Circular law for non-central random matrices

Djalil Chafai

Circular law for non-central random matrices

Djalil Chafaï

Preprint, September 2007

Abstract

Let $(X_{i,j})_{1\leqslant i,j<\infty}$ be an infinite array of i.i.d. complex random variables, with mean $m=0$, variance $\sigma^2=1$, and say with finite fourth moment. The famous circular law theorem states that the empirical spectral distribution $\frac{1}{n}(\delta_{\lambda_1(\mathbf{X})}+\cdots+\delta_{\lambda_n(\mathbf{X})})$ of $\mathbf{X}=(n^{-1/2}X_{i,j})_{1\leqslant i,j\leqslant n}$ converges almost surely, as $n\to\infty$, to the uniform law over the unit disc $\{z\in\mathbb{C};|z|\leqslant 1\}$. For now, most efforts where focused on the improvement of moments hypotheses for the centered case $m=0$. Regarding the non-central case $m\neq 0$, Silverstein has already observed that almost surely, the eigenvalue of $\mathbf{X}$ of largest module goes to infinity as $n\to\infty$, while the rest of the spectrum remains bounded. We show that the circular law theorem remains valid when $m\neq 0$, by using logarithmic potentials and bounds on extremal singular values.

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

Keywords: Random matrices; Circle or Circular law; rank-one deformations; singular values; eigenvalues; non-Hermitian matrices; logarithmic potential.

1 Introduction

For any square $n\times n$ matrix $\mathbf{A}$ with complex entries, let the complex eigenvalues $\lambda_1(\mathbf{A}),\ldots,\lambda_n(\mathbf{A})$ of $\mathbf{A}$ be labeled so that

$$|\lambda_1(\mathbf{A})|\geqslant\cdots\geqslant|\lambda_n(\mathbf{A})|.$$

Similarly, we denote by $s_1(\mathbf{A})\geqslant\cdots\geqslant s_n(\mathbf{A})$ the *singular values* of $\mathbf{A}$, i.e. the eigenvalues of the positive semi-definite Hermitian matrix $(\mathbf{A}\mathbf{A}^*)^{1/2}$. Namely,

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

for every $1 \leqslant k \leqslant n$. In particular, we have the variational formulas

$$s_1(\mathbf{A}) = \max_{\|x\|_2=1} \|\mathbf{A}x\|_2 \quad \text{and} \quad s_n(\mathbf{A}) = \min_{\|x\|_2=1} \|\mathbf{A}x\|_2.$$

Notice that $s_1(\mathbf{A}) = \|\mathbf{A}\|_2$ and $s_n(\mathbf{A}) = \|\mathbf{A}^{-1}\|_2^{-1}$ if $\mathbf{A}^{-1}$ exists. We always have

$$0 \leqslant s_n(\mathbf{A}) \leqslant |\lambda_n(\mathbf{A})| \leqslant |\lambda_1(\mathbf{A})| \leqslant s_1(\mathbf{A}). \tag{1}$$

When $\mathbf{A}$ is a normal matrix, i.e. $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$, then $|\lambda_k(\mathbf{A})| = s_k(\mathbf{A})$ for every $1 \leqslant k \leqslant n$. The *empirical spectral distribution* (ESD) of $\mathbf{A}$ is defined by

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k(\mathbf{A})}.$$

The ESD of $\mathbf{A}$ is a discrete probability distribution on $\mathbb{C}$ with at most n atoms. Let $(\mathbf{X}_n)$ be a sequence of random matrices, defined on the same probability space, where $\mathbf{X}_n$ is an $n \times n$ matrix for each n. For every n, the ESD of $\mathbf{X}_n$ is a random probability distribution on $\mathbb{C}$. Here the distribution of the entries of $\mathbf{X}_n$ may depend on n. Usually, the sequence $(\mathbf{X}_n)$ is constructed by setting $\mathbf{X}_n = (X_{i,j})_{1 \leqslant i,j \leqslant n}$ where $(X_{i,j})_{1 \leqslant i,j < \infty}$ is an infinite array of random variables. A typical topic in random matrix theory is to ask about the almost sure convergence as $n \to \infty$ of the ESD of $\mathbf{X}_n$ to some non-random probability distribution on $\mathbb{C}$, called the *limiting spectral distribution* (LSD). In the present article, we will write most of the time $\mathbf{X}$ instead of $\mathbf{X}_n$, omitting the index n in the notation.

The spectral properties of large dimensional random matrices where explored after the seminal works of the statistician Wishart in the 1930's and of the physicists Wigner and Dyson in the 1950's. The reader may find recent accounts in[1] [3], [20], [23], [5], [11], [30]. One of the most famous result regarding non-normal random matrices is known as the circle (or circular) law theorem, illustrated by figure 1. The version given below is taken from [24].

Theorem 1.1 (Circular law). *Let $(X_{i,j})_{1 \leqslant i,j < \infty}$ be an infinite array of i.i.d. complex random variables of common law L, with mean $m = 0$, variance $\sigma^2 = 1$, and finite fourth moment. Then almost surely, the ESD of $\mathbf{X} = (n^{-1/2}X_{i,j})_{1 \leqslant i,j \leqslant n}$ tends, as $n \to \infty$, to the uniform distribution on the unit disc $D(0,1) = \{z \in \mathbb{C}; |z| \leqslant 1\}$.*

The extension of theorem 1.1 to the case $\sigma^2 \in (0, \infty)$ is immediate by rescaling. The LSD is the uniform law on the unit disc of $\mathbb{C}$, often referred as the *circle or circular law*. If $Z = re^{\sqrt{-1}\theta}$ is a complex random variable distributed according

[1] For the links with Voiculescu's free probability theory, see for instance [20] and [7].

to the uniform law on $D(0, 2\sigma)$, then the module r and the argument θ of Z are independent with joint law of density $(r, \theta) \mapsto (4\pi\sigma^2)^{-1}\, r\, \mathrm{I}_{[0,2\pi]}(\theta)\mathrm{I}_{[0,2\sigma]}(r)$. Both $\mathfrak{Re}(Z)$ and $\mathfrak{Im}(Z)$ follow the Wigner semi-circle law on $\mathbb{R}$ of density

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

More generally, for any angle $\alpha \in [0, 2\pi)$, the random variables $\mathfrak{Re}(e^{\sqrt{-1}\alpha}Z)$ and $\mathfrak{Im}(e^{\sqrt{-1}\alpha}Z)$ follow the Wigner semi-circle law mentioned above. Additionally, if a real random variable S follows the Wigner semi-circle law mentioned above, then its square S^2 follows the Marchenko-Pastur law on $\mathbb{R}_+$ of density

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

The apparison of the circular law as a limiting spectral distribution of large dimensional non-normal random matrices goes back at least to Dyson in the 1960's. The phenomenon was then studied by many authors, see [10] and [2] for a review. Mehta considered in [22] the case of random matrices with i.i.d. entries following the standard complex Gaussian by using the explicit expression of the density of the spectrum provided by Ginibre [12]. The real Gaussian case is more "complex" in a way and was investigated by Edelman [10] in the 1990's. The universal case was attacked by Girko [13] in 1984. However, Girko's proof has been criticised by Bai who pointed out mathematical gaps. In 1997 Bai [2] provided a rigorous proof by using the Fourier-Stieltjes approach introduced by Girko, see also [5, ch. 10] for an improved version. The reader may find a discussion on Girko's contribution in [2, 3] and [19, 18] for instance. More recently, following a suggestion made by Khoruzhenko in 2001, Götze and Tikhomirov [19] used an alternative approach based on logarithmic potentials, an idea already explored formally by Girko. Later, Pan & Zhou in [24] used this method to provide the version of the circular law theorem given above (theorem 1.1). Very recently, Tao & Vu [33] have shown that the finite fourth moment condition can be relaxed for instance to a finite $2 + \varepsilon$ moment where $\varepsilon > 0$. Götze and Tikhomirov have also a very similar improved result in [18]. Actually, it is conjectured that theorem 1.1 is true under the sole condition that L has finite positive variance. At the time of writing, this conjecture, which is confirmed by numerical experiments, still remains an open problem for mathematicians.

Both the Fourier-Stieltjes method and the logarithmic potential method reduce the problem to the control of the singular values of the matrix $\mathbf{X} - z\mathbf{I}$ where $z \in \mathbb{C}$, i.e. to the control of the spectra of the familly $(\mathbf{H}(z))_{z\in\mathbb{C}}$ of positive semi-definite Hermitian matrices where $\mathbf{H}(z) = ((\mathbf{X} - z\mathbf{I})(\mathbf{X} - z\mathbf{I})^*)^{1/2}$. Namely, the Stieltjes

transform S_n of the ESD μ_n of $\mathbf{X}$ is given by

$$S_n(z) = \int_{\mathbb{C}} \frac{1}{z' - z}\, d\mu_{\mathbf{X}}(z') = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{\lambda_k(\mathbf{X}) - z} = \frac{1}{n}\mathrm{Trace}((\mathbf{X} - z\mathbf{I})^{-1}).$$

Since S_n is analytic on $\mathbb{C}$ except on a finite number of poles $\lambda_1(\mathbf{X}), \ldots, \lambda_n(\mathbf{X})$, the function $\mathfrak{Re}(S_n)$ fully determines the spectrum of $\mathbf{X}$. We have

$$\begin{aligned}
\mathfrak{Re}(S_n(z)) &= \frac{1}{n}\sum_{k=1}^{n} \frac{\mathfrak{Re}(\lambda_k(\mathbf{X}) - z)}{|\lambda_k(\mathbf{X}) - z|^2} \\
&= -\frac{1}{2n}\sum_{k=1}^{n} \partial_{\mathfrak{Re}(z)} \log(|\lambda_k(\mathbf{X}) - z|^2) \\
&= \partial_{\mathfrak{Re}(z)} U_n(z),
\end{aligned}$$

where U_n is the logarithmic potential of the ESD μ_n of $\mathbf{X}$

$$\begin{aligned}
U_n(z) &= -\int_{\mathbb{C}} \log|z - z'|\, d\mu_n(z') \\
&= -\frac{1}{n}\sum_{k=1}^{n} \log|\lambda_k(\mathbf{X} - z\mathbf{I})| \\
&= -\frac{1}{n}\log|\det(\mathbf{X} - z\mathbf{I})| \\
&= -\frac{1}{n}\log\det(\mathbf{H}(z)) \\
&= -\int_0^{\infty} \log(x)\, d\nu_{n,z}(x)
\end{aligned}$$

where $\nu_{n,z}$ is the ESD of $\mathbf{H}(z)$. The Fourier-Stieltjes approach used by Girko [13], Bai [2], and more recently by Tao & Vu [33] is based on the convergence of the Fourier transform of $\mathfrak{Re}(S_n(z)) = \partial_{\mathfrak{Re}(z)} U_n(z)$, whereas the logarithmic potential approach used by Götze and Tikhomirov [19] and Pan & Zhou [24] is based directly on the convergence of the logarithmic potential $U_n(z)$. In both cases, one of the main difficulty is the control of the extremal singular values of $\mathbf{X} - z\mathbf{I}$, which correspond to the singularities at 0 and ∞ in the integral above involving $\nu_{n,z}$.

While the largest singular value can be relatively easily bounded under moments hypotheses, the smallest singular value of random matrices is hard to bound, and the reader may find some recent developments based on the Littlewood-Offord problem in the works of Rudelson [27], Rudelson & Vershynin [28], Pan & Zhou [24], Götze & Tikhomirov [18], Tao & Vu [33], and references therein. The bound on the smallest singular value used in [19] is based on [27], whereas the one used

in [24] and [18] improves the result of [28]. However, the one used in [33] is of different nature. We need this bound since the preceding ones are not suitable for the non-central case due to the norm of the deformation.

Notice that if $\mathbf{X}$ is an $n \times n$ random matrix with i.i.d. entries of absolutely continuous distribution L, then almost surely, $\mathbf{X}$ is not normal, has distinct eigenvalues (and is hence diagonalizable), and is invertible. The situation is more intricate when L has atoms, for instance when L is a purely discrete distribution, and this contributes to make difficult the derivation of generic lower bounds on smallest singular values.

Non-central case

Let $\mathbf{A}$ be some deterministic $n \times n$ complex matrix, and $\mathbf{1}$ be the $n \times n$ matrix with all entries equal to 1. The rank one positive semi-definite symmetric matrix $\mathbf{1}$ has spectrum $\lambda_1(\alpha\mathbf{1}) = n$ and $\lambda_2(\alpha\mathbf{1}) = \cdots = \lambda_n(\alpha\mathbf{1}) = 0$. Following Andrew [1, p. 167] or Silverstein [32], for every complex number $\alpha \in \mathbb{C}$, we have by denoting $\widetilde{\mathbf{A}} = \mathbf{A} - \alpha\mathbf{1}$ and $r = s_1(\widetilde{\mathbf{A}})$,

$$\{\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})\} \subset D(0, r) \cup D(n\alpha, r)$$

where $D(z, r) = \{z' \in \mathbb{C}; |z - z'| \leqslant r\}$. Moreover, if $|n\alpha| > r$ then

$$\{\lambda_2(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})\} \subset D(0, r) \quad \text{and} \quad \lambda_1(\mathbf{A}) \in D(n\alpha, r).$$

This spectrum localization follows from classical perturbation theory[2], by seeing $\mathbf{A} = \widetilde{\mathbf{A}} + \alpha\mathbf{1}$ as an additive perturbation of $\alpha\mathbf{1}$ by $\widetilde{\mathbf{A}}$. Now, let $(X_{i,j})_{1 \leqslant i,j < \infty}$ be an infinite array of i.i.d. complex random variables with common law L of mean m. Define the random matrices

$$\mathbf{X} = (n^{-1/2} X_{i,j})_{1 \leqslant i,j \leqslant n} \quad \text{and} \quad \widetilde{\mathbf{X}} = (n^{-1/2}(X_{i,j} - m))_{1 \leqslant i,j \leqslant n}.$$

We have $\mathbf{X} - \widetilde{\mathbf{X}} = \alpha\mathbf{1}$ where $\alpha = n^{-1/2}m$. Hence, if $m \neq 0$, the random matrix $\mathbf{X}$ is an additive rank one deformation of the centered random matrix $\widetilde{\mathbf{X}}$. However, the norm of this deformation does not collapse as n tends to infiniy, since

$$\|\mathbf{X} - \widetilde{\mathbf{X}}\|_2 = \|\alpha\mathbf{1}\|_2 = |m|\sqrt{n} \underset{n\to\infty}{\longrightarrow} \infty. \tag{2}$$

It is shown in [37, 6] that

$$\mathbb{E}[|X_{1,1} - m|^4] < \infty \quad \Rightarrow \quad \lim_{n\to\infty} s_1(\widetilde{\mathbf{X}}) = 2 \quad \text{a.s.} \tag{3}$$

[2]One can for example make use the Bauer-Fike theorem and the Gerschgorin theorem, see for instance [36, p. 87-88], [17, th. 7.2.2 p. 321] and [21, ch. 6].

whereas

$$\mathbb{E}[|X_{1,1} - m|^4] = \infty \quad \Rightarrow \quad \limsup_{n\to\infty} s_1(\widetilde{\mathbf{X}}) = \infty \quad \text{a.s.} \tag{4}$$

Thus, if L has finite fourth moment, then almost surely, for large enough n,

$$\{\lambda_2(\mathbf{X}), \ldots, \lambda_n(\mathbf{X})\} \subset D(0, s_1(\widetilde{\mathbf{X}})) \quad \text{and} \quad \lambda_1(\mathbf{X}) \in D(m\sqrt{n}, s_1(\widetilde{\mathbf{X}})). \tag{5}$$

This shows that the eigenvalue $\lambda_1(\mathbf{X})$ of largest module and the rest of the spectrum $\lambda_2(\mathbf{X}), \ldots, \lambda_n(\mathbf{X})$ localize on different discs when $m \neq 0$. Relaxing the centering condition creates an isolated largest eigenvalue, which goes to infinity, whereas the rest of the spectrum remains roughly near the centered unit disc. A single eigenvalue plays no role in the asymptotic behavior of the ESD of $\mathbf{X}$, due to the empirical scaling $1/n$. For that reason, one can expect that the circular law theorem remains valid when $m \neq 0$, despite the fact that the largest eigenvalue $\lambda_1(\mathbf{X})$ goes to infinity when $n \to \infty$. This is confirmed by the following theorem, which is our main result, illustrated by figure 2.

Theorem 1.2 (Circular law for non-central entries). *Let $(X_{i,j})_{1\leqslant i,j<\infty}$ be an infinite array of i.i.d. complex random variables of law L, with mean m, variance $\sigma^2 = 1$, and finite fourth moment. Then the ESD of $\mathbf{X} = (n^{-1/2}X_{i,j})_{1\leqslant i,j\leqslant n}$ tends, as $n \to \infty$, to the uniform distribution on the unit disc $D(0,1) = \{z \in \mathbb{C}; |z| \leqslant 1\}$.*

Note that Silverstein has shown in [32] that if L is real with mean $m > 0$, variance $\sigma^2 = 1$, and finite fourth moment $\mathbb{E}[|X_{1,1}|^4] < \infty$, then almost surely, $\lambda_1(\mathbf{X})$ is real for large enough n, and

$$\lambda_1(\mathbf{X}) \quad \xrightarrow[n\to\infty]{\text{a.s.}} \quad +\infty \quad \text{and} \quad s_1(\mathbf{X}) \quad \xrightarrow[n\to\infty]{\text{a.s.}} \quad +\infty. \tag{6}$$

Moreover, by denoting $\mathcal{N}(0,1)$ the standard Gaussian on $\mathbb{R}$

$$\sqrt{n}\big(\lambda_1(\mathbf{X}) - m\sqrt{n}\big) \quad \xrightarrow[n\to\infty]{D} \quad \mathcal{N}(0,1).$$

These results can be easily adapted to the case where L is complex with mean $m \neq 0$. For any complex number $z \in \mathbb{C}$ with $|z| = 1$, and any $n \times n$ matrix $\mathbf{A}$, we have $s_k(z\mathbf{A}) = s_k(\mathbf{A})$ whereas $\lambda_k(z\mathbf{A}) = z\lambda_k(\mathbf{A})$ for every $1 \leqslant k \leqslant n$.

Example 1.3 (Adjacency matrices of Erdős-Rényi random oriented graphs). *An oriented finite simple graph is a set of oriented edges (arrows) between a finite number of vertices (points), with at most one arrow between two vertices. The adjacency matrix of such a graph is defined by $\mathbf{A}_{i,j} = 1$ if there is an arrow from i to j and $\mathbf{A}_{i,j} = 0$ otherwise. The random version corresponds to fix some parameter $p \in (0,1)$, and to put an arrow between i and j with probability p, independently for each couple (i,j). Let n be the number of vertices. By theorem 1.2, almost*

surely, the ESD of $\mathbf{X} = (p(1-p)n)^{-1/2}\mathbf{A}$ *tends to the uniform distribution on the unit disc* $D(0,1)$ *as* $n \to \infty$*. Here* L *is the Bernoulli distribution* $p\delta_1 + (1-p)\delta_0$ *and* $m = p$ *and* $\sigma^2 = p(1-p)$*. However, by (5), almost surely, the largest eigenvalue* $\lambda_1(\mathbf{X})$ *goes to* $+\infty$ *at speed* $\sqrt{np/(1-p)}$*. One can drop the simple graph hypothesis by allowing at most* r *arrows between two nodes. In that case,* L *is a discrete probability distribution supported by* $\{0, 1, \ldots, r\}$*.*

The non-central version of the Marchenko-Pastur theorem can be easily deduced from the centered version by using the rank lemma 3.5. The same approach works for the non-central version of the Wigner theorem, see [3] and [5, ch. 11]. Such rank lemmas are based on interlacing inequalities derived from variational formulas for the eigenvalues of Hermitian matrices. In constrast, no such rank lemmas or variational formulas exist for the eigenvalues of non-Hermitian matrices. However, rank lemmas are of course available for the singular values. The logarithmic potential approach used in the proof of the circular law theorem needs the control of the extremal singular values, which is not provided by rank lemmas.

The spectrum of non-normal matrices is very sensitive to small perturbations, even of finite rank. Let us recall a striking example taken from [5, p. 292]. Consider the $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ defined by

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

where $\varepsilon_n = n^{-\alpha}$ for some $\alpha > 0$. We have $\mathrm{rank}(\mathbf{A} - \mathbf{B}) = 1$ and $\|\mathbf{A} - \mathbf{B}\|_2 = \varepsilon_n$. However, $\lambda_k(\mathbf{A}) = 0$ while $\lambda_k(\mathbf{B}) = (\varepsilon_n e^{2(k-1)\pi\sqrt{-1}})^{1/n}$ for every $1 \leqslant k \leqslant n$. In particular, the ESD of $\mathbf{A}$ is δ_0 while the ESD of $\mathbf{B}$ is supported on the centered circle of radius $\varepsilon_n^{1/n}$ of the complex plane. Moreover, $\varepsilon_n^{1/n} \to 1$ and the LSD of $\mathbf{B}$ 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* [35]. If we denote by $\Lambda(\mathbf{A}) = \{\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})\}$ the spectrum of $\mathbf{A}$, then for any $\varepsilon > 0$, the $(\|\cdot\|_2, \varepsilon)$-pseudo-spectrum $\Lambda_\varepsilon(\mathbf{A})$ of $\mathbf{A}$ is given by

$$\Lambda_\varepsilon(\mathbf{A}) = \bigcup_{\|\mathbf{A}-\mathbf{B}\|_2 \leqslant \varepsilon} \Lambda(\mathbf{B}) = \{z \in \mathbb{C} \text{ such that } s_n(\mathbf{A} - z\mathbf{I}) \leqslant \varepsilon\}.$$

For a normal matrix, the pseudo-spectrum is just the ε-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 $\Lambda_\varepsilon(\mathbf{A}) \to \Lambda(\mathbf{A})$ as $\varepsilon \to 0$, for any fixed $n \times n$ matrix $\mathbf{A}$. However, this says nothing on the behavior of $\Lambda_{\varepsilon_n}(\mathbf{A})$ as $n \to \infty$ where $\mathbf{A}$ is the $n \times n$ truncation of an infinite table.

Outline

The proof of theorem 1.2 is given in the next section of this article. More precisely, we show that theorem 1.1 implies theorem 1.2. The last section collects some crucial tools used in the proof of theorem 1.2.

Extensions and open problems

We ignore if the method used by Ginibre [12] and Edelman [10] for the computation of the spectral density in the Gaussian centered case can be adapted to the Gaussian non-central case. An additional log-linear term forbids some crucial cancellations in the integral.

Tao & Vu [33] has shown that theorem 1.1 remains valid if one replaces the finite fourth moment hypothesis $\mathbb{E}[|X_{1,1} - m|^4] < \infty$ by $\mathbb{E}[|X_{1,1} - m|^{2+\varepsilon}] < \infty$ for some $\varepsilon > 0$. In our reduction to the centered case, the finite fourth moment hypothesis is used to bound $s_1(\widetilde{\mathbf{X}})$ via (3) and then $\lambda_1(\mathbf{X})$ via (5). Under the weaker moment condition of Tao & Vu, one can hopefully use instead the probabilistic bound for the condition number [33, cor. 2.10 p. 8]. Recall that the condition number of $\mathbf{A}$ is $\kappa(\mathbf{A}) = \|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2 = s_1(\mathbf{A})/s_n(\mathbf{A})$. However, the spectrum localization given by (5) is not meaningful without the finite fourth moment condition.

One can also try to obtain a sparse version of the non-central circular law theorem by using the techniques used for the centered case in [33] and [18]. Another possible extension is to drop the identical distribution of the entries, as already mentioned by Pan & Zhou [24] and Tao & Vu [33] for the centered case. See also the discussion in section 10.8 of Bai-Silverstein's book [5].

Rider [26, 25] has shown recently that when L is a centered complex Gaussian, the fluctuation of the largest singular value is given by a Gumbel type extreme distribution rather that by a Tracy-Widom distribution as for classical Hermitian random matrices models. The asymptotic behavior (convergence and fluctuations) of the extremal eigenvalues and singular values remains untouched when L is generic. It seems heuristically that the non-Hermitian nature of the matrices gives more space to the spectrum, which can thus uncorrelate as $n \to \infty$ and finally behave roughly like i.i.d. random variables. Quoting Girko [14], "*if all entries of a random matrix which lie on the main diagonal and above are independent and belong to the domain of attraction of an infinitely divisible law then, under some conditions, the normalized real and imaginary parts of eigenvalues behave like some order statistics when s, n tend to infinity*". Note that it has been

shown by Capitaine & Donati-Martin & Féral [9] that the fluctuations of λ_1 are non-universal in certain Hermitian models with low rank deformation. A related natural question for non-central entries is to ask about the asymptotic behavior of the second largest eigenvalue $\lambda_2(\mathbf{X})$ or singular value $s_2(\mathbf{X})$ in the case where $m \neq 0$, the behavior of $\lambda_1(\mathbf{X})$ and $s_1(\mathbf{X})$ being given already by (5) (see [32] for more details). The behavior of the subdominant eigenvalue λ_2 of certain random matrices is considered in [15, 16]. The study of the asymptotic behavior of $s_n(\mathbf{X})$ and $\lambda_n(\mathbf{X})$ is notoriously more difficult than the study of the asymptotic behavior of $s_1(\mathbf{X})$ and $\lambda_1(\mathbf{X})$. Another interesting question is the asymptotic behavior of the eigenvectors, as in [4].

Notations and conventions

Matrices names are written in bold faced upper case letters. The identity matrix is denoted $\mathbf{I}$. The notation $\mathbf{A}^* = \overline{\mathbf{A}^\top}$ stands for the conjugate transpose of $\mathbf{A}$. The notation $\mathbf{1}$ stands for a matrix with all entries equal to 1. The real and imaginary parts of the complex number z are written $\mathfrak{Re}(z)$ and $\mathfrak{Im}(z)$, and in particular $z\overline{z} = |z|^2 = \mathfrak{Re}(z)^2 + \mathfrak{Im}(z)^2$. The letter i is used for integer indexing and is never used for $\sqrt{-1}$. The letter m is used to denote a mean and is never used for indexing. The symbol $\mathbb{E}$ denotes the mathematical expectation. If Z is a complex random variable of mean $m = \mathbb{E}[Z] \in \mathbb{C}$, we define its variance by $\sigma^2 = \mathbb{E}[|Z-m|^2] = \mathbb{E}[|Z|^2] - |m|^2$, which is the trace of the covariance matrix of Z viewed as a random vector of $\mathbb{R}^2$. The random variables are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and the notation "*a.s.*" stands for "$\mathbb{P}$-*almost-surely*".

2 Proof of theorem 1.2

This section is devoted to the proof that theorem 1.1 implies theorem 1.2. We will adopt the logarithmic potential approach. Lower bounds for the smallest singular value of $\mathbf{X} - z\mathbf{I}$ are crucial. The ones used by Götze & Tikhomirov [19, 18] and by Pan & Zhou [24] do not allow to deal with our non-central random matrices, due to the norm of the deformation (2). We will circumvent this difficulty by using instead the polynomial-polynomial bounds obtained recently by Tao & Vu [33].

In this section, $\mathbf{X}$ is the $n \times n$ random matrix of theorem 1.2, and we assume from now that $m \neq 0$. The ESD of $\mathbf{X}$ is denoted

$$\mu_n = \frac{1}{n}(\delta_{\lambda_1(\mathbf{X})} + \cdots + \delta_{\lambda_n(\mathbf{X})}).$$

We also define, for every complex number $z \in \mathbb{C}$,

$$\nu_{n,z} = \frac{1}{n}(\delta_{s_1(\mathbf{X}-z\mathbf{I})} + \cdots + \delta_{s_n(\mathbf{X}-z\mathbf{I})}),$$

which is the ESD of the positive semi-definite Hermitian matrix

$$\mathbf{H}(z) = \sqrt{(\mathbf{X} - z\mathbf{I})(\mathbf{X} - z\mathbf{I})^*}.$$

The logarithmic potential U_n of the ESD μ_n of $\mathbf{X}$ is given by (see section 3.1)

$$\begin{aligned} U_n(z) &= -\int_{\mathbb{C}} \log|z - z'|\, d\mu_n(z') \\ &= -\frac{1}{n}\log|\det(\mathbf{X} - z\mathbf{I})| \\ &= -\int_0^\infty \log(x)\, d\nu_{n,z}(x). \end{aligned}$$

Define the $n \times n$ random matrix $\widetilde{\mathbf{X}}$ as in (5) by

$$\widetilde{\mathbf{X}} = \mathbf{X} - n^{-1/2} m\mathbf{1} = n^{-1/2}(X_{i,j} - m)_{1 \leqslant i,j \leqslant n}$$

where $\mathbf{1}$ is the $n \times n$ matrix with all entries equal to 1. Define also similarly

$$\widetilde{\mu}_n = \frac{1}{n}(\delta_{\lambda_1(\widetilde{\mathbf{X}})} + \cdots + \delta_{\lambda_n(\widetilde{\mathbf{X}})})$$

and

$$\widetilde{\nu}_{n,z} = \frac{1}{n}(\delta_{s_1(\widetilde{\mathbf{X}} - z\mathbf{I})} + \cdots + \delta_{s_n(\widetilde{\mathbf{X}} - z\mathbf{I})})$$

and

$$\begin{aligned} \widetilde{U}_n(z) &= -\int_{\mathbb{C}} \log|z - z'|\, d\widetilde{\mu}_n(z') \\ &= -\frac{1}{n}\log\left|\det(\widetilde{\mathbf{X}} - z\mathbf{I})\right| \\ &= -\int_0^\infty \log(x)\, d\widetilde{\nu}_{n,z}(x). \end{aligned}$$

Let μ be the circular law, i.e the uniform law on $D(0,1)$, and let U be its logarithmic potential (see [19] and [24]). We consider first the behavior of $\widetilde{U}_n$. By (1) and (3),

$$|\lambda_1(\widetilde{\mathbf{X}})| \leqslant s_1(\widetilde{\mathbf{X}}) \xrightarrow[n\to\infty]{a.s.} 2 < \infty.$$

Thus, almost surely, the sequence of probability distributions $(\widetilde{\mu}_n)$ are supported in a common compact subset of $\mathbb{C}$. Since by the centered circular law (theorem 1.1), $\mu_n \to \mu$ weakly as $n \to \infty$, the Lower Envelope Theorem 3.3 gives

$$\liminf_{n\to\infty} \widetilde{U}_n(z) = U(z)$$

for quasi-every $z \in \mathbb{C}$. Set now

$$\Delta_n(z) := U_n(z) - \widetilde{U}_n(z) = \int_0^\infty \log(x)\, d\nu_{n,z}(x) - \int_0^\infty \log(x)\, d\widetilde{\nu}_{n,z}(x).$$

Let us denote by $F_{n,z}$ and $\widetilde{F}_{n,z}$ the cumulative distribution functions of $\nu_{n,z}$ and $\widetilde{\nu}_{n,z}$ respectively defined by

$$F_{n,z}(x) = \nu_{n,z}((-\infty, x]) \quad \text{and} \quad \widetilde{F}_{n,z}(x) = \widetilde{\nu}_{n,z}((-\infty, x])$$

for every $x \in \mathbb{R}$. Let (α_n) and (β_n) be as in lemma 2.1 below. By an integration by parts (lemma 3.1), almost surely, for every $z \in \mathbb{C}$ and large enough n,

$$\Delta_n = \int_{\alpha_n}^{\beta_n} \log(x)(d\nu_{n,z}(x) - d\widetilde{\nu}_{n,z}(x)) = \int_{\alpha_n}^{\beta_n} \frac{1}{x}(F_{n,z}(x) - \widetilde{F}_{n,z}(x))\, dx$$

and hence

$$|\Delta_n| \leqslant (\log(\beta_n) - \log(\alpha_n))\left\|F_{n,z} - \widetilde{F}_{n,z}\right\|_\infty.$$

Now, the Bai-Thomson rank Lemma 3.5 gives

$$\left\|F_{n,z} - \widetilde{F}_{n,z}\right\|_\infty \leqslant \frac{1}{n}\mathrm{rank}(\mathbf{X} - \widetilde{\mathbf{X}}) = \frac{1}{n}\mathrm{rank}(n^{-1/2} m\mathbf{1}) = \frac{1}{n}$$

and therefore, almost surely, for every $z \in \mathbb{C}$, and for large enough n,

$$|\Delta_n| \leqslant \frac{1}{n}(\log(\beta_n) - \log(\alpha_n)) \underset{n\to\infty}{\longrightarrow} 0.$$

Consequently, almost surely, for quasi-every $z \in \mathbb{C}$,

$$\liminf_{n\to\infty} U_n(z) = \liminf_{n\to\infty} \widetilde{U}_n(z) = U(z). \tag{7}$$

The probability distributions (μ_n) are not supported in a common compact set, due to the behavior of the atom $a_n = \lambda_1(\mathbf{X})$ in μ_n. One can however make use of (7) with theorem 3.2. Since $m \neq 0$, (5) gives $|a_n| \to \infty$ a.s. and

$$\max(|\lambda_2(\mathbf{X})|, \ldots, |\lambda_n(\mathbf{X})|) \leqslant s_1(\widetilde{\mathbf{X}}) \to 2 < \infty \quad \text{a.s.}$$

Additionally, $p_n \log|a_n| \to 0$ a.s. Finally, theorem 3.2 used with $p_n = 1/n$ implies that almost surely, $\mu_n \to \mu$ weakly as $n \to \infty$. This achieves the proof of theorem 1.2.

Lemma 2.1. *Almost surely, for every $z \in \mathbb{C}$, there exist real numbers α, β and positive real numbers $c_\alpha > 0$ and $c_\beta > 0$ such that for large enough n,*

$$\alpha_n := c_\alpha n^\alpha \leqslant \min(s_n(\mathbf{X} - z\mathbf{I}), s_n(\widetilde{\mathbf{X}} - z\mathbf{I}))$$

and

$$\beta_n := c_\beta n^\beta \geqslant \max(s_1(\mathbf{X} - z\mathbf{I}), s_1(\widetilde{\mathbf{X}} - z\mathbf{I})).$$

Proof. First, by (3), almost surely, $s_1(\widetilde{\mathbf{X}}) \to 2$ as $n \to \infty$. Consequently, almost surely, for every $z \in \mathbb{C}$,

$$s_1(\widetilde{\mathbf{X}} - z\mathbf{I}) \leqslant s_1(\widetilde{\mathbf{X}}) + s_1(z\mathbf{I}) = s_1(\widetilde{\mathbf{X}}) + |z|^2 = \mathcal{O}_z(1)$$

and also almost surely, for every $z \in \mathbb{C}$,

$$s_1(\mathbf{X} - z\mathbf{I}) \leqslant s_1(\widetilde{\mathbf{X}} + z\mathbf{I}) + s_1(n^{-1/2} m \mathbf{1}) = s_1(\widetilde{\mathbf{X}} + z\mathbf{I}) + n^{1/2}|m| = \mathcal{O}_z(1) + \mathcal{O}(n^{1/2}).$$

On the other hand, by the Tao-Vu theorem 3.6 and the Borel-Cantelli lemma, there exist $B > 0$ and $C_2 > 0$ such that for every $z \in \mathbb{C}$, almost surely,

$$\min(s_n(\mathbf{X} - z\mathbf{I}), s_n(\widetilde{\mathbf{X}} - z\mathbf{I})) \geqslant C_2\, n^B$$

for large enough n. Since $s_n(\cdot)$ is continuous, it is locally uniformly continous, and one can then commute the universal quantifiers on z and ω. □

Remark 2.2. *Since $m \neq 0$, we have for every $z \in \mathbb{C}$, by using* (1) *and* (5),

$$s_1(\mathbf{X} - z\mathbf{I}) \geqslant |\lambda_1(\mathbf{X} - z\mathbf{I})| \xrightarrow[n\to\infty]{a.s.} +\infty$$

Thompson's interlacing inequalities (8) *and* (3) *give*

$$s_2(\mathbf{X} - z\mathbf{I}) \leqslant s_1(\widetilde{\mathbf{X}} - z\mathbf{I}) \leqslant s_1(\widetilde{\mathbf{X}}) + s_1(z\mathbf{I}) = s_1(\widetilde{\mathbf{X}}) + |z|^2 \xrightarrow[n\to\infty]{a.s.} 2 + |z|^2 < \infty.$$

3 Tools

This section gathers some tools used in our proof of the non-central version of the circular law theorem (theorem 1.2). The following lemma is a special case of the integration by parts formula for the Lebesgue-Stieltjes integral (with atoms).

Lemma 3.1 (Integration by parts). *If $a_1, \ldots, a_n, b_1, \ldots, b_n \in [\alpha, \beta] \subset \mathbb{R}$, and F_μ and F_ν are the cumulative distribution functions of $\mu = \frac{1}{n}(\delta_{a_1} + \cdots + \delta_{a_n})$ and $\nu = \frac{1}{n}(\delta_{b_1} + \cdots + \delta_{b_n})$ respectively, then for any smooth $f : [\alpha, \beta] \to \mathbb{R}$,*

$$\int_\alpha^\beta f(x)\, d\mu(x) - \int_\alpha^\beta f(x)\, d\nu(x) = \int_\alpha^\beta f'(x)(F_\mu(x) - F_\nu(x))\, dx.$$

In particular, when f is non decreasing,

$$\left|\int_\alpha^\beta f(x)\,d\mu(x) - \int_\alpha^\beta f(x)\,d\nu(x)\right| \leqslant (f(\beta)-f(\alpha))\|F_\mu - F_\nu\|_\infty.$$

Proof. One can assume by continuity that $a_1,\dots,a_n,b_1,\dots,b_n$ are all different. We reorder $a_1,\dots,a_n,b_1,\dots,b_n$ into $c_1 \leqslant \cdots \leqslant c_{2n}$. For every $1 \leqslant k \leqslant 2n$, set $\varepsilon_k = +1$ if $c_k \in \{a_1,\dots,a_n\}$ and $\varepsilon_k = -1$ if $c_k \in \{b_1,\dots,b_n\}$. We have

$$\int_\alpha^\beta f(x)\,d\mu(x) - \int_\alpha^\beta f(x)\,d\nu(x) = \frac{1}{n}\sum_{k=1}^{n}(f(a_i)-f(b_i)) = \frac{1}{n}\sum_{k=1}^{2n}\varepsilon_k f(c_k).$$

By an Abel transform, we get by denoting $S_k = \varepsilon_1 + \cdots + \varepsilon_k$,

$$\sum_{k=1}^{2n}\varepsilon_k f(c_k) = -\sum_{k=1}^{2n-1} S_k(f(c_{k+1})-f(c_k)) + S_{2n}f(c_{2n}).$$

Since $F_\mu - F_\nu$ is constant and equal to S_k on $[c_k, c_{k+1}[$,

$$S_k(f(c_{k+1})-f(c_k)) = \int_{c_k}^{c_{k+1}} f'(x)(F_\mu(x)-F_\nu(x))\,dx.$$

It remains to notice that $S_{2n} = F_\mu(c_{2n}) - F_\nu(c_{2n}) = 0$. □

3.1 Logarithmic potential tools

We give here some logarithmic potentials tools, already used by Götze & Tikhomirov [19] and Pan & Zhou [24] in order to prove centered versions of the circular law theorem, following a suggestion made by Khoruzhenko.

1. The *logarithmic potential* U_μ of a probability measure μ on $\mathbb{C}$ is the function defined for every $z \in \mathbb{C}$ by

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

2. The *logarithmic energy*[3] $\mathcal{E}(\mu)$ of a probability distribution μ on $\mathbb{C}$ is

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

[3]Up to a sign, it is the *free entropy* in Voiculescu's free probability theory, see [20, sec. 5.3].

3. The *energy* $\mathcal{E}(K)$ of a compact subset K of $\mathbb{C}$ is

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

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

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

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

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

If $\mathrm{Cap}(B) = 0$ then B has zero Lebesgue measure. A property is said to hold *quasi-everywhere* on B if the set where it does not hold has zero capacity.

Theorem 3.2. *Let (ν_n) be a sequence of probability distributions on $\mathbb{C}$ with support in a common compact subset of $\mathbb{C}$, (a_n) be a sequence of complex numbers such that $|a_n| \to +\infty$, and (p_n) be a sequence of real numbers in $[0,1]$ such that $p_n \to 0$. Define $\mu_n = (1-p_n)\nu_n + p_n\delta_{a_n}$. If $p_n \log|a_n| \to 0$ and $\liminf_{n\to\infty} U_{\mu_n} = U_\mu$ quasi-everywhere on $\mathbb{C}$ for some compactly supported probability distribution μ on $\mathbb{C}$, then $\mu_n \to \mu$ and $\nu_n \to \mu$ weakly as $n \to \infty$.*

Proof. We have 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,

$$p_n \log|z - a_n| = p_n \log|a_n| + p_n \log\left|\frac{z}{a_n} - 1\right| \to 0.$$

Consequently, $\liminf_{n\to\infty} U_{\nu_n} = \liminf_{n\to\infty} U_{\mu_n} = U_\mu$ quasi-everywhere on $\mathbb{C}$. The sequence (ν_n) is tight (i.e. weakly retatively compact) since it is supported in a common compact subset of $\mathbb{C}$. Let $\mu_{n_k} \to \eta$ be a weakly converging subsequence, to some probability distribution η (necessarily supported in the same compact subset). By the Lower Envelope Theorem 3.3, $\liminf_{n\to\infty} U_{\nu_n} = U_\eta$ 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 3.4, we get $\eta = \mu$. Therefore, μ is the unique weak adherence value of (ν_n). By a standard theorem of Prohorov [8], the set of probability distributions on a compact space equipped with the weak topology is metrizable and compact. In follows then that $\nu_n \to \mu$ weakly. Since $p_n \to 0$, we get also that $\mu_n = (1-p_n)\nu_n + p_n\delta_{a_n} \to \mu$ weakly. □

Theorems 3.3 and 3.4 can be found in [29, th. 6.9 p. 73 and cor. 2.2 p. 98].

Theorem 3.3 (Lower Envelope Theorem). *Let (μ_n) be a sequence of probability distributions on $\mathbb{C}$ with support in a common compact subset of $\mathbb{C}$. If $\mu_n \to \mu$ weakly as $n \to \infty$, then $\liminf_{n\to\infty} U_{\mu_n} = U_\mu$ quasi-everywhere on $\mathbb{C}$.*

Theorem 3.4 (Unicity Theorem). *If μ and ν are two compactly supported probability distributions on $\mathbb{C}$ and if their logarithmic potential U_μ and U_ν coincide almost everywhere with respect to the Lebesgue measure on $\mathbb{R}^2$, then $\mu = \nu$.*

3.2 Singular values tools

For an $n \times n$ Hermitian matrix $\mathbf{H}$, we denote by $F_{\mathbf{H}} : \mathbb{R} \to [0,1]$ the cumulative distribution function of the ESD $\mu_{\mathbf{H}}$ of $\mathbf{H}$ defined by

$$F_{\mathbf{H}}(x) = \mu_{\mathbf{H}}((-\infty, x]) = \frac{1}{n}\text{card}\{k \in \{1,\dots,n\} \text{ s.t. } \lambda_k(\mathbf{H}) - x \leqslant 0\}$$

for any $x \in \mathbb{R}$. The following lemma, due to Bai, shows that the rank of the deformation controls the effect on singular values.

Lemma 3.5 (Rank inequality). *For any $n \times m$ complex matrices $\mathbf{A}$ and $\mathbf{B}$,*

$$\left\|F_{\sqrt{\mathbf{AA}^*}} - F_{\sqrt{\mathbf{BB}^*}}\right\|_\infty \leqslant \frac{1}{n}\text{rank}(\mathbf{A} - \mathbf{B}).$$

See [3, Lemma 2.6 page 621] or [5, Theorem 11.43 page 362] for a proof based on interlacing inequalities derived from the Courant-Fischer min-max variational formulas for singular values. Bai states this lemma with $F_{\mathbf{AA}^*}$ and $F_{\mathbf{BB}^*}$ instead of $F_{\sqrt{\mathbf{AA}^*}}$ and $F_{\sqrt{\mathbf{BB}^*}}$ but the proof are identical since $\sqrt{\cdot}$ is non decreasing. The rank lemma 3.5 appears actually in a different form in the article [34, Theorem 3 page 76] published in 1976 by Thompson, who showed that for any $n \times n$ complex matrices $\mathbf{A}$ and $\mathbf{B}$ with $\text{rank}(\mathbf{A} - \mathbf{B}) \leqslant k$,

$$s_{i+k}(\mathbf{A}) \leqslant s_i(\mathbf{B}) \leqslant s_{i-k}(\mathbf{A}) \tag{8}$$

for any $1 \leqslant i \leqslant n$, where $s_r = +\infty$ if $r \leqslant 0$ and $s_r = 0$ if $r \geqslant n+1$. In particular,

$$[s_{n-k}(\mathbf{B})\,,\, s_{k+1}(\mathbf{B})] \subset [s_n(\mathbf{A})\,,\, s_1(\mathbf{A})].$$

Conversely, Thompson showed additionally that any couple of sequences of non negative real numbers which satisfy to the interlacing inequalities (8) are the singular values of two respective matrices $\mathbf{A}$ and $\mathbf{B}$ with $\text{rank}(\mathbf{A} - \mathbf{B}) \leqslant k$.

It is worthwhile to mention that the Bai-Thompson inequalities give neither an upper bound for the largest singular values $s_1(\mathbf{B}), \dots, s_k(\mathbf{B})$ nor a lower bound for the smallest singular values $s_{n-k+1}(\mathbf{B}), \dots, s_n(\mathbf{B})$, even in the case $k = 1$.

The following theorem is nothing else but Theorem 2.1 by Tao & Vu in [33, p. 6]. It gives a probabilistic polynomial bound for the smallest singular value of a random matrix deformed by a deterministic matrix.

Theorem 3.6 (Polynomial bounds for smallest singular values). *Let L be a probability distribution on $\mathbb{C}$ with finite and non-zero variance. For every constants $A > 0$ and $C_1 > 0$, there exists constants $B > 0$ and $C_2 > 0$ such that for every $n \times n$ random matrix $\mathbf{X}$ with i.i.d. entries of law L and every $n \times n$ deterministic matrix $\mathbf{C}$ with $s_1(\mathbf{C}) \leqslant n^{C_1}$, we have*

$$\mathbb{P}(s_n(\mathbf{X} + \mathbf{C}) \leqslant n^{-B}) \leqslant C_2 n^{-A}. \tag{9}$$

Notice that in theorem 3.6, the probability distribution L in not necessarily centered. We make use of this theorem in our proof of the non-central version of the circular theorem (theorem 1.2), with L of mean $m \in \mathbb{C}$ and $\mathbf{C} = -z\mathbf{I}$.

Remark 3.7 (The matrix inverse of a low rank deformation). *Concerning the behavior of the smallest eigenvalue and singular value under a rank 1 deformation, it is natural to ask about the expression of the inverse of the deformed matrix. The Sherman-Morrison formula [31] states that if $\mathbf{A}$ is an $n \times n$ invertible matrix and if u and v are two column n-vectors such that $1 + v^\top \mathbf{A}^{-1} u \neq 0$, then the rank 1 deformation $\mathbf{A} + uv^\top$ of $\mathbf{A}$ is invertible and*

$$(\mathbf{A} + uv^\top)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} uv^\top \mathbf{A}^{-1}}{1 + v^\top \mathbf{A}^{-1} u}.$$

In particular, it shows that the inverse of a rank one deformation is a rank one deformation of the inverse. In the same spirit, we have under the same assumptions

$$\det(\mathbf{A} + uv^\top) = (1 + v^\top \mathbf{A}^{-1} u) \det(\mathbf{A}).$$

Unfortunately, this gives, to the author knowledge, nothing really useful for the control of logarithmic potentials in the non-central case. The Sherman-Morrison formula is a particular case of the Woodbury formula which states that if $\mathbf{A}, \mathbf{C}, \mathbf{U}, \mathbf{V}$ are matrices of suitable dimensions then

$$(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V}^\top)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{V}^\top \mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^\top \mathbf{A}^{-1}.$$

In the same spirit, and under the same assumptions,

$$\det(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V}^\top) = \det(\mathbf{C}^{-1} + \mathbf{V}^\top \mathbf{A}^{-1}\mathbf{U}) \det(\mathbf{C}) \det(\mathbf{A}).$$

Acknowledgements. Part of this work was done during two visits to the Laboratoire Jean Dieudonné in Nice, France. The author would like to thank Pierre Del Moral and Persi Diaconis for their kind hospitality there.

References

[1] A. L. Andrew, *Eigenvalues and singular values of certain random matrices*, J. Comput. Appl. Math. **30** (1990), no. 2, 165–171.

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

[3] ———, *Methodologies in spectral analysis of large-dimensional random matrices, a review*, Statist. Sinica **9** (1999), no. 3, 611–677, With comments by G. J. Rodgers and J. W. Silverstein; and a rejoinder by the author.

[4] Z. D. Bai, B. Q. Miao, and G. M. Pan, *On asymptotics of eigenvectors of large sample covariance matrix*, Ann. Probab. **35** (2007), no. 4, 1532–1572, `arXiv:0708.1720v1 [math.PR]`.

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

[6] 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.

[7] 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.

[8] 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.

[9] M. Capitaine, C. Donati-Martin, and D. Féral, *The largest eigenvalue of finite rank deformation of large Wigner matrices: convergence and non-universality of the fluctuations*, preprint `arXiv:0706.0136v1 [math.PR]`, 2007.

[10] 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.

[11] A. Edelman and N. R. Rao, *Random matrix theory*, Acta Numer. **14** (2005), 233–297.

[12] J. Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, J. Mathematical Phys. **6** (1965), 440–449.

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

[14] ———, *Theory of random determinants*, Mathematics and its Applications (Soviet Series), vol. 45, Kluwer Academic Publishers Group, Dordrecht, 1990, Translated from the Russian.

[15] 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).

[16] 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.

[17] 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.

[18] F. Götze and A. Tikhomirov, *The Circular Law for Random Matrices*, preprint `arXiv:math/07093995v1 [math.PR]`, 2007.

[19] ———, *On the Circular Law*, preprint `arXiv:math/0702386v1 [math.PR]`, 2007.

[20] 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.

[21] R. A. Horn and Ch. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1990, Corrected reprint of the 1985 original.

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

[23] ———, *Random matrices*, third ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004.

[24] G. Pan and W. Zhou, *Circular law, extreme singular values and potential theory*, `arXiv:0705.3773v1 [math.PR]`, 2007.

[25] 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.

[26] ———, *Order statistics and Ginibre's ensembles*, J. Statist. Phys. **114** (2004), no. 3-4, 1139–1148.

[27] M. Rudelson, *Invertibility of random matrices: norm of the inverse*, preprint `arXiv:math/0507024v1 [math.FA]`, to appear in Annals of Mathematics, 2005.

[28] M. Rudelson and R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, preprint `arXiv:math/0703503v1 [math.PR]`, 2007.

[29] 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.

[30] A. Sen and A. Bose, *Another look at the moment method for large dimensional random matrices*, preprint `http://www.isical.ac.in/~statmath/html/publication/volume25.pdf`, May 2007.

[31] J. Sherman and W. J. Morrison, *Adjustment of an inverse matrix corresponding to a change in one element of a given matrix*, Ann. Math. Statistics **21** (1950), 124–127.

[32] 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.

[33] T. Tao and V. Vu, *Random Matrices: The circular law*, preprint `arXiv:0708.2895v1 [math.PR]`, 2007.

[34] 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.

[35] L. N. Trefethen and M. Embree, *Spectra and pseudospectra*, Princeton University Press, Princeton, NJ, 2005, The behavior of nonnormal matrices and operators.

[36] J. H. Wilkinson, *The algebraic eigenvalue problem*, Monographs on Numerical Analysis, The Clarendon Press Oxford University Press, New York, 1988, , Oxford Science Publications.

[37] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, *On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix*, Probab. Theory Related Fields **78** (1988), no. 4, 509–521.

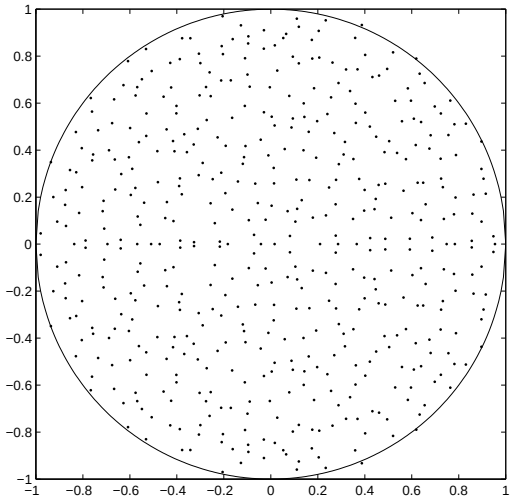

Figure 1: The spectrum of a single realization of $\mathbf{X}$. Here $n = 500$ and L is a standard Gaussian law $\mathcal{N}(0, 1)$ of mean $m = 0$ and variance $\sigma^2 = 1$. The plot illustrate well the circular law theorem in the central case $m = 0$ (theorem 1.1).

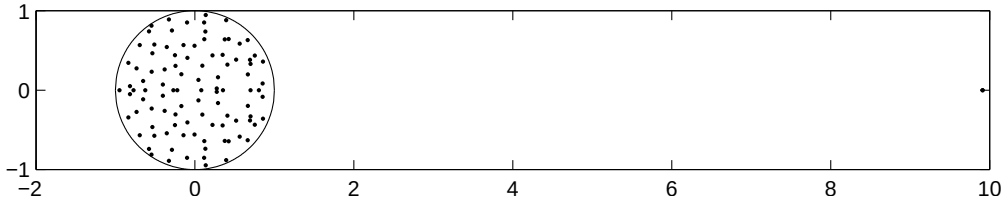

Figure 2: The typical behavior of the non-central case. The graphic shows the spectrum of a single realization of $\mathbf{X}$, in the case where $n = 100$ and L is the Gaussian distribution $\mathcal{N}(1, 1)$ of mean $m = 1$ and variance $\sigma^2 = 1$. The plot is in accordance with (5), which gives $\lambda_1(\mathbf{X}) \approx m\sqrt{n} = 10$, whereas the rest of the spectrum remains localized near the unit disc. The largest eigenvalue $\lambda_1(\mathbf{X})$ has weight $1/n \to 0$ in the ESD of $\mathbf{X}$, which explains why the non-central circular theorem is possible (theorem 1.2).

Djalil Chafaï **e-mail address:** chafai[@]math.ups-tlse.fr

Institut de Mathématiques de Toulouse (UMR CNRS 5219)
Université Paul Sabatier Toulouse III
118 route de Narbonne, F-31062 Toulouse Cedex 9, France.

Unité de Physiologie et Toxicologie Expérimentales (UMR INRA 181)
École Nationale Vétérinaire de Toulouse
23 Chemin des Capelles, B.P. 87614, F-31076 Toulouse Cedex 3, France.