

The Circular Law

Van H. Vu

Department of Mathematics

Rutgers

vanvu@math.rutgers.edu

(joint work with T. Tao, UCLA)

First story: Circular Law

The models.

M_n (symmetric or non-symmetric): n by n matrix with i.i.d (in many considerations this assumption can be significantly weakened) entries: ξ_{ij} .

Continuous models: ξ_{ij} have continuous distribution. Representative example: Gaussian.

Discrete models: ξ_{ij} have discrete distribution. Representative example: Bernoulli (± 1 with probability $1/2$).

Eigenvalues: $\lambda_1, \dots, \lambda_n$.

The problem: Limiting distribution of the spectra

Hermitian Case:

Wigner's semi-circle law (1950s): M_n *symmetric* random matrix. Then the eigenvalues follow the semi-circle law. (Of course, after a proper normalization by a factor $\Theta(\sqrt{n})$).

Wigner proof introduced the Trace method:

For all fixed k

$$E(\lambda_1^k + \cdots + \lambda_n^k) \rightarrow \int x^k d\mu.$$

But

$$\lambda_1^k + \cdots + \lambda_n^k = \text{Trace}(M_n^k).$$

So it is sufficient to show

$$E(\text{Trace}(M_n^k)) \rightarrow \int x^k d\mu,$$

which is in essence a combinatorial problem about counting certain kind of paths in the complete graph K_n :

$$\text{Trace}(M^k) = \sum_{i_1, \dots, i_k} \xi_{i_1 i_2} \xi_{i_2 i_3} \cdots \xi_{i_k i_1}.$$

Wigner proof works for both Gaussian and Bernoulli models. Extensions by Arnold, Grenander. The most general form of the circular law was proved later by Pastur (1960s):

Theorem. (Pastur) Let $\xi_{ij}, 1 \leq i \leq j \leq n$ be i.i.d random variables with mean 0 and variance 1. Then the eigenvalues of the symmetric random matrix $M_n = \{\xi_{ij}\}$ follows the semi-circle law.

(Moment method+ Truncation to control terms like ξ_{ij}^4 .)

Non-hermittian case:

Conjecture. (Circular Law Conjecture) Let $\xi_{ij}, 1 \leq i, j \leq n$ be i.i.d random variables with mean 0 and variance 1. Then the eigenvalues of the random matrix $M_n = \{\xi_{ij}\}$ follows the circular law.

Weak Circular Law: weak convergence. (Roughly speaking, error term $\epsilon(n)$ goes to zero with n .)

Strong Circular Law: Almost surely convergence. (Roughly speaking, error term $\epsilon(n)$ is summable in n , i.e., tends to n as fast as n^{-1-c} .)

The conjecture was verified by Mehta (and also Silvestein) for the Gaussian case thanks to Ginibre (1962) formula of the joint density of the eigenvalues:

$$p(\lambda_1, \dots, \lambda_n) = c_n \prod_{[i < j]} |\lambda_i - \lambda_j|^2 \prod_{i=1}^n e^{-n|\lambda_i|^2}.$$

Proved for many continuous models (Girko 1984, Bai 1997, Edelman1997, Bai-Silvestein 2006)

Bai, Bai-Silverstein proved Strong CL under two additional assumptions:

Moment: $E(|\xi_{ij}|^{2+\epsilon}) = O(1)$.

Boundedness: *Joint distribution of the real and imaginary part to have bounded density.*

Good for complex gaussian but not real gaussian. The later case was done by Edelman (1997).

Moment method is not useful, as all moments of the circular distribution are zero, and they do not determine the distribution uniquely.

Truncation is dangerous. In the symmetric case, the eigenvalues distribution does not change significantly if we change the matrix a little bit.

But the situation for the non-symmetric case is very different.
(Example).

Recent progress: No boundedness assumption.

Götze-Tikhomirov (07): Weak CL under all moments assumption (sub-gaussian entries).

Pan-Zhou (07): Strong CL under 4th moment assumption.

Tao-Vu (07): [Strong CL under \$2 + \epsilon\$ -moment assumption](#).

Götze-Tikhomirov (07): Weak CL under $2 + \epsilon$ -moment assumption.

The last two results also extend to sparse matrices of density n^{-1+c} (each row has roughly n^c non-zero elements). Instead of ξ_{ij} consider $\xi_{ij}I_{ij}$, where $I = 1$ with probability n^{-1+c} zero otherwise.

Stieltjes transform: $s_n(z) = \frac{1}{n} \sum \frac{1}{\lambda_i - z}$.

Need to show this goes to the right limit $s(z)$. Set $z = s + it$.

$s_n(z) = S + iT$.

$$\begin{aligned} S &= \frac{1}{n} \sum \frac{\lambda_i(r) + s}{|\lambda_i - z|^2} \\ &= -\frac{1}{2n} \sum \frac{\partial}{\partial s} \log |\lambda_i - z|^2 \\ &= -\frac{1}{2} \frac{\partial}{\partial s} \int_0^\infty \log x \, \partial \eta_n \end{aligned}$$

where η_n is the counting measure of the (squares of the) singular values of $\frac{1}{\sqrt{n}} M_n - z I_n$.

Main difficulty. $\int_0^\infty \log x \partial\eta_n$ has a pole at 0.

Bai needs the boundedness assumption to overcome this problem, namely showing that with high probability, the least singular value of $M_n - zI$ is not too small (larger than n^{-100} , say).

Question. Let M_n be a random matrix with iid entries with mean 0 and variance 1. What can one say about the least singular value of $M_n - zI$? (How well is it bounded from zero?)

More general

Question. Let M_n be a random matrix with iid entries with mean 0 and variance 1. and A be a fixed matrix of size n . What can one say about the least singular value of $M_n + A$? (How well is it bounded away from zero?)

Let $\sigma_1 \geq \dots \geq \sigma_n$ be the singular values.

M_n is real gaussian:

Edelman showed that typically $\sigma_n(M_n) = \Theta(n^{-1/2})$ (he computed the limiting distribution of $\sqrt{n}\sigma_n$.)

Spielman and Teng shows that typically $\sigma_n(M_n + A) = \Omega(n^{-1/2})$.

$$\mathbf{P}(\sigma_n(M_n + A) \leq \epsilon n^{-1/2}) \leq \text{constant} \times \epsilon,$$

for any A .

Geometrical interpretation: σ_n smallest distance.

Universality. Any reasonable random matrix ensemble behaves like Gaussian one.

Question. Can one expect

$$\mathbf{P}(\sigma_n(M_n + A) \leq \epsilon n^{-1/2}) \leq \text{constant} \times \epsilon,$$

holds for other models, such as Bernoulli, at least in some range of (say $\epsilon = n^{-\Theta(1)}$) ?

Tao-Vu (06): If ξ_{ij} and A are discrete (entries are integers), $\|A\| \leq n^{-C_1}$, ξ_{ij} has constant positive variance, then for any constant C_2 there is a constant B depending on C_1, C_2 such that

$$\mathbf{P}(\sigma_n(M_n + A) \leq n^{-B}) \leq n^{-C_2}.$$

Rudelson-Vershynin (07): If ξ_{ij} has mean zero and bounded fourth moment, then

$$\mathbf{P}(\sigma_n(M_n) \leq \epsilon n^{-1/2}) \leq \text{constant} \times \epsilon.$$

The fourth moment is critical here.

G-T and P-Z used an extension of R-V; T-V used an extension of T-V to prove CL.

Unlike the gaussian case, it has turned out that in other cases (such as Bernoulli) the role of the additional matrix A is significant. In particular $\|A\| = \sigma_1(A)$ influences $\sigma_n(M_n + A)$.

We can construct a matrix A with $\|A\| = L$ such that

$$\mathbf{P}(\sigma_n(M_n + A) \leq L^{-1}n^{3/2}) \geq n^{-1/2}.$$

This shows that the universality conjecture does not hold.

Question. What is the relation between $\|A\|$ and the bound on σ_n ?

Theorem. (Tao-Vu 08) Let ξ be a variable with mean 0 and variance 1. Let $\gamma \geq 1/2$ be a real number. Let M_n be the random matrix whose entries are iid copy of ξ and A a fixed matrix with $\|A\| \leq n^\gamma$. Then for any constant $C > 0$

$$\mathbf{P}(\sigma_n(M_n + A) \leq n^{-(2C+1)\gamma}) \leq n^{-C+o(1)} + \mathbf{P}(\|M_n\| \geq n^\gamma).$$

In application, choose γ such that $\mathbf{P}(\|M_n\| \geq n^\gamma)$ is negligible (super polynomially small)

$$\mathbf{P}(\sigma_n(M_n + A) \leq n^{-(2C+1)\gamma}) \leq n^{-C+o(1)}.$$

Example. ξ has bounded fourth moment, then
 $\mathbf{P}(\|M_n\| \geq 100n^{1/2}) \leq \exp(-n)$. Let A satisfy $\|A\| = O(n^{1/2})$. Thus,
with $\gamma = 1/2 + o(1)$

$$\mathbf{P}(\sigma_n(M_n + A) \leq n^{-C-1/2}) \leq n^{-C+o(1)}.$$

For $A = 0$, one gets back the Rudelson-Vershynin's bound

$$\mathbf{P}(\sigma_n(M_n) \leq n^{-C-1/2}) \leq n^{-C+o(1)}.$$

Example. ξ Bernoulli. $\|A\| = L = n^\gamma \gg n^{1/2}$.

$$\mathbf{P}(\sigma_n(M_n + A) \leq L^{-2C-1}) \leq n^{-C+o(1)}.$$

With $C = 1/2$, $\mathbf{P}(\sigma_n(M_n + A) \leq L^{-2}) \leq n^{-1/2}$.

Our earlier construction shows that -2 in the power of L cannot be replaced by $-1 + \epsilon$.

Main tool behind the study of the least singular value:

Concentration function: $v = (a_1, \dots, a_n)$

$$P_v := \max_x \mathbf{P}\left(\sum_{i=1}^n a_i \xi_i = x\right).$$

We consider a toy version, where ξ_i are iid Bernoulli (the result extends to very general random variables)

Littlewood-Offord (1943), Erdős (1945) If $a_i \neq 0$, then $P_v = O(n^{-1/2})$.

This is sharp: $a_i = 1$.

Many generalizations, improvements: Katona (66), Kleitman (70), Grigg et. al. (83), Halász (75), Erdos-Moser (63), Sarkozi-Szemerédi (65), Stanley (80), Frankl-Furedi (88)etc.

For example:

Theorem. (Erdős-Moser (1963), Sárközi-Szemerédi (1965), Halász (1975), Stanley (1980)) *If the a_i are different , then $P_v = O(n^{-3/2})$.*

This is sharp: $a_i = i$ (or elements of an arithmetic progression).

Inverse problem. Assume that P_v large (say at least n^{-C}). What can one say about the a_i ? **In other words, what make P_v large?**

Example: a_i are elements of a symmetric integral box B of dimension d and volume V . Then with high probability the random sum is contained in $100\sqrt{n}B$. By the pigeon hole principle:

$$P_v \geq \frac{1}{\text{Vol}(100\sqrt{n}B)} = cn^{-d/2}V^{-1}.$$

Theorem. (Tao-V. 2005–2008) **The inverse is (essentially) true.**

(2005) If $\mathbf{P}_v \geq n^{-C_1}$, then most of v is contained in a box of fixed dimension d of volume n^{C_2} , where C_2, d depends on C_1 .

(2008) Sharp dependence: $C_1 = d/2 + C_2$.