Girko's Method, the Smallest Singular Value & Circular Laws

Kevin Cong, Bright Liu

November 25, 2024

The Empirical Spectral Distribution

Definition (Empirical Spectral Distribution (ESD))

Let M be an $n \times n$ matrix. The *empirical spectral distribution* of M is given by

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

If M is Hermitian, all the eigenvalues are real, so we instead consider the spectral distribution over $\mathbb{R}$ and not $\mathbb{C}$.

The Empirical Spectral Distribution

Definition (Empirical Spectral Distribution (ESD))

Let M be an $n \times n$ matrix. The *empirical spectral distribution* of M is given by

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

If M is Hermitian, all the eigenvalues are real, so we instead consider the spectral distribution over $\mathbb{R}$ and not $\mathbb{C}$.

Question

Suppose M_n are matrices with IID entries which are normalized appropriately. If $n \to \infty$, what is the asymptotic ESD of the matrices M_n?

(Semi)Circular Laws

Theorem (Semicircle Law: Wigner 1957b)

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 A_n converge:

$$\mu_{A_n} \xrightarrow[P]{a.s.} \mu,$$

where μ is the 'semicircular law'.

(Semi)Circular Laws

Theorem (Semicircle Law: Wigner 1957b)

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 A_n converge:

$$\mu_{A_n} \xrightarrow[P]{a.s.} \mu,$$

where μ is the 'semicircular law'.

Theorem (Circular Law: Tao-Vu 2008)

Let A_n be $n \times n$ random matrices with IID complex entries with mean 0 and variance 1. Then the ESDs of $\frac{1}{\sqrt{n}}A_n$ converge:

$$\mu_{\frac{1}{\sqrt{n}}A_n} \xrightarrow[P]{a.s.} \mu,$$

where μ is the uniform distribution in the unit disk.

Historical Progress from Semicircle to Circular Law

Timeline For Least Singular Value:

- **1988:** Rudelson develops early bounds for the smallest singular value using geometric functional analysis techniques.
- **2005:** Rudelson and Vershynin apply epsilon-net arguments to achieve probabilistic bounds on the least singular value.
- **2007:** Tao and Vu connect these bounds to universality, showing that least singular value control is critical for non-Hermitian spectra.

Historical Progress from Semicircle to Circular Law

Timeline For Least Singular Value:

- **1988:** Rudelson develops early bounds for the smallest singular value using geometric functional analysis techniques.
- **2005:** Rudelson and Vershynin apply epsilon-net arguments to achieve probabilistic bounds on the least singular value.
- **2007:** Tao and Vu connect these bounds to universality, showing that least singular value control is critical for non-Hermitian spectra.

Timeline For Circular Law:

- **1957:** Wigner proves the semicircle law for Hermitian matrices.
  - Hermitian structure ensures eigenvalues are real, enabling simpler analysis.
- **1960s:** Ginibre established circular law for Gaussian entries.
- **1984:** Girko proposes the logarithmic potential method for the circular law.
- **1997:** Further progress by Zhidong Bai, establishing circular law under certain smoothness assumptions.
- **2000s:** Tao and Vu prove universality of the circular law.
  - Relaxed assumptions: Removed normality and reduced moment conditions to $2 + \epsilon$.

Girko's Method

Definition (Logarithmic Potential)

We define the *logarithmic potential* of $\frac{1}{\sqrt{n}}A_n$ as

$$\begin{aligned} f_n(z) &= \int_{\mathbb{C}} \log|z-w| \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n}(w) \\ &= \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{\lambda_j(A_n)}{\sqrt{n}} - z\right) \\ &= \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right|. \end{aligned}$$

Girko's Method

Definition (Logarithmic Potential)

We define the *logarithmic potential* of $\frac{1}{\sqrt{n}}A_n$ as

$$\begin{aligned} f_n(z) &= \int_{\mathbb{C}} \log|z-w| \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n}(w) \\ &= \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{\lambda_j(A_n)}{\sqrt{n}} - z\right) \\ &= \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right|. \end{aligned}$$

Outline (Girko 1984)

Suppose we want to show that $\mu_{\frac{1}{\sqrt{n}}A_n} \to \mu$.

1. *Show that the least singular value of* $\frac{1}{\sqrt{n}}A_n$ *is bounded below.*
2. *Show that* $f_n(z) \to f(z)$, *where* f_n *are appropriate log potentials.*
3. *Show that log-potential convergence implies distributional convergence.*

Log-Potential Continuity Theorem

The third bullet point is technically quite difficult, but we will explain the basic result here.

Theorem (Logarithmic Potential Continuity Theorem)

Let A_n be a sequence of random matrices. Suppose that for almost all $z \in \mathbb{C}$, we have the convergence result

$$f_n(z) = \int_{\mathbb{C}} \log|z-w| \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n}(w) \xrightarrow{a.s.} \int_C \log|z-w| \mathrm{d}\mu(w) = f(z).$$

Then we have

$$\mu_{\frac{1}{\sqrt{n}}A_n} \xrightarrow{a.s.} \mu.$$

Remark

This formalizes the idea that the log potential 'characterizes' random distributions. In some sense, it is an analogue of the characteristic function for random variables.

Universality of Log Potential

Theorem (Universality of Log Potential)

Suppose that $\frac{1}{n^2}\|A_n\|_2^2 + \frac{1}{n^2}\|B_n\|_2^2$ *is bounded a.s., and for almost all* $z \in \mathbb{C}$,

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

Then

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

Remark

With this result, one can use the circular law for Gaussians to show the circular law for general distributions: indeed, we can let A_n *be an ensemble of Gaussian random matrices and* B_n *be the ensemble of matrices with IID entries of the same mean and variance.*

Universality of Log Potential, Cont'd

Theorem (Universality of Log Potential)

Under regularity conditions and convergence of the difference of the log potential of A_n, B_n, we have convergence in probability of difference of the ESDs to 0.

Universality of Log Potential, Cont'd

Theorem (Universality of Log Potential)

Under regularity conditions and convergence of the difference of the log potential of A_n, B_n, we have convergence in probability of difference of the ESDs to 0.

Proof Sketch.

1. Take characteristic fn: $m_{\frac{1}{\sqrt{n}}A_n}(u,v) = \int_{\mathbb{C}} e^{iu\Re(z)+iv\Im(z)} \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n}(u,v)$.
2. Show that convergence of characteristic fns is equivalent to convergence in distribution [Fourier analytic methods].

Universality of Log Potential, Cont'd

Theorem (Universality of Log Potential)

Under regularity conditions and convergence of the difference of the log potential of A_n, B_n, we have convergence in probability of difference of the ESDs to 0.

Proof Sketch.

1. Take characteristic fn: $m_{\frac{1}{\sqrt{n}}A_n}(u,v) = \int_{\mathbb{C}} e^{iu\Re(z)+iv\Im(z)} \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n}(u,v)$.
2. Show that convergence of characteristic fns is equivalent to convergence in distribution [Fourier analytic methods].
3. Use 'Girko's identity': if $g_{\frac{1}{\sqrt{n}}A_n}(z) = \frac{2}{n}\Re\sum_{j=1}^{n} \frac{z-\frac{1}{\sqrt{n}}\lambda_j}{|z-\frac{1}{\sqrt{n}}\lambda_j|^2}$, then

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

Universality of Log Potential, Cont'd

Theorem (Universality of Log Potential)

Under regularity conditions and convergence of the difference of the log potential of A_n, B_n, we have convergence in probability of difference of the ESDs to 0.

Proof Sketch.

1. Take characteristic fn: $m_{\frac{1}{\sqrt{n}}A_n}(u,v) = \int_{\mathbb{C}} e^{iu\Re(z)+iv\Im(z)} \mathrm{d}\mu_{\frac{1}{\sqrt{n}}A_n}(u,v)$.
2. Show that convergence of characteristic fns is equivalent to convergence in distribution [Fourier analytic methods].
3. Use 'Girko's identity': if $g_{\frac{1}{\sqrt{n}}A_n}(z) = \frac{2}{n}\Re\sum_{j=1}^{n} \frac{z-\frac{1}{\sqrt{n}}\lambda_j}{|z-\frac{1}{\sqrt{n}}\lambda_j|^2}$, then

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

4. For large s,t, the inner integral is small because g is small.
5. For small s,t, the remainder of the integral is bounded; one can use BCT type of results.

Outline

Recall our outline:

Outline (Girko 1984)

Suppose we want to show that $\mu_{\frac{1}{\sqrt{n}}A_n} \to \mu$.

1. *Show that the least singular value of* $\frac{1}{\sqrt{n}}A_n$ *is bounded below.*
2. *Show that* $f_n(z) \to f(z)$, *where* f_n *are appropriate log potentials.*
3. *Show that log-potential convergence implies distributional convergence.*

We have discussed items (2) and (3), related to the log potential. We now just need to analyze the log potentials.

Bounding the Difference in Log-Potentials: Outline

1. Rewrite $\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n\right)\right| = \frac{1}{n}\sum_{i=1}^{n}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right)$, where A_n has rows X_i and V_i is the space spanned by $(X_1,\ldots,X_{n-1})$. This holds because both represent the volume of a parallelopiped. Analogously define Y_i, W_i for B_n.
2. Bound $\frac{1}{n}\displaystyle\sum_{i=n-n^{0.99}}^{n}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right) - \frac{1}{n}\sum_{i=n-n^{0.99}}^{n}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}Y_i, W_i\right)$.
3. Bound $\frac{1}{n}\displaystyle\sum_{i=n(1-\delta)}^{n-n^{0.99}}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right) - \frac{1}{n}\sum_{i=n(1-\delta)}^{n-n^{0.99}}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}Y_i, W_i\right)$. '
4. Bound $\frac{1}{n}\displaystyle\sum_{i=1}^{n(1-\delta)}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right) - \frac{1}{n}\sum_{i=1}^{n(1-\delta)}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}Y_i, W_i\right)$.

Remark

1. *Note that the latter three steps are analyzing different parts of this sum.*
2. *Moreover, the intuition behind looking at the distances is essentially that these are some measure of how 'degenerate' the parallelopiped is, which is closely related to the singular values*

Main Proof: Part I, the top terms

Proof Sketch: Part (1).

1. To bound $\frac{1}{n}\sum_{i=n-n^{0.99}}^{n} \log \operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right)$, note that there are not many terms. One can appeal to the following linear algebra fact.

Lemma

We have $\sum_{j=1}^{n'} \sigma_j(A_n)^{-2} = \sum_{j=1}^{n'} \operatorname{dist}(\frac{1}{\sqrt{n}}X_j, V_j)^{-2}$.

2. It therefore essentially suffices to bound the singular values $\sigma_j(A_n)$ by inverse-polynomial quantities.

$\square$

Main Proof: Part I, the top terms, cont'd.

Proof Sketch: Part (1).

1. Recall: it suffices to bound the singular values $\sigma_j(A_n)$ by inverse-polynomial quantities.

Main Proof: Part I, the top terms, cont'd.

Proof Sketch: Part (1).

1. Recall: it suffices to bound the singular values $\sigma_j(A_n)$ by inverse-polynomial quantities.

Theorem

We have $\sigma_n(A_n) \geq n^{-O(1)}$ for all but finitely many n almost surely.

2. This is a concluding result in a sequence of works. It is a corollary of a result in [Tao-Vu 2008, 'The Circular Law'], where they show that $\mathbb{P}(\sigma_n(A_n) \geq n^{-O_C(1)}) \leq n^{-C}$, so picking C large enough and using Borel-Cantelli yields the result.

$\square$

An aside: Results on the Least Singular Value

Distribution of Smallest Singular Value:

Theorem (Asymptotic Distribution)

The distribution for Gaussian random matrix can be computed explicitly, and similarly for iid Bernoulli matrix. As $n \to \infty$, the scaled smallest singular value $\sqrt{n}\sigma_n(M)$ converges to the distribution:

$$\mu_E = \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-x/2-\sqrt{x}} dx.$$

An aside: Results on the Least Singular Value, Cont'd.

Definition (Singularity Probability)

The singularity probability, $P(\sigma_n(M) = 0)$, is the probability that M is not invertible.

Theorem (Bounds on Singularity Probability)

Bourgain et al. (2010) showed that for random matrices with i.i.d. entries:

$$P(\sigma_n(M) = 0) \leq \left(\frac{1}{\sqrt{2}} + o(1) \right)^n .$$

In this paper, they proved the version:

$$P(\sigma_n(M) = 0) \ll \frac{1}{\sqrt{n}}.$$

Main Proof: Part II, the middle terms

Proof Sketch: Part (2).

1. The main difficulty is when bounding the distances in the sum $\frac{1}{n}\sum_{i=n(1-\delta)}^{n-n^{0.99}} \log \operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right)$ away from 0.

Main Proof: Part II, the middle terms

Proof Sketch: Part (2).

1. The main difficulty is when bounding the distances in the sum $\frac{1}{n}\sum_{i=n(1-\delta)}^{n-n^{0.99}} \log \operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right)$ away from 0.
2. The main idea is to use Talagrand's concentration inequality!

Lemma

In the setup above, take W to be an arbitrary fixed d-dimensional subspace (to replace V_i). Then

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

Main Proof: Part II, the middle terms

Proof Sketch: Part (2).

1. The main difficulty is when bounding the distances in the sum $\frac{1}{n}\sum_{i=n(1-\delta)}^{n-n^{0.99}} \log \operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right)$ away from 0.
2. The main idea is to use Talagrand's concentration inequality!

Lemma

In the setup above, take W to be an arbitrary fixed d-dimensional subspace (to replace V_i). Then

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

3. To prove the lemma, we can use the Talagrand Concentration Inequality, which roughly states that $\operatorname{dist}(X, W)$ is concentrated near its median, with exponential tail probabilities.

Main Proof: Part II, the middle terms

Proof Sketch: Part (2).

1. The main difficulty is when bounding the distances in the sum $\frac{1}{n} \sum_{i=n(1-\delta)}^{n-n^{0.99}} \log \operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right)$ away from 0.
2. The main idea is to use Talagrand's concentration inequality!

Lemma

In the setup above, take W to be an arbitrary fixed d-dimensional subspace (to replace V_i). Then

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

3. To prove the lemma, we can use the Talagrand Concentration Inequality, which roughly states that $\operatorname{dist}(X, W)$ is concentrated near its median, with exponential tail probabilities.
4. Then one can establish $\mathbb{E}[\operatorname{dist}(X, W)^2] \approx n - d$. The lemma follows.
5. Then the distances are not too small. It is not hard to bound them above.

Main Proof: Part III, the smaller terms

Proof Sketch: Part (3).

1. Recall that $\prod_1^{n'} \mathrm{dist}(\frac{1}{\sqrt{n}} X_i, V_i)$ is a volume of the parallelopiped generated by $X_1, \dots, X_{n'}$; it also equals the product of singular values of $A_{n,n'}$ (the matrix generated by the first n rows of A_n).

Main Proof: Part III, the smaller terms

Proof Sketch: Part (3).

1. Recall that $\prod_1^{n'} \operatorname{dist}(\frac{1}{\sqrt{n}} X_i, V_i)$ is a volume of the parallelopiped generated by $X_1, \ldots, X_{n'}$; it also equals the product of singular values of $A_{n,n'}$ (the matrix generated by the first n rows of A_n).
2. We can replace these singular values via the square root of the eigenvalues of $A_{n,n'} A^*_{n,n'}$; this reduces our problem to looking at the determinant of $A_{n,n'} A^*_{n,n'}$, which is now Hermitian.

Main Proof: Part III, the smaller terms

Proof Sketch: Part (3).

1. Recall that $\prod_1^{n'} \mathrm{dist}(\frac{1}{\sqrt{n}} X_i, V_i)$ is a volume of the parallelopiped generated by $X_1, \ldots, X_{n'}$; it also equals the product of singular values of $A_{n,n'}$ (the matrix generated by the first n rows of A_n).
2. We can replace these singular values via the square root of the eigenvalues of $A_{n,n'} A^*_{n,n'}$; this reduces our problem to looking at the determinant of $A_{n,n'} A^*_{n,n'}$, which is now Hermitian.
3. One can then use results and ideas from the Hermitian case, as well as the concentration result in Part (2), to do analysis on the terms in the summation of different magnitudes.

□

Conclusion

1. We have sketched the key ideas in the circular law, which describes the asymptotic distribution of the spectra of Hermitian matrices under minimal assumptions!
2. The least singular value question itself is quite interesting. There has been work on bounding the probability that $\sigma_n = 0$, as well as stronger probabilistic lower bounds.
3. Another natural question is in the case of matrices with dependent, non-IID entries, etc.

References

1. Tao, T. and Vu, V. (2008). Random matrices: Universality of ESDs and the circular law.
2. Wigner, E. P. (1957). On the distribution of the roots of certain symmetric matrices.
3. Girko, V. L. (1984). Theory of random determinants.