with independent columns. C. R. Math. Acad. Sci. Paris **346**(15–16), 853–856 (2008)
2. Akemann, G., Kanzieper, E.: Integrable structure of Ginibre's ensemble of real random matrices and a Pfaffian integration theorem. J. Stat. Phys. **129**(5–6), 1159–1231 (2007)

3. Aubrun, G.: Random points in the unit ball of l_p^n. Positivity **10**(4), 755–759 (2006)
4. Bai, Z.D.: Circular law. Ann. Probab. **25**(1), 494–529 (1997)
5. Bai, Z.D., Silverstein, J.W.: Spectral Analysis of Large Dimensional Random Matrices. Mathematics Monograph Series 2. Science Press, Beijing (2006)
6. Bai, Z.D., Silverstein, J.W., Yin, Y.Q.: A note on the largest eigenvalue of a large-dimensional sample covariance matrix. J. Multivar. Anal. **26**(2), 166–168 (1988)
7. Bai, Z.D., Yin, Y.Q.: Limiting behavior of the norm of products of random matrices and two problems of Geman–Hwang. Probab. Theory Relat. Fields **73**(4), 555–569 (1986)
8. Bai, Z.D., Yin, Y.Q.: Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. Ann. Probab. **21**(3), 1275–1294 (1993)
9. Biane, Ph., Lehner, F.: Computation of some examples of Brown's spectral measure in free probability. Colloq. Math. **90**(2), 181–211 (2001)
10. Bollobás, B.: Random graphs. Cambridge Studies in Advanced Mathematics, vol. 73, 2nd edn. Cambridge University Press, Cambridge (2001)
11. Bordenave, Ch., Caputo, P., Chafaï, D.: Spectrum of large random reversible Markov chains: Heavy-tailed weigths on the complete graph. Ann. Probab. (2010). Preprint (accepted). arXiv:0903.3528 [math. PR]
12. Bordenave, Ch., Caputo, P., Chafaï, D.: Spectrum of large random reversible markov chains: two examples. ALEA Latin Am. J. Probab. Math. Stat. (7), 41–64 (2010)
13. Bordenave, Ch., Caputo, P., Chafaï, D.: Spectrum of non-Hermitian heavy tailed random matrices (2010, preprint)
14. Brown, L.G.: Lidskiĭ's theorem in the type II case. Geometric Methods in Operator Algebras (Kyoto, 1983). Pitman Res. Notes Math. Ser., vol. 123, pp. 1–35. Longman Sci. Tech. Harlow (1986)
15. Chafaï, D.: Aspects of large random Markov kernels. Stochastics **81**(3–4), 415–429 (2009)
16. Chafaï, D.: Circular law for noncentral random matrices. J. Theor. Probab. **23**(4), 945–950 (2010)
17. Chafaï, D.: The Dirichlet Markov ensemble. J. Multivar. Anal. **101**, 555–567 (2010)
18. Dozier, R.B., Silverstein, J.W.: On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. J. Multivar. Anal. **98**(4), 678–694 (2007)
19. Edelman, A.: The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. J. Multivar. Anal. **60**(2), 203–232 (1997)
20. Erdős, P., Rényi, A.: On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutató Int. Közl **5**, 17–61 (1960)
21. Girko, V.L.: The circular law. Teor. Veroyatnost. i Primenen. **29**(4), 669–679 (1984)
22. Girko, V.L.: Theory of random determinants. Mathematics and its Applications (Soviet Series), vol. 45. Kluwer, Dordrecht (1990, Translated from the Russian)
23. Girko, V.L.: Strong circular law. Random Oper. Stoch. Equ. **5**(2), 173–196 (1997)
24. Girko, V.L.: The circular law. Twenty years later. III. Random Oper. Stoch. Equ. **13**(1), 53–109 (2005)
25. Goldberg, G., Neumann, M.: Distribution of subdominant eigenvalues of matrices with random rows. SIAM J. Matrix Anal. Appl. **24**(3), 747–761 (2003, electronic)
26. Götze, F., Tikhomirov, A.: The circular law for random matrices. Ann. Probab. **38**(4), 1444–1491 (2010)
27. Horn, A.: On the eigenvalues of a matrix with prescribed singular values. Proc. Am. Math. Soc. **5**, 4–7 (1954)
28. Horn, R.A., Johnson, Ch.R.: Topics in matrix analysis. Cambridge University Press, Cambridge (1994, Corrected reprint of the 1991 original)
29. Hwang, C.-R.: 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. American Mathematical Society, Providence, RI, pp. 145–152 (1986)
30. Ledoux, M.: The concentration of measure phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence, RI (2001)
31. Marchenko, V.A., Pastur, L.A: The distribution of eigenvalues in sertain sets of random matrices. Mat. Sb. **72**, 507–536 (1967)
32. Mehta, M.L.: Random Matrices and the Statistical Theory of Energy Levels. Academic Press, New York (1967)
33. Pan, G.M., Zhou, W.: Circular law, extreme singular values and potential theory. J. Multivar. Anal **101**(3), 645–656 (2010)
34. Rudelson, M., Vershynin, R.: The Littlewood–Offord problem and invertibility of random matrices. Adv. Math. **218**(2), 600–633 (2008)

35. Saff, E.B., Totik, V.: Logarithmic potentials with external fields, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 316. Springer, Berlin (1997, Appendix B by Thomas Bloom)
36. Seneta, E.: Non-negative matrices and Markov chains. Springer Series in Statistics. Springer, New York (2006, Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544])
37. Silverstein, J.W.: The spectral radii and norms of large-dimensional non-central random matrices. Commun. Stat. Stoch. Models **10**(3), 525–532 (1994)
38. Śniady, P.: Random regularization of Brown spectral measure. J. Funct. Anal **193**(2), 291–313 (2002)
39. Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Études Sci. Publ. Math (81) 73–205 (1995)
40. Tao, T., Vu, V.: Random matrices: the circular law. Commun. Contemp. Math. **10**(2), 261–307 (2008)
41. Tao, T., Vu, V.: Random matrices: universality of ESDs and the circular law. Ann. Probab. 38(5), 2023–2065 (2010, With an appendix by Manjunath Krishnapur)
42. Thompson, R.C.: The behavior of eigenvalues and singular values under perturbations of restricted rank. Linear Algebra Appl. **13** (1/2), 69–78 (1976. Collection of articles dedicated to Olga Taussky Todd)
43. Trefethen, L.N., Embree, M.: Spectra and pseudospectra. The Behavior of Nonnormal Matrices and Operators. Princeton University Press, Princeton (2005)
44. Voiculescu, D.: Free entropy. Bull. Lond. Math. Soc **34**(3), 257–278 (2002)
45. Wachter, K.W.: The strong limits of random matrix spectra for sample matrices of independent elements. Ann. Probab. **6**(1), 1–18 (1978)
46. Weyl, H.: Inequalities between the two kinds of eigenvalues of a linear transformation. Proc. Nat. Acad. Sci. USA **35**, 408–411 (1949)
47. Yin, Y.Q.: Limiting spectral distribution for a class of random matrices. J. Multivar. Anal. **20**(1), 50–68 (1986)