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 / ...~~ 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 $\mathbb{E} H'_{ij} H'_{i'j'} = \cancel{(1+\delta_{ij})\delta_{ii'}\delta_{jj'}}$ $\delta_{ii'}\delta_{jj'} + \delta_{ij'}\delta_{ji'}$ ~~for $1 \le i,j \le N$, $1 \le i',j' \le N$~~,

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

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

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

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

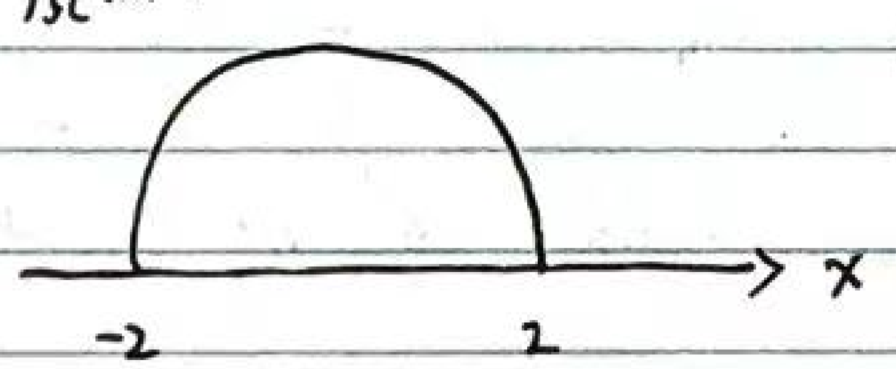
Wigner proved an "LLN" for the empirical spectral density (ESD) of $\frac{1}{\sqrt{N}} H_N$:

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

$\mu_{\frac{1}{\sqrt{N}} H_N} \to \mu_{sc}$ 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 $|I| = \varepsilon$,

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

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

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

$\uparrow$

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

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

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 ~~i.i.d.~~ 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.

~~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 X are i.i.d. Gaussian, then Q_N is called the Wishart ensemble (1928):

$$Q_N \sim W_M(I, N)$$

M: data dimension; I: covariance; N: degrees of freedom

> **Exercise:** Check that UXV^* ~~is~~ has the same distribution as 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 $X = \Sigma^{1/2} Y$, 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.

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

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

$\boxed{\text{Q3:}}$ Edge universality?

$\boxed{\text{Q4:}}$ Eigenvectors?

3. non-Hermitian random matrices

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

We want to study the asymptotic behavior of the eigenvalues & eigenvectors of X.

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

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

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

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?

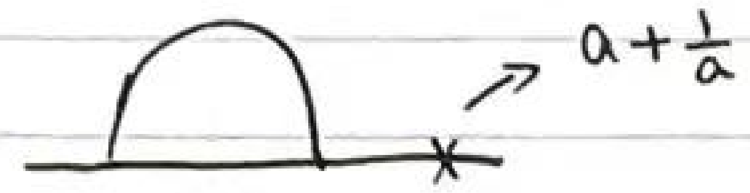
5. Deformed random matrices

<u>Deformed Wigner</u>

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

WLOG, let $a>0$. A <u>BBP</u> 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 $a = \sqrt{\frac{M}{N}}$.

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

Let H_N be a Wigner matrices. We have

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

$\Rightarrow$ $\frac{1}{N}\mathbb{E}\big(\sum_i \lambda_i^2\big) = N$, 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~~
$$\|H_N\| := \sup_{x\in\mathbb{C}^n:\,|x|=1} |H_N x|, \quad |\cdot| \text{ means the } L^2\text{-norm.}$$

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 a $\overbrace{\text{random}}^{N\times N}$ 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 $A \ge 2C$,
$$\mathbb{P}(\|H_N\| > A\sqrt{N}) \le \mathbb{P}(\|U_N\| > AN/2) + \mathbb{P}(\|L_N\| > AN/2)$$
$$\le 2C\exp(-cAN/2). \qquad \square$$

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

$$\mathbb{E}_Z \exp(\lambda(Z - \mathbb{E}Z)) = \mathbb{E}_Z \exp(\lambda(Z - \mathbb{E}_{Z'}(Z'))) \underset{\uparrow}{\leq} \mathbb{E}_Z \mathbb{E}_{Z'} \exp(\lambda(Z - Z'))$$

Jensen's ineq.

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

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

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

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

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

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

~~$\Rightarrow \mathbb{P}(\sum_{i=1}^N |X_i \cdot x|^2 \ldots)$~~ 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, $\mathbb{E}\exp(c'|Mx|^2) = \prod_{i=1}^{N} \exp(c'|X_i\cdot x|^2) \le (C')^N$.

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

How to extend Lem 1.5 to a bound on

$$\mathbb{P}(\|M_N\| \ge A\sqrt{N}) \cancel{\le \mathbb{P}(\sup_{x\in S}|M_Nx| \ge A\sqrt{N})}$$
$$= \mathbb{P}\Big(\bigcup_{x\in S^N}\{|M_Nx| \ge A\sqrt{N}\}\Big).$$

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

<u>Def</u> (ε-net) An ~~A maximal~~ 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 an ~~a maximal~~ ε-net. Then $|\Sigma| \le (3/\varepsilon)^N$.

<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. □

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

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

for large enough $A>0$.

Next, we show that $\mathbb{P}(\|M\| > \lambda) \le \mathbb{P}(\max_{x\in\Sigma}|Mx| > \lambda/2)$ $(*)$ 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., ~~$\|My\| \ge \|Mx\|$~~ $|My| \ge |Mx| - |M(x-y)| > \|M\| - \frac{1}{2}\|M\| = \frac{1}{2}\|M\|$.

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

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

* 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, $\forall\ t\geq 0$,
$$\mathbb{P}\Big\{\Big|\sum_{i=1}^N X_i\Big|\geq t\Big\}\leq 2\exp\Big(\frac{-ct^2}{\sum_{i=1}^N\|X_i\|_{\psi_2}^2}\Big).$$

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
$$\mathrm{tr}(H_N^k)=\sum_{i=1}^n\lambda_i^k\geq\max_i|\lambda_i|^k=\|H_N\|^k.$$
Hence,
$$\mathbb{E}\|H_N\|^k\leq\mathbb{E}\,\mathrm{tr}(H_N^k)\Rightarrow\mathbb{P}(\|H_N\|^k\geq(2+\varepsilon)\sqrt{N})\leq[(2+\varepsilon)\sqrt{N}]^{-k}\,\mathbb{E}\,\mathrm{tr}(H_N^k).$$
The moment method aims to control $\mathbb{E}\,\mathrm{tr}(H_N^k)$. One can show that
$$\mathbb{E}\,\mathrm{tr}(H_N^k)=[2+o_k(1)]^k N^{\frac{k}{2}+1}\quad(*)\quad\text{for } k \text{ as large as } C\log N.$$
Then, we have
$$\mathbb{P}(\|H_N\|^k\geq(2+\varepsilon)\sqrt{N})\leq\Big(1-\frac{\varepsilon}{4}\Big)^k N\ll 1 \quad\text{for } k=C\log N \text{ if } C \text{ is large enough.}$$
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 $\mathbb{E}\,\mathrm{tr}(H_N^k)$ 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.