# From smooth analysis to circular law: A journey through additive combinatorics

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: $\xi_{ij}$.

*Continuous models*: $\xi_{ij}$ have continuous distribution. Representative example: Gaussian.

*Discrete models*: $\xi_{ij}$ have discrete distribution. Representative example: Bernoulli ($\pm 1$ with probability $1/2$).

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

*The problem: Limiting distribution of the spectra*

Symmetric (Hermittian) 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 $k$

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

But

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

So it is sufficient to show

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

which is in essence a combinatorial problem about counting certain kind of paths in a complete graph.

Wigner proved works for both Gaussian and Bernoulli models (and many others). The most general form of the circular law was proved by Pastur (1960s):

**Theorem.** (Pastur) Let $\xi_{ij}, 1 \le i \le j \le n$ 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.

Non-symmetric case (non-hermittian):

**Conjecture.** (Circular Law Conjecture)Let $\xi_{ij}, 1 \le i, j \le n$ 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.

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

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

Basically solved for most continuous models (Girko 1984, Bai 1997):

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

The discrete case was open until very recently.

*Main difficulty.*

Trace method: $\int x^k d\mu = 0$ for all $k$. (Problem with Uniqueness).

Stieltjes transform method: $\frac{1}{\lambda_i - z}$ can be arbitrary large.

## Second story: Simplex Method

Linear programming: maximize $c \cdot x$, under the constrain that $Ax \leq b$.

($A$ is a matrix, $b, c$ are vectors, $\leq$ means coordinate-wise)

For a long time there was no polynomial algorithms (until Karmakar's and GLS, 1980s), but in practice problems still got solved quickly.

The most popular algorithm is Simplex method, which can be exponential in the worst case (there are BAD polytopes), but runs wonderfully in practice.

There have been lots of attempts to explain this, but a reasonable one came only very recently.

Spielman-Teng smooth analysis: Noise is Helpful.

The real data is $A$, but the computer will work with $A + Noise$. The equality $(A + Noise)x \leq b$ with very high probability, determines a GOOD polytope.

*Key parameter.* $\kappa(A) = \|A\| \|A^{-1}\|$ (condition number of $A$).

GOOD (well-conditioned) if $\kappa$ small: $n^{O(1)}$.

*Fact.* $\|A\|$ is typically small. But $\|A^{-1}\|$ can be very large. There is a $\pm 1$ matrix of size $n$ whose inverse has norm $n^{(1/2+o(1))n}$ (Alon-V. 96).

**Theorem.** (Spielman-Teng 2000) Assume $\|A\| = n^{O(1)}$, then with high probability $\kappa(A + Gaussian) = n^{O(1)}$.

**Problem.** What happens with discrete noise ?

## Third story: Concentration function.

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

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

Littlewood-Offord (1943), Erdös (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-Szemeredi (65), Stanley (80), Frankl-Furedi (88)etc.

For example:

**Theorem.** (Erdös-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 the random sum is contained in $nB$. By the pigeon hole principle:

$$P_v \geq \frac{1}{Vol(nB)} = n^{-d}V^{-1}.$$

Fact: If $a_i$ are elements of a low dimensional box with small volume, then $P_v$ is large.

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

**Main Ingredient.** Harmonic analysis, Dissociated Sets (Sidon), Random walks.

## Simplex Method revisited

*Geometrical meaning of* $|M^{-1}\|$. Let $d_i$ be the distance from the $i$th row vector to the hyperplane spanned by the other $n-1$ rows and define $d := \min d_i$. Then

$$\|M^{-1}\| \approx \frac{1}{d}.$$

We need to show that for $M = A + Noise$, $d$ is typically large.

*Toy case.* $A = 0$. $M = Gaussian$. One can fix the hyperplane. $d_i$ is well understood.

*Main difficulty with the discrete case.* There are BAD hyperplanes. For example: if $H$ has normal vector $(1, 1, 0, \cdots, 0)$, then half of the vertices of the $\pm 1$ hypercube have distance zero to $H$.

$H$ is BAD if it contains many (say at least a $n^{-C}$ fraction) vertices of the hypercube.

*Key point.* Understand the BAD planes.

Let $v = (a_1 \cdots, a_n)$ be the normal vector of $H$. *$H$ is BAD then $P_v$ is large*, since the number of vertices on $H$ is exactly

$$2^n P(a_1\xi_1 + \cdots + a_n\xi_n = 0).$$

Using the Inverse Theorem (and many other things), we settle Spielman-Teng question

**Theorem.** (Tao-V. 2006) Assume $\|A\| = n^{O(1)}$, then with high probability $\kappa(A + Bernoulli) = n^{O(1)}$.

**Main ingredients.** Inverse Theorem. Discretization of generalized arithmetic progression. A variant of the singularity bound.

**Remark.** Bernoulli can be replaced by very general models, which capture some important features in real-life situations.

(1) The entries do not need to be i.i.d: The variance of noise occurring to a large entry is usually larger than that of the noise occurring to a small entry.

(2) We can allow many entries be "frozen", i.e., noise is zero. For instance, 0 usually has no noise.

## Circular Law revisited

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 $\eta_n$ 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.

But the condition number theorem addresses this point completely (take $A = -zI_n$).

**Theorem.** (Tao-V. 2007) Bernoulli random matrix satisfies the circular law.

Gotze-Tikhomirov (2007) proved the same result.

*Generalizations:*

G-T: entries have sub-gaussian distribution: $P(|\xi \geq T|) \leq exp(-cT^2)$.

T-V.: $(2+\epsilon)$ moment may already satisfy.