杨帆 (Fan Yang) 、 宁斋 W02

Topics in Random Matrix Theory

* A dynamical approach to random matrix theory. by L. Erdős and H.-T. Yau.

* Topics in random matrix theory. by Terence Tao.

Random matrix theory aims to study "properties of large random matrices," such as: the operator norm, eigenvalue / eigenvector distributions, condition number, the singular probability, characteristic ~~polynomials~~ polynomials .... Many of these properties reduce to studying the asymptotic behaviors of the "eigenvalues and eigenvectors" as the matrix dimension tends to ∞.

The grand principle

The key concept of RMT is the "~~univer~~ universality phenomenon": the "asymptotic eigenvalue & eigenvector statistics" are independent of the law of matrix elements, but only depend on the symmetry class (i.e., symmetric / hermitian). (Same spirit as LLN and CLT.)

We will illustrate this principle with ~~theree~~ three standard examples.

1. Winger ensemble

Wigner's pioneering work in 1955 marks the birth of RMT. He proposed to use a large real symmetric / complex Hermitian random matrix with independent entries to model the ~~Hatmi~~ Hamiltonian of large nuclei. This simple-minded model surprisingly produce the correct gap statistics between energy levels of large nuclei, indicating the "universality principle" behind the model.

Wigner matices: $H = (h_{ij})_{1 \le i,j \le N}$ is an $N \times N$ self-adjoint matrix with matrix elements having mean 0, variance 1 and independent up to symmetry: $h_{ij} = \bar{h}_{ji}$.

Gaussian orthogonal ensemble (GOE): The entries h_{ij}, $1 \le i \le j \le N$, are real Gaussian random variables, and

$$\mathbb{E} h_{ij} = 0, \quad \mathbb{E} h_{ij}^2 = 1 + \delta_{ij}.$$

Gaussian unitary ensemble (GUE): The upper-triangular entries are i.i.d. $N(0,1)_{\mathbb{C}}$ random variables with $\mathbb{E} h_{ij} = 0$, $\mathbb{E}|h_{ij}|^2 = 1$, $\mathbb{E} h_{ij}^2 = 0$ $(1 \le i < j \le N)$. The diagonal entries are $N(0,1)$ random variables.

The GOE / GUE is ~~orthogonal / ST~~ invariant under orthogonal / unitary transformations.

Prop:

Let H be a GOE, and O be an ~~orto~~ orthogonal matrix. Then, $\overbrace{O^T H O}^{H'} \overset{d}{=} H$.

Pf: We only need to check that $\mathbb{E} H'_{ij} H'_{i'j'} = \delta_{ii'}\delta_{jj'} + \delta_{ij'}\delta_{ji'}$ ~~for $1\le i,j\le N$, $1\le i',j'\le N$~~,

$$\mathbb{E} \sum_{\substack{k,l,\\ k',l'}} H_{kl} O_{ki} O_{lj} \; H_{k'l'} O_{k'i'} O_{l'j'}$$

$$= \sum_k 2 O_{ki} O_{kj} O_{ki'} O_{kj'} + \sum_{k\neq l} (\delta_{kk'}\delta_{ll'} + \delta_{kl'}\delta_{lk'}) O_{ki} O_{lj} O_{k'i'} O_{l'j'}$$

$$= 2\sum_k O_{ki} O_{kj} O_{ki'} O_{kj'} + \sum_{k\neq l} (O_{ki} O_{ki'} O_{lj} O_{lj'} + O_{ki} O_{kj'} O_{lj} O_{li'})$$

$$= \sum_{k,l} (O_{ki} O_{ki'} O_{lj} O_{lj'} + O_{ki} O_{kj'} O_{lj} O_{li'}) = \delta_{ii'}\delta_{jj'} + \delta_{ij'}\delta_{ji'} \qquad \square$$

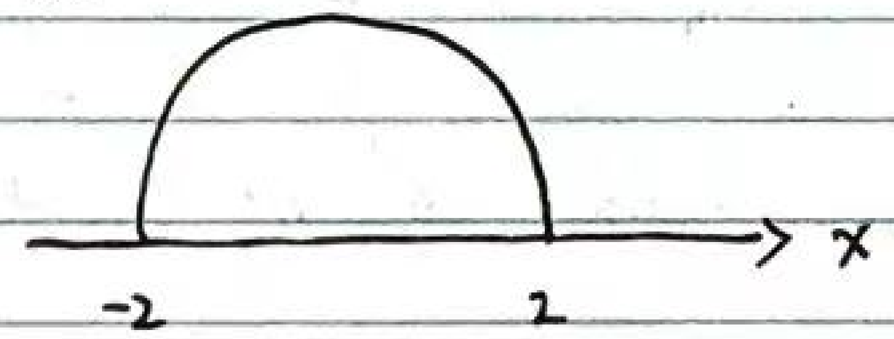
Wigner proved an "LLN" for the empirical spectral density (ESD) of $\frac{1}{\sqrt{N}} H_N$:

$$\mu_{\frac{1}{\sqrt{N}}H_N} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\frac{\lambda_i}{\sqrt{N}}}, \quad \lambda_1 \le \lambda_2 \le \cdots \le \lambda_N \text{ are the eigenvalues of } H_N.$$

$\mu_{\frac{1}{\sqrt{N}}H_N} \to \mu_{sc}$ weakly, where μ_{sc} has density

$$\rho_{sc}(x) = \frac{1}{2\pi}\sqrt{4-x^2}, \quad -2 \le x \le 2.$$

$\rho_{sc}(x)$:



Rmk: The above result implies that for any small constant $\varepsilon>0$, interval $I \subseteq \mathbb{R}$ with $|I| = \varepsilon$,

$$\frac{1}{N}\#\{i : \tfrac{1}{\sqrt{N}}\lambda_i \in I\} = \int_I d\mu_{\frac{1}{\sqrt{N}}H_N}(x) \xrightarrow[N\to+\infty]{} \int_I \rho_{sc}(x)\,dx$$

$\boxed{\text{Q1:}}$ Does the SC law holds in a stronger sense, i.e., ~~$\frac{1}{|I|}\int_I d\mu_{\frac{1}{\sqrt{N}}H_N}(x) \to \frac{1}{|I|}\int_I \rho_{sc}(x)\,dx$~~

for $\frac{1}{N} \ll a_N \ll 1$, $\displaystyle\frac{1}{a_N}\int_{[E-a_N, E+a_N]} d\mu_{\frac{1}{\sqrt{N}}H_N}(x) \xrightarrow[N\to+\infty]{} \rho_{sc}(E)$?

$\uparrow$

("Local" semicircle law)

In the bulk, ~~around $E\in(-2,2)$~~ what is the typical gap between $\frac{\lambda_i}{\sqrt{N}}$ & $\frac{\lambda_{i+1}}{\sqrt{N}}$ for $\varepsilon N \le i \le (1-\varepsilon)N$?

$$\int_{\frac{\lambda_i}{\sqrt{N}}}^{\frac{\lambda_{i+1}}{\sqrt{N}}} \rho_{sc}(x)\,dx = \frac{1}{N} \Rightarrow \frac{\lambda_{i+1}}{\sqrt{N}} - \frac{\lambda_i}{\sqrt{N}} \sim \frac{1}{N}$$

Q2: Does $\sqrt{N}(\lambda_{i+1}-\lambda_i)$ has a limiting distribution in the bulk? Does this distribution depends on the distribution of h_{ij}? (Does bulk universality holds?)

Near the edge: $\int_{-2}^{\frac{\lambda_1}{\sqrt{N}}} \rho_{sc}(x)\,dx = \frac{1}{N} \Rightarrow \int_{-2}^{\frac{\lambda_1}{\sqrt{N}}} \sqrt{2+x}\,dx \sim \frac{1}{N}$

$$\Rightarrow \left(\frac{\lambda_1}{\sqrt{N}}+2\right)^{3/2} \sim \frac{1}{N} \Rightarrow N^{1/6}(\lambda_1+2\sqrt{N}) \sim 1$$

Q3: Does $N^{1/6}(\lambda_1+2\sqrt{N})$ has a limiting distribution? Is this distribution universal?

Every eigenvector of H (GOE) is uniformly distributed on the unit sphere $S(N-1)$. ☆
(Think about why?)

Rmk: A uniformly distributed unit vector can be generated as $\frac{\vec{g}}{\|\vec{g}\|}$, where $\vec{g}=(g_1,\cdots,g_N)$ is a Gaussian vector with i.i.d. $N(0,1)$ entries.

Q4: What is the asymptotic behavior of the eigenvectors of a (non-invariant) Wigner matrix?

We expect that an eigenvector $\vec{u}_k$ is "asyptotically uniform" on $S(N-1)$. But defining this concept is already very non-trivial.

2. Sample covariance matrices

$X=(x_{ij})_{1\le i\le M,\ 1\le j\le N}$, the entries of X are independent random variables of mean 0, variance 1.

① To study the SVD of $X = UDV^*$, $\left.\begin{array}{l} U:\ M\times M \\ V:\ N\times N\end{array}\right\}$ orthogonal/unitary, $D=\begin{pmatrix} d_1 & & & \\ & \ddots & & 0 \\ & & d_M & \end{pmatrix}$,

rectangular diagonal

it reduces to studying the eigendecomposition of XX^* and X^*X.

② Let $\vec{x}=\begin{pmatrix}x_1\\ \vdots \\ x_M\end{pmatrix}\in\mathbb{R}^M$ be a random vector with independent entries of mean 0, variance 1

The covariance matrix of $\vec{x}$ is given by $\mathbb{E}(\vec{x}\vec{x}^*) = I_{M\times M}$.
Suppose we draw N i.i.d. copies of $\vec{x}$: $\vec{x}_1,\cdots,\vec{x}_N$. Then we form the sample covariance matrix

$$Q_N=\frac{1}{N}\sum_{i=1}^{N}\vec{x}_i\vec{x}_i^* = \frac{1}{N}XX^*,\quad X=(\vec{x}_1,\cdots,\vec{x}_N).$$

Rmk: By LLN, if M is fixed, letting $N\to\infty$ we have: Q_N converges a.s. to the true covariance I_M. This is called the "low-dimensional" setting.

~~$\boxed{Q1:}$ Does the ESD of Q_N also converge? What is its limit?~~

~~A: The limit~~

Rmk: The "high-dimension" setting considers $c_N = \frac{M}{N} \to c \in (0, +\infty)$, where M and N are of the same order. Then LLN fails, and the behavior of Q_N is very different from that in the low-d setting. This is related to the so-called "curse of dimensionality" in statistics.

When the entries of $\mathbb{X}$ are i.i.d. Gaussian, then Q_N is called the Wishart ensemble (1928):

$$Q_N \sim W_{\underset{\text{data dimension}}{\underbrace{M}}}(\underset{\text{covariance}}{\underbrace{I}}, \overset{\text{degrees of freedom}}{N})$$

Exercise: Check that $U \mathbb{X} V^*$ has the same distribution as $\mathbb{X}$ in the Wishart case for any unitary U & V.

A useful extension is: $\vec{x} \sim \mathcal{N}_M(0, \Sigma)$, p-variate normal with covariance Σ. Then $Q_N \sim W_M(\Sigma, N)$. We can write that $\mathbb{X} = \Sigma^{1/2} Y$, where the entries of Y are i.i.d. $N(0,1)$. Then $Q_N = \frac{1}{N} \Sigma^{1/2} Y Y^* \Sigma^{1/2}$.

As a further extension, the entries of Y are not necessarily Gaussian. We only require that they are independent, of mean 0 & variance 1.

$\boxed{Q1:}$ Does the ESD of Q_N also converge? What is the limit?
(We will see that the limit is called the Marchenko – Pastur law.)

$\boxed{Q2:}$ Bulk universality?

$\boxed{Q3:}$ Edge universality?

$\boxed{Q4:}$ Eigenvectors?

3. non-Hermitian random matrices

$$\mathbb{X} = (x_{ij})_{1 \le i,j \le N}, \quad \text{the entries of } \mathbb{X} \text{ are i.i.d., mean } 0, \text{ variance } 1.$$

We want to study the asymptotic behavior of the eigenvalues & eigenvectors of $\mathbb{X}$.

Note that: almost surely, $\mathbb{X}$ has N different eigenvalues. For general $\mathbb{X}$, $\mathbb{P}(\mathbb{X} \text{ is sigular}) \to$ as $N \to$

Unlike Hermitian matrices, the eigenvalues of a non-Hermitian can be complex.
People find that the ESD of $\mathbb{X}$ satisfies a circular law:

$$\mu_{\frac{1}{\sqrt{N}}\mathbb{X}_N}(dxdy) \to \frac{1}{\pi} \mathbb{1}_{\{z \in \mathbb{C} : |z| \le 1\}} \, dxdy \quad \text{weakly.}$$

The bulk universality & edge universality are still open. The study of eigenvectors is even harder.

4. Invariant ensembles

For GOE,
$$\mu_{H_N} = C_N \prod_{1\le i\le N} e^{-h_{ii}^2/4} \prod_{1\le i<j\le N} e^{-h_{ij}^2/2}\, dH_N$$
$$= C_N\, e^{-\sum_i h_{ii}^2/4 - \sum_{1\le i<j\le N} h_{ij}^2/2}\, dH_N$$
$$= C_N\, e^{-\mathrm{tr}(H_N^2)/4}\, dH_N .$$

For GUE, $\mu_{H_N} = C_N\, e^{-\mathrm{tr}(H_N^2)/2}\, dH_N$.

Under the conjugation by any unitary matrix $\mathcal{U}$, $H_N \to \mathcal{U} H_N \mathcal{U}^{-1}$, we have that $\mathrm{tr}(H_N^2)$ is invariant.

In general, we can define a density function on the set of random matrices as $\mathbb{P}(H_N)\, dH_N = \frac{1}{Z_N} \exp(-\mathrm{Tr}\, V(H_N))\, dH_N$, where $dH = \prod_{1\le i\le j\le N} dH_{ij}$ is the Lebesgue measure, V is a "potential function" that grows mildly at ∞ (to guarantee integrability), Z_N is the normalization factor (partition function).

Note: $\mathrm{Tr}\ V(\mathcal{U} H_N \mathcal{U}^{-1}) = \mathrm{Tr}[\mathcal{U}\, V(H_N)\, \mathcal{U}^{-1}] = \mathrm{Tr} V(H_N)$, i.e. orthogonal/unitary conjugation leaves the distribution $\mathbb{P}(H_N)\, dH_N$ invariant. So we call it "invariant ensemble".

Invariant ensembles are very different from Wigner ensembles: Gaussian ensembles are the only invariant Wigner ensembles.

As discussed before, the eigenvectors of invariant ensembles are uniformly distributed on unit sphere.

$\boxed{\text{Q1:}}$ What is the prob. density function for all the N eigenvalues only?

$\boxed{\text{Q2:}}$ Bulk universality? Edge universality?

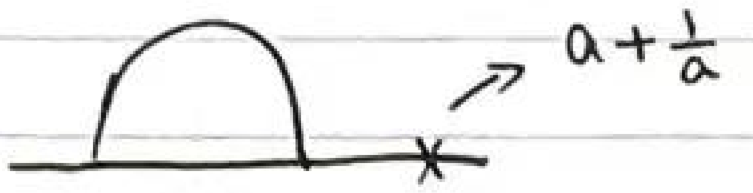
5. Deformed random matrices

Deformed Wigner

$H(a) := \frac{1}{\sqrt{N}} H_N + a u u^*$, H_N: Wigner matrix, $\overset{a\in\mathbb{R}}{\cancel{a>0}}$, u is an arbitrary unit vector.

WLOG, let $a>0$. A BBP transition as a crosses 1:

- If $a<1$, semicircle law still holds.
- If $a>1$, we have semicircle law + an outlier:



$\lambda_N - (a+\frac{1}{a})\sqrt{N}$ is asymptotically normal as $N \to +\infty$.

Spiked covariance

$$Q_N = \frac{1}{N}\Sigma^{1/2} Y Y^* \Sigma^{1/2}, \quad \Sigma = I + a u u^*, \ a>0, \ u: \text{unit vector}.$$

A Similar BBP transition occurs at $a = \sqrt{\frac{M}{N}}$.

Section 1 Why $\frac{1}{\sqrt{N}}$ is the correct saling for Wigner?

Let H_N be a Wigner matrices. We have

$$\mathbb{E}(\sum_i \lambda_i^2) = \mathbb{E}\,\mathrm{Tr}(H^2) = \mathbb{E}\sum_{i,j} H_{ij}H_{ji} = N^2$$

$\Rightarrow$ $\frac{1}{N}\mathbb{E}(\sum_i \lambda_i^2) = N$, i.e. the averaged size of λ_i^2 is of order N. So the eigenvalues of $\frac{1}{\sqrt{N}}H_N$ are of order 1.

Next, we aim to show the following bound on the operator norm of H_N:

~~there exists a~~ $\|H_N\| := \sup_{x\in\mathbb{C}^n : |x|=1} |H_N x|$, $|\cdot|$ means the L^2-norm.

Thm 1.1: Suppose the upper-triangular entries of H_N are independent, have mean zero, and uniformly bounded by 1 (i.e., $|h_{ij}| \le 1$ a.s.). Then, there exists absolute constants $c, C>0$ such that

$$\mathbb{P}(\|H_N\| > A\sqrt{N}) \le C\exp(-cAN) \quad \text{for } A \ge C.$$

(In words, $\|H_N\| = O(\sqrt{N})$ with very high probability.)

Lemma 1.2: Suppose M_N is $\overset{N\times N}{\text{a}}$ random matrix whose entries are independent, have mean zero, and uniformly bounded by 1. Then, there exist absolute constants $c, C>0$ such that

$$\mathbb{P}(\|M_N\| > A\sqrt{N}) \le C\exp(-cAN) \quad \text{for } A \ge C.$$

Pf of Thm 1.1: We write $H_N = U_N + L_N$, U_N consists of the upper-triangular entries, L_N consists of strict lower-triangular entries.

By Lemma 1.2, $\mathbb{P}(\|U_N\| > A\sqrt{N}) \le C\exp(-cAN)$, $\mathbb{P}(\|L_N\| > A\sqrt{N}) \le C\exp(-cAN)$ for $A \ge C$.

Then for $A \ge 2C$,

$$\mathbb{P}(\|H_N\| > A\sqrt{N}) \le \mathbb{P}(\|U_N\| > AN/2) + \mathbb{P}(\|L_N\| > AN/2) \le 2C\exp(-cAN/2). \quad \square$$

The proof of Lemma 1.2 uses some "standard" concentration inequalities & ε-net argument.

Thm 1.3 (Hoeffding's inequality) Let $X_1, \cdots, X_N$ be independent bounded random variables with $X_i \in [a_i, b_i]$ a.s. Let $S_N := X_1 + \cdots + X_N$. Then $\forall \lambda > 0$,

$$\mathbb{P}(|S_N| \geq \lambda\sigma) \leq C\exp(-c\lambda^2), \qquad \sigma^2 := \sum_{i=1}^{N} |b_i - a_i|^2.$$

Lem 1.4 (Hoeffding's lemma) For $Z \in [a,b]$, $\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} \leq \exp\left(\frac{\lambda^2(b-a)^2}{2}\right)$.

Rmk: The RHS can be improved to $\exp\left(\frac{\lambda^2(b-a)^2}{8}\right)$.

Pf of Lem 1.4: Let Z' be an independent copy of Z. Then

$$\mathbb{E}_Z \exp(\lambda(Z - \mathbb{E}Z)) = \mathbb{E}_Z \exp(\lambda(Z - \mathbb{E}_{Z'}(Z'))) \underset{\substack{\uparrow \\ \text{Jensen's ineq.}}}{\leq} \mathbb{E}_Z \mathbb{E}_{Z'} \exp(\lambda(Z - Z'))$$

Since $Z - Z'$ is symmetric about 0, for a random sign s, $\mathbb{P}(s=1) = \mathbb{P}(s=-1) = \frac{1}{2}$, $s(Z-Z') \overset{d}{=} Z - Z'$. So

$$\mathbb{E}_Z\mathbb{E}_{Z'} \exp(\lambda(Z-Z')) = \mathbb{E}_{Z,Z'}\mathbb{E}_s \exp(\lambda s(Z-Z')) = \mathbb{E}_{Z,Z'}\left[\tfrac{1}{2}e^{\lambda(Z-Z')} + \tfrac{1}{2}e^{-\lambda(Z-Z')}\right]$$

$$\underset{\uparrow}{\leq} \mathbb{E}\exp\left(\frac{\lambda^2}{2}(Z-Z')^2\right) \leq \exp\left(\frac{\lambda^2}{2}(b-a)^2\right). \qquad \square$$

$(\cosh(x) \leq \exp(\frac{x^2}{2}))$

Pf of Thm 1.3: $\forall t > 0$, $\mathbb{E}\exp(tS_N) = \prod_{i=1}^{N} \mathbb{E}\exp(tX_i) \leq \prod_{i=1}^{N} \exp\left(\frac{\overset{t^2}{\cancel{\lambda^2}}}{2}(b_i - a_i)^2\right) = \exp\left(\frac{\overset{t^2}{\cancel{\lambda^2}}}{2}\sigma^2\right)$.

So $\mathbb{P}(S_N > \lambda\sigma) \overset{\text{Markov}}{\leq} \exp(-t\lambda\sigma)\exp\left(\frac{\overset{t^2}{\cancel{\lambda^2}}}{2}\sigma^2\right) = \exp\left(\frac{\overset{t^2}{\cancel{\lambda^2}}}{2}\sigma^2 - t\lambda\sigma\right)$

~~Taking $t = \lambda$, $\mathbb{P}(S_N > \lambda\sigma) \leq \exp(-\frac{1}{2}\lambda^2\sigma^2)$.~~

$$= \exp\left(\frac{1}{2}(t\sigma - \lambda)^2 - \frac{\lambda^2}{2}\right).$$

Taking $t = \lambda/\sigma$ gives $\mathbb{P}(S_N > \lambda\sigma) \leq \exp(-\lambda^2/2)$. Can get a similar bound for $\mathbb{P}(S_N < -\lambda\sigma)$. $\square$

Lem 1.5 Under the setting of Lemma 1.2, for any fixed unit vector $x \in \mathbb{R}^N$,

$$\mathbb{P}(|M_N x| \geq A\sqrt{N}) \leq C\exp(-cAN) \quad \text{for } A \geq C.$$

Pf: Let $M_N = \begin{pmatrix} - X_1 - \\ - X_2 - \\ \vdots \\ - X_N - \end{pmatrix}$, X_i are the row vectors of M_N.

Then, $M_N x = \begin{pmatrix} X_1 \cdot x \\ X_2 \cdot x \\ \vdots \\ X_N \cdot x \end{pmatrix}$. For each $X_i \cdot x = \sum_{j=1}^{N} X_{ij} x_j$, applying Hoeffding, we get

$$\mathbb{P}(|X_i \cdot x| \geq \lambda\sigma) \leq C\exp(-c\lambda^2), \quad \text{where } \sigma^2 = \sum_{j=1}^{N} 4x_j^2 = 4.$$

~~$\Rightarrow \mathbb{P}(\sum_{i=1}^{N} |X_i \cdot x|^2 \geq$ [illegible]$)$~~ For any $c' < c$, $\mathbb{E}\exp(c'|X_i \cdot x|^2) \leq C'$ for a constant $C' > 0$

[Use the tail-probability formula,

$$\mathbb{E}f(X) = \int_0^{+\infty} \mathbb{P}(X \geq t) f'(t)\, dt,$$ X positive, f increasing on $[0, +\infty)$, & $f(0) = 0$.]

Thus, $\mathbb{E} \exp(c'|Mx|^2) = \prod_{i=1}^{N} \exp(c'|X_i \cdot x|^2) \leq (C')^N$.

$\Rightarrow \mathbb{P}(|Mx| \geq A\sqrt{N}) \leq \exp(-c'A^2N)(C')^N \leq C\exp(-cAN)$ for A large enough. $\square$

How to extend Lem 1.5 to a bound on

$$\mathbb{P}(\|M_N\| \geq A\sqrt{N})$$

$$= \mathbb{P}\Big(\bigcup_{x \in S^N} \{|M_N x| \geq A\sqrt{N}\}\Big).$$

Of course, we cannot take a union bound over a uncountable set. The idea is to "discretize" S^N.

<u>Def</u> (ε-net) ~~A maximal~~ An ε-net of the sphere S^N denotes a set of points in S^N that are separated from each other by a distance of at least ε, and which is <u>maximal</u> with respect to set inclusion.

~~By~~ let Σ be such an ~~maximal~~ ε-net. By maximality, for any point $x \in S^N$, there exists a point $y \in \Sigma$ such that $|x-y| < \varepsilon$.

<u>Lemma 1.6</u> (Volume packing) Let $0 < \varepsilon < 1$, and Σ be ~~a maximal~~ an ε-net. Then $|\Sigma| \leq (3/\varepsilon)^N$.

<u>Pf</u>: Consider the collection of balls of radius $\varepsilon/2$ centered around each point in Σ. Then these balls are disjoint. On the other hand, they are also contained in the ball of radius 3/2 centered at the origin. The volume of the larger ball is $(3/\varepsilon)^N$ times the volume of each small ball. $\square$

<u>Proof of Lemma 1.2</u>: Let Σ be a $\frac{1}{2}$-net of S^N. Then $|\Sigma| \leq 6^N$. Taking a union bound, we get ~~$\mathbb{P}(\max_{x\in\Sigma} |Mx|) \leq \sum_{x\in\Sigma}$~~

$$\mathbb{P}(\max_{x \in \Sigma} |Mx| > A\sqrt{N}) \leq \sum_{x \in \Sigma} \mathbb{P}(|Mx| > A\sqrt{N}) \leq C\exp(-cAN) \cdot 6^N \leq C\exp(-\tfrac{c}{2}AN) \quad (\dagger)$$

for large enough $A > 0$.

Next, we show that $\mathbb{P}(\|M\| > \lambda) \leq \mathbb{P}(\max_{x \in \Sigma} |Mx| > \lambda/2)$ $(*)$ for any $\lambda > 0$.

To show $(*)$, let $x \in S^N$ be such that

$$\|M\| = |Mx|.$$

Then we can find $y \in \Sigma$ so that $|x-y| < \frac{1}{2}$. Then $|M(x-y)| < \frac{1}{2}\|M\|$.

By triangle ineq., ~~$\|My\| \geq \|Mx\|$~~ $|My| \geq |Mx| - |M(x-y)| > \|M\| - \frac{1}{2}\|M\| = \frac{1}{2}\|M\|$.

Combining $(*)$ and $(\dagger)$ completes the proof. $\square$

Rmk: The above proofs can be extended to Wigner matrices with sub-gaussian entries.
A random variable X is said to be sub-gaussian if there exists an absolute constants $c>0$, so that

$$\mathbb{P}(|X|\geq t) \leq 2\exp(-ct^2) \quad \forall\ t\geq 0.$$

- * Gaussian r.v.s are sub-gaussian
- * If a random variable is bounded by a const, then it is subgaussian.

The sub-gaussian norm of X, $\|X\|_{\psi_2}$, is defined as

$$\|X\|_{\psi_2} := \inf\{t>0 : \mathbb{E}\exp(X^2/t^2) \leq 2\}.$$

Then we have the general Hoeffding's inequality

Thm 1.7 Let $X_1, \cdots, X_N$ be independent, mean-zero, sub-gaussian r.v.s. Then, $\forall\ t\geq 0$,

$$\mathbb{P}\left\{ \left|\sum_{i=1}^{N} X_i\right| \geq t\right\} \leq 2\exp\left(\frac{-ct^2}{\sum_{i=1}^{N}\|X_i\|_{\psi_2}^2}\right).$$

Rmk: In accordance with the semicircle law, we should have that $\forall$ constant $\varepsilon>0$,
$\mathbb{P}(\|H_N\| \geq (2+\varepsilon)\sqrt{N})$ with high probability.
One slick way to prove this result is the important "moment method in RMT":
for any $k\in 2\mathbb{N}$, note

$$\operatorname{tr}(H_N^k) = \sum_{i=1}^{n}\lambda_i^k \geq \max_i |\lambda_i|^k = \|H_N\|^k.$$

Hence,

$$\mathbb{E}\|H_N\|^k \leq \mathbb{E}\operatorname{tr}(H_N^k) \Rightarrow \mathbb{P}(\|H_N\|^k \geq (2+\varepsilon)\sqrt{N}) \leq [(2+\varepsilon)\sqrt{N}]^{-k}\,\mathbb{E}\operatorname{tr}(H_N^k).$$

The moment method aims to control $\mathbb{E}\operatorname{tr}(H_N^k)$. One can show that

$$\mathbb{E}\operatorname{tr}(H_N^k) = [2+o_k(1)]^k N^{\frac{k}{2}+1} \quad (*) \quad \text{for } k \text{ as large as } C\log N.$$

Then, we have

$$\mathbb{P}(\|H_N\|^k \geq (2+\varepsilon)\sqrt{N}) \leq (1-\tfrac{\varepsilon}{4})^k N \ll 1 \quad \text{for } k=C\log N \text{ if } C \text{ is large enough.}$$

For details, see Tao, Section 2.3.4.
We will give a proof using a different method. In fact, we will show a much stronger result: $\|H_N\| \leq 2+N^{-\frac{2}{3}+\varepsilon}$ w.h.p. for any const. $\varepsilon>0$.

But, we will use the moment method to prove the first important RMT result, i.e., the Wigner semicircle law. It requires us to calculate $\mathbb{E}\operatorname{tr}(H_N^k)$ for large but finite $k\in\mathbb{N}$.

~~Rmk:~~ Rmk: Moment method together with a truncation argument gives the operator norm bound for Wigner matrices whose entries have finite fourth moment.

Section 2 Wigner Semicircle Law

For the rest of this course, we rescale H_N to $\frac{1}{\sqrt{N}} H_N$, so that the eigenvalues of H_N are typically of order 1.

Thm 2.1 (Semicircle law) Let H be a real Wigner matrix whose entries have finite moments up to any order, i.e., $\forall k \in \mathbb{N}$, $\exists\ C_k > 0$ so that

$$\max_{i,j} \mathbb{E} |\sqrt{N} h_{ij}|^k \leq C_k .$$

Then, the ESD μ_{H_N} converges in distribution to μ_{sc} almost surely.

① Moment method

We will prove a weaker convergence in expectation of μ_{H_N} under a stronger sub-gaussian assumption on the entries of H_N. ↓

$$\forall\ \varphi \in C_c(\mathbb{R}), \quad \mathbb{E} \int \varphi(x)\, d\mu_{H_N}(x) \to \int \varphi(x) \cancel{\mathbb{E}}\, d\mu_{sc}(x) .$$

Define a sequence of measures $\mathbb{E}\mu_{H_N}(A) := \mathbb{E} \int \mathbb{1}(x \in A)\, d\mu_{H_N}(x)$.
The tightness of $\{\mathbb{E}\mu_{H_N}\}$ follows from the operator norm bounds.

To show the convergence, it suffices to show the convergence of moments, i.e., $\forall k \in \mathbb{N}$,

(*) $\mathbb{E} \int x^k d\mu_{H_N}(x) \to \int x^k d\mu_{sc}(x)$. [This follows from an application of Taylor.]

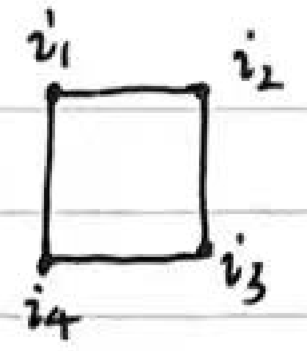
Rmk: To show convergence in prob. of μ_{H_N}, we need to show concentration of measure, i.e., μ_{H_N} concentrates around $\mathbb{E}\mu_{H_N}$. For that purpose, we need to show that $\mathrm{Var}[\int x^k d\mu_N(x)] \to 0$, $\forall k \in \mathbb{N}$. To show ~~convergence almost surely~~, almost sure convergence, we need to control $\mathbb{P}(|\int x^k d\mu_N(x) - \mathbb{E}\int x^k d\mu_N(x)| > \varepsilon)$ and use Borel-Cantelli.
While these can be done, we refrain from doing that and will prove the semicircle law using another " ~~possibly~~ more powerful" method — the Stieltjes transform method.

By definition, $\int x^k d\mu_{H_N}(x) = \int x^k \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i}(x)\, dx = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^k = \frac{1}{N}\mathrm{Tr}(H_N^k)$.

To show (*), we need to show that $\mathbb{E}\frac{1}{N}\mathrm{Tr}(H_N^k) \to \int x^k d\mu_{sc}(x)$ for any fixed k.

$\frac{1}{N}\mathbb{E}\,\mathrm{Tr}(H_N^k) = \frac{1}{N}\mathbb{E}\sum_{i_1,\cdots,i_k} h_{i_1 i_2} h_{i_2 i_3} \cdots h_{i_k i_1}$. For the expectation to be non-zero, every h_{xy} must be paired with anothe $h_{yx} = h_{xy}$.

Let's starts with $k=4$ case: $\frac{1}{N}\mathbb{E}$



We have the following cases to have non-zero expectation (up to cyclic permutation of vertices

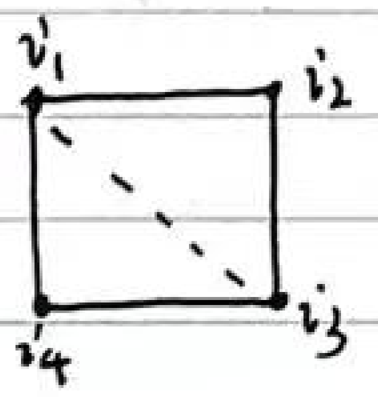
(i) $i_1 = i_3 \neq i_2 \neq i_4$ Typical order $\left(\frac{1}{N} \times N^3 \times \frac{1}{N^2} = 1\right)$

(ii) $i_1 = i_3 \neq i_2 = i_4$ $\left(\frac{1}{N} \times N^2 \times \frac{1}{N^2} = \frac{1}{N}\right)$

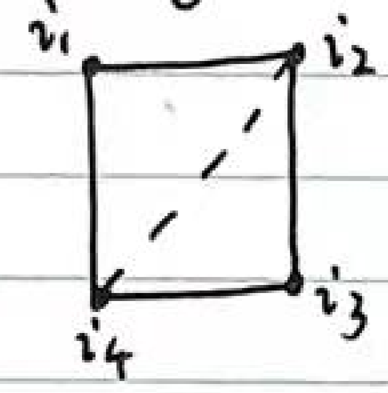
(iii) $i_1 = i_2 = i_3 \neq i_4$ $\left(\frac{1}{N} \times N^2 \times \frac{1}{N^2} = \frac{1}{N}\right)$

(iv) $i_1 = i_2 = i_3 = i_4$ $\left(\frac{1}{N} \times N \times \frac{1}{N^2} = \frac{1}{N^2}\right)$

Case (i) is dominating, and there are two such graphs: $i_1 = i_3 \neq i_2 \neq i_4$



$i_2 = i_4 \neq i_1 \neq i_3$



. With the fact that $\mathbb{E} h_{ij}^2 = \frac{1}{N}$, we get

$$\mathbb{E} \frac{1}{N} \mathrm{Tr}(H_N^4) = 2 \times \frac{1}{N} \times N^2 \times \frac{1}{N^2} + O(\tfrac{1}{N}) = 2 + O(\tfrac{1}{N}).$$

Let's turn to $k=6$. There are four types of graphs to deal with:

(i) There are 3 distinct edges, each occuring twice, and hence 4 distinct vertices. $\left(\frac{1}{N} \times N^4 \times \frac{1}{N^3} = 1\right)$

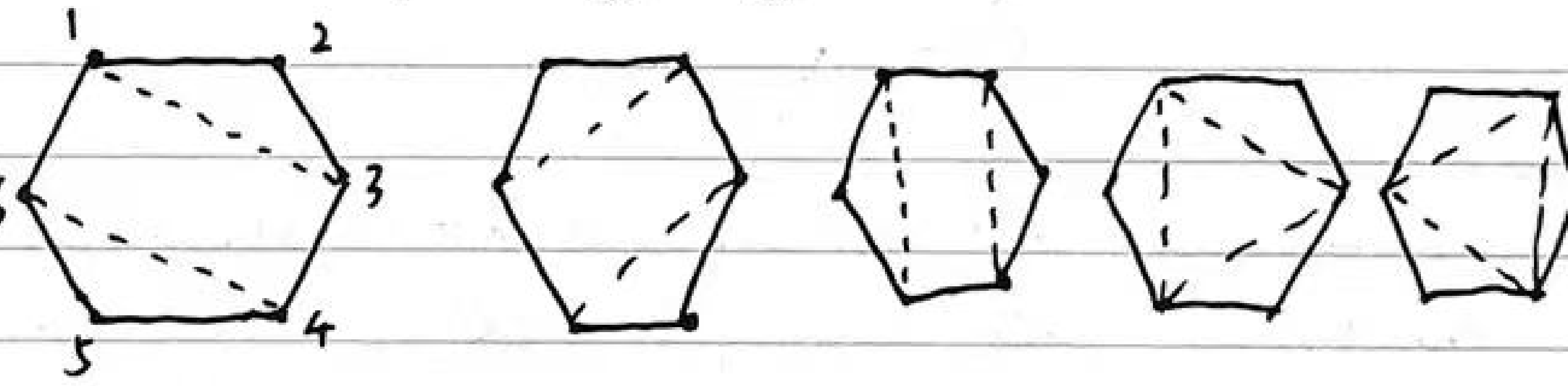
(ii) 2 distinct edges, one occuring twice & one occuring four times, and hence 3 distinct vertice $\left(\frac{1}{N} \times N^3 \times \frac{1}{N^3} = \frac{1}{N}\right)$

(iii) 2 distinct edges, each occuring three times, and hence 3 distinct vertices.

(iv) Only one distinct edge, occuring 6 times. $\left(\frac{1}{N} \times N \times \frac{1}{N^3} = \frac{1}{N^3}\right)$

How many graphs in case (i)?

There are "5" non-crossing planar graphs with 6 edges.



Hence, $\mathbb{E} \frac{1}{N} \mathrm{Tr}(H_N^6) = 5 \times \frac{1}{N} \times N^4 \times \frac{1}{N^3} + O(\frac{1}{N}) = 5 + O(\frac{1}{N})$.

In general, we consider $\frac{1}{N}\mathbb{E} \sum_{i_1, \cdots, i_k} h_{i_1 i_2} h_{i_2 i_3} \cdots h_{i_k i_1}$. The sequence $(i_1, i_2, \cdots, i_k, i_1)$ can be regarded as a cycle with at most k vertices and over all possible labellings of $i_j \in \{1, \cdots, n\}$. Since each ~~edge~~ distinct edge is traversed by two times, there are at most $k/2$ distinct edges and $(\frac{k}{2}+1)$ vertices traversed by the cycle.

For the cycles with at most $k/2$ distinct vertices, their order is $O(\frac{1}{N} \times N^{\frac{k}{2}} \times \frac{1}{N^{k/2}}) = O(\frac{1}{N})$.

So we only need to consider cycles which traverse exactly $(\frac{k}{2}+1)$ vertices and has $\frac{k}{2}$ distinct edges. We call such cycles non-crossing cycles of length k.

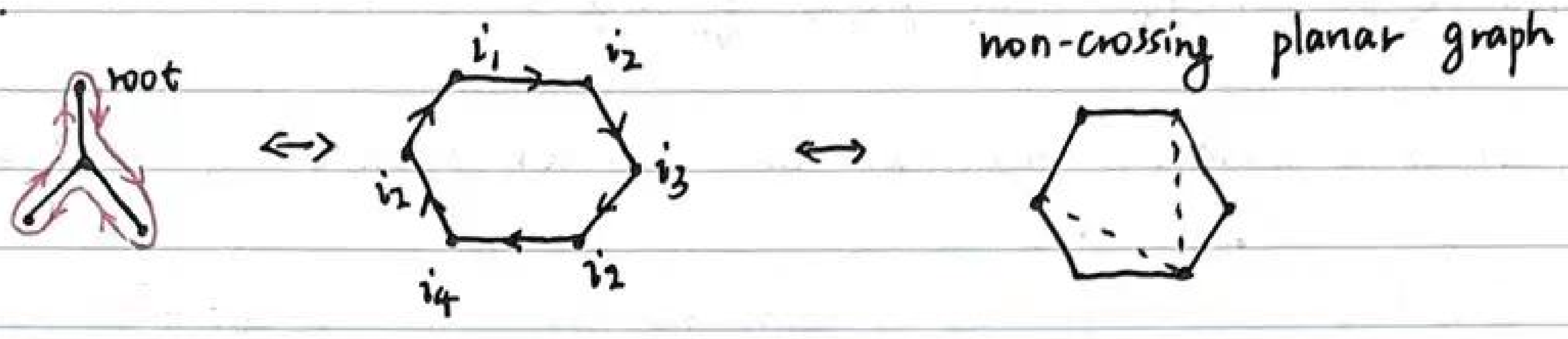
We need to count the number of non-crossing cycles.

Rmk: If k is odd, then ~~∃~~ each cycle has at most $\frac{k+1}{2}$ distinct vertices. Hence its order is $O(\frac{1}{N} \times N^{\frac{k+1}{2}} \times \frac{1}{N^{k/2}}) = O(\frac{1}{\sqrt{N}})$. Hence, $\mathbb{E}\,\frac{1}{N}\mathrm{Tr}(H_N^k) \to 0$ for any odd k.

Lemma 2.2 There is a one-to-one correspondance between non-crossing cycles of length k and rooted trees of $k/2$ edges and $(\frac{k}{2}+1)$ edges.

* The cycle lies in the corresponding tree and traverses each edge in the tree exactly twice.

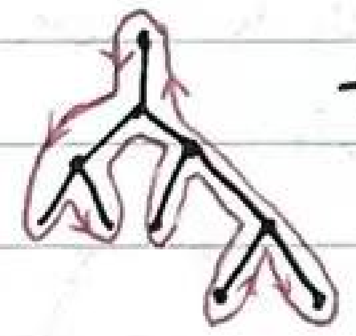
For example,



With lemma 2.2, we can get the following corollary:

Exercise: Let $i_1, \cdots, i_k$ be a cycle of length k. Arrange the integers $1, 2, \cdots, k$ around a circle. Whenever $1 \le a < b \le k$ s.t. $i_a = i_b$ with no c ~~bett~~ between a, b for which $i_a = i_c = i_b$, draw a dashed line between a & b. Then the cycle is non-crossing if and only if the number of dashed lines is exacty $\frac{k}{2}-1$ and the dashed lines do not cross each other.

Pf. of Lemma 2.2: Given an unlabelled rooted tree, starting from the root, traverse the tree from left to right gives a non-crossing cycle. For example:

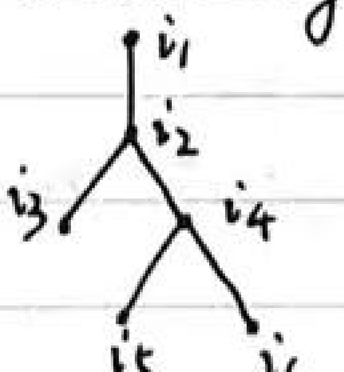


$\to (i_1, i_2, i_3, i_4, i_3, i_5, i_3, i_2, i_6, i_7, i_6, i_8, i_9, i_8, i_{10}, i_8, i_6, i_2, i_1)$

We now show that we can constuct a unique tree from a non-crossing cycle: $(i_1, i_2, \cdots, i_k)$. We traverse this cycle from i_1 to i_2, then from i_2 to i_3, and so on. At a step, say, from i_j to i_{j+1}, we either use an edge that we have not seen before, or else we are using an edge for the second time. We call a step ~~an edge~~ of former type an "innovative (I)" step ~~edge~~, and a step ~~an edge~~ of the latter type an "returning (R)" step ~~edge~~. Then there are $k/2$ (I) steps ~~edges~~, and $k/2$ (R) steps ~~edges~~. It is obvious that only the (I) ~~edge~~ step can bring us to new vertices we have not seen before. On the other hand, since we have to visit $(\frac{k}{2}+1)$ vertices starting from i_1, each (I) step must take us to a new vertex.

Then, traversing the cycle $(i_1, i_2, \cdots, i_k, i_1)$, we constuct a graph as follows. Let i_1 be the root. For each (I) step, we add a new vertex and a new edge. For example:

$(i_1, i_2, i_3, i_2, i_4, i_5, i_4, i_6, i_4, i_2, i_1) \to$

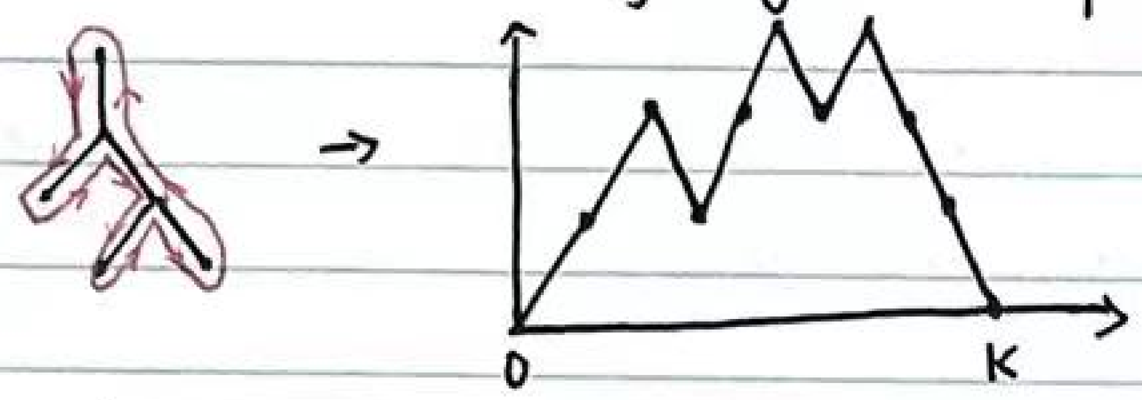


This clearly gives a rooted tree. □

Fact: The number of unlabelled rooted trees with $\frac{k}{2}+1$ vertices ~~is~~ is the ~~Catalan~~ ~~numb~~ Catalan number $C_{k/2}$.

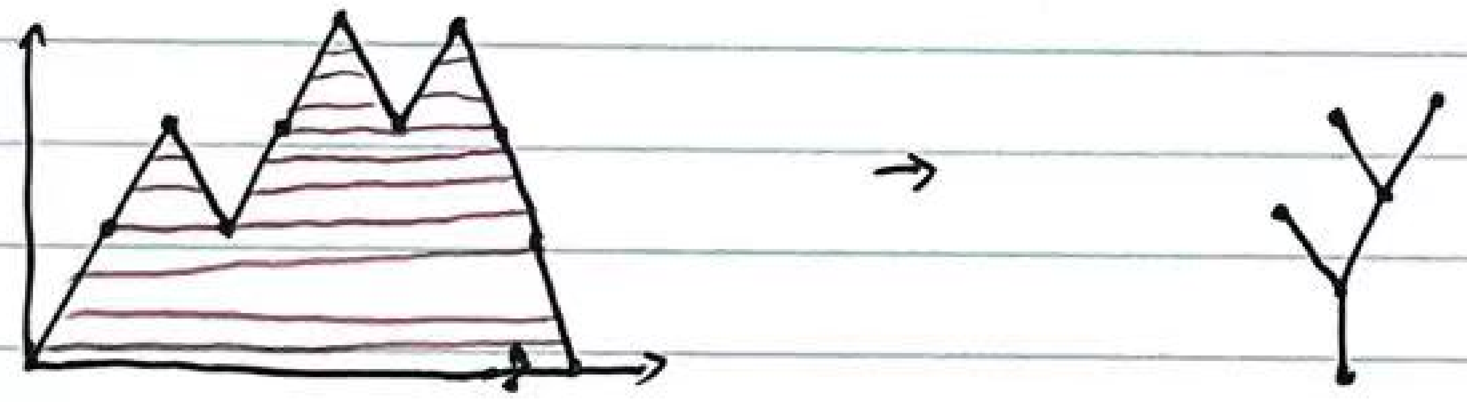
Catalan number $C_n := \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!\,(n+1)!}$.

One version of poof of the Fact: We further construct a 1-1 correspondance between the rooted trees and random walks on the positive half line.
~~Trans~~ Traversing a rooted tree, if we travese a (I) step, then walk to the right; otherwise, walk to the left by one step. For example:



It gives a RW from 0 to 0 and stays to the right of 0.

From ~~the~~ graph of a random walk, say f, on $[0,k]$. Define a quotient graph under the equivalence relation: $(a, f(a)) \sim (b, f(b))$ if $f(a) = f(b) = \min_{t\in[a,b]} f(t)$.



How many different k-step random walks from 0 to 0 and stays positive?

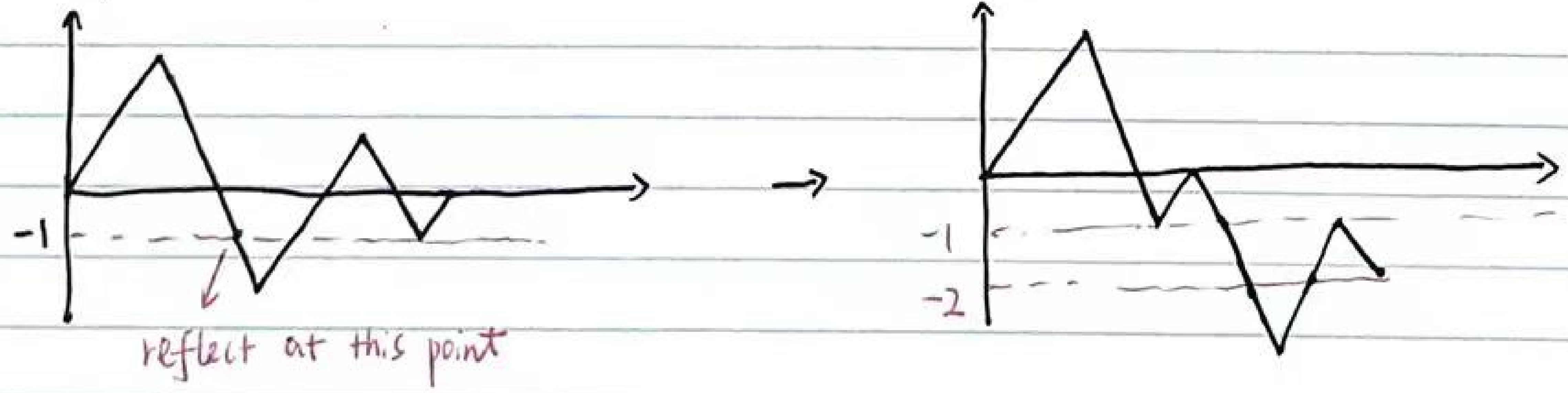
$\#\{$ ~~Random~~ k-step ~~walks~~ simple walks from 0 to 0$\}$ − $\#\{$k-step simple walks from 0 to 0 and hit $-1\}$

$\|$ ~~$\binom{2n}{n}$~~ $\binom{k}{k/2}$

↑ Use reflection principle

$= \#\{$k-step simple walks from 0 to $-2\}$ $= \binom{k}{k/2+1}$

For example:



$\binom{k}{k/2} - \binom{k}{k/2+1} = C_{k/2}$. □

It remains to show that: $\int x^k \, d\mu_{sc}(x) = C_{k/2} \cdot \mathbb{1}(k \text{ is even})$.
This is trivial for $k =$ odd. For k even,

$$I_k = \int_{-2}^{2} x^k \frac{\sqrt{4-x^2}}{2\pi} dx = \int_0^\pi (2\cos\theta)^k \frac{2\sin\theta}{2\pi} \cdot 2\sin\theta \, d\theta = \frac{2^{k+1}}{\pi}\int_0^\pi (\cos\theta)^k \sin^2\theta \, d\theta$$

$$= \frac{2^{k+1}}{\pi}\int_0^\pi (\cos\theta)^{k-1} \sin^2\theta \, d\sin\theta = -\frac{2^{k+1}}{\pi}\int_0^\pi \sin\theta \cdot \left[2\sin\theta\cdot(\cos\theta)^k - (k-1)\sin^3\theta \, (\cos\theta)^{k-2}\right] d\theta$$

$$= -2 I_k + \frac{2^{k+1}}{\pi}\int_0^\pi (k-1)\sin^2\theta \, (1-\cos^2\theta)(\cos\theta)^{k-2} \, d\theta$$

$$= -(k+1) I_k + 4(k-1)\frac{2^{k-1}}{\pi}\int_0^\pi \sin^2\theta \, (\cos\theta)^{k-2} d\theta = -(k+1) I_k + 4(k-1) I_{k-2}.$$

$$\Rightarrow I_k = \frac{4(k-1)}{k+2} I_{k-2}.$$

On the other hand, it is simple to check: $C_{k/2} = \frac{4(k-1)}{k+2} C_{(k-2)/2}$.

Sometimes, it is very challenging to recover a measure from its moments. We now use a different method to "derive" the semicircle law directly.

② The Stieltjes transform method

The Stieltjes transform of a measure μ on the real line $\mathbb{R}$ is defined as

$$S_\mu(z) := \int_{\mathbb{R}} \frac{1}{x-z} d\mu(x), \quad \text{for } z \text{ not in the support of } \mu.$$

For the ESD μ_{H_N}, we have $S_{\mu_{H_N}}(z) := \frac{1}{N}\int_{\mathbb{R}} \frac{1}{x-z} \sum_{i=1}^N \delta_{\lambda_i}(dx) = \frac{1}{N}\sum_{i=1}^N \frac{1}{\lambda_i - z} = \frac{1}{N}\mathrm{Tr}(H_N - z)^{-1}$.

Def: (Green's function / resolvent) For $z \in \mathbb{C}$, $G(z) = (H-z)^{-1}$ is called the Green's function (or resolvent) of H. Moreover, we denote its normalized trace by $m_N(z) \equiv S_N(z) = \frac{1}{N}\mathrm{Tr}\, G(z)$.

Rmk: The Stieltjes transform $m(z)$ can be regarded as a generating function of the moments:

$$m(z) = -\frac{1}{z} - \frac{1}{z^2}\int x \, d\mu_{H_N}(x) - \frac{1}{z^3}\int x^2 d\mu_{H_N}(x) - \cdots$$

This is a point of view that will be taken by free probability.

Prop 2.3 (Properties of $S_\mu(z)$) Let μ be a probability measure supported on the real line.

(i) $\overline{S_\mu(z)} = S_\mu(\bar{z})$.

(ii) $|S_\mu(z)| \le |\mathrm{Im}\, z|^{-1}$. $S_\mu(z)$ is complex analytic on the upper and lower half complex plane.

(iii) $\lim_{\eta \to +\infty} i\eta \, S_\mu(i\eta) = -1$. (by Dominated convergence thm)

Prop 2.4 Let H be a real symmetric / complex Hermitian matrix, we have

(i) $G(z)^* = G(\bar{z})$.

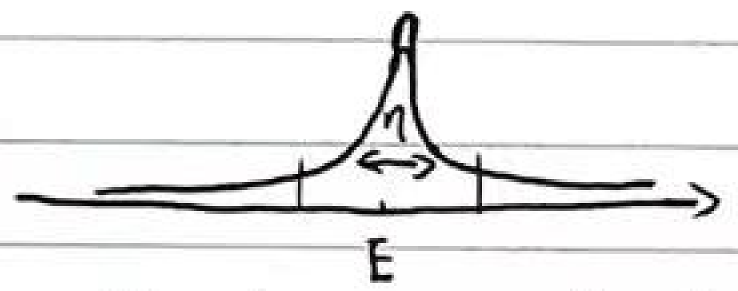
(ii) $\|G(z)\| \leq |\operatorname{Im} z|^{-1}$.

Really interesting things happen near the real axis, i.e., $|\operatorname{Im} z| \to 0$. For $z = E + i\eta \in \underset{\text{upper-half complex plane}}{\mathbb{C}_+}$, we can calculate the imaginary part of $S_\mu(z)$:

$$(*) \quad \operatorname{Im} S_\mu(z) = \int \frac{\eta}{(x-E)^2+\eta^2}\, d\mu(x) > 0$$

$$= \pi(\mu * P_\eta)(E), \qquad P_\eta(x) = \frac{1}{\pi}\frac{\eta}{x^2+\eta^2} = \frac{1}{\eta}\left[\frac{1}{\pi}\frac{1}{1+(x/\eta)^2}\right].$$

P_η :



Poisson kernel.

As $\eta \to 0$, P_η forms a family of <u>approximations to the identity (δ function)</u> in the following sence: $\int_{-\infty}^{+\infty} P_\eta(x) = 1$; $\lim_{\eta \to 0} \int_{|x| \geq \delta} P_\eta(x) = 0 \quad \forall\, \delta > 0$.

Then, we have $\mu * P_\eta$ converges weakly to μ. Together with $(*)$, it gives

$$\frac{1}{\pi} \operatorname{Im} S_\mu(\cdot + i\eta) \xrightarrow{\eta \to 0} \mu(\cdot). \quad ☆$$

<u>KEY</u>: A probability measure μ can be recovered from the limiting behavior of $\operatorname{Im} S_\mu$ down to the real axis.

<u>Lemma 2.5</u> Let μ_n be a sequence of random probability measures on $\mathbb{R}$, and let μ be a deterministic probability measure. Suppose $\{\mu_n\}$ is tight. Then μ_n converges almost surely (in probability) ~~if and only if S_{μ_n}~~ to μ in distribution if and only if $S_{\mu_n}(z)$ converges almost surely (in probability) to $S_\mu(z)$ for every $z \in \mathbb{C} \setminus (-\infty, 0)$.

Thus, to prove the semicircle law, we only need to show that $\forall\, x \in \mathbb{C}_+$, $S_{\mu_{H_n}}(z) \equiv m_N(z)$ converges almost surely (in probability) to $S_\mu(z) \equiv m_{sc}(z)$:

$$m_{sc}(z) = \int_{-2}^{2} \frac{1}{x-z} \sqrt{4-x^2}\, dx = \frac{-z + \sqrt{z^2-4}}{2},$$

where we take the branch that $\sqrt{z} \in \mathbb{C}_+$ for $z \in \mathbb{C}_+$. Note $\lim_{z \in \mathbb{C}_+, |z| \to +\infty} z S_\mu(z) = -1$, and

$$\frac{1}{\pi} \lim_{\eta \downarrow 0} m_{sc}(x+i\eta) = \frac{\sqrt{4-x^2}}{2\pi} \mathbb{1}_{x \in [-2,2]}, \quad \text{i.e. the semicircle density.}$$

<u>Lemma 2.6</u> (Schur complement) Let A, B, C, D be $n \times n$, $n \times m$, $m \times n$, $m \times m$ matrices. If D is invertible, the inverse of the block matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is given by

$$\begin{bmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$ if $A - BD^{-1}C$ is invertible.

A heuristic derivation of the semicircle law :

<u>Definition</u> (Resolvent minors) ① For any $1 \le i \le N$, let $H^{(i)}$ be the $(N-1)\times(N-1)$ matrix defined as : $(H^{(i)})_{ab} = h_{ab}$, $a, b \in \{1, \cdots, N\} \setminus \{i\}$. In other words, $H^{(i)}$ is the $(N-1)\times(N-1)$ minor of H with i-th row and column removed. (Note the row / column indices are kept in the new matrix. For example, the rows and columns of $H^{(1)}$ are labelled by $2, 3, \cdots, N$ instead of $1, 2, \cdots, N-1$.)

② Then, we define the resolvent minors $G^{(i)} = (H^{(i)} - z)^{-1}$.

To simplify notations, we will adopt the covention $G^{(i)}_{ab} = 0$ and $H^{(i)}_{ab} = 0$ if $a = i$ or $b = i$.

③ We can define $G^{(ij)}$, $G^{(ijk)}$ etc. in a similar way. The superscript in parenthesis $(\cdot)$ always means "removing the corresponding row and column of H".

<u>Def</u>: (Partial expectation) For $1 \le i \le N$, we define partial expectation $\mathbb{E}_i$ with respect to the i-th row and column of H as $\mathbb{E}_i(X) = \mathbb{E}(X \mid H^{(i)})$. We say a random variable X is independent of a set $S \subseteq \{1, \cdots, N\}$ if $\mathbb{E}_i X = X$ for all $i \in S$.

<u>A key remark</u>: $G^{(i)}$ is independent of i (i.e., the i-th row / column) of H.

We aim to show that $m_N(z) = \frac{1}{N}\mathrm{Tr}\, G(z) \approx m_{sc}(z)$ for N large enough. For any $1 \le i \le N$, using the Schur complement formula, we get

$$G_{ii} = \frac{1}{h_{ii} - z - \sum_{k,l \neq i} h_{ik} h_{il} G^{(i)}_{kl}} .$$

~~Fist~~ First, we observe that h_{ii} is of typical order $N^{-1/2}$. Second, we observe that $G^{(i)}$ is <u>independent</u> of h_{ik} entries. The partial expectation of $\sum_{k,l} h_{ik} h_{il} G^{(i)}_{kl}$ is given by

$$\mathbb{E}_i \sum_{k,l} h_{ik} h_{il} G^{(i)}_{kl} = \frac{1}{N} \sum_{k} G^{(i)}_{kk} =: m_N^{(i)}(z).$$

We expect that there is a concentration phenomenon where $\sum_{k,l} h_{ik} h_{il} G^{(i)}_{kl}$ concentrates around its partial expectation $m_N^{(i)}(z)$. $\Leftarrow$ (Will justify it later.)

Furthermore, heuristically, $m_N^{(i)}(z)$ is very close to $m_N(z)$ for large N. So, we ought to have that

$$G_{ii} \approx \frac{1}{-z - m_N(z)} .$$

Taking average over i : $m_N = \frac{1}{N}\sum_i G_{ii} \approx \frac{1}{-z - m_N(z)} \Rightarrow m_N^2 + z m_N + 1 \approx 0$.

Solving the equation and requiring that $\operatorname{Im} m_N \ge 0$, we get that

$$m_N(z) \approx \frac{-z + \sqrt{z^2 - 4}}{2} = m_{sc}(z) .$$

This is the self-consistent equation for m_N.

Rigorous proof of the semicircle law

To change the above argument to a rigorous proof, we need to show:

① Concentration of quadratic forms of independent random variables.

② ~~Closem~~ Closeness between m_N and $m_N^{(i)}$.

③ Stability of the self-consistent equation (m_N satisfies the equation for m_{sc} approximately $\Rightarrow$ m_N is close to m_{sc}).

① Concentration inequalities

Def: (Stochastic domination) Let X_N and Y_N be two sequences of random variables depending on N. We say X_N is stochastically dominated by Y_N, denoted as "$X_N \prec Y_N$", if for any small $\varepsilon > 0$ and large $D > 0$, we have

$$\mathbb{P}(|X_N| > N^{\varepsilon}|Y_N|) \leq N^{-D}$$

for large enough $N \geq N_0(\varepsilon, D)$. We will also use $X_N = O_{\prec}(Y_N)$ to mean $X_N \prec Y_N$.

Def: (High-probability event) An event Ξ is said to holds with high probability (w.h.p.) if $\mathbb{P}(\Xi) \geq 1 - N^{-D}$ for large enough N. (Hence, $X_N \prec Y_N$ iff for any small constant $\varepsilon > 0$, $|X_N| \leq N^{\varepsilon}|Y_N|$ w.h.p.)

"$\prec$" can be treated as "$\leq$" in some sense.

Prop: "$\prec$" satisfies the following properties.

(i) If $X \prec Y$ and $Y \prec Z$, then $X \prec Z$.

(ii) If $X_1 \prec Y_1$ and $X_2 \prec Y_2$, then $|X_1| + |X_2| \prec |Y_1| + |Y_2|$, and $X_1 X_2 \prec Y_1 Y_2$.

(iii) If $X \prec Y + N^{-\varepsilon} X$ for some constant $\varepsilon > 0$, then $X \prec Y$.

(iv) If $X \prec Y$, $\mathbb{E}|Y| \geq N^{-C}$ and $|X| \leq N^{C}$ almost surely for some constant $C > 0$. Then we have $\mathbb{E}X \prec \mathbb{E}|Y|$. (By definition, two deterministic numbers a, b ~~satisfying~~ satisfy $a \prec b$ means that $\forall$ constant $\varepsilon > 0$, $|a| \leq N^{\varepsilon}|b|$.)

Proof of (iv): For any $\varepsilon > 0$, let $\Omega := \{|X| \leq N^{\varepsilon}|Y|\}$. Then,

$$|\mathbb{E}X| \leq \mathbb{E}|X| = \mathbb{E}(|X| 1_{\Omega}) + \mathbb{E}(|X| 1_{\Omega^c})$$

$$\leq N^{\varepsilon}\mathbb{E}|Y| + N^{C}\mathbb{P}(\Omega^c) \leq N^{\varepsilon}\mathbb{E}|Y| + N^{C-D}.$$

Taking D large enough so that $N^{C-D} \leq N^{-C} \leq \mathbb{E}|Y|$, we get $|\mathbb{E}X| \leq (N^{\varepsilon}+1)\mathbb{E}|Y|$. □

However, all these properties can be applied for at most $N^{O(1)}$ many times.

Let $X_N(u)$, $Y_N(u)$ be two families of random variables, where $u \in \mathcal{U}^{(N)}$ is a parameter and $\mathcal{U}^{(N)}$ is certain parameter set. (For example, $\mathcal{U}^{(N)}$ can be the set of matrix indices, e.g., G_{ij}, $(i,j) \in \{1, \cdots, N\}^2$.)

We say $X_N \prec Y_N$ uniformly in $u \in \mathcal{U}^{(N)}$ if $\forall\, \varepsilon, D > 0$,

$$\sup_{u \in \mathcal{U}^{(N)}} \mathbb{P}\left[\, |X_N(u)| > N^{\varepsilon} |Y_N(u)| \,\right] \leq N^{-D} \quad \text{for large enough } N \geq N_0(\varepsilon, D).$$

<u>Prop</u>: Suppose $X_N \prec Y_N$ uniformly in $u \in \mathcal{U}^{(N)}$, and $|\mathcal{U}^{(N)}| \leq N^C$ for a constant $C > 0$. Then

$$\sum_{u \in |\mathcal{U}^{(N)}|} X_N(u) \prec \sum_{u \in |\mathcal{U}^{(N)}|} |Y_N(u)| .$$

A key tool that will be used frequently in this course is the following large deviation estimates (concentration inequalities) for linear and quadratic forms of independent random variables.

<u>Thm 2.7</u> (Large deviation bounds) Let $\{X_i : 1 \leq i \leq N\}$, $\{Y_i : 1 \leq i \leq N\}$ be independent families of random variables, and let $\{a_{ij} : 1 \leq i, j \leq N\}$, $\{b_i : 1 \leq i \leq N\}$ be deterministic. Suppose all entries X_i and Y_i are independent and satisfy

$$\mathbb{E} X = 0, \quad \mathbb{E}|X|^2 \leq 1, \quad (\mathbb{E}|X|^k)^{1/k} \leq C_k \quad \forall\, k \in \mathbb{N},$$

where C_k is some constant depending on k. Then, we have

$$(1) \quad \sum_{i=1}^{N} b_i X_i \prec \Big(\sum_{i=1}^{N} |b_i|^2 \Big)^{1/2},$$

$$(2) \quad \sum_{i,j} a_{ij} X_i Y_j \prec \Big(\sum_{i,j} |a_{ij}|^2 \Big)^{1/2},$$

$$(3) \quad \sum_{i \neq j} a_{ij} X_i X_j \prec \Big(\sum_{i \neq j} |a_{ij}|^2 \Big)^{1/2} . \qquad (*)$$

<u>Rmk</u>: Note that the second moment calculation gives the correct order of magnitude. For example,

$$\mathbb{E}\Big| \sum_{i \neq j} a_{ij} X_i X_j \Big|^2 = \mathbb{E} \sum_{i \neq j,\, i' \neq j'} a_{ij}\, a_{i'j'}\, X_i X_j \overline{X_{i'}}\, \overline{X_{j'}}$$

$$= \mathbb{E} \sum_{i \neq j} \Big[a_{ij} \overline{a_{ij}} |X_i|^2 |X_j|^2 + \overset{a_{ij}\overline{a_{ji}}}{\cancel{(a_{ij})^2}} |X_i|^2 |X_j|^2 \Big]$$

$$= \sum_{i \neq j} \big[|a_{ij}|^2 + a_{ij} \overline{a_{ji}} \big] \leq 2 \sum_{i \neq j} |a_{ij}|^2 .$$

In other words, the large deviation estimates are optimal up to an N^{ε} factor.

<u>Pf</u>: The most basic idea to prove stochastic domination is to bound the high moments and then use Markov's inequality. Denote $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$.

(1) follows directly from the classical Marcinkiewicz – Zygmund inequality:

$$\mathbb{E}\Big[\Big| \sum_i b_i X_i \Big|^p \Big] \leq C_p \Big\| \Big(\sum_i |b_i|^2 |X_i|^2 \Big)^{1/2} \Big\|_p^p = C_p\, \mathbb{E}\Big[\Big(\sum_i |b_i|^2 |X_i|^2 \Big)^{p/2} \Big].$$

Denoting $B^2 = \sum_i |b_i|^2$, we then have that

$$\mathbb{E}\Big[\big|\sum_i b_i X_i\big|^p\Big] \le C_p B^p\, \mathbb{E}\Big[\Big(\sum_i \frac{|b_i|^2}{B^2}|X_i|^2\Big)^{p/2}\Big] \underset{\uparrow}{\overset{\text{Jensen}}{\le}} C_p B^p\, \mathbb{E}\Big[\sum_i \frac{|b_i|^2}{B^2}|X_i|^p\Big]$$

$$\le C_p' B^p \sum_i \frac{|b_i|^2}{B^2} = C_p' B^p .$$

$\forall \varepsilon > 0$, we have

$$\mathbb{P}\Big[\big|\sum_i b_i X_i\big| \ge N^{\varepsilon} B\Big] \le \frac{\mathbb{E}[|\sum_i b_i X_i|^p]}{(N^{\varepsilon}B)^p} \le C_p' N^{-p\varepsilon}. \quad \text{Take } p > \frac{D}{\varepsilon}.$$

For (2), we write $\sum_{i,j} a_{ij} X_i Y_j = \sum_j b_j Y_j$, $b_j := \sum_i a_{ij} X_i$.

Conditioning on $\{X_i\}$, we use M-Z inequality again to get that

$$\mathbb{E}\Big[\big|\sum_{i,j} a_{ij} X_i Y_j\big|^p\Big] \le C_p\, \mathbb{E}\Big[\Big(\sum_j |b_j|^2 |Y_j|^2\Big)^{p/2}\Big]$$

$$\le C_p'\, \mathbb{E}\Big[\Big(\sum_j |b_j|^2\Big)^{p/2}\Big] = C_p' \Big\| \sum_j |b_j|^2 \Big\|_{p/2}^{p/2}$$

Use the $\overset{\text{(triangle)}}{\text{Minkowski}}$ ineq., $\rightarrow$ $\le C_p' \Big(\sum_j \big[\mathbb{E}(|b_j|^p)\big]^{\frac{2}{p}}\Big)^{\frac{p}{2}}$.

Use the M-Z ineq. again, $\mathbb{E}|b_j|^p = \mathbb{E}\big[|\sum_i a_{ij} X_i|^p\big] \le C_p \big[\sum_i |a_{ij}|^2\big]^{p/2}$.

Hence, $\mathbb{E}\Big[\big|\sum_{i,j} a_{ij} X_i Y_j\big|^p\Big] \le C_p \Big(\sum_j \sum_i |a_{ij}|^2\Big)^{p/2}$. Then apply Markov's ineq.

Finally, for (3), we ~~[illegible]~~ reduce it to ~~the~~ a cases where (2) ~~can~~ can be applied. For $\sum_{i\ne j} a_{ij} X_i X_j$, we use the following identity to decompose $\{X_i\}$ into two sets of independent random variables: $\forall i \ne j$,

$$1 = \frac{1}{2^{N-2}} \sum_{I \sqcup J} \mathbb{1}(i \in I)\,\mathbb{1}(j \in J),$$ where $I \sqcup J$ means I and J is a partition of $\{1, \cdots, N\}$.

Also, notice that the $\overset{\text{total}}{\vee}$ number of partitions is $2^N - 2$.

Then,

$$\Big\| \sum_{i \ne j} a_{ij} X_i X_j \Big\|_p \underset{\underset{\text{Minkowski}}{\uparrow}}{\le} \frac{1}{2^{N-2}} \sum_{I \sqcup J} \Big\| \sum_{i \in I} \sum_{j \in J} a_{ij} X_i X_j \Big\|_p$$

$$\le \frac{1}{2^{N-2}} \sum_{I \sqcup J} C_p \Big(\sum_{i,j} |a_{ij}|^2\Big)^{\frac{1}{2}} \le 4 C_p \Big(\sum_{i,j} |a_{ij}|^2\Big)^{1/2}.$$

In other words, $\mathbb{E}\Big[\big|\sum_{i\ne j} a_{ij} X_i X_j\big|^p\Big] \le C_p \Big(\sum_{i,j} |a_{ij}|^2\Big)^{p/2}$. Then apply Markov's ineq. □

② Eigenvalue interlacing (Bounding $|m_N - m_N^{(i)}|$)

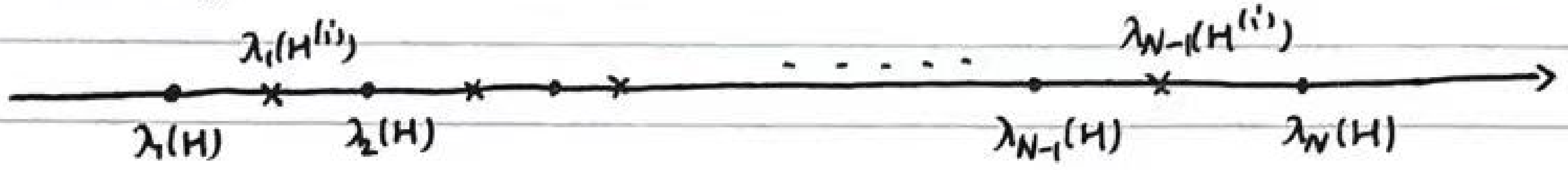
We want to bound

$$m_N - m_N^{(i)} = \frac{1}{N}\big[\mathrm{Tr}\, G(z) - \mathrm{Tr}\, G^{(i)}(z)\big] = \frac{1}{N}\Big(\mathrm{Tr}\frac{1}{H - z} - \mathrm{Tr}\frac{1}{H^{(i)} - z}\Big).$$

~~$= \frac{1}{N}\Big[\sum \frac{1}{\lambda_i - z} - \sum \frac{1}{\lambda_i}$~~

$$m_N - m_N^{(i)} = \frac{1}{N}\Big[\sum_{j=1}^{N} \frac{1}{\lambda_j(H)-z} - \sum_{j=1}^{N-1} \frac{1}{\lambda_j(H^{(i)})-z}\Big]$$

<u>Cauchy interlacing</u>: $\lambda_1(H) \le \lambda_1(H^{(i)}) \le \lambda_2(H) \le \cdots \le \lambda_{N-1}(H^{(i)}) \le \lambda_N(H)$.



Using this fact, we can show that $|\mathrm{Re}\, m_N - \mathrm{Re}\, m_N^{(i)}| \le \frac{C}{N\eta}$, $\eta = \mathrm{Im}\, z$,

$|\mathrm{Im}\, m_N - \mathrm{Im}\, m_N^{(i)}| \le \frac{C}{N\eta}$. (They are alternating sums.)

Hence, $|m_N - m_N^{(i)}| \le \frac{C}{N\eta}$. We now give a different proof.

<u>Prop</u>: (Useful resolvent identities) (i) If $i \neq j$, then

$$G_{ij} = -G_{ii} \sum_{k}^{(i)} h_{ik} G_{kj}^{(i)} = -G_{jj} \sum_{k}^{(j)} G_{ik}^{(j)} h_{kj} . \quad \Big(\sum_{k}^{(i)} := \sum_{k:k\neq i} \Big)$$

(ii) If $k \notin \{i,j\}$, then $G_{ij} = G_{ij}^{(k)} + \frac{G_{ik}G_{kj}}{G_{kk}}$.

(iii) [Ward identity] We have $\sum_j |G_{ij}|^2 = \sum_j |G_{ji}|^2 = \frac{\mathrm{Im}\, G_{ii}}{\eta}$. $(\eta = \mathrm{Im}\, z)$

<u>Proof</u>: (i) follows from Schur complement formula.

For (ii), we use the identity $\frac{1}{A+B} - \frac{1}{A} = -\frac{1}{A+B} B \frac{1}{A} = -\frac{1}{A} B \frac{1}{A+B}$.

Let $\widetilde{H}^{(k)}$ be the $N\times N$ matrix obtained by setting the k-th row and column of H to 0. (In other words, $\widetilde{H}^{(k)}$ is obtained by adding a zero row and column to $H^{(k)}$.)

Let $A = \widetilde{H}^{(k)} - z$, $B = H - \widetilde{H}^{(k)}$:

$$G(z) - \frac{1}{\widetilde{H}^{(k)} - z} = -G(z)\,(H - \widetilde{H}^{(k)})\,\frac{1}{\widetilde{H}^{(k)} - z} .$$

Note $(\widetilde{H}^{(k)} - z)^{-1}_{ij} = G_{ij}^{(k)}$ if $k \notin \{i,j\}$. Then, the above equation gives

$$G_{ij} - G_{ij}^{(k)} = -G_{ik} \sum_{\ell} h_{k\ell} G_{\ell j}^{(k)} \underset{\substack{\uparrow \\ \text{by (i)}}}{=} G_{ik}\Big(\frac{1}{G_{kk}} G_{kj}\Big) .$$

This gives (ii).

For (iii), $G(z) = \sum_k \frac{u_k u_k^*}{\lambda_k - z}$. Then, $\sum_j |G_{ij}|^2 = \sum_j G_{ij}\overline{G_{ij}} = \sum_j G_{ij} G_{ji}^*$

$$= (GG^*)_{ii} = \Big(\sum_k \frac{u_k u_k^*}{|\lambda_k - z|^2} \Big)_{ii} = \sum_k \frac{|u_k(i)|^2}{|\lambda_k - z|^2} = \frac{1}{\eta} \sum_k \frac{\eta}{|\lambda_k - z|^2} |u_k(i)|^2$$

$$= \frac{1}{\eta} \mathrm{Im}\, G_{ii}(z) . \qquad \square$$

Then, we have $|m_N - m_N^{(i)}| = |\frac{1}{N} \operatorname{Tr} G - \frac{1}{N} \operatorname{Tr} G^{(i)}| \leq \frac{1}{N} \sum_j |G_{jj} - G_{jj}^{(i)}|$ (with the convention $G_{kl}^{(i)} = 0$ if $k=i$ or $l=i$)

$$= \frac{1}{N} \sum_j \left| \frac{G_{ji} G_{ij}}{G_{ii}} \right| \leq \frac{1}{N|G_{ii}|} \left(\sum_j |G_{ji}|^2 \right)^{1/2} \left(\sum_j |G_{ij}|^2 \right)^{1/2}$$

$$= \frac{1}{N|G_{ii}|} \cdot \frac{\operatorname{Im} G_{ii}}{\eta} \leq \frac{1}{N\eta}.$$

③ Stability of the self-consistent equation

<u>Prop</u>: Fix any $z = E + i\eta \in \mathbb{C}_+$. Suppose m (deter. / random) satisfies that $|m^2 + zm + 1| \leq \delta$ for some $\delta \lesssim 1$, and $\operatorname{Im} m > 0$. Then,

$$|m - m_{sc}| \leq \frac{C\delta}{\eta}. \quad \left(\text{Recall } m_{sc}(z) = \frac{-z + \sqrt{z^2 - 4}}{2}. \right)$$

<u>Proof</u>: ~~Suppose~~ Let $m^2 + zm + 1 = \Delta$, where $|\Delta| \leq \delta$. Then, m satisfies

$$m = \frac{-z + \sqrt{z^2 - 4(1-\Delta)}}{2}. \quad (*)$$

Then, the conclusion follows from a direct calculation. □

<u>Rmk</u>: We have $(*)$ because η is a fixed value. When $\eta \to 0$, choosing the branch of $\sqrt{\cdot}$ is quite non-trivial.

<u>Proof of the Semicircle law:</u> Recall that we only need to show that $m_N(z) = \frac{1}{N} \operatorname{Tr} G(z) \to m_{sc}(z)$ almost surely for any fixed $z \in \mathbb{C}_+$.

By Schur complement,

$$\frac{1}{G_{ii}} = h_{ii} - z - \sum_{k,l}^{(i)} h_{ik} h_{il} G_{kl}^{(i)}.$$

Since $\mathbb{E}|\sqrt{N} h_{ii}|^p \leq C_p$, we have $|h_{ii}| \prec N^{-\frac{1}{2}}$. Using the large deviation estimate, we have

$$\sum_{k \neq l}^{(i)} h_{ik} h_{il} G_{kl}^{(i)} \prec \frac{1}{N} \left(\sum_{k \neq l}^{(i)} |G_{kl}^{(i)}|^2 \right)^{1/2},$$

$$\sum_k^{(i)} (h_{ik}^2 - \tfrac{1}{N}) G_{kk}^{(i)} \prec \frac{1}{N} \left(\sum_k^{(i)} |G_{kk}^{(i)}|^2 \right)^{1/2}.$$

(Note $h_{ik}^2 - \frac{1}{N}$ are mean 0, independent random variables with variance $O(\frac{1}{N^2})$.)

Note: It is key that $G^{(i)}$ is independent of h_{ik}, h_{il}.

Hence, we get

$$\left| \sum_{k,l}^{(i)} h_{ik} h_{il} G_{kl}^{(i)} - \frac{1}{N} \operatorname{Tr} G^{(i)} \right| \leq \left| \sum_{k \neq l}^{(i)} h_{ik} h_{il} G_{kl}^{(i)} \right| + \left| \sum_k^{(i)} (h_{ik}^2 - \tfrac{1}{N}) G_{kk}^{(i)} \right|$$

$$\prec \frac{1}{N} \left(\sum_{k \neq l}^{(i)} |G_{kl}^{(i)}|^2 \right)^{1/2} + \frac{1}{N} \left(\sum_k^{(i)} |G_{kk}^{(i)}|^2 \right)^{1/2}$$

$$\leq \frac{2}{N} \left(\sum_{k,l}^{(i)} |G_{kl}^{(i)}|^2 \right)^{1/2} = \frac{2}{N} \left(\sum_k^{(i)} \frac{\operatorname{Im} G_{kk}^{(i)}}{\eta} \right)^{1/2} \leq \frac{1}{\sqrt{N}\eta}.$$

In sum, we get $\frac{1}{G_{ii}} = -z - \overset{m_N^{(i)}}{\overset{\parallel}{\frac{1}{N}\operatorname{Tr} G^{(i)}}} + O_\prec\left(\frac{1}{\sqrt{N}\eta}\right)$

$$= -z - m_N(z) + O_\prec\left(\frac{1}{\sqrt{N}\eta}\right)$$

$$\Rightarrow \quad G_{ii} = -\frac{1}{z + m_N + O_\prec\left(\frac{1}{\sqrt{N}\eta}\right)} = -\frac{1}{z+m_N} + O_\prec\left(\frac{1}{\sqrt{N}\eta^3}\right)$$

$$\Rightarrow \quad m_N(z) = \frac{1}{N}\sum_i G_{ii}(z) = -\frac{1}{z+m_N(z)} + O_\prec\left(\frac{1}{\sqrt{N}\eta^3}\right)$$

$$\Rightarrow \quad m_N^2(z) + z m_N(z) + 1 = O_\prec\left(\frac{1}{\sqrt{N}\eta^4}\right). \quad \Rightarrow |m_N(z) - m_{sc}(z)| \prec \frac{1}{\sqrt{N}\eta^5}. \qquad \square$$

Derivation of the Marchenko-Pastur law for sample covariance matrices:

$Q_N = \frac{1}{N} X X^*$, $\quad X$: $M \times N$ matrix with i.i.d. entries of mean 0, variance 1.

We use the scaling $X \to \frac{1}{\sqrt{N}} X$. Then, we write $Q_N = X X^*$, the entries of X are i.i.d., with mean 0, variance $1/N$.

We would like to derive the <u>limiting ESD</u> of Q_N.

We will use a linearization trick. Define $H = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}$.

Let $X = \sum_{k=1}^{M\wedge N} \sqrt{\lambda_k}\, u_k v_k^*$ be a SVD of X. Then, the non-zero eigenvalues of H are given by $\pm\sqrt{\lambda_k}$, $k = 1, \cdots, M \wedge N$, corresponding to the eigenvectors $\begin{pmatrix} u_k \\ \pm v_k \end{pmatrix}$.

The matrix H has an advantage that it is linear in X.

For $z \in \mathbb{C}_+$, choose $\sqrt{z}$ such that $\sqrt{z} \in \mathbb{C}_+$. Then, we define the resolvent

$$G(z) := \frac{1}{\sqrt{z}}(H - \sqrt{z})^{-1} = \frac{1}{\sqrt{z}}\begin{pmatrix} -\sqrt{z} & X \\ X^* & -\sqrt{z} \end{pmatrix}^{-1} = \begin{pmatrix} -z & \sqrt{z} X \\ \sqrt{z} X^* & -z \end{pmatrix}^{-1}.$$

Using Schur complement, we can check that

$$G(z) = \begin{bmatrix} (XX^* - z)^{-1} & (XX^* - z)^{-1} z^{-\frac{1}{2}} X \\ z^{-\frac{1}{2}} X^* (XX^* - z)^{-1} & (X^* X - z)^{-1} \end{bmatrix}.$$

The upper left block $G(z)$ is the resolvent $R(z) := (XX^* - z)^{-1}$ of Q_N we need.

To study the limiting ESD of Q_N, we need to study $m_N(z) = \frac{1}{M}\operatorname{Tr} R(z)$.

We label the rows and columns of H and G as: $\quad i \in I_1 = \{1, \cdots, M\}$,

$\mu \in I_2 = \{M+1, \cdots, M+N\}$.

Define $H^{(i)}$, $H^{(\mu)}$ and $G^{(i)} = \frac{1}{\sqrt{z}}(H^{(i)} - \sqrt{z})^{-1}$, $G^{(\mu)} = \frac{1}{\sqrt{z}}(H^{(\mu)} - \sqrt{z})^{-1}$ is a similar way as before. Denote

$$m_N(z) = \frac{1}{M}\sum_{i\in I_1} G_{ii}(z) = \frac{1}{M}\mathrm{Tr}\, R(z), \quad \tilde{m}_N(z) = \frac{1}{N}\sum_{\mu\in I_2} G_{\mu\mu}(z) = \frac{1}{N}\mathrm{Tr}\,(\mathbb{X}^*\mathbb{X} - z)^{-1}.$$

Note that : $\mathbb{X}^*\mathbb{X}$ has the same nonzero eigenvalues as $\mathbb{X}\mathbb{X}^*$, and N-M more (or fewer) zero eigenvalues. Hence,

$$\tilde{m}_N(z) = \frac{1}{N}\sum_{i=1}^{M\wedge N}\frac{1}{\lambda_i - z} - \frac{N - M\wedge N}{Nz} = \frac{M}{N}\left(\frac{1}{M}\sum_{i=1}^{M\wedge N}\frac{1}{\lambda_i - z} - \frac{M - M\wedge N}{Mz}\right) - \frac{N-M}{Nz}$$

$$= c_N m_N(z) - \frac{1-c_N}{z}, \quad \text{where } c_N = \frac{M}{N}.$$

Now, we apply the Schur complement to G_{ii}, $i\in I_1$, and $G_{\mu\mu}$, $\mu \in I_2$. We have

$$\frac{1}{G_{ii}} = -z - z\sum_{\mu,\nu\in I_2} X_{i\mu}X_{i\nu}G^{(i)}_{\mu\nu}$$

$$\approx -z - \frac{z}{N}\sum_{\mu} G_{\mu\mu} = -z(1+\tilde{m}_N(z)).$$

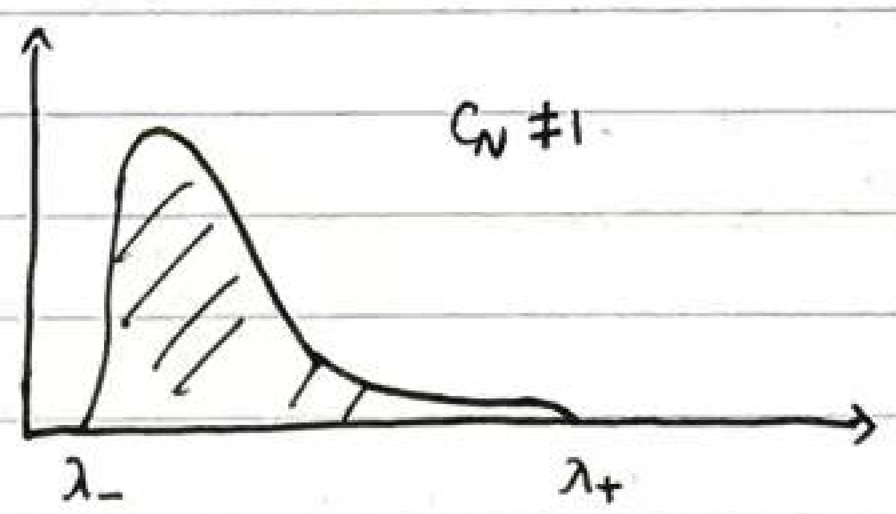
The above equation implies that $m_N(z) = \frac{1}{M}\sum_{i\in I_1} G_{ii}(z) \approx \frac{1}{-z(1+\tilde{m}_N(z))} = \frac{1}{-z(1+c_N m_N) + 1 - c_N}$

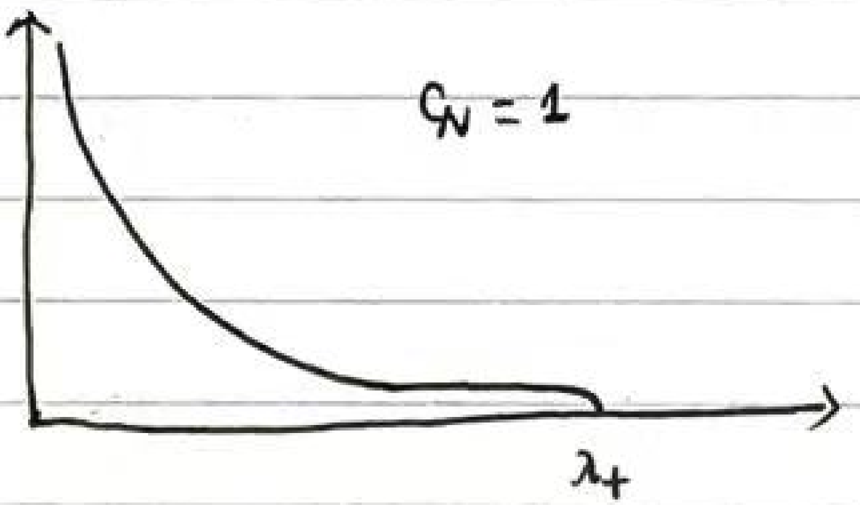
$$\Rightarrow \quad z c_N m_N^2 + (z - 1 + c_N)m_N + 1 = 0 \quad \Rightarrow \quad m_N(z) \approx \frac{1 - c_N - z + \sqrt{(z-1+c_N)^2 - 4zc_N}}{2zc_N}$$

$$= \boxed{\frac{1 - c_N - z + \sqrt{(z-\lambda_+)(z-\lambda_-)}}{2zc_N}} \to m_{MP}(z),$$

$$\lambda_+ := (1+\sqrt{c_N})^2, \quad \lambda_- := (1-\sqrt{c_N})^2.$$

Then, $\rho_{MP}(x) = \frac{1}{\pi}\lim_{\eta\downarrow 0} m_{MP}(x+i\eta) = \frac{1}{2\pi c_N x}\sqrt{(\lambda_+ - x)(x-\lambda_-)}\,\mathbb{1}_{\lambda_- \le x \le \lambda_+}$





* If $c_N > 1$ (i.e., $M > N$), there is a δ-mass at $x=0$: $\frac{M-N}{M}\delta_0$.