

### 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  $\infty$ .

The grand principle [ The key concept of RMT is the "~~wider~~ 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 ~~three~~ three standard examples.

#### 1. Wigner ensemble

Wigner's <sup>large</sup> pioneering work in 1955 marks the birth of RMT. He proposed to use a real symmetric/ complex Hermitian random matrix with independent entries to model the ~~station~~ 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 matrices :  $H = (h_{ij})_{1 \leq i, j \leq N}$  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  $h_{ij}$ ,  $1 \leq i \leq j \leq N$ , are <sup>real</sup> Gaussian random variables, and  $Eh_{ij} = 0$ ,  $Eh_{ij}^2 = 1 + \delta_{ij}$ .

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

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

Prop:

Let  $H$  be a GOE, and  $O$  be an orthogonal matrix. Then,  $\underbrace{O^T H O}_H = H$ .

Pf: We only need to check that  $E H_{ij} H_{ij}' = \delta_{ii}' \delta_{jj}' + \delta_{ij}' \delta_{ji}''$  for  $1 \leq i, j \leq N$ ,  
 $i' \leq i, j' \leq N$ ,

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

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

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

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

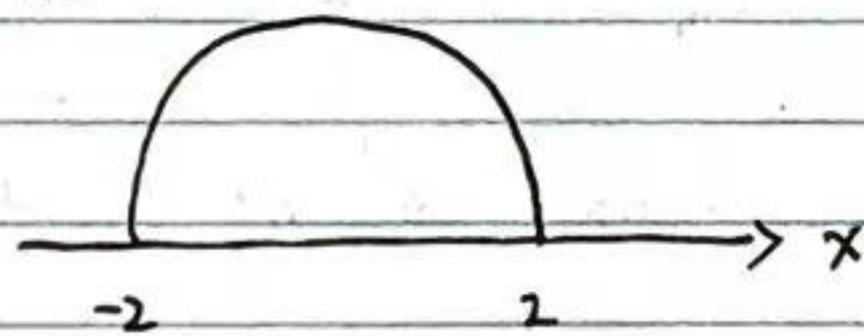
Wigner proved an "LLN" for the empirical spectral density (ESD) of  $\frac{1}{\sqrt{N}} H_N$ :

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

$\mu_{\frac{1}{\sqrt{N}} H_N} \rightarrow \mu_{SC}$  weakly, where  $\mu_{SC}$  has density

$$p_{SC}(x) = \frac{1}{2\pi} \sqrt{4-x^2} , \quad -2 \leq x \leq 2 .$$

$p_{SC}(x)$ :



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

$$\frac{1}{N} \#\{i : \frac{1}{\sqrt{N}} \lambda_i \in I\} = \int_I d\mu_{\frac{1}{\sqrt{N}} H_N}(x) \xrightarrow[N \rightarrow \infty]{} \int_I p_{SC}(x) dx$$

Q1: Does the SC law holds in a stronger sense, i.e.,  $\int_I d\mu_{\frac{1}{\sqrt{N}} H_N}(x) \xrightarrow[N \rightarrow \infty]{} \int_I p_{SC}(x) dx$  for  $\frac{1}{N} \ll a_N \ll 1$ ,  $\int_{[E-a_N, E+a_N]} d\mu_{\frac{1}{\sqrt{N}} H_N}(x) \xrightarrow[N \rightarrow \infty]{} \int_{[E-a_N, E+a_N]} p_{SC}(x) dx$  ("Local" semicircle law)

In the bulk, around  $E \approx \lambda_i$ , what is the typical gap between  $\frac{\lambda_i}{\sqrt{N}}$  &  $\frac{\lambda_{i+1}}{\sqrt{N}}$  for  $EN \leq i \leq (1-\epsilon)N$ ?

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

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

Near the edge:  $\int_{-2}^{\frac{\lambda_1}{\sqrt{N}}} f_{sc}(x) dx = \frac{1}{\pi} \frac{1}{N} \Rightarrow \int_{-2}^{\frac{\lambda_1}{\sqrt{N}}} \sqrt{4+x} dx \sim \frac{1}{N}$

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

[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  $\vec{g} = (g_1, \dots, g_N)$  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 "asymptotically uniform" on  $S(N-1)$ . But defining this concept is already very non-trivial.

## 2. Sample covariance matrices

$\mathbf{X} = (X_{ij})$   $1 \leq i \leq M, 1 \leq j \leq N$ , the entries of  $\mathbf{X}$  are ~~independent~~ independent random variables of mean 0, variance 1.

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

it reduces to studying the eigendecomposition of  $\mathbf{X}\mathbf{X}^*$  and  $\mathbf{X}^*\mathbf{X}$ . rectangular diagonal

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  $\vec{x}$ :  $\vec{x}_1, \dots, \vec{x}_N$ . Then we form the sample covariance matrix

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

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

**Q1:** Does the ESD of  $Q_N$  also converge? What is its limit?  
**A:** The limit

Rmk: The "high-dimension" setting considers  $C_N = \frac{M}{N} \rightarrow C \in (0, +\infty)$ , 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  $\mathbf{X}$  are i.i.d. Gaussian, then  $Q_N$  is called the Wishart ensemble (1928):

$$Q_N \sim W_M(I, N) \xrightarrow{\substack{\text{degrees of freedom} \\ \downarrow \text{covariance} \\ \text{data dimension}}} \mathbf{X}$$

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

A useful extension is:  $\tilde{\mathbf{X}} \sim \mathcal{N}_M(\mathbf{0}, \Sigma)$ ,  $p$ -variate normal with covariance  $\Sigma$ .

Then  $Q_N \sim W_M(\Sigma, N)$ . We can write that  $\mathbf{X} = \Sigma^{1/2} \mathbf{Y}$ , where the entries of  $\mathbf{Y}$  are i.i.d.  $\mathcal{N}(0, 1)$ . Then  $Q_N = \frac{1}{N} \Sigma^{1/2} \mathbf{Y} \mathbf{Y}^* \Sigma^{1/2}$ .

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

**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.)

**Q2:** Bulk universality?

**Q3:** Edge universality?

**Q4:** Eigenvectors?

### 3. non-Hermitian random matrices

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

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

Note that: almost surely,  $\mathbf{X}$  has  $N$  different eigenvalues. For general  $\mathbf{X}$ ,  $\text{IP}(\mathbf{X} \text{ is singular}) \rightarrow 0$  as  $N \rightarrow \infty$ .

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

$$\mu_{\frac{1}{N} \mathbf{X} \mathbf{X}^*} \xrightarrow{(dxdy)} \frac{1}{\pi} \mathbb{1}_{\{z \in \mathbb{C}: |z| \leq 1\}} \text{ dxdy weakly.}$$

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

#### 4. Invariant ensembles

$$\begin{aligned} \text{For GOE, } \mu_{H_N} &= C_N \prod_{1 \leq i \leq N} e^{-\frac{h_{ii}^2}{4}} \prod_{1 \leq i < j \leq N} e^{-\frac{h_{ij}^2}{2}} dH_N \\ &= C_N e^{-\sum_i \frac{h_{ii}^2}{4} - \sum_{1 \leq i < j \leq N} \frac{h_{ij}^2}{2}} dH_N \\ &= C_N e^{-\text{tr}(H_N^2)/4} dH_N. \end{aligned}$$

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

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

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

Note:  $\text{Tr } V(U H_N U^{-1}) = \text{Tr}[U^* V(H_N) U] = \text{Tr } V(H_N)$ , i.e. orthogonal/unitary conjugation leaves the distribution  $\rho(H_N) dH_N$  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.

**Q1:** What is the prob. density function for all the  $N$  eigenvalues only?

**Q2:** Bulk universality? Edge universality?

#### 5. Deformed random matrices

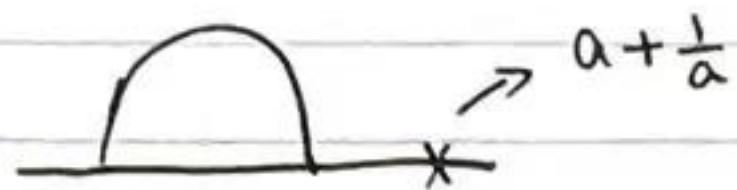
##### Deformed Wigner

$H(a) := \frac{1}{\sqrt{N}} H_N + a u u^*$ ,  $H_N$ : Wigner matrix,  $a \gg 0$ ,  $u$  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:



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

### Spiked covariance

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

A similar BBP transition occurs at  $\alpha = \sqrt{\frac{M}{N}}$ .

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

Let  $H_N$  be a Wigner matrices. We have

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

$\Rightarrow \frac{1}{N} \mathbb{E} \left( \sum_i \lambda_i^2 \right) = N$ , i.e. the averaged size of  $\lambda_i^2$  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.,  $|h_{ij}| \leq 1$  a.s.). Then, there exists absolute constants  $c, C > 0$  such that

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

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

**Lemma 1.2:** Suppose  $M_N$  is 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

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

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,  $\mathbb{P}(\|U_N\| > A\sqrt{N}) \leq C \exp(-cAN)$ ,  $\mathbb{P}(\|L_N\| > A\sqrt{N}) \leq C \exp(-cAN)$  for  $A \geq C$ .

Then for  $A \geq 2C$ ,

$$\begin{aligned} \mathbb{P}(\|H_N\| > A\sqrt{N}) &\leq \mathbb{P}(\|U_N\| > AN/2) + \mathbb{P}(\|L_N\| > AN/2) \\ &\leq 2C \exp(-cAN/2). \end{aligned}$$

□

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

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

$$\Pr(|S_N| \geq \lambda b) \leq C \exp(-C\lambda^2), \quad b^2 := \sum_{i=1}^N (b_i - a_i)^2.$$

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

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

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

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

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

$$\begin{aligned} \mathbb{E}_Z \mathbb{E}_{Z'} \exp(\lambda(Z - Z')) &= \mathbb{E}_{Z, Z'} \mathbb{E}_s \exp(\lambda s(Z - Z')) = \mathbb{E}_{Z, Z'} \left[ \frac{1}{2} e^{\lambda(Z - Z')} + \frac{1}{2} e^{-\lambda(Z - Z')} \right] \\ &\stackrel{\uparrow}{=} \mathbb{E} \exp\left(\frac{\lambda^2}{2}(Z - Z')^2\right) \leq \exp\left(\frac{\lambda^2}{2}(b-a)^2\right). \end{aligned} \quad \square$$

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

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

$$\text{So } \Pr(S_N > \lambda b) \leq \exp(-t\lambda b) \exp\left(\frac{t^2}{2}b^2\right) = \exp\left(\frac{t^2}{2}b^2 - t\lambda b\right).$$

$$\text{Taking } t = \lambda, \quad \Pr(S_N > \lambda b) \leq \exp\left(-\frac{1}{2}\lambda^2 b^2\right) = \exp\left(\frac{1}{2}(t\delta - \lambda)^2 - \frac{\lambda^2}{2}\right).$$

Taking  $t = \lambda/b$  gives  $\Pr(S_N > \lambda b) \leq \exp(-\lambda^2/2)$ . Can get a similar bound for  $\Pr(S_N < -\lambda b)$ .  $\square$

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

$$\Pr(|M_N x| \geq A\sqrt{N}) \leq C \exp(-CAN) \text{ for } A \geq C.$$

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

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

$\Rightarrow \Pr\left(\sum_{i=1}^N |X_i \cdot x|^2 \geq CN\right)$  For any  $c' < c$ ,  $\mathbb{E} \exp(c' |\sum_i X_i \cdot x|^2) \leq c'$  for a constant  $c' > 0$ .

[Use the tail-probability formula,

$$\mathbb{E} f(X) = \int_0^{+\infty} \Pr(X \geq t) f(t) dt, \quad X \text{ positive, } f \text{ increasing on } [0, +\infty), \quad \& \quad f(0) = 0.$$

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

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

How to extend Lem 1.5 to a bound on

$$\begin{aligned} \mathbb{P}(\|M_N\| \geq A\sqrt{N}) &\leq \mathbb{P}\left(\bigcup_{x \in S^N} \{M_N x \geq A\sqrt{N}\}\right) \\ &= \mathbb{P}\left(\bigcup_{x \in S^N} \{\|M_N x\| \geq A\sqrt{N}\}\right). \end{aligned}$$

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

Def ( $\varepsilon$ -net) An maximal  $\varepsilon$ -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  $\varepsilon$ , and which is maximal with respect to set inclusion.

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

Lemma 1.6 (Volume packing) Let  $0 < \varepsilon < 1$ , and  $\Sigma$  be an maximal  $\varepsilon$ -net. Then  $|\Sigma| \leq (3/\varepsilon)^{EN}$ .

Pf: Consider the collection of balls of radius  $\varepsilon/2$  centered around each point in  $\Sigma$ . 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.  $\square$

Proof of Lemma 1.2: Let  $\Sigma$  be a  $\frac{1}{2}$ -net of  $S^N$ . Then  $|\Sigma| \leq 6^N$ .

Taking a union bound, we get  $\mathbb{P}(\max_{x \in \Sigma} \|Mx\|) \leq |\Sigma|$

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

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

Combining (\*) and (t) 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 constant  $C > 0$ , so that  
 $\Pr(|X| \geq t) \leq 2\exp(-ct^2) \quad \forall t \geq 0.$

\* 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  $X_1, \dots, X_N$  be independent, mean-zero, sub-gaussian r.v.s. Then,  $\forall t \geq 0$ ,

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

Rmk: In accordance with the semicircle law, we should have that  $\forall$  constant  $\epsilon > 0$ ,  
 $\Pr(\|H_N\| \geq (2+\epsilon)\sqrt{N})$  with high probability.

One slick way to prove this result is the important "moment method in RMT":  
for any  $k \in \mathbb{Z}/N$ , note

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

Hence,

$$\mathbb{E} \|H_N\|^k \leq \mathbb{E} \text{tr}(H_N^k) \Rightarrow \Pr(\|H_N\|^k \geq (2+\epsilon)\sqrt{N}) \leq [(2+\epsilon)\sqrt{N}]^{-k} \mathbb{E} \text{tr}(H_N^k).$$

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

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

Then, we have

$$\Pr(\|H_N\|^k \geq (2+\epsilon)\sqrt{N}) \leq \left(1 - \frac{\epsilon}{4}\right)^k N \ll 1 \quad \text{for } k = C \log N \text{ if } C \text{ is large enough.}$$

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:  $\|H_N\| \leq 2 + N^{-\frac{2}{3} + \epsilon}$  w.h.p. for any const.  $\epsilon > 0$ .

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  $\mathbb{E} \text{tr}(H_N^k)$  for  $k$  large but finite  $k \in \mathbb{N}$ .

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 <sup>real</sup> 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  $\mu_{H_N}$  converges in distribution to  $\mu_{sc}$  almost surely.

① Moment method

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

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

Define a sequence of measures  $\{\mathbb{E}\mu_{H_N}\}(A) := \mathbb{E} \int \mathbf{1}(x \in A) d\mu_{H_N}(x)$ .

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.,  $\forall k \in \mathbb{N}$ ,

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

Rmk: To show convergence in prob. of  $\mu_{H_N}$ , we need to show concentration of measure, i.e.,  $\mu_{H_N}$  concentrates around  $\mathbb{E}\mu_{H_N}$ . For that purpose, we need to show that  $\text{Var}[\int x^k d\mu_N(x)] \rightarrow 0$ ,  $\forall k \in \mathbb{N}$ . 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.

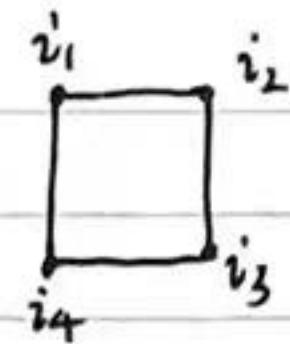
While these can be done, we refrain from doing that and will prove the semicircle law using another "~~more powerful~~" method — the Stieltjes transform method.

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

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

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

Let's start with  $k=4$  case:  $\frac{1}{N} \text{IE}$

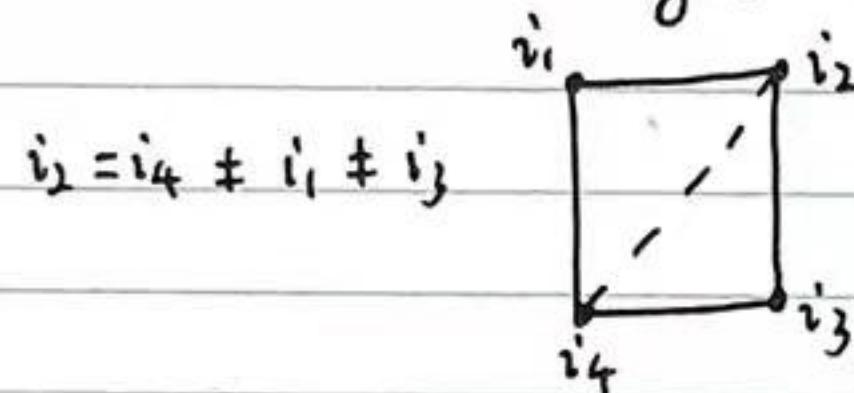


vertices

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

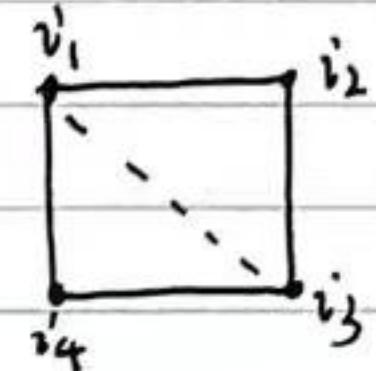
- (i)  $i_1 = i_3 \neq i_2 \neq i_4$  Typical order ( $\frac{1}{N} \times N^3 \times \frac{1}{N^2} = 1$ )
- (ii)  $i_1 = i_3 \neq i_2 = i_4$  ( $\frac{1}{N} \times N^2 \times \frac{1}{N^2} = \frac{1}{N}$ )
- (iii)  $i_1 = i_2 = i_3 \neq i_4$  ( $\frac{1}{N} \times N^2 \times \frac{1}{N^2} = \frac{1}{N}$ )
- (iv)  $i_1 = i_2 = i_3 = i_4$  ( $\frac{1}{N} \times N \times \frac{1}{N^2} = \frac{1}{N^2}$ )

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



With the fact that  $\text{IE } h_{ij}^2 = \frac{1}{N}$ , we get

$$\text{IE } \frac{1}{N} \text{Tr}(H_N^4) = 2 \times \frac{1}{N} \times N^2 \times \frac{1}{N^2} + O\left(\frac{1}{N}\right) = 2 + O\left(\frac{1}{N}\right).$$



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

- (i) There are 3 distinct edges, each occurring twice, and hence 4 distinct vertices.

$$\left( \frac{1}{N} \times N^4 \times \frac{1}{N^3} = 1 \right)$$

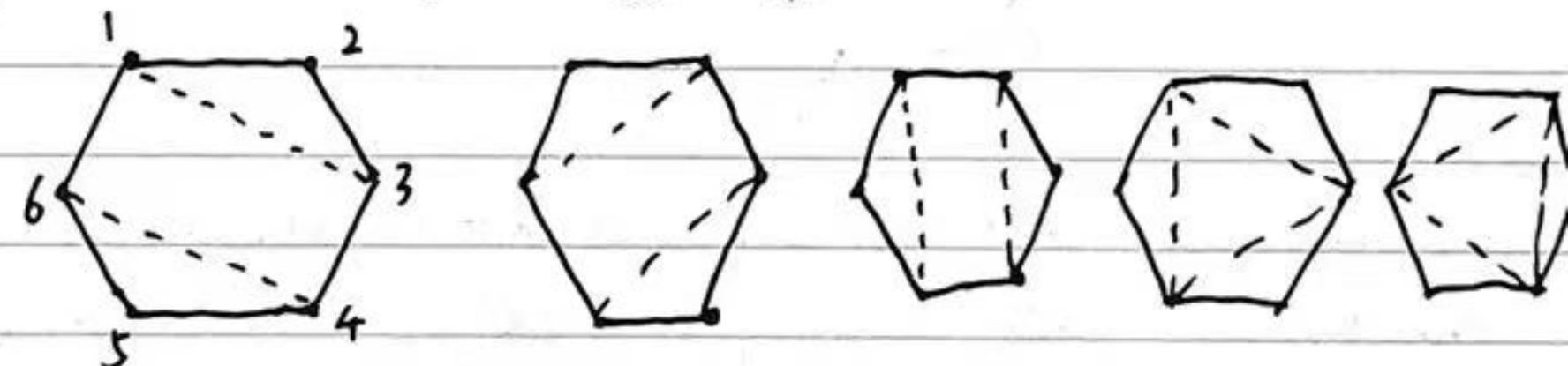
- (ii) 2 distinct edges, one occurring twice & one occurring four times, and hence 3 distinct vertices.

$$\left( \frac{1}{N} \times N^3 \times \frac{1}{N^3} = \frac{1}{N} \right)$$

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

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

How many graphs in case (i)?



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

$$\text{Hence, } \text{IE } \frac{1}{N} \text{Tr}(H_N^6) = 5 \times \frac{1}{N} \times N^4 \times \frac{1}{N^3} + O\left(\frac{1}{N}\right) = 5 + O\left(\frac{1}{N}\right).$$

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

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

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  $k$ .

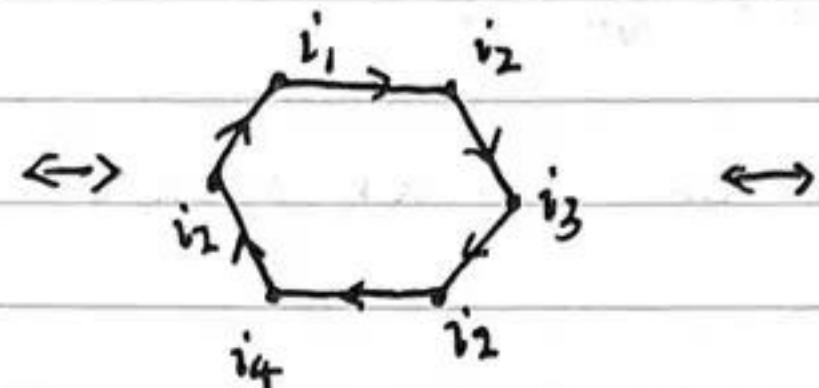
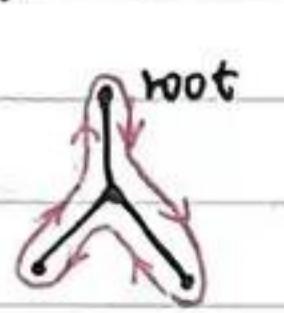
We need to count the number of non-crossing cycles.

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,  $\mathbb{E} \frac{1}{N} \text{Tr}(H_N^k) \rightarrow 0$  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  $\frac{k}{2}$  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,



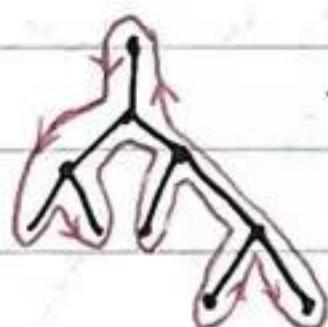
non-crossing planar graph



With lemma 2.2, we can get the following corollary:

Exercise: Let  $i_1, \dots, i_k$  be a cycle of length  $k$ . Arrange the integers  $1, 2, \dots, k$  around a circle. Whenever  $1 \leq a < b \leq k$  s.t.  $i_a = i_b$  with no  $c$  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 exactly  $\frac{k}{2} - 1$  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:



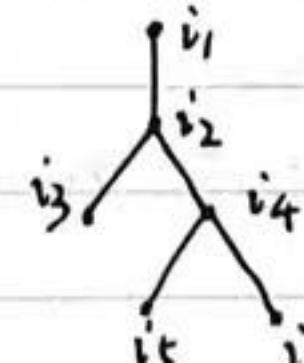
$\rightarrow (i_1, i_2, i_3, i_4, i_3, i_5, i_3, i_2, i_6, i_7, i_6, i_8, i_9, i_8, i_{10}, i_8, i_6, i_2, i_1)$

We now show that we can construct a unique tree from a non-crossing cycle:  $(i_1, i_2, \dots, i_k)$ .

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 of former type an "innovative (I) edge", and a step of the latter type an "returning (R) edge". Then there are  $\frac{k}{2}$  (I) edges, and  $\frac{k}{2}$  (R) 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  $(i_1, i_2, \dots, i_k, i_1)$ , we construct 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:

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



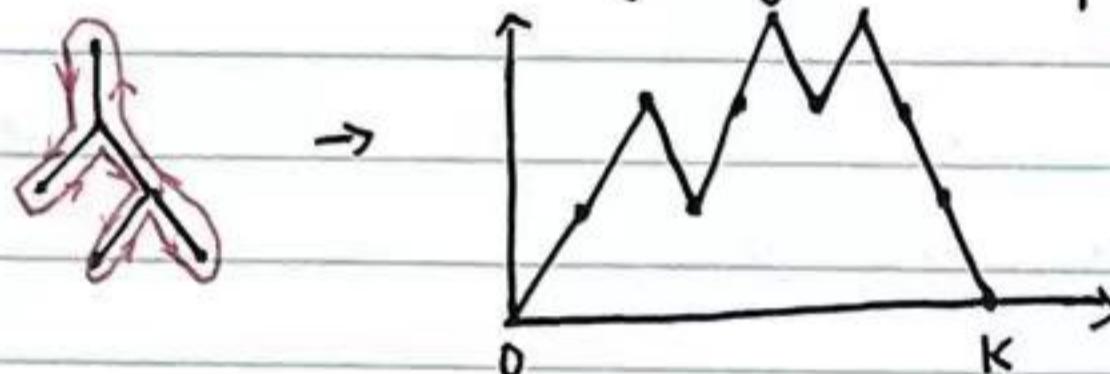
This clearly gives a rooted tree. □

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

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

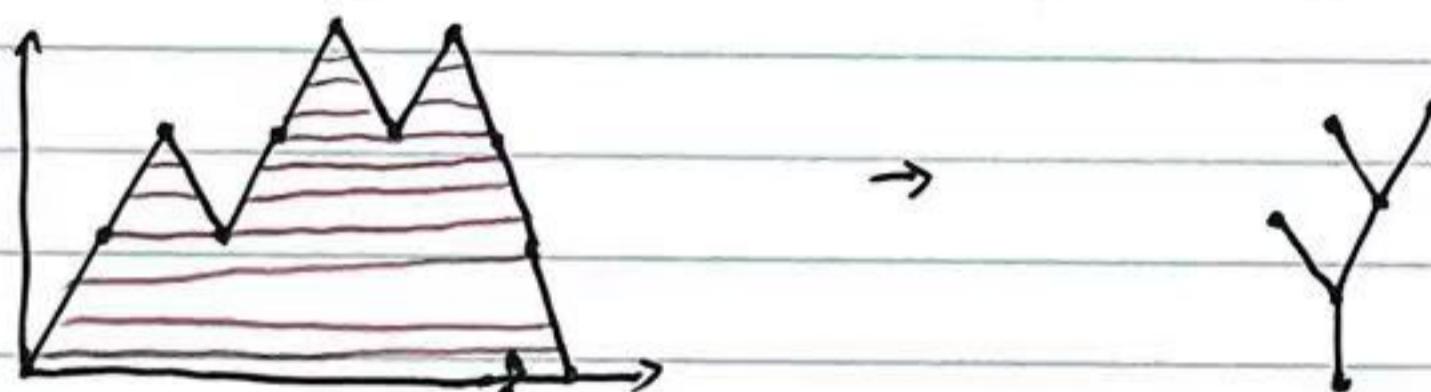
One version of proof of the Fact: We further construct a 1-1 correspondance between the rooted trees and random walks on the positive half line.

Trans: Traversing a rooted tree, if we traverse 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 <sup>k-step</sup> random walks from 0 to 0 and stays positive?

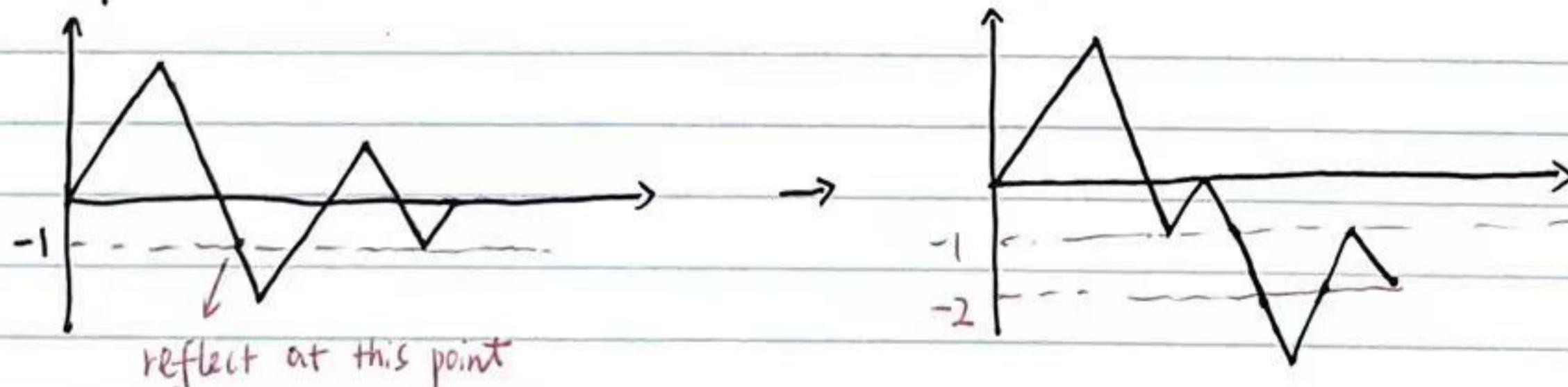
$$\# \{ \text{random } k\text{-step walks simple walks from 0 to 0} \} - \# \{ \begin{matrix} k\text{-step simple walks from 0 to 0} \\ \text{and hit } -1 \end{matrix} \}$$

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

↑  
Use reflection principle

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

For example:



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

□

It remains to show that:  $\int x^k d\mu_{S^2}(x) = C_{k/2} \cdot 1$  ( $k$  is even).

This is trivial for  $k=$  odd. For  $k$  even,

$$\begin{aligned} I_k &= \int_{-2}^2 x^k \frac{\sqrt{4-x^2}}{2\pi} dx = \int_0^\pi (2\cos\theta)^k \frac{2\sin\theta}{2\pi} \cdot 2\sin\theta d\theta \stackrel{\text{Integrate by parts}}{=} \frac{2^{k+1}}{\pi} \int_0^\pi (\cos\theta)^k \sin^2\theta d\theta \\ &= \frac{2^{k+1}}{\pi} \int_0^\pi (\cos\theta)^{k-1} \sin^2\theta ds\sin\theta = -\frac{2^{k+1}}{\pi} \int_0^\pi \sin\theta \cdot [2\sin\theta \cdot (\cos\theta)^k - (k-1)\sin^3\theta (\cos\theta)^{k-2}] d\theta \\ &= -2I_k + \frac{2^{k+1}}{\pi} \int_0^\pi (k-1) \sin^2\theta (1-\cos^2\theta) (\cos\theta)^{k-2} d\theta \\ &= -(k+1)I_k + 4(k-1) \frac{2^{k-1}}{\pi} \int_0^\pi \sin^2\theta (\cos\theta)^{k-2} d\theta = -(k+1)I_k + 4(k-1)I_{k-2}. \end{aligned}$$

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

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  $\mu$  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  $M_{H_N}$ , we have  $S_{M_{H_N}}(z) := \frac{1}{N} \int_{\mathbb{R}} \frac{1}{x-z} \sum_{i=1}^N \delta_{\lambda_i}(dx) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i - z} = \frac{1}{N} \text{Tr}(H_N - z)^{-1}$ .

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  $m_N(z) \equiv S_N(z) = \frac{1}{N} \text{Tr}(G(z))$ .

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

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

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

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

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

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

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

Prop 2.4 Let  $H$  be a <sup>real</sup> symmetric / complex Hermitian matrix, we have

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

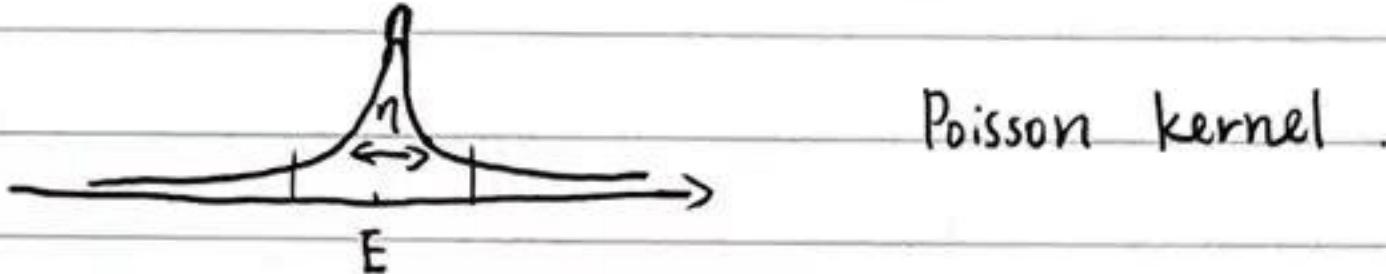
$$(ii) \quad \|G(z)\| \leq |\operatorname{Im} z|^{-1}.$$

Upper-half complex plane

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

$$(*) \quad \operatorname{Im} S_\mu(z) = \int \frac{\eta}{(x-E)^2 + \eta^2} d\mu(x) > 0 \\ = \pi(\mu * P_\eta)(E), \quad P_\eta(x) = \frac{1}{\pi} \frac{\eta}{x^2 + \eta^2} = \frac{1}{\eta} \left[ \frac{1}{\pi} \frac{1}{1 + (x/\eta)^2} \right].$$

$P_\eta$ :



Poisson kernel.

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

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

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

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

Lemma 2.5 Let  $\mu_n$  be a sequence of random probability measures on  $\mathbb{R}$ , and let  $\mu$  be a deterministic probability measure. Suppose  $\{\mu_n\}$  is tight. Then  $\mu_n$  converges almost surely (in probability) if and only if  $\mu_n$  converges 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  $\lim_{z \in \mathbb{C}_+, |z| \rightarrow +\infty} z S_\mu(z) = -1$ , and

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

Lemma 2.6 (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

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

## A heuristic derivation of the semicircle law:

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

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

To simplify notations, we will adopt the convention  $G_{ab}^{(i)} = 0$  and  $H_{ab}^{(i)} = 0$  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  $H$ ".

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

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

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

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

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

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

We expect that there is a concentration phenomenon where  $\sum_{k,l} h_{ik} h_{il} G_{kl}^{(i)}$  concentrates around its partial expectation  $m_N^{(i)}(z)$  (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

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

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

Solving the equation and requiring that  $\text{Im } m_N \geq 0$ , 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:

- Fix  $\epsilon$
- ① Concentration of quadratic forms of independent random variables.
  - ② 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  $\epsilon > 0$  and large  $D > 0$ , we have

$$\Pr(|X_N| > N^\epsilon |Y_N|) \leq N^{-D}$$

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

Def: (High-probability event) An event  $E$  is said to hold with high probability (w.h.p.) if  $\Pr(E) \geq 1 - N^{-D}$  for large enough  $N$ . (Hence,  $X_N \prec Y_N$  iff for any small constant  $\epsilon > 0$ ,  $|X_N| \leq N^\epsilon |Y_N|$  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  $X \prec Y + N^{-\epsilon} X$  for some constant  $\epsilon > 0$ , then  $X \prec Y$ .

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

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

$$|EX| \leq |E[X|_{\Omega}|] = |E(|X| \mathbf{1}_{\Omega})| + |E(|X| \mathbf{1}_{\Omega^c})|$$

$$\leq N^\epsilon |E[Y]| + N^\epsilon \Pr(\Omega^c) \leq N^\epsilon |E[Y]| + N^{C-D}.$$

Taking  $D$  large enough so that  $N^{C-D} \leq N^{-C} \leq |E[Y]|$ , we get  $|EX| \leq (N^\epsilon + 1) |E[Y]|$ .  $\square$

However, ~~" $\prec$ "~~ 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  $u \in U^{(N)}$  is a parameter and  $U^{(N)}$  is certain parameter set. (For example,  $U^{(N)}$  can be the set of matrix indices, e.g.,  $G_{ij}, (i, j) \in \{1, \dots, N\}^2$ .)

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

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

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

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

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.

Thm 2.1 (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

$$(1) \quad \sum_{i=1}^N b_i x_i \prec \left( \sum_{i=1}^N |b_i|^2 \right)^{1/2},$$

$$(2) \quad \sum_{i,j} a_{ij} x_i y_j \prec \left( \sum_{i,j} |a_{ij}|^2 \right)^{1/2},$$

$$(3) \quad \sum_{i \neq j} a_{ij} x_i x_j \prec \left( \sum_{i \neq j} |a_{ij}|^2 \right)^{1/2}. \quad (*)$$

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

$$\begin{aligned} \mathbb{E} \left| \sum_{i \neq j} a_{ij} x_i x_j \right|^2 &= \mathbb{E} \sum_{i \neq j, i' \neq j'} a_{ij} a_{i'j'} x_i x_j \bar{x}_{i'} \bar{x}_{j'} \\ &= \mathbb{E} \sum_{i \neq j} [a_{ij} \bar{a}_{ij} |x_i|^2 |x_j|^2 + a_{ij} \bar{a}_{ji} |x_i|^2 |x_j|^2] \\ &= \sum_{i \neq j} [a_{ij} \bar{a}_{ij} + a_{ij} \bar{a}_{ji}] \leq 2 \sum_{i \neq j} |a_{ij}|^2. \end{aligned}$$

In other words, the large deviation estimates are optimal up to an  $N^\epsilon$  factor.

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

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

$$\mathbb{E} \left[ \left| \sum_i b_i x_i \right|^p \right] \leq C_p \left\| \left( \sum_i |b_i|^2 |x_i|^2 \right)^{1/2} \right\|_p^p = C_p \mathbb{E} \left[ \left( \sum_i |b_i|^2 |x_i|^2 \right)^{p/2} \right].$$

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

$$\begin{aligned} \mathbb{E}\left[\left|\sum_i b_i x_i\right|^p\right] &\leq C_p B^p \mathbb{E}\left[\left(\sum_i \frac{|b_i|^2}{B^2} |x_i|^2\right)^{p/2}\right] \stackrel{\text{Jensen}}{\leq} C_p B^p \mathbb{E}\left[\sum_i \frac{|b_i|^2}{B^2} |x_i|^p\right] \\ &\leq C_p' B^p \sum_i \frac{|b_i|^2}{B^2} = C_p' B^p. \end{aligned}$$

$\forall \varepsilon > 0$ , we have

$$\mathbb{P}\left[\left|\sum_i b_i x_i\right| \geq N^\varepsilon B\right] \leq \frac{\mathbb{E}\left[\left|\sum_i b_i x_i\right|^p\right]}{(N^\varepsilon B)^p} \leq C_p' N^{-p\varepsilon}. \text{ Take } p > \frac{D}{2}.$$

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

$$\begin{aligned} \mathbb{E}\left[\left|\sum_{i,j} a_{ij} x_i Y_j\right|^p\right] &\leq C_p \mathbb{E}\left[\left(\sum_j |b_j|^2 |x_j|^2\right)^{p/2}\right] \\ &\leq C_p' \mathbb{E}\left[\left(\sum_j |b_j|^2\right)^{p/2}\right] = C_p' \left\|\sum_j |b_j|^2\right\|_{p/2}^{p/2} \\ (\text{triangle}) \quad \text{use the Minkowski ineq.}, \quad \rightarrow \quad &\leq C_p' \left(\sum_j [\mathbb{E}(|b_j|^p)]^{\frac{2}{p}}\right)^{\frac{p}{2}}. \end{aligned}$$

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

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

Finally, for (3), we ~~need to~~ reduce it to ~~the~~ cases where (2) ~~can~~ be applied.

For  $\sum_{i \neq j} a_{ij} x_i x_j$ , we use the following identity to decompose  $\{x_i\}$  into two sets of independent random variables:  $\forall i \neq j$ ,

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

Also, notice that the <sup>total</sup> number of partitions is  $2^{N-2}$ .

Then,

$$\left\|\sum_{i \neq j} a_{ij} x_i x_j\right\|_p \leq \frac{1}{2^{N-2}} \sum_{I \sqcup J} \left\|\sum_{i \in I} \sum_{j \in J} a_{ij} x_i x_j\right\|_p$$

Minkowski:

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

In other words,  $\mathbb{E}\left[\left|\sum_{i \neq j} a_{ij} x_i x_j\right|^p\right] \leq C_p \left(\sum_{i,j} |a_{ij}|^2\right)^{p/2}$ . Then apply Markov's ineq.  $\square$

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

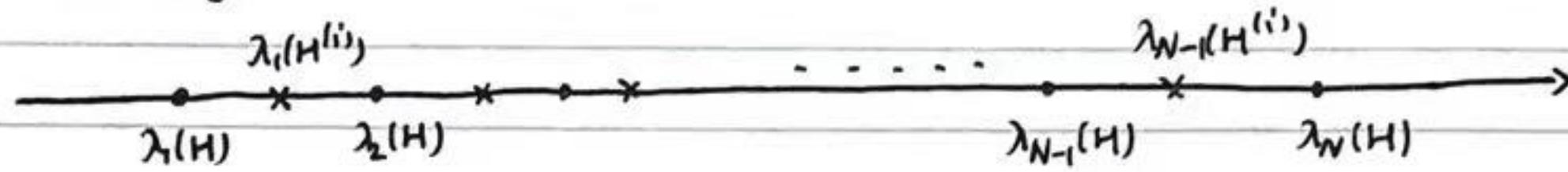
We want to bound

$$m_N - m_N^{(i)} = \frac{1}{N} [\text{Tr } G(\beta) - \text{Tr } G^{(i)}(\beta)] = \frac{1}{N} \left( \text{Tr } \frac{1}{H-\beta} - \text{Tr } \frac{1}{H^{(i)}-\beta} \right),$$

$$= \frac{1}{N} \left[ \sum_{i \neq j} \frac{1}{\lambda_{i-j}} \right] - \left[ \sum_{i \neq j} \frac{1}{\lambda_{i-j}^{(i)}} \right]$$

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

Cauchy interlacing:  $\lambda_1(H) \leq \lambda_1(H^{(i)}) \leq \lambda_2(H) \leq \dots \leq \lambda_{N-1}(H^{(i)}) \leq \lambda_N(H)$ .



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

$$|\operatorname{Im} m_N - \operatorname{Im} m_N^{(i)}| \leq \frac{C}{N\eta}. \quad (\text{They are alternating sums.})$$

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

Prop: (Useful resolvent identities) (i) If  $i \neq j$ , then

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

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

$$(iii) [\text{Ward identity}] \text{ we have } \sum_j |G_{ij}|^2 = \sum_j |G_{ji}|^2 = \frac{\operatorname{Im} G_{ii}}{\eta}. \quad (\eta = \operatorname{Im} z)$$

Proof: (i) follows from Schur complement formula.

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

Let  $\tilde{H}^{(k)}$  be the  $N \times N$  matrix obtained by setting the  $k$ -th row and column of  $H$  to 0.

(In other words,  $\tilde{H}^{(k)}$  is obtained by adding a zero row and column to  $H^{(k)}$ .)

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

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

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

$$G_{ij} - G_{ij}^{(k)} = -G_{ik} \sum_k h_{kj} G_{kj}^{(k)} \stackrel{\substack{\uparrow \\ \text{by (i)}}}{=} G_{ik} \left( \frac{1}{G_{kk}} G_{kj} \right).$$

This gives (ii).

$$\begin{aligned} \text{For (iii), } G(z) &= \sum_k \frac{u_k u_k^*}{\lambda_k - z}. \text{ Then, } \sum_j |G_{ij}|^2 = \sum_j G_{ij} \overline{G_{ij}} = \sum_j G_{ij} G_{ji}^*. \\ &= (GG^*)_{ii} = \left( \sum_k \frac{u_k u_k^*}{|\lambda_k - z|^2} \right)_{ii} = \sum_k \frac{|u_k(i)|^2}{|\lambda_k - z|^2} = \frac{1}{\eta} \sum_k \frac{\eta}{|\lambda_k - z|^2} |u_k(i)|^2 \\ &= \frac{1}{\eta} \operatorname{Im} G_{ii}(z). \quad \square \end{aligned}$$

$$\begin{aligned}
\text{Then, we have } |m_N - m_N^{(i)}| &= \left| \frac{1}{N} \operatorname{Tr} G - \frac{1}{N} \operatorname{Tr} G^{(i)} \right| \leq \frac{1}{N} \sum_j |G_{jj} - G_{jj}^{(i)}| \quad (\text{with the convention } \\
&= \frac{1}{N} \sum_j \left| \frac{G_{ji} G_{ij}}{G_{ii}} \right| \leq \frac{1}{N|G_{ii}|} \left( \sum_j |G_{ji}|^2 \right)^{1/2} \left( \sum_j |G_{ij}|^2 \right)^{1/2} \\
&= \frac{1}{N|G_{ii}|} \cdot \frac{\operatorname{Im} G_{ii}}{\eta} \leq \frac{1}{N\eta}.
\end{aligned}$$

### ③ Stability of the self-consistent equation

Prop: Fix any  $\beta = E + i\eta \in \mathbb{C}_+$ . Suppose  $m$  (deter./random) satisfies that  $|m^2 + 3m + 1| \leq \delta$  for some  $\delta \ll 1$ , and  $\operatorname{Im} m > 0$ . Then,

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

Proof: Suppose let  $m^2 + 3m + 1 = \Delta$ , where  $|\Delta| \leq \delta$ . Then,  $m$  satisfies

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

Then, the conclusion follows from a direct calculation.  $\square$

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

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

By Schur complement,

$$\frac{1}{G_{ii}} = h_{ii} - \beta - \sum_{k \neq i}^{(i)} h_{ik} h_{il} G_{kk}^{(i)}.$$

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

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

$$\sum_k^{(i)} (h_{ik}^2 - \frac{1}{N}) G_{kk}^{(i)} \leq \frac{1}{N} \left( \sum_k^{(i)} |G_{kk}^{(i)}|^2 \right)^{1/2}. \quad (\text{Note } h_{ik} - \frac{1}{N} \text{ are mean 0, independent random variables with variance } O(\frac{1}{N^2}).)$$

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

Hence, we get

$$\begin{aligned}
\left| \sum_{k \neq i}^{(i)} h_{ik} h_{il} G_{kk}^{(i)} - \frac{1}{N} \operatorname{Tr} G^{(i)} \right| &\leq \left| \sum_{k \neq i}^{(i)} h_{ik} h_{il} G_{kk}^{(i)} \right| + \left| \sum_k^{(i)} (h_{ik}^2 - \frac{1}{N}) G_{kk}^{(i)} \right| \\
&\leq \frac{1}{N} \left( \sum_{k \neq i}^{(i)} |G_{kk}^{(i)}|^2 \right)^{1/2} + \frac{1}{N} \left( \sum_k^{(i)} |G_{kk}^{(i)}|^2 \right)^{1/2} \\
&\leq \frac{2}{N} \left( \sum_{k \neq i}^{(i)} |G_{kk}^{(i)}|^2 \right)^{1/2} = \frac{2}{N} \left( \sum_k^{(i)} \frac{\operatorname{Im} G_{kk}^{(i)}}{\eta} \right)^{1/2} \leq \frac{1}{\sqrt{N}\eta}.
\end{aligned}$$

In sum, we get  $\frac{1}{G_{ii}} = \cancel{\alpha + O_N(\frac{1}{\sqrt{N}\eta})} - \gamma - \frac{1}{N} \text{Tr } G^{(i)} + O_\gamma(\frac{1}{\sqrt{N}\eta})$

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

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

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

$$\Rightarrow m_N^2(z) + 3m_N(z) + 1 = O_\gamma(\frac{1}{\sqrt{N}\eta^4}). \Rightarrow |m_N(z) - m_{SC}(z)| < \frac{1}{\sqrt{N}\eta^5}. \quad \square$$