EMPIRICAL SPECTRAL DISTRIBUTIONS OF RANDOM MATRICES: THE GIRKO METHOD, SMALLEST SINGULAR VALUE, AND UNIVERSALITY

KEVIN CONG, BRIGHT LIU

1. Introduction

One of the most classical problems in random matrix theory is to determine the distribution of eigenvalues of a 'nice' random matrix. This line of work began with that of Wigner [Wig57], in which he stated the beautiful *Wigner Semicircle Law*, Theorem 1.5. In order to state the question and theorems more formally, we need to establish the objects we are working with.

Definition 1.1 (Empirical Spectral Distribution)**.** Let M be a random matrix. The *empirical spectral distribution* (ESD) of M is the empirical distribution of the eigenvalues in $\mathbb{C}$; that is,

$$\mu_M(s,t) = \frac{1}{n}\left|\{1 \le i \le n, \Re\lambda_i \le s, \Im\lambda_i \le t\}\right|.$$

In the special case where $M \subset \mathcal{C}$ for $\mathcal{C}$ a class of Hermitian random matrices, all the eigenvalues are real. Therefore, it is better to consider the ESD as a subset of the real line.

Definition 1.2 (Empirical Spectral Distributions for Hermitian RMs)**.** Let $M \subset \mathcal{C}$ be a Hermitian random matrix. The *empirical spectral distribution* (ESD) of M is the empirical distribution of the eigenvalues in $\mathbb{R}$; that is,

$$\mu_M(s,t) = \frac{1}{n}\left|\{1 \le i \le n, \Re\lambda_i \le s, \Im\lambda_i \le t\}\right|.$$

Usually, it will be clear which regime we are discussing, so we will use the same terminology for both. In general, the spectrum of a matrix carries many important properties. For instance, in the symmetric Hermitian setting, the top eigenvalue equals to the operator norm, and the bottom eigenvalue is the smallest singular value; this characterizes the robust invertibility of the matrix and its condition number. In our case, we are interested in the distribution of these eigenvalues, and in particular how they behave asymptotically as n is large:

Question 1.3. What is the limiting distribution of eigenvalues of a general random matrix M? In particular, does there exist a distribution μ, such that

$$\mu_M \overset{n\to\infty}{\longrightarrow} \mu$$

in the sense of convergence in distribution?

Usually, one imposes some constraints and appropriate scaling on M, so that at least the distribution is bounded. Moreover, we should specify the sense of convergence.

Definition 1.4 (Convergence of Distributions)**.** Let μ_n be a sequence of distributions.

Date: Fall 2024.

(1) We say that $\mu_n \to \mu$ *in probability* if for any continuous and compactly supported test function $f \in C_c^1$, we have

$$\int f \mathrm{d}\mu_n \xrightarrow{P} \int f \mathrm{d}\mu,$$

where the integral convergence is in the probability sense.

(2) We say that $\mu_n \to \mu$ *almost surely* if for any continuous and compactly supported test function $f \in C_c^1$, we have

$$\int f \mathrm{d}\mu_n \xrightarrow{P} \int f \mathrm{d}\mu,$$

where the integral convergence is in the almost sure sense.

Intuitively, in order to characterize the 'distance' between measures μ_n, μ, we should look at how close the integrals of 'nice' functions with respect to μ_n and μ are; this motivates the above definition.

Now, we can finally state the semicircle law, alluded to previously.

Theorem 1.5. [Wigner Semicircle Law] Let the *semicircle law* μ be such that

$$\mu(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & |x| \leq 2 \\ 0 & |x| > 2 \end{cases}.$$

Let A_n be $n \times n$ Hermitian random matrices with independent entries, such that the diagonal entries are $(A_n)_{(ii)} \sim \mathcal{N}(0,2)$, and the off-diagonal entries are $(A_n)_{ij} \sim \mathcal{N}(0,1)$. Then the ESDs of M_n converge to the semicircle law:

$$\mu_{M_n} \overset{P,a.s.}{\to} \mu,$$

where the convergence can be taken to be either in probability or almost surely.

In the non-Hermitian setting, one has an analogous *circular law.*

Theorem 1.6. [Circular Law] Suppose that A_n are $n \times n$ random matrices with independent $\mathcal{N}(0,1)$ entries. Let μ denote the uniform distribution on the unit disk $\|z\| \leq 1$. Then

$$\mu_{M_n} \xrightarrow{P,a.s.} \mu,$$

where the convergence can be taken to either in probability or almost surely.

These were, in some sense, the first formal statements in the task to answer Question 1.3.[1] However, it requires many strong conditions; first, normality of the matrix entries; and second, the Hermitian property of the matrix. Since then, there has been much work to remove these constraints. First, in 1984 Girko [Gir84] provided a general program using the so-called 'logarithmic potential' to establish universality of the circular law. There followed a series of works to obtain the circular law under successively weaker assumptions. In 1997, Bai [Bai97] proved the circular law under the existence of a sixth moment and smoothness conditions. Around a decade later, Gotze-Tikhimirov [GT07] removed the assumption that the entries of the random matrix were continuous random variables, and then Pan-Zhou [PZ07] reduced the moment assumption to bounded fourth moments. In 2008, Tao and Vu [Tao08] reduced

[1]The Wigner law was really the first; the circular law was the first statement in the non-Hermitian setting.

the moment assumptions even further, all the way to the odd $\mathbb{E}[|x|^2]\text{polylog}(1+|x|)] < \infty$, which is stronger than even $2+\epsilon$ moments for any ϵ.

Removal of the first constraint reflects a general principle of *universality*. In particular, one needs to prove that the dependence of the measures μ_n on the distributions of the matrix entries is limited beyond the first and second moments. This phenomenon turns out to hold frequently in random matrix theory results, and in fact appears in other areas of high-dimensional probability as well, and is known as *universality*.[2] Proving formal universality statements is, however, very difficult, and at the heart of many seminal papers in the field; it turns out that there is a deep connection between ESD universality and lower bounds on the least singular value.

Removal of the second constraint is also difficult. It turns out that post-reduction to bounding the least singular value, one can consider matrices of the form A^*A, which are then Hermitian.

This line of work culminated in the following beautiful theorem of Tao and Vu [TVK10].

Theorem 1.7 (Theorem 1.7 of [TVK10])**.** Suppose that X_n, Y_n are $n \times n$ random matrices such that $(X_n)_{ij} \overset{iid}{\sim} \mathcal{F}_x$, $(Y_n)_{ij} \overset{iid}{\sim} \mathcal{F}_y$, where $\mathcal{F}_x, \mathcal{F}_y$ are distributions with zero mean and unit variance. For each n, let M_n be a sequence of deterministic $n \times n$ 'offset' matrices such that

$$\sup_n \frac{1}{n^2} \|M_n\|_2^2 < \infty.$$

Let $A_n = M_n + X_n$ and $B_n = M_n + Y_n$. Then

$$\mu_{\frac{1}{\sqrt{n}} A_n} - \mu_{\frac{1}{\sqrt{n}} B_n} \xrightarrow{P} 0.$$

Remark 1.8. To interpret the theorem, note the following. First, M_n is essentially an 'offset' matrix which allows our random matrix to not be centered. Second, suppose we take $M_n = \mathbf{0}_{n\times n}$ and $Y_n \sim \mathcal{N}(0,1)$. Via Theorem 1.6 and applications of the triangle inequality, we can find that

$$\mu_{\frac{1}{\sqrt{n}} A_n} \xrightarrow{P} \mu$$

for any ensemble of matrices A_n, with iid entries of mean 0 and variance 1. This theorem thus proves the 'universal' nature of the circular law, and in particular establishes Theorem 1.6 in full generality, i.e. for all distributions with two moments.

Corollary 1.9 (Circular Law, General)**.** Suppose that A_n are $n \times n$ random matrices with mean 0 and variance 1. Let μ denote the uniform distribution on the unit disk $\|z\| \leq 1$. Then

$$\mu_{M_n} \xrightarrow{P, a.s.} \mu,$$

where the convergence can be taken to either in probability or almost surely.

[2]For instance, in the SK model, etc; see Lectures 14, 15 of this course (Statistics 216) thus far.

1.1. **Organization.** The aim of this exposition is to discuss Girko's program, provide some historical discussion, and lastly provide a detailed explanation of the history of the program, ending with a proof of Theorem 1.7. We largely follow ideas and exposition in [TVK10] and [Tao10], though we omit a few proofs of basic lemmas and provide extra details in other proofs.[3] The remainder of the exposition is organized as follows. In section 2, we discuss the general principles and intuition of Girko's program; in particular, the connection between the smallest singular value and ESDs. We then discuss the history and results leading up to Theorem 1.7. In section 3, we rigorize the intuition in section 2. At this point, we will also state several classical linear-algebra lemmas that will be frequently used. In section 4, we detail singular value bounds and the remainder of the proof of Theorem 1.7. Lastly, in section 5, we provide a short conclusion.

2. Girko's Method: ESDs, Log Determinants, and Singular Values

We now establish the general principles and intuition of Girko's program. The main idea is a beautiful connection between three objects: the ESDs, a quantity known as the 'log potential', and the least singular value.

2.1. **Logarithmic Potential.** The first piece of motivation comes from the difficulty of the circular law over the semicircle law. Classically, there are proofs of the semicircle law via the trace moment method, and via Fourier-analytic methods (the Stieltjes transform)[4]. The main idea of the moment method is to note that the *moments*

$$\frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}}A_n\right)^k = \int z^k \mathrm{d}\mu_{\left(\frac{1}{\sqrt{n}}A_n\right)}$$

uniquely determine the spectral distribution $\mu_{\left(\frac{1}{\sqrt{n}}A_n\right)}$. Interestingly, this wholly fails in the non-Hermitian setting, because there exists distinct matrices $M_1 = I, M_2$ such that $M_1^n = M_2^n = I$; indeed, this is possible due to the presence of complex eigenvalues in M_2. Thus, one has to appeal to Fourier-analytic methods. This induces the following definition.

Definition 2.1 (Logarithmic Potential)**.** For a measure μ, its *logarithmic potential* is given by

$$s(\mu) = \int \log|z-w|\mathrm{d}\mu(w).$$

In particular, for the spectral measure $\mu_{\frac{1}{\sqrt{n}}A_n}$, we have

$$s(\mu_{\frac{1}{\sqrt{n}}A_n}) = \frac{1}{n}\sum_{i=1}^{n}\log\left|\frac{\lambda_i(A_n)}{\sqrt{n}} - z\right| = \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right|.$$

It turns out that the log potential is a 'characterizing' function of a distribution. In particular, there is the following result.

[3]In particular, we write out some of the lengthier computations and provide some extra details for some claims such as 'standard Chernoff/Hoeffding yield a certain concentration bound', when we feel that they are not completely trivial. On the other hand, we omit the very end case in the proof, due to the large amount of casework involved which relies largely on the already-presented ideas. We encourage the reader to read [TVK10], which is presented very nicely.

[4]See, for example, [Kem16] for a reference

Theorem 2.2 (Log-Potential Continuity Theorem)**.** If $s(\mu_n) \to s(\mu)$ almost surely, then $\mu_n \to \mu$ almost surely. The result also holds for convergence in probability.

We will not prove this here, in part because it is not directly necessary for proving the circular law via universality. However, it will be an important motivating idea (see the next section). For some intuition, one can think of this as a variant of the usual *characteristic function* of measures, which is used to prove other convergence-in-distribution type of results, e.g. the classical CLT.

2.2. **Least Singular Values.** The second piece of motivation comes from looking at stability results on the spectrum. In particular, we have the following.

Fact. Let M be a square matrix and $z \in \mathbb{C}$. Then $\|(M - zI)^{-1}\|_{\text{op}} \geq R$ if and only if there exists E such that $\|E\|_{\text{op}} \leq \frac{1}{R}$ and $M + E$ has z as an eigenvalue.

Proof. The intuition is that the appropriate eigenvector of $M + E$ will give a singular value to $M + E - zI = (M - zI) + E$, which then relates the operator norm of $M - zI$ to that of E.

First suppose such E exists; let v be a unit eigenvector of $M + E$ corresponding to the eigenvalue z, then $(M + E - zI)v = \mathbf{0}$. We thus find that

$$\|(M - zI)(v)\| = \|Ev\| \leq \|E\|_{\text{op}} \leq \frac{1}{R}.$$

Now using the fact that for any matrix A, $\|A\|_{\text{op}}\|A^{-1}\|_{\text{op}} \geq 1$, it follows that

$$\|(M - zI)^{-1}\|_{\text{op}} \geq \frac{1}{R}.$$

This proves the 'if' direction.

Now assume $\|(M - zI)^{-1}\|_{\text{op}} \geq R$. Pick a unit vector v such that $w = (M - zI)^{-1}v$ satisfies $\|w\| \geq R$. Then $(M - zI)w = v$. Set $E = v\frac{w^*}{\|w\|^2}$. Then since the operator norm is submultiplicative and the operator norm of a vector[5] is its norm by Cauchy-Schwarz, we see that

$$\|E\|_{\text{op}} \leq \|v\| \cdot \frac{\|w^*\|}{\|w\|^2} \leq \frac{1}{R}.$$

Moreover,

$$(M + E - zI)w = (M - zI)w - Ew = v - v\frac{w^*w}{\|w\|^2} = v - v = 0.$$

Hence E satisfies the requirements. This completes the proof of both directions. □

Morally, this suggests that if we do not have guarantees on the least singular value of $M - zI$, then the spectrum of M is wild to small perturbations. Thus, in order to show asymptotic results on the ESD, one should expect connections to least singular value bounds. Indeed, the logarithmic potential for the spectral measure can be written as a sum of $\log s_i$, where s_i are the singular values of $\frac{1}{\sqrt{n}}A_n - zI$. Hence, the second main idea of the Girko program is to establish lower bounds on the least singular value $\sigma_n\left(\frac{1}{\sqrt{n}}A_n - zI\right)$, and then apply these in bounding the logarithmic potential. As we will see, the necessary ingredient will be a probabilistic inverse-polynomial bound on the least singular value, which controls the necessary terms.

[5]considered as an $1 \times n$ or $n \times 1$ matrix

3. Rigorous Formulation of Girko's Method: Sufficiency of Log Potentials

3.1. **The Replacement Principle: Universality of Log Potentials Suffices.** We will now make rigorous the first idea of Girko. In the previous section, we established the intuition that the log-potentials roughly characterize distributions. The work of [TVK10] focuses on proving a universality result. Recall that the aim is to show Theorem 1.7. Analogously to Theorem 2.2, it should suffice to prove that

$$s(\mu_{\frac{1}{\sqrt{n}}A_n}) - s(\mu_{\frac{1}{\sqrt{n}}B_n}) \xrightarrow{P} 0.$$

Formally, we have the following result.

Theorem 3.1 (Theorem 2.1 of [TVK10], 'Replacement Principle')**.** Suppose that A_n, B_n are random matrix ensembles. Moreover, suppose that:

(1) The quantity $\frac{1}{n^2}\|A_n\|_2^2 + \frac{1}{n^2}\|B_n\|_2^2$ is bounded in probability, and
(2) For almost every $z \in \mathbb{C}$, the difference in log potentials of $\frac{1}{\sqrt{n}}A_n$ and $\frac{1}{\sqrt{n}}B_n$ converges to 0; namely,

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right| - \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI\right)\right| \xrightarrow{P} 0.$$

Then we have the convergence in ESDs

$$\mu_{\frac{1}{\sqrt{n}}A_n} - \mu_{\frac{1}{\sqrt{n}}B_n} \xrightarrow{P} 0.$$

The remainder of this section will largely be dedicated to proving this result. However, at this point, we find it useful to state several lemmas which will be repeatedly used, both in this section and throughout the remainder of this exposition.

3.2. **Quick Preliminaries: Relevant Linear-Algebra Tools, a DCT variant.** We will state these tools consecutively and largely without proof; one can find detailed proofs in the appendices of [TVK10]. We first have four linear-algebra lemmas; they are largely used to bound eigenvalues and singular values.

Lemma 3.2 (Cauchy's Interlacing Law for Rectangular Submatrices)**.** Let A be an $n \times n$ matrix and A' be the $(n-k) \times n$ matrix consisting of the first $n-k$ rows of A. Let $\sigma_i(M)$ denote the singular values of a matrix M in descending order. Then

$$\sigma_i(A) \geq \sigma_i(A') \geq \sigma_{i+k}(A).$$

Lemma 3.3 (Weyl Comparison Inequality for Second Moments)**.** Let A be an $n \times n$ matrix with generalized eigenvalues $\lambda_i(A)$ and singular values $\sigma_i(A)$. Then we have

$$\sum_i |\lambda_i(A)|^2 \leq \sum_i \sigma_i(A)^2 = \|A\|_2^2.$$

Lemma 3.4 (Weyl Comparison Inequality for Products)**.** Let A be an $n \times n$ matrix with generalized eigenvalues $\lambda_i(A)$ and singular values $\sigma_i(A)$. Then we have $\forall J \in [n]$ that

$$\prod_{i=1}^{J} |\lambda_i| \leq \prod_{i=1}^{J} \sigma_i(A)$$

and as a corollary

$$\prod_{i=J}^{n} \sigma_i(A) \leq \prod_{i=J}^{n} |\lambda_i|.$$

The next result is a bit odd at first glance.

Lemma 3.5 (Negative Second Moment of Singular Values and Distances)**.** Let $1 \leq n' \leq n$ and let A be an $n' \times n$ matrix with full rank and singular values σ_i. Let X_i denote the rows and $W_i = \mathrm{span}(X_{-i})$ denote the span of all rows excluding X_i. Then we have the identity

$$\sum_i \sigma_i^{-2} = \sum_i \mathrm{dist}(X_i, W_i)^{-2}.$$

Proof Sketch. Let Y_i be the i-th column of A^{-1}. First, note that $AY_i = e_i$ by definition. In particular, Y_i is a normal vector to $\mathrm{span}(X_{-i})$; hence, $\frac{Y_i}{\|Y_i\|}$ is the unit normal. It follows from the distance formula that

$$\mathrm{dist}(X_i, W_i) = X_i \cdot \left(\frac{Y_i}{\|Y_i\|} \right) = \frac{1}{\|Y_i\|}.$$

Summing across all i, we know that

$$\begin{aligned}
\sum_{i=1}^{n} \mathrm{dist}(X_i, W_i)^{-1} &= \sum_{i=1}^{n} \|Y_j\|^2 = \sum_{i,j} (A^{-1})_{ij}^2 \\
&= \mathrm{tr}((A^{-1})^* A^{-1}) = \sum_{i=1}^{n} \lambda_j((A^{-1})^* A^{-1}) \\
&= \sum_{i=1}^{n} \sigma_j(A)^{-2}.
\end{aligned}$$

This completes the proof. □

We will discuss this lemma more later on in section 4, because it hints at an important perspective shift, in using the distances $\mathrm{dist}(X_j, W_j)$ to analyze the least singular value and log-determinants which appear in the log-potential. This will be an important idea in the analysis of the log-potential itself.

We also have the following variant of DCT, where we replace domination by a UI-type condition. The proof is analogous to that of the standard truncation-style proof of the usual DCT; we will omit it.

Lemma 3.6 (Lemma 3.1 of [TVK10], 'DCT variant')**.** Suppose we have measurable functions f_n on a measure μ such that $f_n \xrightarrow{P} 0$ and the f_n are UI in the sense that there is some $\delta > 0$ for which the moments $F_n = \int |f_n(x)|^{1+\delta} d\mu$ are bounded in probability. Then $\int f_n(x) d\mu \xrightarrow{P} 0$.

3.3. **A Note on Notation and Simplifications.** To reconcile all of the notation used commonly above, we will just note that in general, unless otherwise stated, $\lambda_i(A)$ refers to the eigenvalues of A in decreasing order of magnitude, and $\sigma_i(A)$ refers to the singular values of A in decreasing order. If the matrix is obvious, we will sometimes omit A. Moreover, we will frequently use big-O notation in the standard way. We will be a bit loose about this; for instance, we will often find statement such as: $X_n \leq \epsilon$ with probability $1 - \wr(1) - \mathcal{O}(\epsilon)$, and note that this implies $X_n \xrightarrow{P} 0$. These statements can be made rigorous, but for the sake of simplicity of exposition, we will be more rough in making such assertions.

3.4. **Proof of Theorem 3.1.** In the remainder of this section, we aim to prove this result. The proof is relatively technical, and relies largely on Fourier analytic methods. The main point is to first relate convergence in distribution to convergence of the usual complex characteristic function $m_\mu(u, v)$. Then, $m_\mu(u, v)$ can be explicitly computed by an integral via an identity of Girko. The remainder of the argument is a relatively standard analytic argument by truncating the integrand and analyzing different parts of the integral.

3.4.1. *Boundedness of the Spectrum.* We need to first establish that the spectrum is generally not very big with high probability; this will be important in establishing later bounds and also using Lemma 3.6.

Lemma 3.7 (Spectrum is Usually Bounded)**.** Fix some $\epsilon > 0$. Then we have with 'high probability' $1 - \mathcal{O}(\epsilon) - \mathcal{O}(1)$, we have the bound

$$\mu_{\frac{1}{\sqrt{n}}A_n}(z \in \mathbb{C}, |Z| \geq R) \leq \mathcal{O}_\epsilon\left(\frac{1}{R^2}\right)$$

and similarly

$$\mu_{\frac{1}{\sqrt{n}}B_n}(z \in \mathbb{C}, |Z| \geq R) \leq \mathcal{O}_\epsilon\left(\frac{1}{R^2}\right).$$

Proof. Let λ_i denote the eigenvalues of A_n y Lemma 3.3 and condition (1) of Theorem 3.1, □

3.4.2. *Reduction to Characteristic Functions.* We define the characteristic function

$$m_\mu(u, v) = \int_{\mathbb{C}} e^{iu\Re(z)+iv\Im(z)} \mathrm{d}\mu(z).$$

We then have the following equivalence.

Lemma 3.8. The difference in measures converges to 0 in probability i.e.

$$\mu_{\frac{1}{\sqrt{n}}A_n} - \mu_{\frac{1}{\sqrt{n}}B_n} \xrightarrow{P} 0,$$

if and only if the difference in characteristic functions converges to 0 in probability, i.e.

$$m_{\frac{1}{\sqrt{n}}A_n} - m_{\frac{1}{\sqrt{n}}B_n} \xrightarrow{P} 0$$

Proof Sketch. We will provide only a proof sketch, because again this is a fairly standard type of result.

For the 'if' direction: first note that one only needs to show that for compactly-supported smooth test functions $f \in C_c^\infty$, one has

$$\int_{\mathbb{C}} f \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n} - \int_{\mathbb{C}} f \mathrm{d}\mu_{\frac{1}{\sqrt{n}}B_n} \xrightarrow{P} 0.$$

To do this, one can rewrite by standard Fourier-analytic results that

$$\int_{\mathbb{C}} f \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n} - \int_{\mathbb{C}} f \mathrm{d}\mu_{\frac{1}{\sqrt{n}}B_n} \overset{P}{=} \int_{\mathbb{R}\times\mathbb{R}} \hat{f}(u, v)(m_{\frac{1}{\sqrt{n}}A_n} - m_{\frac{1}{\sqrt{n}}B_n})(u, v) \mathrm{d}u\mathrm{d}v,$$

where $\hat{f}$ is the Fourier transform.[6] Applying the DCT variant given in Lemma 3.6 yields the result.

For the 'only if' direction: first by Lemma 3.7, for large R, with probability $1-\mathcal{O}(\epsilon)-o(1)$, the measures $\mu_{\frac{1}{\sqrt{n}}A_n}$, $\mu_{\frac{1}{\sqrt{n}}B_n}$ have low probability mass $\mathcal{O}(\epsilon)$ on $|z| \geq R$; that is,

$$\mu_{\frac{1}{\sqrt{n}}A_n}(\{|z| \geq R\}) + \mu_{\frac{1}{\sqrt{n}}B_n}(\{|z| \geq R\}) \leq \epsilon.$$

This is somewhat important, so we justify it: one can write $\sum \frac{1}{n^2}|\lambda_i(A_n)|^2 \leq C_\epsilon$ via Lemma 3.3, implying that $\int_{\mathbb{C}} |z|^2 \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n}(z) \leq C_\epsilon$, and similarly for B_n; the result follows via Markov. Then the difference in characteristic functions can be written as

$$m_{\frac{1}{\sqrt{n}}A_n} - m_{\frac{1}{\sqrt{n}}A_n}(u,v) = \int \mathbb{I}(|z| \leq R) e^{iu\Re(z)+iv\Im(z)} \left(\mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n} - d\mu_{\frac{1}{\sqrt{n}}B_n}\right)(z) + \mathcal{O}(\epsilon)$$

because the integrand is always at most 1; then, replacing $\mathbb{I}(|z| \leq R)$ with a C_c^∞-approximation,[7] the assumption $\mu_{\frac{1}{\sqrt{n}}A_n} - \mu_{\frac{1}{\sqrt{n}}B_n} \xrightarrow{P} 0$ implies that the integral approaches 0.

□

3.4.3. *Girko's Identity.* Now the key point is to try to compute the characteristic functions $m_{\frac{1}{\sqrt{n}}A_n}(u,v)$ in a convenient way for the proof. This was already done by Girko [Gir84], who proved the following nice eponymous identity. First, we define a function 'g' which is given by

$$\begin{aligned} g_{\frac{1}{\sqrt{n}}A_n}(z) &= 2\Re \int_{\mathbb{C}} \frac{z-w}{|z-w|^2} \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n}(w) \\ &= \frac{2}{n} \Re \sum_{i=1}^{n} \frac{z - \frac{1}{\sqrt{n}}\lambda_i(A_n)}{\left|z - \frac{1}{\sqrt{n}}\lambda_i(A_n)\right|^2}. \end{aligned}$$

Remark that this is defined almost everywhere. Intuitively, this is the real-partial derivative of the scaled log-potential, since

$$\frac{\partial}{\partial \Re(z)} \log |z - \frac{1}{\sqrt{n}}\lambda_i(A_n)| = \Re \frac{z - \frac{1}{\sqrt{n}}\lambda_i(A_n)}{|z - \frac{1}{\sqrt{n}}\lambda_i(A_n)|^2}.$$ [8]

Theorem 3.9 (Lemma 3.3 in [TVK10], 'Girko's Identity')**.** We have the following equality:

$$m_{\frac{1}{\sqrt{n}}A_n}(u,v) = \frac{u^2+v^2}{4\pi i u} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g_{\frac{1}{\sqrt{n}}A_n}(s+it) e^{ius+ivt} \mathrm{d}t \right) \mathrm{d}s.$$

Moreover, the inner integral is absolutely integrable a.e., and the outer integral is absolutely convergent.

Remark 3.10. To relate this to log-potential convergence, one can try to apply integration by parts on the integrand, and note that the log-potential will pop out. We will see this after sketching the proof of the above result.

[6] This is a standard Fourier identity; as a sketch, we can use Fourier inversion to write $\int f \mathrm{d}\mu = \int (\int \hat{f} e^{iu\Re(z)+iv\Im(z)} \mathrm{d}u \mathrm{d}v) \mathrm{d}\mu(z)$, swap the integration order to obtain $\int f \mathrm{d}\mu = \int (\hat{f}(e^{iu\Re(z)+iv\Im(z)} \mathrm{d}\mu(z)) \mathrm{d}u \mathrm{d}v$, and recognize the resulting appearance of m_μ; one then obtains $\int f \mathrm{d}\mu = \int \hat{f} m_\mu \mathrm{d}u \mathrm{d}v$ under mild assumptions which are satisfied.

[7] For instance, a standard nonnegative 'bump function' which equals 1 on the ball $|z| \leq R$ will suffice.

[8] In general, $\frac{\partial}{\partial \Re z} \log |z| = \Re \frac{z}{|z|^2}$; this can be checked easily.

Proof. Recall that

$$m_{\frac{1}{\sqrt{n}}A_n}(u,v) = \frac{1}{n}\sum_{j=1}^{n} e^{i(u\Re(\frac{1}{\sqrt{n}}\lambda_j(A_n)+v\Im(\frac{1}{\sqrt{n}}\lambda_j(A_n))},$$

so it is sufficient to check the equality for corresponding terms in the expressions for m and g; in particular, writing $w = \frac{1}{\sqrt{n}}\lambda_j(A_n)$, it suffices to show that

$$e^{i(u\Re(w)+v\Im(w))} = \frac{u^2+v^2}{2\pi iu}\int_{\mathbb{R}}\left(\int_{\mathbb{R}}\frac{\Re(s+it-w)}{|s+it-w|^2}e^{ius+ivt}\mathrm{d}t\right)\mathrm{d}s. \tag{1}$$

We first evaluate the inner integral. Note that it is clearly absolutely integrable, because the integrand is of order $\mathcal{O}\left(\frac{1}{t^2}\right)$. Moreover, this boundedness property allows us to employ standard contour integration techniques. In particular, we claim the following

Fact. The inner integral evaluates to

$$\int_{\mathbb{R}} f(t)\mathrm{d}t = \int_{\mathbb{R}}\frac{\Re a}{(a+it)(\overline{a}-it)}e^{ius+ivt}\mathrm{d}t = \int_{\mathbb{R}}\frac{\Re a}{|a+it|^2}e^{ius+ivt}\mathrm{d}t = \pi\mathrm{sgn}(\Re a)e^{-(\Re a)v}e^{ius}e^{i(\Im a)v}.$$

Proof of Fact: Evaluation of Inner Integral. WLOG assume $\Re a < 0$.[9] It suffices to evaluate the integral $\int_{-R}^{R} f(t)\mathrm{d}t$ and take the limit as $R\to\infty$, because the tails $\int_{|t|>R} f(t)\mathrm{d}t \xrightarrow{R\to\infty} 0$. Consider the contour consisting of the closed semicircle of radius R in the upper half-plane; call the full contour $\mathcal{C}$. Note that $\mathcal{C}$ is the union of $[-R,R]$ and the circular part, call it $\mathcal{C}'$. Then note that

$$\int_{-R}^{R} f(t)\mathrm{d}t + \int_{\mathcal{C}'} f(t)\mathrm{d}t = \int_{\mathcal{C}} f(t)\mathrm{d}t,$$

and the second term is of order $\mathcal{O}\left(\frac{1}{R}\right)$. Thus, since we are taking $R\to\infty$, it suffices to compute $\lim_{R\to\infty}\int_{\mathcal{C}} f(t)\mathrm{d}t$. But the function $f(t)$ is holomorphic except at $t_0 = ai$ where it has a pole of order 1; moreover, t_0 is in the region enclosed by $\mathcal{C}$ for all sufficiently large R. Let $a = p+qi$; then $t_0 = -q+pi$. We find via basic residue calculus that

$$\begin{aligned}\int_{\mathcal{C}} f(t)\mathrm{d}t &= 2\pi i\mathrm{Res}_2(-q+pi)\\ &= 2\pi i\lim_{t\to -q+pi}(t-(-q+pi))\frac{p}{(p+qi+ti)(p-qi-ti)}e^{iuv}e^{ivt}\\ &= 2\pi i e^{iuv}e^{iv(q-pi)}\lim_{a+bi\to -q+pi}\frac{p(a+q+(b-p)i)}{((a+q)i+(p-b))((p+b)-(q+a)i)}\\ &= 2\pi i e^{iuv}e^{iv(q-pi)}\lim_{a+bi\to -q+pi}\frac{pi}{(p+b)-(q+a)i}\\ &= -\pi e^{iuv}e^{(-\Re a)v}e^{i(\Im a)v},\end{aligned}$$

exactly as desired. This completes the proof of the fact. ∎

[9] This WLOG is fine; we will otherwise take a reflection of the closed semicircle contour we consider, such that the unique pole of the integrand lies within the contour. The signum term $\mathrm{sgn}(\Re a)$ comes from the reversal in direction of the contour, which contributes an extra sign.

Now, we can return to the main part of the proof. Using the Fact, it suffices to check that

$$e^{i(u\Re(w)+v\Im(w))} = \frac{u^2+v^2}{2\pi iu}\int_{\mathbb{R}}\left(\pi\mathrm{sgn}(s-\Re(w))e^{-v|s-\Re(w)|}e^{ius}e^{iv\Im(w)}\right)\mathrm{d}s.$$

However, we note that by letting $t = s-\Re(w)$, we have

$$\begin{aligned}\int_{\mathbb{R}}\left(\pi\mathrm{sgn}(s-\Re(w))e^{-v|s-\Re(w)|}e^{ius}e^{iv\Im(w)}\right)\mathrm{d}s &= \pi e^{iu\Re(w)+iv\Im(w)}\int_{\mathbb{R}}\mathrm{sgn}(t)e^{iut-v|t|}\mathrm{d}t\\ &= \pi e^{iu\Re(w)+iv\Im(w)}\left(\int_{t>0}e^{iut-vt}\mathrm{d}t+\int_{t<0}e^{iut+vt}\mathrm{d}t\right)\\ &= \pi e^{iu\Re(w)+iv\Im(w)}\left(-\frac{1}{iu-v}-\frac{1}{iu+v}\right)\\ &= \frac{2\pi iu}{u^2+v^2}e^{iu\Re(w)+iv\Im(w)}.\end{aligned}$$

This immediately implies Equation (1) for all w, which as remarked earlier completes the proof. $\square$

This now gives a concrete form of the characteristic function, in terms of an integral related to partials of the log potential. We can now perform a relatively direct analysis of this quantity.

3.5. **Completing the Proof of Theorem 3.1.** We are now positioned to apply Girko's identity and the above reduction to complete the proof.

Proof of Theorem 3.1. Define $g_{\frac{1}{\sqrt{n}}B_n}$ analogously to $g_{\frac{1}{\sqrt{n}}A_n}$. By Lemma 3.8 and Theorem 3.9, it suffices to show that with high probability[10] $1-\mathcal{O}(\epsilon)-o(1)$, we have

$$\int_{\mathbb{R}}\left(\int_{\mathbb{R}}(g_{\frac{1}{\sqrt{n}}A_n}(s+it)-g_{\frac{1}{\sqrt{n}}B_n}(s+it))e^{ius+ivt}\mathrm{d}t\right)\mathrm{d}s = \mathcal{O}(\epsilon).$$

Recall from Lemma 3.7 that with high probability,

$$\mu_{\frac{1}{\sqrt{n}}A_n}(\{|z|\geq R\})+\mu_{\frac{1}{\sqrt{n}}B_n}(\{|z|\geq R\})\leq\epsilon. \tag{2}$$

We can of course assume that we are in this regime. Now, the idea is to truncate the double integral to the range $s,t<R^2$. Let $\psi\in C_c^\infty$ be 1 on $[-1,1]$.[11]

Lemma 3.11 (Truncation Lemma)**.** For fixed w, we have the following:

(1) Truncation in s: the integral

$$\int_{\mathbb{R}}\left|\int_{\mathbb{R}}\frac{\Re(w-(s+it))}{|w-(s+it)|^2}e^{ius+ivt}\mathrm{d}t\right|(1-\psi(s/R^2))\mathrm{d}s = \mathcal{O}(1)$$

and for large enough R is $\mathcal{O}(\epsilon)$ whenever $|w|\leq R$.

(2) Truncation in t: the integral

$$\int_{\mathbb{R}}\left|\int_{\mathbb{R}}\frac{\Re(w-(s+it))}{|w-(s+it)|^2}e^{ius+ivt}(1-\psi(t/R^2)\mathrm{d}t\right|\psi(s/R^2)\mathrm{d}s = \mathcal{O}(1)$$

and for large enough R is $\mathcal{O}(\epsilon)$ whenever $|w|\leq R$.

[10]In this rest of this proof, we will note write the term $1-\mathcal{O}(\epsilon)-o(1)$ and just say 'with high probability'; the full details can be made explicit easily.

[11]Again, take a standard bump function.

Proof Sketch. This is a purely technical lemma, and once it is stated the proof is not difficult, so we will not write it out in full.

The first point can be established easily because the inner integral has been computed explicitly in the proof of Theorem 3.9; then, the term $1-\psi(s/R^2)$ implies that the integral at hand is roughly equivalent to integrating over $s \notin [-R^2, R^2]$. This is easily seen to be bounded and approaching 0 for $R \to \infty$.[12]

The second point can be established through repeated integration by parts applied to the inner integral, where the e^{ivt} term is integrated and the remaining terms are differentiated; roughly, differentiating $1-\psi(t/R^2)$ gives a factor R^{-2}, and differentiating the fractional term gives factor of $|s-\Re(w)|^{-1}$. Enough repetitions immediately yield both results. ☐

Intuitively, this lemma simply states that the integral is generally small if s, t are large, so this term can be ignored. Return to the main proof; applying the above truncations for $\frac{1}{\sqrt{n}}A_n$ and $\frac{1}{\sqrt{n}}B_n$, as well as Equation (2), it is sufficient to prove that

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} (g_{\frac{1}{\sqrt{n}}A_n}(s+it) - g_{\frac{1}{\sqrt{n}}B_n}(s+it)) e^{ius+ivt} \psi(t/R^2)\psi(s/R^2) \mathrm{d}t \right) \mathrm{d}s \xrightarrow{P} 0.^{13} \tag{3}$$

Now, the point is to apply integration by parts in s, where we integrate the g terms and differentiate the ψ terms. Let

$$f_n = \frac{1}{n} \log \left| \det \left(\frac{1}{\sqrt{n}} A_n - zI \right) \right| - \frac{1}{n} \log \left| \det \left(\frac{1}{\sqrt{n}} B_n - zI \right) \right|$$

and

$$\Phi_{u,v,R}(s,t) = -\frac{\partial}{\partial s} \left(e^{ius+ivt} \psi(t/R^2)\psi(s/R^2) \right).$$

Then the left side of Equation (3) equals via the aforementioned integration by parts the expression

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_n(s,t) \Phi_{u,v,r}(s,t) \mathrm{d}s \mathrm{d}t.$$

It suffices to show that this converges to 0 in probability; however, we know that $f_n(s,t) \xrightarrow{P} 0$ by the theorem assumption, convergence in log-potentials! Hence, the integrand converges to 0 in probability. Moreover, we claim that the integral

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |f_n(s,t) \Phi_{u,v,R}(s,t)|^2 \mathrm{d}s \mathrm{d}t = \mathcal{O}(1).$$

Indeed, look first at the inner integral; the Φ term is bounded and has compact support. Recalling that the determinant equals the product of all eigenvalues, the other term can be

[12] After evaluating the inner integral, the main term is $e^{-v|s-\Re w|}$; for $s > R > w$ this is at most e^{-v} and decays exponentially, so it is not hard to prove either above claim.

[13] One simply uses the triangle inequality along with the previous inequalities. Moreover, the reason why $\mathcal{O}(\epsilon)$ can be attained is that when applying this to g, we are taking $w = \lambda_i$ and averaging (recall the form of g!). Moreover, we are in the regime where the proportion of $w = \lambda_i$ with $w > R$ is $\mathcal{O}(\epsilon)$. For such $w = \lambda_i > R$, the contribution to the tail integrals is $\mathcal{O}(1)$; for all other $w = \lambda_i$, the contribution is $\mathcal{O}(\epsilon)$. Therefore, the tails are at most $\mathcal{O}(\epsilon) \cdot \mathcal{O}(1) + \mathcal{O}(1) \cdot \mathcal{O}(\epsilon) = \mathcal{O}(\epsilon)$.

written as

$$\begin{aligned} f_n(s,t)^2 &= \left[\frac{1}{n}\sum_{j=1}^{n}\log|\frac{1}{\sqrt{n}}\lambda_j(A_n)-(s+it)| - \frac{1}{n}\sum_{j=1}^{n}\log|\frac{1}{\sqrt{n}}\lambda_j(B_n)-(s+it)|\right]^2 \\ &\leq \frac{1}{n}\sum_{j=1}^{n}\left[(|\log|\frac{1}{\sqrt{n}}\lambda_j(A_n)-(s+it)|)^2 + (\log|\frac{1}{\sqrt{n}}\lambda_j(B_n)-(s+it)|)^2\right], \end{aligned}$$

via Cauchy-Schwarz. Then the entire integral is

$$\mathcal{O}(1) + \mathcal{O}\left(\frac{1}{n^2}\sum[\lambda_j(A_n)^2 + \lambda_j(B_n)^2]\right) = \mathcal{O}(1)$$

by Lemma 3.3 and the first assumption in the theorem statement. The desired convergence Equation (3) follows from Lemma 3.6. This completes the proof. □

4. Log-Potentials, Singular Values, and The Remains of the Proof

We have now established that in order to prove that the circular law (in probability) is universal, it suffices to prove that the difference in log-potentials for different distributions with mean 0 and variance 1 converges in probability to 0. In particular, the key result is the following.

Theorem 4.1 (Proposition 2.2 of [TVK10]; 'Convergence of Log-Potentials')**.** Let x, y, M_n, X_n, Y_n, A_n, B_n be as in Theorem 1.7. Then for all $z \in \mathbb{C}$, we have the convergence result

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right| - \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI\right)\right| \xrightarrow{P} 0.$$

This is sufficient and key to proving Theorem 1.7. In this remainder of this section, we will sketch the proof of this result, and then apply it to finish the proof of Theorem 1.7.

4.1. **Reformulations of the Log-Potential.** The first interesting idea is that the log-potential can be written in several ways. In particular, let $d_i(A)$ denote the distance from the i-th row of A to the span of the first $i-1$ rows. Then

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

This is a standard linear algebra fact; the inclusion of $\prod d_i$ follows from recalling that $\det A$ equals the volume of the parallelopiped formed by the rows of A.

We can, of course, therefore rewrite the log potential as a sum of expressions involving eigenvectors, singular values, or distances. In particular, we have the fundamental identity

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n\right)\right| = \frac{1}{n}\sum_{j=1}^{n}\log|\lambda_j(\frac{1}{\sqrt{n}}A_n)| \tag{4}$$

$$= \frac{1}{n}\sum_{j=1}^{n}\log|\sigma_j(\frac{1}{\sqrt{n}}A_n)| \tag{5}$$

$$= \frac{1}{n}\sum_{j=1}^{n}\log|d_j(\frac{1}{\sqrt{n}}A_n)|. \tag{6}$$

In the subsequent parts of the proof, all three formulations will be used.

4.2. **Sketch of the Main Idea.** For concreteness, let X_i be the rows of A_n and V_i be the span of the first $i-1$ rows of A_n; similarly, let Y_i be the rows of B_n and W_i be the span of the first $i-1$ rows of B_n. Recall that we want to prove that

$$\frac{1}{n}\sum_{j=1}^{n}\log|\mathrm{dist}(\frac{1}{\sqrt{n}}X_i,V_i)| - \frac{1}{n}\sum_{j=1}^{n}\log|\mathrm{dist}(\frac{1}{\sqrt{n}}Y_i,W_i)| \xrightarrow{P} 0. \tag{7}$$

The idea is to handle the parts of this sum for different j differently:

(1) For the terms $j > n - n^{0.99}$, employ an inverse-polynomial lower bound for the least singular value. Each term is then bounded logarithmically.
(2) For the terms $j \in [(1-\delta)n, n-n^{0.99}]$, we will employ Talagrand's concentration inequality to provide a probabilistic bound on the lower tail of the distances.
(3) For the terms $j \in [1,(1-\delta)n]$, we will employ results of Dozier and Silverstein [DS07].

Intuitively, the point is that for the 'very-high-dimensional' (larger) terms, where $j > n - n^{0.99}$, their analysis is tricky, but can be immediately bypassed by a high-probability, worst-case least singular value bound. For the 'high-dimensional' terms, where $j \in [(1-\delta)n, n-n^{0.99}]$, employing Talagrand will be a nice trick. Then, the 'low-dimensional' terms restrict the problem to essentially one about sufficiently rectangular matrices consisting of the first $(1-\delta)n$ rows of A_n and B_n, at which point one can employ results about ESDs of AA^* for rectangular A, which then gives information about the singular values of A. [14]

We will now discuss the proofs of these three items to varying detail.[15] For full proofs, again consult [TVK10].

4.3. **Proof of Theorem 4.1.** We now aim to prove Theorem 4.1, by explicating the sketch above. We first aim to analyze the contributions of each class of terms to the expression Equation (7).

4.3.1. *The very-high-dimensional terms.* This is probably the 'easiest' set of terms to bound, but only because of the vast literature on least singular value bounds. The result, of course, is the following:

Lemma 4.2 (Very-High-Dimensional Contribution)**.** We have the convergence result

$$\frac{1}{n}\sum_{j=n-n^{0.99}}\log|\mathrm{dist}(\frac{1}{\sqrt{n}}X_i,V_i)| - \frac{1}{n}\sum_{j=n-n^{0.99}}^{n}\log|\mathrm{dist}(\frac{1}{\sqrt{n}}Y_i,W_i)| \xrightarrow{P} 0.$$

Proof. The main point is to recall the following inverse-polynomial singular value bound from the vast and deep literature of analysis on the least singular values of a random matrix.

Fact (Least Singular Value Bound)**.** The bounds

$$\sigma_n(A_n), \sigma_n(B_n) \geq n^{-\mathcal{O}(1)}$$

hold for all but finitely many n with probability 1.

---

[14]One can imagine that singular value bounds are easier for rectangular matrices, because they are more likely to have robustly linearly independent rows. This might provide some intuition for why ESD results are more known for AA^* with A rectangular.

[15]Note to the reader: for the purpose of this project, we focus on the use of Talagrand because of the interesting invocation of concentration. Some parts will be fully proven and others sketched

This is a corollary of a result from [Tao08], that inverse-polynomial bounds exist on the least singular value with probability $(1-n^{-C})$ for any fixed C; then, one can apply Borel-Cantelli with any choice $C>1$.

In the reverse direction, we of course have the bound

$$\sigma_1(A_n), \sigma_1(A_n) \leq n^{\mathcal{O}(1)}$$

for all but finitely many n with probability 1. Indeed, we have the bound $\sigma_1(A_n)^2 \leq \|A_n\|_2^2$ as a corollary of Lemma 3.3. Now, we have assumed that $\sup \|M_n\|_2^2 < \infty$; then because x, y have mean 0 and variance 1; hence, their contributions to $\|A_n\|_2^2$ are small with high probability. One can then apply Borel-Cantelli to complete the proof.[16]

Given this two-sided polynomial bound on the singular values, we know from Lemma 3.5 that

$$\operatorname{dist}(X_i, V_i)^{-2}, \operatorname{dist}(Y_i, W_i)^{-2} \leq \operatorname{poly}(n)$$

and hence

$$|\log|\operatorname{dist}(X_i, V_i)||, |\log|\operatorname{dist}(Y_i, W_i)|| = \mathcal{O}(\log n).$$

It immediately follows that

$$\frac{1}{n}\sum_{j=n-n^{0.99}} \log|\operatorname{dist}(\frac{1}{\sqrt{n}}X_i, V_i)| - \frac{1}{n}\sum_{j=n-n^{0.99}}^{n} \log|\operatorname{dist}(\frac{1}{\sqrt{n}}Y_i, W_i)| = \mathcal{O}\left(\frac{\log n}{n^{0.01}}\right) \xrightarrow{P} 0.$$

This completes the proof of this case. $\square$

4.3.2. *The high-dimensional terms.* This set of terms will be bounded via Talagrand's concentration inequality. The result, of course, is the following:

Lemma 4.3 (High-Dimensional Contribution)**.** We have the convergence result

$$\frac{1}{n}\sum_{j=(1-\delta)n}^{n-n^{0.99}} \log|\operatorname{dist}(\frac{1}{\sqrt{n}}X_i, V_i)| - \frac{1}{n}\sum_{j=(1-\delta)n}^{n-n^{0.99}} \log|\operatorname{dist}(\frac{1}{\sqrt{n}}Y_i, W_i)| \xrightarrow{P} 0.$$

Proof Sketch. Fix ϵ. We split the proof into two parts: bounding the positive part and then the negative parts of the logarithms. By Borel-Cantelli, it suffices to prove that

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{1}{n}\sum_{i=(1-\delta)n}^{n-n^{-0.99}} \max\left(\log \operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right), 0\right) \geq \epsilon\right) < \infty$$

and

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{1}{n}\sum_{i=(1-\delta)n}^{n-n^{-0.99}} \max\left(-\log \operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right), 0\right) \geq \epsilon\right) < \infty.$$

We first sketch the proof of the first inequality (the 'positive' part). Here, the distance term is not fundamentally important; one can upper bound it by $\|X_i\|$. It is easy to see by

[16] Formally, one just needs to bound $\sum X_{ij}^2$ and $\sum X_{ij}$ polynomially; it therefore suffices to bound the former. However, $\mathbb{P}(X_{ij} > n^C) \leq \mathcal{O}(n^{-C})$ by Markov's inequality; therefore, $\mathbb{P}(\exists i, j : X_{ij} > n^C) \leq \mathcal{O}(n^{2-C})$. Taking C large enough, this quantity is summable, so Borel-Cantelli proves the necessary result.

'discretizing' the logarithm that

$$\frac{1}{n}\sum_{i=(1-\delta)n}^{n-n^{-0.99}} \max\left(\log \operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right), 0\right) \le \mathcal{O}\left(\sum_{m=0}^{\infty}\frac{1}{n}\sum_{i=(1-\delta)n}^{n-n^{0.99}} \mathbb{I}(\|X_i\| \ge 2^m\sqrt{n})\right).$$

Then, employing Chebyshev's and Hoeffding's inequalities, one can obtain bounds on the above expression (which is essentially a sum of tail indicators) which hold for probability $1-\mathcal{O}(-\exp(\operatorname{poly}(n,m)))$, at which point the desired result will follow.

We provide these details below. First, the RHS is $\mathcal{O}(\epsilon)$ as long as

$$\frac{1}{n}\sum_{i=(1-\delta)n}^{n} \mathbb{I}(\|X_i\| \ge 2^m\sqrt{n}) = \mathcal{O}\left(\frac{\epsilon}{(1+m)^2}\right)$$

for all m. But

$$\mathbb{P}(\|X_i\| \ge 2^m\sqrt{n}) \le \frac{\mathbb{E}\|X_i\|^2}{2^{2m}n} = \mathcal{O}(2^{-2m}),$$

since $\|X_i\|$ has second moment $n\mathbb{E}(x_{i1}^2)$. Hence,

$$\frac{1}{n}\sum_{i=(1-\delta)n}^{n} \mathbb{P}(\|X_i\| \ge 2^m\sqrt{n}) = \mathcal{O}(2^{-2m}).$$

Applying Hoeffding's inequality, we find that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=(1-\delta)n}^{n} \mathbb{I}(\|X_i\| \ge 2^m\sqrt{n}) > \mathcal{O}\left(\frac{\epsilon}{(1+m)^2}\right)\right) \le \mathcal{O}\left(\exp\left(-\frac{n}{(1+m)^4}\right)\right).$$

This is not quite summable for large m; however, the point is that when $m \gg \operatorname{poly}(n)$, the summation of indicators is essentially always 0; using the above inequality loses too much strength. In particular,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=(1-\delta)n}^{n} \mathbb{I}(\|X_i\| \ge 2^m\sqrt{n}) = 0\right) \ge 1 - n2^{-2m}.$$

It follows from combining the above that

$$\begin{aligned}\sum_{n=1}^{\infty}\mathbb{P}\left(\frac{1}{n}\sum_{i=(1-\delta)n}^{n-n^{-0.99}} \max\left(\log \operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right), 0\right) \ge \epsilon\right) &\le \sum_{m,n}\min\left(\mathcal{O}\left(\exp\left(-\frac{n}{(1+m)^4}\right)\right), n2^{-2m}\right)\\ &\le \sum_{n}\mathcal{O}\left(n2^{-n^{0.01}} + n^{0.01}\exp\left(-n^{0.96}\right)\right)\\ &\le \sum_{n}\mathcal{O}(\exp(-n^{0.01}))\\ &< \infty,\end{aligned}$$

as desired. This completes the proof of the first inequality.

We now sketch the proof of the second inequality (the 'negative part'). This is the much more difficult term, because it depends on showing that the parallelopiped generated by the rows is not too degenerate. Actually, however, we can show that the distances are generally large. In particular, we have the following lemma, which suggests that at least for *generally random* rows and subspaces, the distances are bounded below.

Lemma 4.4 (Lower Tail of Distances)**.** Let $1 \leq d \leq n - n^{0.99}$ and $c \in (0,1)$. Let W be a *fixed* subspace of $\mathbb{C}^n$ of dimension d. Then for any row X of A_n, we have

$$\mathbb{P}(\mathrm{dist}(X, W) \leq c\sqrt{n-d}) = \mathcal{O}\left(\exp(-n^{0.01})\right).$$

Proof Sketch of Lemma. We will sketch the proof, since there are some technical details which are largely simple. The main idea is to use Talagrand's inequality:

Fact (Talagrand's Inequality)**.** For any convex 1-Lipschitz function $F : \mathbb{C}^n \to \mathbb{R}$, if the X_i are IID random variables with $|X_i| \leq 1$, then

$$\mathbb{P}(|F(X_1, \ldots, X_n) - M(F(X_1, \ldots, X_n))| \geq r) \leq 4\exp\left(-\frac{r^2}{8}\right).$$

Here, $M(F)$ denotes the median of F.

In order to use this result, however, we need X to be bounded. Thus, the idea is to first address the tails of the coordinates x_i of X; in particular, the x_i have finite second moment and are therefore 'usually' small. Then, assuming that the x_i are bounded, we can restrict to the case where Section 4.3.2 applies, in which case the distance term is well-concentrated and hence the tail bound can be explicitly computed. This is essentially a combination of a truncation-type argument with Section 4.3.2.

We provide the details below. It is not hard to see via standard concentration inequalities (Chebyshev, Chernoff) that with high probability $1 - \mathcal{O}(\exp(-n^{\mathcal{O}(1)}))$, we have $|x_i| \leq n^{0.1}$ for all but at most $n^{0.9}$ indices i.[17] One can consider all possible sets $S = \{i : |x_i| > n^{0.1}\}$; by a simple union bound, it suffices to show that

$$\mathbb{P}(\mathrm{dist}(X, W) \leq c\sqrt{n-d} \mid S = \hat{S}) = \mathcal{O}(\exp(-n^{\mathcal{O}(1)}))$$

for all $|\hat{S}| \leq n^{0.9}$. Now, roughly, project X and W onto the coordinates $[n]\ S$, replace n with $n - |S|$, X with $X' = X \mid (|x_i| < n^{0.1} \forall i)$; it suffices to show that

$$\mathbb{P}(\mathrm{dist}(X', W) \leq c\sqrt{n-d}) = \mathcal{O}(\exp(-n^{\mathcal{O}(1)})).$$

Now we come to the key point: by Section 4.3.2 applied to $\frac{x_i}{n^{0.1}}$, we know that

$$\mathbb{P}(|\mathrm{dist}(X', W) - M(\mathrm{dist}(X', W))| \geq n^{0.1} r) \leq 4\exp\left(-\frac{r^2}{8}\right).$$

We are almost done: it suffices to show that $m = M(\mathrm{dist}(X', W)) \lesssim \sqrt{n-d}$. However, in fact we can estimate m, because we can exactly compute the second moment of $\mathrm{dist}(X', W)$.

[17] In particular, x_i has bounded second moment, so $\mathbb{P}(|x_i| \geq n^{0.1}) = \mathcal{O}(n^{-0.2})$; then the expected number of indices i is at most around $n^{0.8}$, so Chernoff's concentration inequality will yield the result.

Indeed, if $H = W(W^\intercal W)^{-1}W^\intercal$ is the projection matrix onto W, then

$$\begin{aligned}
\mathbb{E}[\mathrm{dist}(X', W)^2] &= \mathbb{E}[(HX')^\intercal(HX')] = \mathbb{E}[X'^\intercal HX'] \\
&= \mathbb{E}[\sum_{i,j} H_{ij}x'_i x'_j] = \mathbb{E}(x'_1)^2 \sum_i H_{ii} \\
&= \mathbb{E}(x'_1)^2 \mathrm{tr} H = \mathbb{E}(x'_1)^2 \mathrm{tr}(W(W^\intercal W)^{-1}W^\intercal \\
&= \mathbb{E}(x'_1)^2 \mathrm{tr}(W^\intercal W(W^\intercal W)^{-1}) = \mathbb{E}(x'_1)^2 \mathrm{tr}(\mathbf{I}_{n-d}) \\
&= \mathcal{O}(n-d).
\end{aligned}$$

Lastly, note that the median of any random variable can be bounded by its moments: indeed,

$$\frac{1}{2} = \mathbb{P}(X > m) \leq \frac{\mathbb{E}X^2}{m^2} \implies m \lesssim \sqrt{\mathbb{E}X^2}.$$

Applying this to the above computation, we find that $M(F) = \mathcal{O}(n-d)$. Combining everything, we obtain exactly the left-tail bound as desired. This completes the proof of the lemma. $\square$

Returning to the problem, note that X_i is independent of V_i. Therefore, Lemma 4.4 applies (conditioning on each fixed V_i, X_i is still distributed as a fixed row of A_n, then integrating across all V_i, one obtains the same bound); we find that

$$\mathrm{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right) \geq \frac{1}{2\sqrt{n}}\sqrt{n-i+1}$$

for all $n \in [(1-\delta)n, n - n^{0.99}]$ with high probability $1 - \mathcal{O}(n^{-100})$. Let $\delta \ll \epsilon$ be picked appropriately. Summing in i, we find that with probability $1 - \mathcal{O}(n^{-100})$, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=(1-\delta)n}^{n-n^{-0.99}} \max\left(-\log \mathrm{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i)\right), 0\right) &\leq \frac{1}{n}\sum_{k=n^{0.99}}^{\delta n} \log\frac{n}{k} + \mathcal{O}(\delta) \\
&\lesssim \frac{1}{n}\int_{n^{0.99}}^{\delta n} \log\frac{n}{k} \\
&= \frac{1}{n}[(\delta n - n^{0.99})\log n] - \frac{1}{n}(x\log x - x)\Big|_{n^{0.99}}^{\delta n} \\
&= |\delta \log \delta| + |\delta| + o(1) \\
&\leq \epsilon.
\end{aligned}$$

This shows the second inequality, and completes the proof of this case. $\square$

4.3.3. *The low-dimensional terms.* This set of terms will be bounded via employing results of Dozier-Silverstein. We will not go into much detail here, and largely refer the reader to [TVK10].

Lemma 4.5 (Low-Dimensional Contribution)**.** We have the convergence result

$$\frac{1}{n}\sum_{j=1}^{(1-\delta)n} \log|\mathrm{dist}(\frac{1}{\sqrt{n}}X_i, V_i)| - \frac{1}{n}\sum_{j=1}^{(1-\delta)n} \log|\mathrm{dist}(\frac{1}{\sqrt{n}}Y_i, W_i)| \xrightarrow{P} 0.$$

Very Rough Proof Sketch. Very roughly, the idea is to note that

$$\prod_{j=1}^{r} \operatorname{dist}(\frac{1}{\sqrt{n}}X_i, V_i) = \sqrt{\det\left(\frac{1}{n}A_{n,r}A_{n,r}^*\right)},$$

where $r = \lfloor(1-\delta)n\rfloor$, because both equal the volume of the parallelopiped spanned by the first r rows of $\frac{1}{\sqrt{n}}A_n$. It then suffices to prove that

$$\int_0^\infty (\log t) \mathrm{d}\mu_{\frac{1}{r}A_{n,r}A_{n,r}^*} - \mu_{\frac{1}{r}B_{n,r}B_{n,r}^*} \xrightarrow{P} 0,$$

since the determinant $\det\left(\frac{1}{n}A_{n,r}A_{n,r}^*\right) = \int_0^\infty \log t \mathrm{d}\nu_{n,r}$ can be written as a product of eigenvalues.

The remainder of the proof follows similar ideas as those already seen: we apply multiple truncations, analyzing the integral for five ranges of t.

When t is of medium size, then $|\log t|$ is bounded. A result of Dozier-Silverstein states that $\nu_{n,r} \to 0$[18]; applying dominated convergence completes the proof in this case. If t is small or large, $|\log t|$ cannot be easily bounded; therefore, we handle these cases separately.

When t is large, the result follows from boundedness properties of the singular values of $A_{n,r}$ and $B_{n,r}$[19] and $t \gg \log t$, to write $\int |\log t| \mathrm{d}\nu_{n,r} \le \int |t| \mathrm{d}\nu_{n,r} \le \epsilon$.

When t is small, more careful analysis using the Lemmas in Section 3.2 and the Talagrand bound on distances will suffice, but we will omit these (rather lengthy) details. □

The proof of Theorem 4.1 is now immediate by combining the contributions of the above three sets of terms.

Proof of Theorem 4.1. Summing the results of Lemma 4.2, Lemma 4.3, and Lemma 4.5, we immediately find that Equation (7) holds. □

4.4. **Proof of Theorem 1.7.** Finally, we can prove Theorem 1.7.

Proof of Theorem 1.7. This is now immediate; by Theorem 3.1, it suffices to show that

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right| - \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI\right)\right| \xrightarrow{P} 0.$$

But this is true by Theorem 4.1. This completes the proof! □

5. Conclusion

In conclusion, we have presented an exposition based on [Tao10] and [TVK10] of the proof of the general circular law, Corollary 1.9. The proofs demonstrated a number of ideas we have seen throughout the course, most notably concentration (applications of Talagrand and Chernoff-type bounds) and universality (in showing that the ESDs of any two distributions with the same mean and variance are asymptotically identical). There are many natural further questions to ask. For instance, mostly naturally, one can consider the case of matrices with non-IID or even dependent entries. There has even been some recent work along these

[18] This result requires $r \approx cn$ for $c < 1$ constant. The proof of the result is highly nontrivial, but for motivation, singular values of highly rectangular matrices are often easier to analyze.

[19] In particular, we have $\frac{1}{n}\sum_{i=1}^{r}(\frac{1}{\sqrt{n}}\sigma_i(A_{n,r}))^2 \le \frac{1}{n^2}\|A_{n,r}\|_2^2 \le \frac{1}{n^2}\|A_n\|_2^2$ is bounded in probability, and similarly for B_n.

lines; for instance, in 2021, Alt et. al. [Alt21] considered non-Hermitian random matrices with correlated entries, and derived an 'Inhomogeneous Circular Law', showing that in this regime the ESDs converge to a deterministic, radially symmetric density. Ultimately, the circular law and its variants document a line of work with an abundance of both deep theoretical ideas and fundamental, beautiful theorems.

References

[Gir84] V. L. Girko, 'Circular law'.

[Kem16] T. Kemp, 'Introduction to Random Matrix Theory'. Link. x

[Bai97] Z. D. Bai, 'Circular Law'. *Annals of Probability*, vol. 25, no. 1, pp. 494–529, 1997.

[DS07] R. Dozier, J. Silverstein, On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices, J. Multivar. Anal. 98 (2007), 678–694

[PZ07] G. Pan and W. Zhou, 'Circular Law, Extreme Singular Values, and Potential Theory'. arXiv:0705.3773, 2007.

[GT07] F. Götze and A. Tikhomirov, 'On the Circular Law'. arXiv:math/0702386, 2007.

[Tao08] T. Tao and V. Vu, 'Random Matrices: The Circular Law'. arXiv:0708.2895, 2008.

[Tao10] T. Tao, 'Least singular value, circular law, and Lindeberg exchange'. Link.

[Tao11] T. Tao, 'From the Littlewood-Offord problem to the Circular Law'. Link.

[TVK10] T. Tao, V. Vu, and M. Krishnapur, 2010. 'Random matrices: Universality of ESDs and the circular law.' Link

[Wig57] E.P. Wigner, On the distribution of the roots of certain symmetric matrices. Ann. of Math. 67 (1958), 325–327. Link

[Alt21] Alt, J., and Krüger, T., "Inhomogeneous Circular Law for Correlated Matrices." Journal of Functional Analysis 281, no. 7 (2021): 109120. Link.