Summer School in Peking University
Random Matrix

Professor: Joel A. Tropp
Scribe: Alex

Contents

1 Overview

(Q1:) What is a random matrix? It is a matrix where entries are random variables, maybe dependent. We denote it as

$$Z = \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1j} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2j} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{i1} & Z_{i2} & \cdots & Z_{ij} & \cdots & Z_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mj} & \cdots & Z_{mn} \end{pmatrix}$$

Examples:

- Independent matrix, i.e. noise matrix.
- Independent columns, i.e. $(Z_1, ..., Z_n)$.
- Independent rows. We can use it to do dimension reduction.

We focus on symmetric random matrix. Here are some examples:

- Wigner Matrix: $X_{jk} = X_{kj}$ for all j, k, and (X_{jk}) are independent.
- Covariance Matrix: $X = \sum_{i=1}^{n} Z_i Z_i^*$, where (Z_i) are independent.

(Q2:) Why is random matrix theory a thing? A matrix acts on vectors: $u \mapsto Zu$, we need to understand this random linear map.
Examples:

- Spectral norm: $\|Z\| = \max_{\|u\|=1} \|Zu\|_2 = \sigma_{\max}(Z)$. (How far the ellipsoid of $Z(B(0,1))$ from zero or by what factor can Z dilate a vector?)
- Minimaum singular value: $\sigma_{\min}(Z) = \min_{\|u\|_2=1} \|Zu\|_2$. (How much can we contract?) Z has a null space iff $\sigma_{\min}(Z) = 0$.
- For symmetric situation, we can focus on the eigenvalues. We can claim $\lambda_{\max}(X) = \max_{\|u\|_2=1} u^* X u$ and $\lambda_{\min}(X) = \min_{\|u\|_2=1} u^* X u$.

(Q3:) Where do the random matrices come from? - History?

- Hurwitz(1890s): Averaging over orthogonal group (Random orthogonal matrices).
- Wishart(1927s): Sample covariance of a normal population.
- Goldstein + Von Neumann(1951s): Random matrix model for roundoff errors in numerical linear algebra.
- Wigner(1952s): Models for a slow nonclear reaction. They study the distribution of the eigenvalues of large Wigner matrix.

My Interest:

- Compressed sensing.
- Randomized linear algebra + optimization, such as RSVD
- Quantum information theory.
- How did you encounter RMT?

My Perspective: This lecture may be very different from standard textbook of RMT.

- Flexible models.
- Nonasymptotic.
- Resources.

Schedule:

- Scalar concentration.
- Exp matrix concentration, Gaussian.

- Matrix Chernoff and Bernstein + examples.
- Matrix Khinchin theory.
- Gaussian Lipchitz concentration.
- GOE matrices and a proof of Wigner's theorem (not so easy).
- Universality + staff.

2 Scalar Concentration

For a scalar random variable X takes values in $\mathbb{R}$. We expect bounds for $\mathbb{E}X$ and the tail $\mathbb{P}(X \geq t)$.
We define MGF (moment generating function) and the log-mgf are:

$$m_X(\theta) = \mathbb{E}e^{\theta X} \quad \text{and} \quad \xi_X(\theta) = \log \mathbb{E}e^{\theta X}.$$

Example (Gaussian):

$$m_Z(\theta) = \mathbb{E}e^{\theta Z} = \int_{\mathbb{R}} e^{\theta z} \frac{e^{-z^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} = e^{\theta^2\sigma^2/2},$$

and:

$$\xi_X(\theta) = \theta^2\sigma^2/2.$$

Let X be a real rv, then we have

$$\mathbb{E}X \leq \inf_{\theta>0} \theta^{-1}\xi_X(\theta) \leq \sup X,$$

and

$$\mathbb{E}X \geq \sup_{\theta<0} \theta^{-1}\xi_X(\theta) \geq \inf X.$$

Example: Let $Z_1, ..., Z_n \sim \mathcal{N}(0, \sigma^2)$, what is $\mathbb{E}\max_i Z_i$?

$$\mathbb{E}\max_i Z_i \leq \theta^{-1} \log \mathbb{E}e^{\theta \max_i Z_i} \leq \theta^{-1} \log \sum_i \mathbb{E}e^{\theta Z_i} \leq \theta^{-1}(\log n + \log \mathbb{E}e^{\theta X_i}) = \theta^{-1}(\log n + \theta^2\sigma^2/2) = \sqrt{2\sigma^2 \log n}.$$

Problem:

$$\lim_{n\to\infty} \frac{\mathbb{E}\max_i Z_i}{\sqrt{2\sigma^2 \log n}} = 1$$

Let X be a real rv. Then for $t \in \mathbb{R}$:

$$\mathbb{P}(X \geq t) \leq \inf_{\theta>0} e^{-\theta t + \xi_X(\theta)}$$

$$\mathbb{P}(X \leq t) \leq \inf_{\theta<0} e^{-\theta t + \xi_X(\theta)}$$

if the $\xi_X(\theta) \sim \theta^2$, then $\mathbb{P}(X \geq t) \lesssim e^{-ct^2}$.
Example: Let $Z \sim \mathcal{N}(0, \sigma^2)$. For $t \geq 0$,

$$P(X \geq t) \leq \inf_{\theta>0} e^{-\theta t + \xi_Z(\theta)} = \inf_{\theta>0} e^{-\theta t + \sigma^2\theta^2/2} = e^{-t^2/2\sigma^2}.$$

Problem:

$$\mathbb{P}(Z \geq t) \asymp t^{-1} e^{-t^2/2}/\sqrt{2\pi}.$$

If (Y_i) are independent, $X = \sum_{i=1}^n Y_i$, then $\xi_X(\theta) = \sum_{i=1}^n \xi_{Y_i}(\theta)$.
Example: random series: $X = \sum_{i=1}^n \epsilon_i a_i$, where $\epsilon_i \sim Unif[\pm 1]$, $a_i \in \mathbb{R}$.

$$\xi_{Y_i}(\theta) = \log(\frac{1}{2}e^{\theta a_i} + \frac{1}{2}e^{-\theta a_i}) \leq \log\exp(\frac{1}{2}\theta^2 a_i^2) = \frac{1}{2}\theta^2 a_i^2,$$

and:

$$\xi_X(\theta) \leq \sum_{i=1}^n \frac{1}{2}\theta^2 a_i^2 = \frac{1}{2}\theta^2 \nu,$$

where $\nu = \sum_{i=1}^n a_i^2 =$ variance of X. So:

$$\mathbb{P}(X \leq t) \leq \exp(-t^2/2\nu).$$

3 Exponential Matrix Concentration

We focus on $Z = \sum_{i=1}^n S_i$, or $X = \sum_{i=1}^n Y_i$, where $Y_i \in \mathbb{H}_d$. Our goal is to study the behaviour of the $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$

First we define the standard matrix function f. For

$$A = U\mathrm{diag}\{\lambda_1, ..., \lambda_d\}U^*,$$

we define:

$$f(A) = U\mathrm{diag}\{f(\lambda_1), ..., (\lambda_d)\}U^*.$$

For instance, we can define $\exp(A)$ and $\log(A)$. So we can generalize the mgf from last lecture:

$$M_X(\theta) := \mathbb{E}e^{\theta X} \quad \text{and} \quad \Xi_X(\theta) := \log\mathbb{E}e^{\theta X}.$$

And we have a proposition if $X \in \mathbb{H}_d$ is random and symmetric:

$$\mathbb{E}\lambda_{\max}(X) \leq \inf_{\theta>0}\frac{1}{\theta}\log\mathbb{E}\,\mathrm{Tr}e^{\theta X} = \inf_{\theta>0}\frac{1}{\theta}\log\mathrm{Tr}(\exp(\Xi_X(\theta))).$$

$$\mathbb{E}\lambda_{\min}(X) \geq \sup_{\theta<0}\frac{1}{\theta}\log\mathbb{E}\,\mathrm{Tr}e^{\theta X}$$

Proof:

$$\begin{aligned}
\mathbb{E}\lambda_{\max}(X) &\leq \frac{1}{\theta}\mathbb{E}\lambda_{\max}\lambda X && (1)\\
&= \frac{1}{\theta}\log\exp(\mathbb{E}\lambda_{\max}(\theta X)) && (2)\\
&\leq \frac{1}{\theta}\log\mathbb{E}\exp(\lambda_{\max}(\theta X)) && (3)\\
&= \frac{1}{\theta}\log\mathbb{E}(\lambda_{\max}(e^{\theta X})) && (4)\\
&\leq \frac{1}{\theta}\log\mathbb{E}(\mathrm{Tr}(e^{\theta X})) && (5)
\end{aligned}$$

And we immediately have:

$$\mathbb{P}(\lambda_{\max}(X) \geq t) \leq \inf_{\theta>0} e^{-\theta t}\mathbb{E}\,\mathrm{Tr}(e^{\theta X}),$$

$$\mathbb{P}(\lambda_{\max}(X) \leq t) \leq \inf_{\theta<0} e^{-\theta t}\mathbb{E}\,\mathrm{Tr}(e^{\theta X}).$$

However the associative property doesn't hold for matrices A, B unless $[A, B] = 0$. We introduce a strong weapon to deal with:

Theorem 3.1 (Lieb 1973)

Let $H \in \mathbb{H}_d$. For positive definite A, we have the map: $A \mapsto \mathrm{Tr}e^{H+\log A}$ is concave.

We will omit the proof and pay more attention to the application. For $X = \sum_{i=1}^n Y_i$

$$\mathrm{Tr}\exp(\Xi_X(\theta)) = \mathbb{E}\,\mathrm{Tr}(e^{\theta X}) \le \mathrm{Tr}\exp\left(\sum_{i=1}^n \Xi_{Y_i}(\theta)\right),$$

where $\Xi_Y(\theta) = \log \mathbb{E}e^{\theta Y}$. Then we have the main result:

$$\mathbb{E}\lambda_{\max}(X) \le \inf_{\theta>0} \frac{1}{\theta}\left[\log d + \lambda_{\max}(\sum_{i=1}^n \Xi_{Y_i}(\theta))\right].$$

$$\mathbb{P}(\lambda_{\max}(X) \ge t) \le d \inf_{\theta>0} e^{-\theta t + \lambda_{\max}(\sum_{i=1}^n \Xi_{Y_i}(\theta))}.$$

Proof:

$$\begin{aligned}
\mathbb{E}\lambda_{\max}(X) &\le \frac{1}{\theta}\log \mathbb{E}\,\mathrm{Tr}(e^{\theta X}) && (6)\\
&\le \frac{1}{\theta}\log \mathrm{Tr}\exp\left(\sum_{i=1}^n \Xi_{Y_i}(\theta)\right) && (7)\\
&\le \frac{1}{\theta}\log \mathrm{Tr}\exp\left(\sum_{i=1}^n \Xi_{Y_i}(\theta)\right) && (8)\\
&\le \frac{1}{\theta}\log\left[d\lambda_{\max}(e^{\sum_{i=1}^n \Xi_{Y_i}(\theta)})\right] && (9)\\
&\le \frac{1}{\theta}\left[\log d + \lambda_{\max}(\sum_{i=1}^n \Xi_{Y_i}(\theta))\right]. && (10)
\end{aligned}$$

We now consider some important examples. Let $A_1, ..., A_n \in \mathbb{H}_d$, and $\gamma_i \in \mathcal{N}(0,1)$ r.v.s. We have $X = \sum_{i=1}^n \gamma_i A_i \in \mathbb{H}_d$. Every symmetric random matrix with jointly Gaussian entries can be written as above, moreover we can assume $\mathrm{Tr}(A_i A_j) = 0 (i \ne j)$.

Here we calculate the $Y = \gamma A$, where $\gamma \sim \mathcal{N}(0,1)$. We write $A = U\mathrm{diag}\{\lambda_1, ..., \lambda_d\}U^*$.We have $M_Y(\theta) = \mathbb{E}e^{\theta Y} = \mathbb{E}e^{\theta\gamma U\mathrm{diag}\{\lambda_1,...,\lambda_d\}U^*} = U\mathrm{diag}\{e^{\theta^2\lambda_i^2/2}\}U^* = \exp(\frac{\theta^2}{2}A^2)$, and thus $\Xi_X(\theta) = \frac{\theta^2}{2}A^2$.

Naturally, we generalize the results to the Gaussian series. Let $X = \sum_{i=1}^n \gamma_i A_i$, define $\nu(X) = \|\mathbb{E}X^2\| = \|\sum_{i=1}^n A_i^2\|$. We can easily have:

$$\mathbb{E}\lambda_{\max}(X) \le \sqrt{2\nu(X)\log d}.$$

$$\mathbb{P}(\lambda_{\max}(X) \ge t) \le d\exp(-t^2/2\nu(X)).$$

Another example is the diagonal Gaussian, i.e.:

$$X = \begin{pmatrix} \gamma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_i & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \gamma_n \end{pmatrix},$$

where $\gamma_i \sim \mathcal{N}(0,1)$. We have $\mathbb{E}\lambda_{\max}(X) = \mathbb{E}\max_i \gamma_i \leq \sqrt{2\log d}$ (max-inequality), while $\nu(X) = \|\sum_{i=1}^d E_{ii}^2\| = \|I_d\| = 1$. This is compatible.

Another example is Gaussian orthogonal ensemble: $X_{goe} = \frac{1}{\sqrt{2d}}(G+G^*)$, where $G = [X_{ij}] \in \mathbb{H}_d$. We know $\nu(X) = \frac{d+1}{d}$ and $\mathbb{E}\lambda_{\max}(X_{goe}) \leq \sqrt{2(1+d^{-1})\log d} \approx \sqrt{2\log d}$. But the accurate answer is 2 !

If $Z = \sum_{i=1}^n \gamma_i B_i \in \mathbb{R}^{d_1\times d_2}$, then $\nu(Z) = \|\mathbb{E}(Z^*Z)\| \wedge \|\mathbb{E}(ZZ^*)\|$, and $\mathbb{E}\|Z\| \leq \sqrt{2\nu(Z)\log(d_1+d_2)}$.

We can check that: $\mathbb{E}(Z) \gtrsim \sqrt{\nu(Z)}$, and the gap is $\sqrt{d}$.

4 Matrix Chernoff Inequality

We now ask whan is a random sub-matrix is nonsingular? We write $A = [a_1, ..., a_n] \in \mathbb{R}^{n\times d}$, A is surjective $\iff \lambda_{\min}(AA^*) > 0$. And consider $Z = [\delta_1 a_1, ..., \delta_n a_n]$, where $\delta_i \in \mathrm{Bern}(p)$. Z surjective $\iff \lambda_{\min}(\sum_{i=1}^n \delta_i a_i a_i^*) > 0$.

We recall the Chernoff Theorem in scalar form:

Theorem 4.1 (Scalar Chernoff Inequality)

Let $X = \sum_{i=1}^n Y_i$, where (Y_i) independent in $\mathbb{R}$, $0 \leq Y_i \leq R$.

$$\mathbb{P}(X \geq (1+t)\mathbb{E}X) \leq \left(\frac{e^t}{(1+t)^{1+t}}\right)^{\mathbb{E}X/R} \leq \left(\frac{e}{(1+t)}\right)^{(1+t)\mathbb{E}X/R}$$

$$\mathbb{P}(X \leq (1-t)\mathbb{E}X) \leq \left(\frac{e^{-t}}{(1-t)^{1-t}}\right)^{\mathbb{E}X/R} \leq e^{-t^2\mathbb{E}X/2R}$$

Remark: this bound may not be good for t is small.

We recall the theory over psd matrix order. We say $A \preceq B$ iff $B - A$ is psd. The definition $\iff \lambda_i(A) \leq \lambda_i(B)$. A warning is $A \preceq B \implies f(A) \preceq f(B)$ is usually wrong, e.g wrong for $f(t) = t^2$ and $f(t) = e^t$, but true for $f(t) = \log t$.

We introduce the lemma: For random symmetric matrix Y with $0 \preceq Y \preceq RI$, then:

$$\Xi_Y(\theta) = \log \mathbb{E}e^{\theta Y} \preceq \left(\frac{e^{\theta R}-1}{\theta R}\right)\mathbb{E}Y.$$

Proof: By convexity, we have $e^{\theta a} \leq 1 + \frac{e^{\theta R}-1}{R}a$. Then

$$e^{\theta A} \leq I + \frac{e^{\theta R}-1}{R}A,$$

for $A \in \mathbb{H}_d$ and eigenvalues $\in [0, R]$. So:

$$\mathbb{E}e^{\theta Y} \preceq I + \frac{e^{\theta R}-1}{R}\mathbb{E}Y.$$

We then use the logarithm monotone and:

$$\log \mathbb{E}e^{\theta Y} \preceq \log(I + \frac{e^{\theta R}-1}{R}\mathbb{E}Y) \preceq \frac{e^{\theta R}-1}{R}\mathbb{E}Y.$$

Then we introduce the main theorem:

Theorem 4.2 (Matrix Chernoff Inequality)

Let $X = \sum_{i=1}^n Y_i$, where (Y_i) independent in $\mathbb{H}_d$, $0 \preceq Y_i \preceq RI$.

$$\mathbb{P}\left(\lambda_{\max}(X) \geq (1+t)\lambda_{\max}(\mathbb{E}X)\right) \leq d\left(\frac{e^t}{(1+t)^{1+t}}\right)^{\mathbb{E}\lambda_{\max}(X)/R} \leq d\left(\frac{e}{(1+t)}\right)^{(1+t)\mathbb{E}\lambda_{\max}(X)/R}.$$

$$\mathbb{P}\left(\lambda_{\min}(X) \leq (1-t)\mathbb{E}\lambda_{\min}(X)\right) \leq d\left(\frac{e^{-t}}{(1-t)^{1-t}}\right)^{\mathbb{E}\lambda_{\min}(X)/R} \leq de^{-t^2\mathbb{E}\lambda_{\min}(X)/2R}.$$

5 Matrix Bernstein Inequality

We now want to estimate random multiplication by random sampling. Now we have $A = [a_1, ..., a_n]$ and we want to compute $AA^* = \sum_{i=1}^n a_i a_i^* \in \mathbb{R}^{d\times d}$. The computation cost is $O(nd^2)$.

A probability algorithm is to compute $X = p^{-1}\sum_{i=1}^n \delta_i a_i a_i^*$. We have $\mathbb{E}X = AA^*$, which means it is a unbiased estimation. We denote:

$$X - AA^* = \sum_{i=1}^n (p^{-1}\delta_i - 1)a_i a_i^* = \sum_{i=1}^n S_i = Z.$$

We have $\mathbb{E}S_i = 0$ and $\|S_i\| \leq p^{-1}\|a_i\|^2 \leq R$.

We introduce the Bernstein theorem:

Theorem 5.1 (Scalar Bernstein Inequality)

Let $Z = \sum_{i=1}^n S_i$, where (S_i) independent in $\mathbb{R}$, $|S_i| \leq R$. For $t \geq 0$:

$$\mathbb{P}(|Z| \geq t) \leq 2\exp(-\frac{t^2/2}{\nu + Rt/3}),$$

where $\nu = \mathrm{Var}[Z] = \sum_{i=1}^n \mathbb{E}S_i^2$.

Remark: We can observe the rate is first quadratic and then linear as t increases.

And the matrix version is:

Theorem 5.2 (Matrix Bernstein Inequality)

Let $Z = \sum_{i=1}^n Y_i$, where (Y_i) independent in $\mathbb{R}^{d_1\times d_2}$, $\|Y_i\| \leq R$. For $t \geq 0$:

$$\mathbb{P}(\|Z\| \geq t) \leq (d_1 + d_2)\exp(-\frac{t^2/2}{\nu + Rt/3}),$$

where $\nu = \max\{\mathbb{E}[ZZ^*], \mathbb{E}[Z^*Z]\} = \max\{\sum_{i=1}^n \mathbb{E}[Y_iY_i^*], \mathbb{E}[Y_i^*Y_i]\}$.

$$\mathbb{E}\|Z\| \leq \sqrt{2\nu(Z)\log(d_1+d_2)} + \frac{1}{3}R\log(d_1+d_2).$$

6 Matrix Khinchin Inequality

Tool box we need here:

- $\mathbb{E}[\gamma_i f(\gamma_1, ..., \gamma_n)] = \mathbb{E}[(\partial_i f)(\gamma_1, ...\gamma_d)]$.
- $D(A^p)(H) = \sum_{q=0}^{p-1} A^q H A^{p-1-q}$.
- Matrix consolidation: $|\operatorname{Tr}[HA^q HA^r]| \le \operatorname{Tr}(H^2 |A|^{q+r})$.

Now we focus on the bound on the matrix norm.

Theorem 6.1 (Matrix Khinchin Inequality)

Let $X = \sum_{i=1}^n \gamma_i H_i$, where $H_i \in \mathbb{H}_d$, $\gamma \sim \mathcal{N}(0,1)$ i.i.d. Define $\nu(X) = \|\mathbb{E}X^2\| = \|\sum_{i=1}^n H_i^2\|$. Then for $p \in \mathbb{N}$:

$$\mathbb{E}\operatorname{Tr} X^{2p} \le d(2p-1)!!\nu(X)^p$$

We can prove that: $((2p-1)!!)^{1/2p} \le \sqrt{\frac{2p+1}{e}}$, and then we can deduce the following inequality:

$$\mathbb{E}\|X\| \le (\mathbb{E}\operatorname{Tr} X^{2p})^{1/2p} \le d^{1/2p}\sqrt{\frac{2p+1}{e}}\sqrt{\nu(X)},$$

and choose proper $p \asymp \frac{1}{\log d}$:

$$\mathbb{E}\|X\| \le \sqrt{(2\log d + 2)\nu(X)}.$$

For exponential momentum, we have the bound:

$$\mathbb{E}\operatorname{Tr} e^{\theta X} \le d e^{\theta^2 \nu(X)/2}.$$

That is stronger then exp matrix concentration inequality.

Proof of Theorem 6.1: Use the toolbox, we have:

$$\begin{aligned}
E &:= \mathbb{E}\operatorname{Tr}(X^{2p}) = \mathbb{E}\operatorname{Tr}[X \cdot X^{2p-1}] && (11)\\
&= \mathbb{E}((\sum_{i=1}^n \gamma_i H_i) X^{2p-1}) && (12)\\
&= \sum_{i=1}^n \mathbb{E}[\gamma_i \operatorname{Tr}(H_i X^{2p-1})] && (13)\\
&= \sum_{i=1}^n \mathbb{E}\operatorname{Tr}[H_i(\partial_{\gamma_i} X^{2p-1})] && (14)\\
&= \sum_{i=1}^n \mathbb{E}\operatorname{Tr}[H_i \sum_{q=0}^{2p-2} X^q(\partial_{\gamma_i} X) X^{2p-2-q}] && (15)\\
&= \sum_{p=0}^{2q-2}\sum_{i=1}^n \mathbb{E}\operatorname{Tr}[H_i X^q H_i X^{2p-2-q}] && (16)\\
&\le \sum_{p=0}^{2q-2}\sum_{i=1}^n \mathbb{E}\operatorname{Tr}[H_i^2 X^{2p-2}] && (17)\\
&= (2p-1)\mathbb{E}\operatorname{Tr}(V X^{2p-2}) && (18)\\
&\le (2p-1)\|V\|\mathbb{E}\operatorname{Tr} X^{2p-2} && (19)\\
&= (2p-1)\nu(X)\mathbb{E}\operatorname{Tr} X^{2p-2} && (20)\\
&= (2p-1)!!\nu(X)^p \mathbb{E}\operatorname{Tr}(I) && (21)\\
&= d(2p-1)!!\nu(X)^p. && (22)
\end{aligned}$$

The goe matrix can't get the equality either and the problem lies in the consolidation inequality.

7 Gaussian Lipchitz Concentration

Our main ingredients are:

- Gaussian IBP: $\mathbb{E}[\gamma_i f(\gamma_1,...,\gamma_n)] = \mathbb{E}[(\partial_i f)(\gamma_1,...\gamma_d)]$.
- Gaussian Interpolation: Given independence standard normal $Z = (\gamma_1,...,\gamma_n)$ and $Z' = (\gamma_1',...,\gamma_n')$. For $t \in [0,1]$:
$$Z_t = tZ + \sqrt{1-t^2}Z'.$$

Theorem 7.1 (Gaussian Covariance Identity)

Let Z, Z', Z_t as above. Let $f, g : \mathbb{R}^n \mapsto \mathbb{R}$,

$$\mathrm{Cov}(f(Z), g(Z))) = \mathbb{E}[f(Z)g(Z)] - \mathbb{E}[f(Z)]\mathbb{E}[g(Z)] = \int_0^1 \mathrm{d}t\mathbb{E}\langle \nabla f(Z), \nabla g(Z_t)\rangle.$$

where $\nabla f = (\partial_1 f, ..., \partial_n f)$.

Proof:

$$\begin{aligned}
E := \mathrm{Cov}(f(Z), g(Z)) &= \mathbb{E}_{Z,Z'}[f(Z)g(Z) - f(Z)g(Z')](\text{create a copy}) && (23)\\
&= \mathbb{E}[\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f(Z)g(Z_t)\,\mathrm{d}t] && (24)\\
&= \int_0^1 \mathbb{E}[f(Z)\sum_{i=1}^n (\partial_i g)(Z_t)\frac{\mathrm{d}}{\mathrm{d}t}(Z_t)_i] && (25)
\end{aligned}$$

We notice that:
$$\frac{\mathrm{d}}{\mathrm{d}t}(Z_t)_i = \frac{\mathrm{d}}{\mathrm{d}t}(t\gamma_i + \sqrt{1-t^2}\gamma_i') = \gamma_i - \frac{t}{\sqrt{1-t^2}}\gamma_i'$$

So:

$$\begin{aligned}
E &= \int_0^1 \mathrm{d}t \sum_{i=1}^n \mathbb{E}[f(Z)(\partial_i g)(Z_t)\gamma_i - \frac{t}{\sqrt{1-t^2}} f(Z)(\partial_i g)(Z_t)\gamma_i'] && (26)\\
&= \int_0^1 \mathrm{d}t \sum_{i=1}^n \mathbb{E}[(\partial_i f)(Z)(\partial_i g)(Z_t) + f(Z)(\partial_{ii} g)(Z_t)t - \frac{t}{\sqrt{1-t^2}} f(Z)(\partial_{ii} g)(Z_t)\sqrt{1-t^2}] && (27)\\
&= \int_0^1 \mathrm{d}t \sum_{i=1}^n \mathbb{E}(\partial_i f)(Z)(\partial_i g)(Z_t) && (28)\\
&= \int_0^1 \mathrm{d}t\mathbb{E}\langle \nabla f(Z), \nabla g(Z_t)\rangle. && (29)
\end{aligned}$$

Nice proof!

An important corollary is **Gaussian Poincare Inequality**: For $z \sim \mathcal{N}(0, I_n)$, $f : \mathbb{R}^n \to \mathbb{R}$. Then

$$\mathrm{Var}(f(z)) = \int_0^1 \mathrm{d}t\mathbb{E}\langle \nabla f(Z), \nabla f(Z_t)\rangle \leq \int_0^1 \mathrm{d}t[\mathbb{E}\|\nabla f(Z)\|_2^2]^{\frac{1}{2}}[\mathbb{E}\|\nabla f(Z_t)\|_2^2]^{\frac{1}{2}} = \mathbb{E}\|\nabla f(Z)\|_2^2.$$

Remark: The inequality is a kind of amplifier, $\mathbb{E}f^2$ is large when $\mathbb{E}(f')^2$ is large. The inequality is the foundation of the langevin monte-carlo analysis.

If f is Lipchitz, i.e $\|\nabla f(x)\| \leq L$ almost everywhere. Then the poincare inequality implies:

$$\mathrm{Var} f(Z) \leq L^2.$$

Let's give an interesting example. Let $S \in \mathbb{R}^n$ be compact. $\mathrm{dist}(x, S) = \min_{a \in S} \|x - a\|_2$, it's an Lipchitz function. We have:

$$\mathrm{Var}(\mathrm{dist}(z, S)) \leq 1.$$

However we don't know $\mathbb{E}(\mathrm{dist}(z, S))$ haha.

Theorem 7.2 (Gaussian Lipchitz Concentration)

Let $Z \in \mathcal{N}(0, I_n)$. For $f : \mathbb{R}^n \to \mathbb{R}$, which is L-Lipchitz.

$$\mathbb{P}(|f(Z) - \mathbb{E}f(Z)| \geq t) \leq 2e^{-t^2/2L^2}.$$

For example we have:

$$\mathbb{P}(|\mathrm{dist}(z, S) - \mathbb{E}\mathrm{dist}(z, S)| \geq t) \leq 2e^{-t^2/2}.$$

Proof: We only need to prove:

$$m(\theta) = \mathbb{E}e^{\theta f(z)} \leq e^{\theta^2 L^2/2}.$$

WLOG, assume $\mathbb{E}f(Z) = 0$. We have

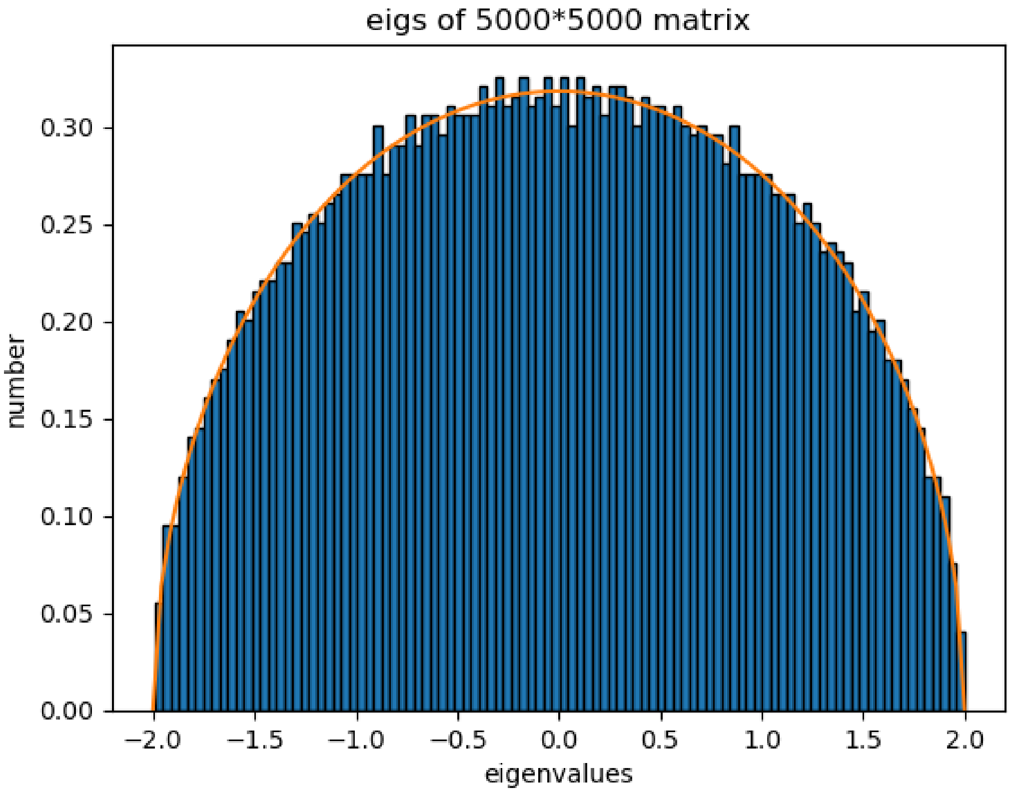
$$\begin{aligned}
m'(\theta) &= \mathbb{E}[f(z)e^{\theta f(z)}] = \mathrm{Cov}(f(z), e^{\theta f(z)}) && (30)\\
&= \int_0^1 \mathrm{d}t \mathbb{E}\langle \nabla f(z), \theta \nabla f(Z_t) e^{\theta f(Z_t)} \rangle && (31)\\
&= \theta \int_0^1 \mathrm{d}t \mathbb{E}\langle \nabla f(z), \theta \nabla f(Z_t) \rangle e^{\theta f(Z_t)} && (32)\\
&\leq \theta \int_0^1 \mathrm{d}t [\mathbb{E}\|\nabla f(Z)\|_2^2]^{1/2} [\mathbb{E}\|\nabla f(Z_t)\|_2^2]^{1/2} e^{\theta f(Z_t)} && (33)\\
&\leq \theta L^2 \mathbb{E}[e^{\theta f(Z)}] && (34)\\
&= \theta L^2 m(\theta) && (35)
\end{aligned}$$

With Gronwall's inequality,

$$m(\theta) \leq \exp(\frac{1}{2}\theta^2 L^2).$$

The whole picture of this theorem is: $\|\cdot\|$, $\lambda_{\max}$, $\lambda_{\min}$ are all 1-Lipchitz function w.r.t. $\|\cdot\|_F$. An example (maybe exercise):

- $\mathbb{E}\lambda_{\max}(X) \leq \sqrt{2(d-1)\log d}$.
- $\mathbb{P}(\lambda_{\max}(X) \geq t) \leq de^{-t^2/2(d-1)}$.



8 GOE and Semicircle Law

We recall that $X_{goe(d)} = \frac{1}{\sqrt{2d}}(G + G^*)$ where $G \in \mathbb{H}_d$ has iid $\mathscr{N}(0, 1)$ entries. We have:

$$\mathbb{E}\|X_{goe}\| \le \sqrt{2(1 + d^{-1})\log(2d)},$$

and

$$\mathbb{P}(\|X_{goe} - \mathbb{E}X_{goe}\| \ge t) \le e^{-t^2 d/4}.$$

Now I wonder what do the eigenvalues of X_{goe} actually look like.

The eigenvalues have a nice profile, which looks like a semicircle:

$$\phi_{sc} = \frac{1}{2\pi}\sqrt{4 - t^2}.$$

We try to model the distribution of eigenvalues using a probability measure on $\mathbb{R}$. The eigenvalues are: $\lambda_1(A) \ge ... \ge \lambda_n(A)$.

We denote:

$$\mu_A = \frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i(A)}. \quad \text{(total mass = 1)}$$

and we have:

$$\mu_A(E) = \frac{\#\{i : \lambda_i(A) \in E\}}{d}.$$

or

$$\mu_A(f) = \int_{\mathbb{R}} f(t)\mu_A(\mathrm{d}t) = \frac{1}{d}\sum_{i=1}^{d} f(\lambda_i(A)).$$

Here we define the empirical spectral distribution (ESD): The spectral measure of μ_X of X:

$$\mu_X = \frac{1}{d}\sum_{i=1}^{d}\delta_{\lambda_i}(X)$$

It is a "random measure". We define:

$$\bar{\mu}_X(f) = \int_{\mathbb{R}} f(t)\mu_X(\mathrm{d}t) = \bar{\mathrm{Tr}}\, f(X).$$

is a random variance for each f.

Theorem 8.1 (Mean spectral distribution)

A random matrix $X \in \mathbb{H}_d$ admits a unique probability measure $\bar{\mu}(X)$ on $\mathbb{R}$ s.t.

$$\int_{\mathbb{R}} f(t)\bar{\mu}_X(\mathrm{d}t) = \mathbb{E}\bar{\mathrm{Tr}}\, f(X),$$

for all $f : \mathbb{R} \to \mathbb{R}$.

We now define μ_{SC} to be the probability measure on $\mathbb{R}$ with density $\phi_{sc} = \frac{1}{2\pi}\sqrt{4-t^2}$. Our goal is to compare the ESD $\mu_{X_{goe}(d)}$ with the semicircle distributions. Our key idea is to compare the moments, and we believe:

$$|\mu(f) - \nu(f)| \approx 0$$

for lots of $f : \mathbb{R} \to \mathbb{R}$, then $\mu \approx \nu$.

By triangle inequality:

$$|\mu_X(f) - \mu_{SC}(f)| \leq |\mu_X(f) - \bar{\mu}_X(f)| + |\bar{\mu}_X(f) - \mu_{SC}(f)|.$$

We'll compare polynomial moments of ESD over the two distributions. (The polynomials can't see the difference.)

We now introduce Catalan numbers: $\mathrm{Cat}_0 = 1$ and $\mathrm{Cat}_p = \frac{1}{p+1}\binom{2p}{p}$. We can prove:

$$\mathrm{Cat}_p \leq 4^p = 2^{2p}.$$

$$\mu_{SC}(t^p) = \int_{\mathbb{R}} t^p \mu_{SC}(\mathrm{d}t) = \mathrm{Cat}_{p/2}.$$

Theorem 8.2 (GOE moments)

Let $X = X_{goe}(d)$. For $p \in \mathbb{Z}_+$, $p \leq \sqrt{d/2}$

- $|\bar{\mu}_X(t^p) - \bar{\mu}_{SC}(t^p)| \leq \frac{4p^2}{d}\mathrm{Cat}_{p/2}$.
- $\mathrm{Var}[\mu_X(t^p)] = \mathbb{E}|\mu_X(t^p) - \bar{\mu}_X(t^p)|^2 \leq \frac{2ep^2}{d}\mathrm{Cat}_{p-1}$.

Let's apply the theorem first. The norm:

$$\mathbb{E}\|X\| \leq 2\left(1 + \frac{e\log d}{\sqrt{d}}\right) \to 2.$$

Proof sketch: By moment method,

$$\mathbb{E}\|X\| \le (d \cdot \mathbb{E}\bar{\mathrm{Tr}} X^{2p})^{1/2p} \le d^{1/2p}\left((1+\frac{cp^2}{d})\mathrm{Cat}_p\right)^{1/2p} \le 2d^{1/2p}(1+\frac{cp^2}{d})^{1/2p}.$$

We minimize it over p.

Another example is: Let $\chi(t) = \sum_{p=0}^{n} c_p t^p$ with $n \le \log d$. We have:

$$|\bar{\mu}_X(\chi) - mu_{SC}(X)| \le \frac{2ne^{4n}}{d} \cdot \max_{|t|\le 2} |\chi(t)|.$$

$$\left[\mathbb{E}|\bar{\mu}_X(\chi) - \mu_{SC}(\chi)|^2\right]^{1/2} \le \frac{4ne^{4n}}{d} \max_{|t|\le 2} |\chi(t)|.$$

Proof sketch: We can prove: $|c_p| \le \frac{(2n)^p}{p!} \max_{|t|\le 2} |\chi(t)|$.

Theorem 8.3 (Eigen's Theorem for GOE)

Let $X_{goe(d)}$ for $d \in \mathbb{N}$. For $f : \mathbb{R} \to \mathbb{R}$ bounded and continuous,

$$\bar{\mu}_{X_{goe(d)}}(f) \to \mu_{SC}(f)$$

$$\mu_{X_{goe(d)}}(f) \to \mu_{SC}(f)$$

Proof sketch:

- $\|X_{(d)}\| \le 3$ for all large d.
- Approximate f by a polynomial on $[-3, 3]$.
- Borel Cantelli.

8.1 GOE Moment Computation

Recall: $X = X_{goe(d)} = \frac{1}{\sqrt{2d}} \sum_{i,j=1}^{d} \gamma_{i,j}(E_{ij} + E_{ji}) =: \sum_k \gamma_k H_k$, where $H_k = \frac{1}{\sqrt{2d}}(E_{ij} + E_{ji})$.

Theorem 8.4 (GOE moments)

For $X = X_{goe}(d)$ and $p \in \mathbb{Z}_+$,

- MSD odd: $\bar{\mu}_X(t^{2p+1}) = 0$
- MSD even: $(1-\epsilon_p)_+ \le \bar{\mu}_X(t^{2p}) \le e^{\epsilon_p}\mathrm{Cat}_p$, where $\epsilon_p = p^2/d + p^4/d^2$.
- ESD: $\mathrm{Var}[\mu_X(t^p)] \le \frac{2e^{\epsilon_{p-1}}p^2}{d^2}\mathrm{Cat}_{p-1}$.

We'll follow the proof of MKI without so many inequalities.

Proof: **Step I:** $p = 0$, $\bar{\mu}_X(t^0) = 1$;
$p = 2k+1$, $X \sim -X$, and $\mathbb{E}\bar{\mathrm{Tr}} X^{2k+1} = \mathbb{E}\bar{\mathrm{Tr}}(-X)^{2k+1}(-1)\mathbb{E}\bar{\mathrm{Tr}} X^{2k+1} = 0$.

Step II: Lemma: For $A \in \mathbb{H}_d$, $\sum_k H_k A H_k = (\bar{\mathrm{Tr}} A)I + d^{-1}A$.

Step III: $\alpha_p = \mathbb{E}\bar{\mathrm{Tr}}X^{2p}$. So:

$$\alpha_{p+1} := \mathbb{E}\bar{\mathrm{Tr}}[X^{2p+2}] = \mathbb{E}\bar{\mathrm{Tr}}[X \cdot X^{2p+1}] = \sum_k \mathbb{E}[\gamma_k \cdot \bar{\mathrm{Tr}}[H_k \cdot X^{2p+1}]] \tag{36}$$

$$= \sum_{q=0}^{2p} \sum_k \mathbb{E}\bar{\mathrm{Tr}}[H_k X^q H_k X^{2p-q}] = \sum_{q=0}^{2p} \mathbb{E}\bar{\mathrm{Tr}}[((\bar{\mathrm{Tr}}X^q)I + d^{-1}X^q) \cdot X^{2p-q}] \tag{37}$$

$$= \sum_{q=0}^{2p} \mathbb{E}\left[(\bar{\mathrm{Tr}}X^q)(\bar{\mathrm{Tr}}X^{2p-q}) + \frac{1}{d}\bar{\mathrm{Tr}}[X^{2p}]\right] \tag{38}$$

$$= \left[\sum_{q=0}^{2p} \mathbb{E}[(\bar{\mathrm{Tr}}X^q)(\bar{\mathrm{Tr}}X^{2p-q})] + \mathrm{Cov}(\bar{\mathrm{Tr}}(X^q), \bar{\mathrm{Tr}}(X^{2p-q}))\right] + \frac{2p+1}{d}\alpha_p \tag{39}$$

$$= \sum_{q=0}^{p} \alpha_q \alpha_{p-q} + \sum_{q=0}^{2p} \mathrm{Cov}(\bar{\mathrm{Tr}}(X^q), \bar{\mathrm{Tr}}(X^{2p-q}) + \frac{2p+1}{d}\alpha_p \tag{40}$$

$$= \sum_{q=0}^{p} \alpha_q \alpha_{p-q} + \mathrm{err}_{p+1}. \tag{41}$$

Notice that: $\mathrm{Cat}_{p+1} = \sum_{q=0}^{p} \mathrm{Cat}_q \mathrm{Cat}_{p-q}$ for $p \in \mathbb{Z}_+$.

Step IV: We now calculate the error:

$$\mathrm{err}_{p+1} = \frac{2p+1}{d}\alpha_p + \sum_{q=0}^{2p} \mathrm{Cov}(\bar{\mathrm{Tr}}(X^q), \bar{\mathrm{Tr}}(X^{2p-q})$$

Recall the lemma:

$$\mathrm{Cov}(f(z), g(z)) = \int_0^1 \mathrm{d}t \mathbb{E}\langle \nabla f(z), \nabla g(z_t)\rangle.$$

Now we let $f(z) = \bar{\mathrm{Tr}}(X^q)$, and $g(z) = \bar{\mathrm{Tr}}(X^{2p-q})$. Preparation: $\gamma'_{ij} \sim \mathscr{N}(0,1)$, $X' = X(\gamma') = \frac{1}{\sqrt{2d}} \sum_{ij} \gamma'_{ij}(E_{ij} + E_{ji})$, $X_t = tX + \sqrt{1-t^2}X'$. Of course $X_t \sim X$ for all $t \in [0,1]$.

The derivation:

$$\partial_{\gamma_{ij}} f(\gamma) = q \cdot \bar{\mathrm{Tr}}[X^{q-1}(\partial_{\gamma_{ij}} X)] = q \cdot \bar{\mathrm{Tr}}[X^{q-1}(\frac{1}{\sqrt{2d}}(E_{ij} + E_{ji}))] \tag{42}$$

$$= \frac{2q}{\sqrt{2}d^{3/2}}(X^{q-1})_{ij}.(\text{Since } X^{q-1} \text{ is symmetric, and } \bar{\mathrm{Tr}} \text{ has a mean term } \frac{1}{d}) \tag{43}$$

Similarly:

$$\partial_{\gamma_{ij}} g(\gamma) = \frac{2(2p-q)}{\sqrt{2}d^{3/2}}(X^{2p-q-1})_{ij}.$$

So:

$$\langle \nabla f(\gamma), \nabla g(\gamma_t)\rangle = \sum_{ij} \frac{\sqrt{2}q}{d^{3/2}}(X^{q-1})_{ij} \frac{\sqrt{2}(2p-q)}{d^{3/2}}(X_t^{2p-q-1})_{ij} = \frac{2q(2p-q)}{d^3} \mathrm{Tr}[X^{q-1}X_t^{2p-q-1}].$$

The expectation is:

$$|\mathbb{E}\langle \nabla f(\gamma), \nabla g(\gamma_t)\rangle| = \frac{2q(2p-q)}{d^2}|\mathbb{E}[X^{q-1}X_t^{2p-q-1}]| \tag{44}$$

$$\leq \frac{2q(2p-q)}{d^2}\left(\mathbb{E}\bar{\mathrm{Tr}}X^{2p-2}\right)^{(q-1)/2p-2}\left(\mathbb{E}\bar{\mathrm{Tr}}X_t^{2p-2}\right)^{(2p-q-1)/2p-2} \tag{45}$$

$$= \frac{2q(2p-q)}{d^2}\alpha_{p-1}. \tag{46}$$

Now we calculate the covariance.

$$\mathrm{Cov}(\bar{\mathrm{Tr}}(X^q), \bar{\mathrm{Tr}}(X^{2p-q}) \leq \frac{2q(2p-q)}{d^2}\alpha_{p-1}. \tag{47}$$

and

$$|\mathrm{err}_{p+1} - \frac{2p+1}{d}\alpha_p| \leq \sum_{q=0}^{2p} \frac{2q(2p-q)}{d^2}\alpha_{p-1} \tag{48}$$

$$\leq \frac{3p^3}{d^2}\alpha_{p-1} \tag{49}$$

which means:

$$\frac{2p+1}{d}\alpha_p - \frac{3p^3}{d^2}\alpha_{p-1} \leq \mathrm{err}_{p+1} \leq \frac{2p+1}{d}\alpha_p - \frac{3p^3}{d^2}\alpha_{p-1}.$$

Step V: Solving the recurence: our recurrence is a perturbation of Cat_p. (Technique!)

Step VI: Variance:

$$\mathrm{Var}[\mu_X(t^p)] = \mathrm{Var}[\bar{\mathrm{Tr}}X^p] \leq \mathbb{E}\|\nabla_\gamma(\bar{\mathrm{Tr}}X(\gamma)^p)\|_2^2 \tag{50}$$

$$= \frac{2p^2}{d^2}\mathbb{E}\,\mathrm{Tr}X^{2p-2} = \frac{2p^2}{d^2}\alpha_{p-1} \leq \frac{2p^2}{d^2}e^{\epsilon_{p-1}}\mathrm{Cat}_{p-1}. \tag{51}$$

Q.E.D.

We'll then explore other random matrices "look like" GOE.

9 Universality

Recall that the proofs on GOE exploited it is a gaussian matrices. What about other random matrix models.

For example, Wigner matrix!

$$W = \frac{1}{\sqrt{d}}\sum_{i\leq j} W_{ij}(E_{ij}+E_{ji}) \in \mathbb{H}_d,$$

where $\mathbb{E}W_{ji} = 0$, $\mathrm{Var}W_{ij} = 1$.

Recall the CLT: Let $Y \in L_2$ be a real r.v. with $\mathbb{E}Y = 0$ and $\mathrm{Var}Y = 1$. $T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$, where $Y_i \sim Y$ i.i.d. We have for h continuous and bounded:

$$|\mathbb{E}h(T_n) - \mathbb{E}h(Z)| = 0 \quad \text{as } n \to \infty.$$

Here we do not require that $Y \sim \mathcal{N}(0,1)$. We want to apply it to matrix models:

$$X = \sum_{i=1}^n Y_i \in \mathbb{H}_d,$$

where (Y_i) is independent and $\mathbb{E}Y_i = 0$.

And the variance transforms into a covariance tensor (the second order statistics):

$$\mathrm{Var}_\otimes[X] = \mathbb{E}[X \otimes X].$$

Here the tensor of the matrix means:

$$(A \otimes A)_{ijkl} = a_{ij} \cdot a_{kl}.$$

Compared with a Gaussian matrix $Z \in \mathbb{H}_d$ with jointly Gaussian entries. (Unique!)

- $\mathbb{E}Z = \mathbb{E}X = 0$.
- $\mathrm{Var}_{\otimes}[Z] = \mathrm{Var}_{\otimes}[X]$.

The second item $\implies \nu(Z) = \|\mathbb{E}Z^2\| = \|\mathbb{E}X^2\| = \nu(X)$.

Recall Multivariate CLT: Suppose $X_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$, where (Y_i) is independent and $\mathbb{E}Y_i = 0$. Let Z be normal, $\mathbb{E}Z = 0$ and $\mathrm{Var}_{\otimes}[Z] = \mathrm{Var}_{\otimes}[Y]$. We have for h continuous and bounded:

$$|\mathbb{E}h(T_n) - \mathbb{E}h(Z)| = 0 \quad \text{as } n \to \infty.$$

Surprisingly, we can compare polynomials of X and Z without i.i.d. assumptions.

Define:

$$\|W\|_p := (\mathbb{E}\bar{\mathrm{Tr}}|W|^p)^{1/p} \quad \text{for } p \geq 1.$$

Theorem 9.1 (2022)

With $X = \sum_{i=1}^n Y_i \in \mathbb{H}_d$ an independent summation and $\mathbb{E}Y_i = 0$. Assume $\|Y_i\| \leq R$, define $\nu(X) = \|\mathbb{E}X^2\|$. Let Z be normal, $\mathbb{E}Z = 0$ and $\mathrm{Var}_{\otimes}[Z] = \mathrm{Var}_{\otimes}[Y] = \mathrm{Var}_{\otimes}[X]$. Then for $p \in \mathbb{N}$,

$$\big|\|X\|_{2p} - \|Z\|_{2p}\big| \lesssim [p^2 R\nu(X)]^{1/3} + pR.$$

Since MKI:

$$\|Z\|_{2p} \lesssim \sqrt{p\nu(Z)},$$

we have the bounds for relative variance:

$$\frac{|\|X\|_{2p} - \|Z\|_{2p}|}{\sqrt{\nu(Z)}} \leq \left(\frac{p^4R^2}{\nu(Z)}\right)^{1/6} + \left(\frac{p^2R^2}{\nu(Z)}\right)^{1/2} (\ll 1).$$

It is nontrivial when $p^4R^2 \ll \nu(Z) = \nu(X)$.

Remark: In Hilbert space: $a_n \to a \iff \|a_n\| \to \|a\|$ and $a_n \rightharpoonup a$.

Now we add another assumption to Wigner matrix: $|W_{ij}| \leq B$ and $|2W_{ij}| \leq B$. We easily have $\nu(W) = \nu(X_{goe}) = 1 + d^{-1} \leq 2$.

Example 1 (Rademacher Entries): $W_{ii} \sim \mathrm{Unif}[\pm\sqrt{2}]$ and $W_{ij} \sim \mathrm{Unif}[\pm 1]$ for $i \neq j$. Here $B = \sqrt{2}$

Example 2 (Sparse Rademacher Entries): It is similar to a adjacent matrix of a sparse graph. $W_{ij} = \frac{1}{\sqrt{s}}\epsilon_{ij}S_{ij}$, $\epsilon_{ij} \sim \mathrm{Unif}\,[\pm 1]$ and $S_{ij} \sim \mathrm{Bern}(s)$. Here $B = \frac{1}{\sqrt{s}}$.

What does the university apply?

$$\frac{p^4B^2}{d} \ll \mathrm{Const}.$$

For example 1, we require $p^4 \ll d$; for example 2, we require $p^4 \ll sd$. When $p \sim \log d$, the inequality holds.

9.1 Concentration

Recall $\|X_{goe(d)}\|_{2p}^{2p} = (1 + \frac{cp^2}{d})\mathrm{Cat}_p \approx \mathrm{Cat}_p$ when $d \to \infty$. Fix $\epsilon > 0$ and let $p \geq \frac{\log d}{\epsilon}$:

$$\begin{aligned}
\mathbb{E}\|w\| &\leq d^{1/2p}(\mathbb{E}\bar{\mathrm{Tr}}W^{2p})^{1/2p} \leq (1+\epsilon)\|W\|_{2p} && (52)\\
&\leq (1+\epsilon)(\|X_{goe}\|_{2p} + \text{universality error}) && (53)\\
&\leq (1+\epsilon)((1 + \text{goe error})\mathrm{Cat}_p^{1/2p} + \text{universality error}) && (54)\\
&\leq 2(1+c\epsilon) && (55)
\end{aligned}$$

Recall the GOE moments, we need to verify:

- MSD odd: $\bar{\mu}_W(t^{2p+1}) = 0$
- MSD even: $(1-\epsilon_p)_+ \le \bar{\mu}_W(t^{2p}) \le e^{\epsilon_p}\mathrm{Cat}_p$, where $\epsilon_p = p^2/d + p^4/d^2$.
- ESD: $\mathrm{Var}[\mu_W(t^p)] \le \frac{2e^{\epsilon_{p-1}}p^2}{d^2}\mathrm{Cat}_{p-1}$.

We notice that $W_{ij} \sim -W_{ij}$, and the first result holds.
For the second claim:

$$\bar{\mu}_W(t^{2p}) = \|W\|_{2p}^{2p} = \|X_{goe(d)}\|_{2p}^{2p}(1 \pm \text{universality error}) \tag{56}$$

$$= \mathrm{Cat}_p\left(1 \pm \frac{cp^2}{d}\right)\left(1 \pm (\frac{B^2p^4}{d})^{1/6}\right)^{2p} \tag{57}$$

$$\to \mathrm{Cat}_p. \tag{58}$$

Still, we need to transfer polynomials to bounded continuous functions.
For the third claim: we need extra concentration results:

$$\mu_W(f) \approx \bar{\mu}_W(f).$$

Since we do not have enough time, the proof is omitted.

10 Lindeberg Exchange

We now focus on CLT.

Theorem 10.1

Let $Y \in L_3$ be a real r.v. with $\mathbb{E}Y = 0$ and $\mathrm{Var}Y = 1$, $\mathbb{E}|Y|^3 \le M_3$. Let $T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$ where (Y_i) are i.i.d. copies of Y. For standardized normal $Z \sim \mathcal{N}(0,1)$ and $h : \mathbb{R} \to \mathbb{R}$:

$$|\mathbb{E}h(T_n) - \mathbb{E}h(Z)| \le \frac{M_3}{\sqrt{n}}\|h'''\|_{\sup}.$$

A corollary is for $h : \mathbb{R} \to \mathbb{R}$ with $|h(t)| \le 1$ and h is 1-Lipchitz.

$$|\mathbb{E}h(T_n) - \mathbb{E}h(Z)| \le \mathrm{Const}\cdot\frac{M_3^{1/3}}{n^{1/6}}.$$

which implies CLT.
The proof of the corollary is to smooth the bounded Lipchitz function.
Now we come back to the theorem.

Lemma 10.1. *Consider $Y, Z \in L_3$ with $\mathbb{E}Y = \mathbb{E}Z$ and $\mathbb{E}Y^2 = \mathbb{E}Z^2$,*

$$|\mathbb{E}h(Y) - \mathbb{E}h(Z)| \le \frac{1}{6}\|h'''\|_{\sup}(\mathbb{E}|Y|^3 + \mathbb{E}|Z|^3).$$

Proof: By Taylor's theorem,

$$|h(Y) - h(0) - h'(0)Y - \frac{1}{2}h''(0)Y^2| \le \frac{1}{6}\|h'''\|_{\sup}|Y|^3$$

$$|h(Z) - h(0) - h'(0)Z - \frac{1}{2}h''(0)Z^2| \le \frac{1}{6}\|h'''\|_{\sup}|Z|^3$$

Combine them together.

Theorem 10.2 (Lindeberg Exchange)

Let $(Y_1, ..., Y_n)$ and $(Z_1, ..., Z_n)$ be independent r.v.s in L_3. Assume $\mathbb{E}Y_i = \mathbb{E}Z_i$ and $\mathbb{E}Y_i^2 = \mathbb{E}Z_i^2$,

$$|\mathbb{E}f(Y_1, ..., Y_n) - \mathbb{E}f(Z_1, ..., Z_n)| \leq \frac{1}{6}\sum_{i=1}^n \|\partial_{iii}f\|_{\sup}(\mathbb{E}|Y_i|^3 + \mathbb{E}|Z_i|^3).$$

Proof: We define: $W_i = (Y_1, ..., Y_i, Z_{i+1}, ..., Z_n)$, $Y = W_n$, $Z = W_0$. Observe the telescope:

$$|\mathbb{E}f(Y_1, ..., Y_n) - \mathbb{E}f(Z_1, ..., Z_n)| \leq \sum_{i=1}^n |\mathbb{E}f(W_i) - \mathbb{E}f(W_{i-1})| \leq \text{RHS}.$$

An important corollary is standard sums. Let $Y, Z \in L_3$ with Y, Z standardized. $M = \mathbb{E}|Y|^3 \vee \mathbb{E}|Z|^3$. (Y_i) and (Z_i) are copies of Y and Z.

$$|\mathbb{E}h(\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i) - \mathbb{E}h(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i)| \leq \frac{\frac{1}{3}M\|h'''\|_{\sup}}{\sqrt{n}}.$$

We now extend the above results to the matrix cases.

Theorem 10.3 (2024)

Let $Y, Z \in \mathbb{H}_d$ symmetric matrices, $\|Y\| \vee \|Z\| \leq R$, s.t. $\mathbb{E}Y = \mathbb{E}Z = 0$ and $\text{Var}_\otimes[Y] = \mathbb{E}[Y \otimes Y] = \mathbb{E}[Z \otimes Z] = \text{Var}_\otimes[Z]$. Consider the i.i.d. sums:

$$W_n = \sum_{i=1}^n Y_i \quad \text{where } (Y_i) \text{ are i.i.d. copies of } Y$$

$$W_0 = \sum_{i=1}^n Z_i \quad \text{where } (Y_i) \text{ are i.i.d. copies of } Y$$

For $p \in \mathbb{N}$

$$\big|\|W_n\|_{2p} - \|W_0\|_{2p}\big| \leq (p^2 R(\nu(W_n) + \nu(W_0)))^{1/3} + pR.$$

Proof idea: Lindeberg!