Random Matrix Theory

Junyi Liao

Contents

1 Wigner Matrices and the Semicircle Law

For a real symmetric matrix $A \in \mathbb{R}^{n\times n}$ or a complex Hermitian matrix $A \in \mathbb{C}^{n\times n}$, we list its eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$ including repetitions according to algebraic multiplicity. We define the *empirical spectral distribution (ESD)* as

$$\mu_A = \frac{1}{n}\sum_{j=1}^{n} \delta_{\lambda_j(A)},$$

which is the (discrete) probability measure on the eigenvalues of A weighted by algebraic multiplicity. For a measurable function $f : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}} f\, d\mu_A = \frac{1}{n}\sum_{j=1}^{n} f(\lambda_j(A)).$$

In particular, the spectral moments of A is the moments of the ESD:

$$\int_{\mathbb{R}} x^p\, d\mu_A(x) = \frac{1}{n}\sum_{j=1}^{n} \lambda_j(A)^p, \quad p \geq 1.$$

If A is a random matrix, the resulting ESD μ_A is a random measure on $\mathbb{R}$. To study the asymptotic law of ESD of random matrices, we need to clarify the convergence mode of random measures.

Definition 1.1 (Convergence of random measures)**.** Let (μ_n) be a sequence of random probability measures on a topological space Ω with Borel σ-algebra $\mathscr{B}$, and let μ be another probability measure.

(i) μ_n *converges weakly* (or *vaguely*) *almost surely* to μ, if with probability 1,

$$\int_{\Omega} f\, d\mu_n \to \int_{\Omega} f\, d\mu \text{ for all } f \in C_b(\Omega) \quad (\text{or for all } f \in C_c(\Omega)).$$

(ii) μ_n *converges weakly* (or *vaguely*) *in probability* to μ, if for every bounded continuous function $f \in C_b(\Omega)$ (or for every compactly supported continuous function $f \in C_c(\Omega)$),

$$\int_{\Omega} f\, d\mu_n \to \int_{\Omega} f\, d\mu \quad \text{in probability.}$$

(iii) μ_n *converges weakly* (or *vaguely*) *in expectation to* μ, if for every bounded continuous function $f \in C_b(\Omega)$ (or for every compactly supported continuous function $f \in C_c(\Omega)$),

$$\mathbb{E}\left[\int_{\Omega} f\, d\mu_n\right] \to \int_{\Omega} f\, d\mu.$$

Remark. (a) By definition, the weak almost sure convergence implies the other two modes of convergence in (ii) and (iii). In particular, the direction (i) $\Rightarrow$ (iii) holds by dominated convergence theorem.

(b) In fact, if μ is a (nonnegative) random measure on $\mathbb{R}$, we can define its expectation $\mathbb{E}\mu$ by

$$\langle \mathbb{E}\mu, f\rangle = \int_{\mathbb{R}} f\, d\mathbb{E}\mu = \mathbb{E}\left[\int_{\mathbb{R}} f\, d\mu\right], \quad f \in C_c(\mathbb{R}).$$

In that sense $\mathbb{E}\mu$ is a positive linear functional on $C_c(\mathbb{R})$. By Riesz-Markov-Kakutani theorem, the Borel measure $\mathbb{E}\mu$ on $\mathbb{R}$ satisfying the above property exists and is unique. According to this notation, the random measures μ_n converges weakly in expectation to μ means that $\mathbb{E}\mu_n \to \mu$ weakly.

(c) Although the weak convergence appears slightly stronger than the vague convergence, they are indeed equivalent for random measures on $\mathbb{R}$.

To see this, let μ_n be a sequence of measures converging vaguely to μ in probability. (The almost sure and in expectation cases can be handled similarly and deterministically.) Note that a probability measure μ on $\mathbb{R}$ is tight, i.e. for each $\epsilon > 0$ there exists compact interval $[-N, N]$ such that $\mu([-N, N]) \geq 1 - \epsilon$. Then given any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$ and any $\epsilon > 0$, we take $\phi \in C_c(\mathbb{R})$ such that

$$\int_{\mathbb{R}} \phi \, d\mu > 1 - \frac{\epsilon}{3\|f\|_\infty}.$$

By vague convergence in probability, there exists N_0 such that for all $n \geq N_0$,

$$\mathbb{P}\left(\int_{\mathbb{R}} \phi \, d\mu_n < 1 - \frac{\epsilon}{3\|f\|_\infty}\right) < \frac{\epsilon}{2}.$$

Also, since $f\phi \in C_c(\mathbb{R})$, there exists N_1 such that for all $n \geq N_1$,

$$\mathbb{P}\left(\left|\int_{\mathbb{R}} f\phi \, d\mu_n - \int_{\mathbb{R}} f\phi \, d\mu\right| \geq \frac{\epsilon}{3}\right) < \frac{\epsilon}{2}.$$

Then for all $n \geq \max\{N_0, N_1\}$, with probability at least $1 - \epsilon$, we have

$$\begin{aligned}
\left|\int_{\mathbb{R}} f \, d\mu_n - \int_{\mathbb{R}} f \, d\mu\right| &\leq \left|\int_{\mathbb{R}} f(1-\phi) \, d\mu_n\right| + \left|\int_{\mathbb{R}} f\phi \, d\mu_n - \int_{\mathbb{R}} f\phi \, d\mu\right| + \left|\int_{\mathbb{R}} f(1-\phi) \, d\mu\right| \\
&\leq \|f\|_\infty \left(\left|1 - \int_{\mathbb{R}} \phi \, d\mu_n\right| + \left|1 - \int_{\mathbb{R}} \phi \, d\mu\right|\right) + \left|\int_{\mathbb{R}} f\phi \, d\mu_n - \int_{\mathbb{R}} f\phi \, d\mu\right| \\
&\leq \|f\|_\infty \cdot \frac{2\epsilon}{3\|f\|_\infty} + \frac{\epsilon}{3} < \epsilon.
\end{aligned}$$

Therefore $\int_{\mathbb{R}} f \, d\mu_n \to \int_{\mathbb{R}} f \, d\mu$ in probability, and $\mu_n \to \mu$ weakly in probability. In our later discussion, we will not distinguish these two modes of convergence.

Theorem 1.2 (Portmanteau lemma, random version)**.** *Let Ω be a metric space equipped with its Borel σ-algebra $\mathscr{B}$. Let μ_n be a sequence of random probability measures on $(\Omega, \mathscr{B})$. The following are equivalent:*

(i) $\mu_n \to \mu$ *weakly in probability;*

(ii) $\int_\Omega f \, d\mu_n \to \int_\Omega f \, d\mu$ *in probability for every bounded Lipschitz continuous function* f*;*

(iii) for every lower semi-continuous function f *bounded from below,*

$$\liminf_{n\to\infty} \int_\Omega f \, \mathrm{d}\mu_n \geq \int_\Omega f \, d\mu \quad \textit{in probability;}$$

(iv) for every upper semi-continuous function f *bounded from above,*

$$\limsup_{n\to\infty} \int_\Omega f \, \mathrm{d}\mu_n \leq \int_\Omega f \, d\mu \quad \textit{in probability.}$$

Furthermore, if the above conditions hold, then

(v) $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ *in probability for every open set* G*;*

(vi) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *in probability for every closed set* F*.*

Proof. It is clear that (i) ⇒ (ii), (iii) ⇒ (iv), and (iii) + (iv) ⇒ (i). Also, (iii) ⇒ (v) since the indicator $\mathbb{1}_G$ for open G is lower semi-continuous. Similarly (iv) ⇒ (vi).

It remains to prove (ii) ⇒ (iii). We assume that $f : \mathbb{R} \to \mathbb{R}_+$ is a nonnegative, lower semi-continuous function. Take

$$f_k(x) = \min\left\{\inf_{y\in\mathbb{R}} \left(f(y) + kd(x,y)\right), k\right\}, \quad k = 1, 2, \cdots,$$

which are nonnegative bounded Lipschitz continuous functions such that $f_k \uparrow f$ pointwise. By the monotone convergence theorem, $\int_{\mathbb{R}} f_k\,d\mu \uparrow \int_{\mathbb{R}} f\,d\mu$ as $k \to \infty$. Then given any $\epsilon > 0$, we take $k \in \mathbb{N}$ such that $\int_{\mathbb{R}} f_k\,d\mu > \int_{\mathbb{R}} f\,d\mu - \epsilon/2$, and take N_0 large enough that

$$\mathbb{P}\left(\left|\int_{\mathbb{R}} f_k\,d\mu - \int_{\mathbb{R}} f_k\,d\mu\right| > \frac{\epsilon}{2}\right) < \epsilon$$

for all $n \geq N_0$. Then for all $n \geq N_0$, with probability at least $1-\epsilon$,

$$\int_{\mathbb{R}} f\,d\mu_n \geq \int_{\mathbb{R}} f_k\,d\mu_n \geq \int_{\mathbb{R}} f_k\,d\mu - \frac{\epsilon}{2} \geq \int_{\mathbb{R}} f\,d\mu - \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have $\liminf_{n\to\infty} \int_{\mathbb{R}} f\,d\mu_n \geq \int f\,d\mu$ in probability. □

1.1 The Wigner Random Matrices

Definition 1.3 (Wigner matrices)**.** Let $(\xi_{ij})_{1\leq i\leq j<\infty}$ be an upper triangular array of jointly independent random variables. Suppose that

(i) the diagonal entries $(\xi_{ii})_{i\geq 1}$ are real-valued i.i.d. random variables, and

(ii) the off-diagonal entries $(\xi_{ij})_{1\leq i<j}$ are real or complex-valued i.i.d. zero-mean random variables with variance $\mathbb{E}|\xi_{12}|^2 = 1$.

We can define the lower-triangular entries $(\xi_{ij})_{1\leq j<i}$ as following:

- If $(\xi_{ij})_{1\leq i<j}$ is real-valued, let $\xi_{ij} = \xi_{ji}$ for $1 \leq j < i$. Then $(\xi_{ij})_{i,j=1}^{\infty}$ is an infinite real symmetric matrix. The top-left $n\times n$ block $W_n = (\xi_{ij})_{i,j=1}^{n}$ is called a *real symmetric Wigner matrix.*
- If $(\xi_{ij})_{1\leq i<j}$ is complex-valued, let $\xi_{ij} = \overline{\xi}_{ji}$ for $1 \leq j < i$. Then $(\xi_{ij})_{i,j=1}^{\infty}$ is an infinite complex Hermitian matrix. The top-left $n\times n$ block $W_n = (\xi_{ij})_{i,j=1}^{n}$ is called a *complex Hermitian Wigner matrix.*

Example 1.4. Following are some examples of Wigner matrices:

(i) *Symmetric Bernoulli Ensemble.* All triangular entries $(\xi_{ij})_{1\leq i\leq j}$ are i.i.d. Rademacher variables, i.e. $\mathbb{P}(\xi_{ij}=1) = \mathbb{P}(\xi_{ij}=-1) = 1/2$.

(ii) *Gaussian Orthogonal Ensemble (GOE).* The diagonal entries $(\xi_{ii})_{i=1}^{\infty}$ are i.i.d. $\mathcal{N}_{\mathbb{R}}(0,2)$ variables, and the off-diagonal entries $(\xi_{ij})_{1\leq i<j}$ are i.i.d. $\mathcal{N}_{\mathbb{R}}(0,1)$ variables.

(iii) *Gaussian Unitary Ensemble (GUE).* The diagonal entries $(\xi_{ii})_{i=1}^{\infty}$ are i.i.d. $\mathcal{N}_{\mathbb{R}}(0,1)$ variables, and the off-diagonal entries $(\xi_{ij})_{1\leq i<j}$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ variables.

1.2 The Moment Method

We are interested in the asymptotic law of the ESD of Wigner matrices.

Theorem 1.5 (Wigner's semicircle law)**.** *Let W_n be an $n\times n$ complex Hermitian Wigner matrix, i.e. W_n is the topleft $n\times n$ block of the infinite matrix $(\xi_{ij})_{i,j=1}^{\infty}$. Then the ESD of $W_n/\sqrt{n}$ satisfies*

(i) $\mu_{\frac{W_n}{\sqrt{n}}} \to \mu_{\mathrm{sc}}$ *weakly in probability;*

(ii) $\mu_{\frac{W_n}{\sqrt{n}}} \to \mu_{\mathrm{sc}}$ *weakly almost surely;*

(iii) $\mathbb{E}\mu_{\frac{W_n}{\sqrt{n}}} \to \mu_{\mathrm{sc}}$ *weakly in expectation.*

where μ_{sc} is ***the semicircle distribution****, whose density function is given by*

$$\rho_{\mathrm{sc}}(x) = \frac{\mathbb{1}_{[-2,2]}(x)}{2\pi}\sqrt{4-x^2}, \quad x \in \mathbb{R}.$$

Moment. We can reformulate the spectral moments of $W_n/\sqrt{n}$ with the trace of its power:

$$M_{n,k} = \frac{1}{n}\sum_{j=1}^{n}\lambda_j^k = \frac{1}{n}\mathrm{tr}\left(\frac{W_n}{\sqrt{n}}\right)^k = \frac{1}{n^{1+k/2}}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n}\xi_{i_1i_2}\cdots\xi_{i_{k-1}i_k}\xi_{i_ki_1}.$$

To simplify our analysis, we additionally impose the following assumption, which will be removed later:

$$\max\left\{\mathbb{E}|\xi_{11}|^k, \mathbb{E}|\xi_{12}|^k\right\} < \infty, \quad \text{for all } k \in \mathbb{N}. \tag{1.1}$$

1.2.1 Combinatorics in Spectral Moments

Each ordered tuple $(i_1, i_2, \cdots, i_k) \in [n]^k$ is called a *cycle* of length k. We use the following terms:

- The m consecutive pairs $(i_1, i_2), \cdots, (i_{k-1}, i_k), (i_k, i_1)$ are called *steps* of the cycle.
- The distinct unordered pairs from $\{i_1, i_2\}, \cdots, \{i_{k-1}, i_k\}, \{i_k, i_1\}$ are called the *edges* of the cycle.
- The distinct indices among $i_1, i_2, \cdots, i_k$ are called the *vertices* of the cycle.
- The mixed moment $T(i_1, i_2, \cdots, i_k) := \mathbb{E}[\xi_{i_1i_2}\cdots\xi_{i_ki_1}]$ is called the *contribution* of the cycle to the spectral moment. In order that the contribution of cycle $(i_1, i_2, \cdots, i_k)$ is nonzero, we require that each edge $\{i_j, i_{j+1}\}$ should be traversed at least twice.
- We say that two cycles $(i_1, i_2, \cdots, i_k), (i'_1, i'_2, \cdots, i'_k) \in [n]^k$ are *equivalent*, if there exists a bijection $\pi : [n] \to [n]$ such that $i'_j = \pi(i_j)$ for every $1 \le j \le k$. In particular, for every cycle $(i_1, i_2, \cdots, i_k)$ we may relabel its each vertice v by the earliest time that v appears in the cycle:

  $$\pi(v) = \min\{j : i_j = v\}, \quad v \in \{i_1, \cdots, i_k\}.$$

  After relabeling, the cycle $(\pi(i_1), \pi(i_2), \cdots, \pi(i_k))$ is equivalent to $(i_1, i_2, \cdots, i_k)$. We call such a tuple the *shape* of $(i_1, i_2, \cdots, i_k)$. For instance, the shape of the cycle $(9, 5, 9, 4, 9, 3)$ is $(1, 2, 1, 3, 1, 4)$.
- Let $S_k \subset [k]^k$ be the set of all shapes of length k, i.e. tuples obeying the above rules. For every shape $s \in S_k$, we denote by $\mathcal{I}_n^s$ the set of all cycles $(i_1, i_2, \cdots, i_k) \in [n]^k$ that have shape s. Indeed, every shape $s \in S_k$ can be viewed as a representative of the equivalent classes $\mathcal{I}_n^s$ in $[n]^k$.
- The *height* $h(s)$ of a shape $s \in S_k$ is the number of distinct elements it has.

Using the above notations, the expected k-th spectral moment is

$$\mathbb{E}[M_{n,k}] = \frac{1}{n^{1+k/2}}\sum_{s\in S_k}\sum_{(i_1,\cdots,i_k)\in\mathcal{I}_n^s} T(i_1, \cdots, i_k).$$

Now we discuss the contribution of each shape $s \in S_k$ to the spectral moment $M_{n,k}$:

- If s is of height $h(s) > 1 + k/2$, there must exists an edge of s that is traversed only once. Then

  $$T(i_1, \cdots, i_k) = 0 \quad \text{for all } (i_1, \cdots, i_k) \in \mathcal{I}_n^s \text{ with } h(s) > 1 + k/2. \tag{1.2}$$

- If s is of height $h(s) < 1 + k/2$, then

  $$\frac{1}{n^{1+k/2}}\sum_{i_1,\cdots,i_k\in\mathcal{I}_n^s}|T(i_1, \cdots, i_k)| \le \frac{n(n-1)\cdots(n-h(s)+1)}{n^{1+k/2}}R_k \le n^{h(s)-1-\frac{k}{2}}R_k \le \frac{R_k}{\sqrt{n}},$$

  where $R_k := \max\{\mathbb{E}|\xi_{11}|^k, \mathbb{E}|\xi_{12}|^k\} < \infty$ for $k \in \mathbb{N}$. Since $|S_k| \le k^k$, we have

  $$\frac{1}{n^{1+k/2}}\sum_{s\in S_k:h(s)<1+\frac{k}{2}}\sum_{(i_1,\cdots,i_k)\in\mathcal{I}_n^s}|T(i_1, \cdots, i_k)| \le \frac{k^kR_k}{\sqrt{n}}. \tag{1.3}$$

- If k is even and s is of height $h(s) = 1 + k/2$, we focus on the cycles of shapes s that have nonzero weights, which requires that each edge of the cycle is traversed at least twice. Since a cycle with $k/2+1$ vertices has at least $k/2$ edges (one can easily prove this by induction), we are reduced to the cycles of shapes $s \in S_k$ for which each of its $k/2$ edge is traversed exactly twice, once in each direction.
  We denote by S_k^* the set of *all shapes of height* $k/2+1$ *that traverse each of its* $k/2$ *edges twice.* Then for all $s \in S_k^*$ and all cycles $(i_1, \cdots, i_k) \in \mathcal{I}_n^s$, we have exactly $k/2$ pairs of conjugate off-diagonal entries $(\xi_{i_j i_{j+1}}, \xi_{i_{j+1} i_j})$, and hence $T(i_1, \cdots, i_k) = 1$. Then

$$\frac{1}{n^{1+k/2}} \sum_{i_1, \cdots, i_k \in \mathcal{I}_n^s} T(i_1, \cdots, i_k) = \frac{n(n-1)\cdots(n-k/2)}{n^{1+k/2}} \leq 1.$$

We need to compute the number of elements of S_k^*.

Bijection to Dyck paths. We take a cycle $(i_1, \cdots, i_k)$ with $k/2+1$ vertices that traverses each of its $k/2$ edges twice. We imagine traversing a cycle from i_1 to i_2, then from i_2 to i_3, and so forth until we finally return to i_1 from i_k. At each step of this journey, 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 say that the step (i_j, i_{j+1}) is *innovative* if it is in the first category, and *returning* otherwise. Clearly, only the innovative steps can bring us to vertices that we have not seen before. Since we have to visit $k/2+1$ distinct vertices (including the vertex i_1 we start at), we conclude that each innovative leg must take us to a new vertex. We thus recover the shape of the cycle from a sequence of these steps by starting from 1, and

- if the current step is innovative, create a new edge and add a new vertex not visited before.
- if the current step is returning, it must close an edge that was already opened earlier; hence, return along the corresponding previously created edge to an already-seen vertex.

Formally, we can associate a shape $(s_1, \ldots, s_k) \in S_k^*$ with a path of the *simple random walk* on $\mathbb{Z}$ by mapping each step (s_j, s_{j+1}) as follows: if the step is *innovative*, we assign it an increment $+1$; if the step is *returning*, we assign it a decrement -1. Since an edge cannot be revisited before it is first discovered, every prefix of the traversal contains at least as many innovative steps as returning steps. Hence the partial sums of the walk are nonnegative. Moreover, because the number of innovative and returning steps equal ($k/2$ of each), the walk starts at 0, ends at 0, and has total length k. Therefore, the traversal of any shape encodes uniquely a *Dyck path* of length k, i.e. a simple random walk on $\mathbb{Z}$ with steps ± 1 that begins at 0, never goes below 0 and returns to 0 at the end. Following are some examples.

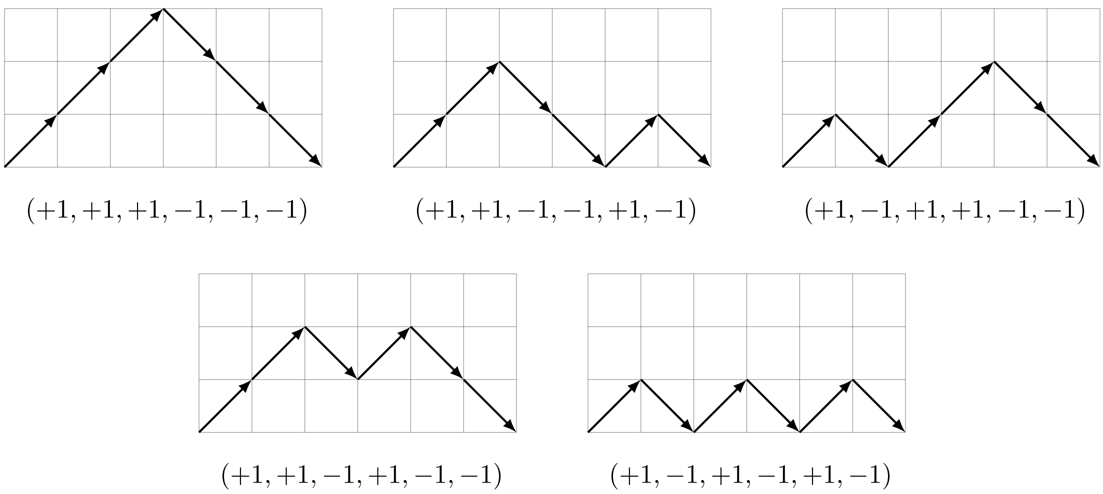


Figure 1: There are 5 Dyck paths with 6 steps

Lemma 1.6. *Let $k \in 2\mathbb{N}$. Then $|S_k^*| = C_{k/2}$, where C_m is the m-th* ***Catalan number****:*

$$C_m = \frac{1}{m+1}\binom{2m}{m}, \quad m \in \mathbb{N}.$$

Proof. Let $k = 2m$ and encode the traversal of a shape $(s_1, \ldots, s_{2m})$ as a walk $(x_1, \cdots, x_{2m})$ with steps $+1$ (innovative) and -1 (returning), as described above. This yields a walk of length $2m$ with exactly m up-steps and m down-steps and never going below 0.

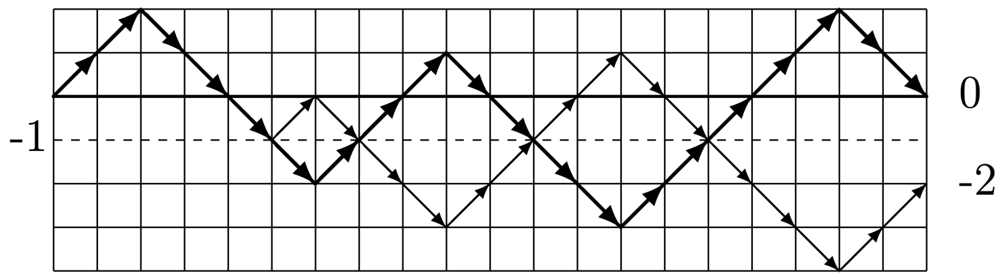
Let $\mathcal{P}$ be the set of all walks from $(0,0)$ to $(2m, 0)$, so $|\mathcal{P}| = \binom{2m}{m}$. Then let $\mathcal{D} \subset \mathcal{P}$ be the subset that never goes below 0 (Dyck paths), and $\mathcal{B} := \mathcal{P} \setminus \mathcal{D}$ the set of "bad" walks that hit -1 at least once. For any walk $x = (x_1, \cdots, x_{2m}) \in \mathcal{B}$, let τ be the first time its partial sum equals -1:

$$\tau = \min\left\{ t \in [2m] : \sum_{j=1}^{t} x_j = -1 \right\}.$$

Reflect the path after time τ across the level -1:

$$Rx = (x_1, \cdots, x_\tau, -x_{\tau+1}, \cdots, -x_{2m}).$$

A graphical illustration is given below:



This *reflection map* sends x to a walk Rx of length $2m$ that still begins at 0 but ends at -2. Conversely, given any walk $y = (y_1, \cdots, y_{2m})$ from $(0,0)$ to $(2m, -2)$, there is a unique inverse operation: reflect the path after its first visit to -1 to obtain a bad walk. Hence the reflection map is a bijection, and the number of walks in $\mathcal{B}$ is the number of walks beginning at 0 and ending at -2, namely $\binom{2m}{m-1}$. Therefore,

$$|\mathcal{D}| = |\mathcal{P}| - |\mathcal{B}| = \binom{2m}{m} - \binom{2m}{m-1} = \frac{1}{m+1}\binom{2m}{m} = C_m.$$

Since the shape-to-walk encoding is a bijection onto Dyck paths, the number of shapes in S_{2m}^* is the m-th Catalan number C_m. □

Recurrence relation and the generating function. We consider the set $\mathcal{D}_m$ of Dyck paths on $[0, 2m]$. For $1 \le k \le m$, we let $\mathcal{D}_m^{(k)}$ be the set of Dyck paths x such that $2k$ is the first time it returns to 0. Then we have two types of paths:

- *Irreducible paths.* Every path in $\in \mathcal{D}_m^{(m)}$ never returns to 0 before step $2m$. That is, it never goes below 1 on $[1, 2m-1]$. Since such a Dyck path always stay at 1 at time 1 and $2m-1$, we can identify it with a Dyck path from $(1,1)$ to $(2m-1, 1)$. Thus we have $|\mathcal{D}_m^{(m)}| = |\mathcal{D}_{m-1}| = C_{m-1}$.
- *Reducible paths.* For $1 \le k < n$, each path in $\mathcal{D}^{(k)}$ can be decomposed into an irreducible path on $[0, 2k]$ and a Dyck path from $(2k, 0)$ to $(2m, 0)$. Thus we have $|\mathcal{D}_m^{(k)}| = |\mathcal{D}_{k-1}| \cdot |\mathcal{D}_{m-k}| = C_{k-1}C_{m-k}$.

Note that $\mathcal{D}_m = \bigcup_{k=1}^m \mathcal{D}_m^{(k)}$. Then we have the following recurrence relation for Catalan numbers:

$$C_0 = C_1 = 1, \qquad C_m = \sum_{k=1}^m C_{k-1}C_{m-k}, \quad m \geq 2.$$

This is known as *Segner's recurrence relation for Catalan numbers.* The *generating function* for the Catalan numbers is defined by the power series

$$C(z) = \sum_{n=0}^\infty C_n z^n.$$

The recurrence relation given above can then be summarized in generating function form by the relation

$$C(z) = 1 + zC(z)^2.$$

Note that $C_0 = \lim_{z\to 0} C(z) = 1$, we select the branch

$$C(z) = \frac{1-\sqrt{1-4z}}{2z} = \sum_{n=0}^\infty \frac{1}{n+1}\binom{2n}{n} z^n.$$

In fact, we can also use this procedure as an alternative proof of Lemma 1.6.

Now we can compute the expectation of spectral moments of Wigner matrices.

Lemma 1.7 (Estimate for spectral moment)**.** *For each $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} \mathbb{E}\left[M_{n,k}\right] = \begin{cases} 0, & k \textit{ is odd,} \\ C_{k/2}, & k \textit{ is even.} \end{cases}$$

Proof. For k odd, the result follows directly from (1.2) and (1.3). For k even, by (1.2), (1.3) and Lemma 1.6,

$$\begin{aligned} \mathbb{E}\left[M_{n,k}\right] &= \frac{1}{n^{1+k/2}} \sum_{s\in S_k^*} \sum_{(i_1,\cdots,i_k)\in\mathcal{I}_n^s} T(i_1,\cdots,i_k) + \frac{1}{n^{1+k/2}} \sum_{s\in S_k : h(s)<1+\frac{k}{2}} \sum_{(i_1,\cdots,i_k)\in\mathcal{I}_n^s} T(i_1,\cdots,i_k) \\ &= C_{k/2}\frac{n(n-1)\cdots(n-k/2)}{n^{1+k/2}} + O(n^{-1/2}). \end{aligned}$$

Letting $n \to \infty$ conclude the proof. □

Lemma 1.8. *For every $k \in \mathbb{N}$, the k-th moment of the Wigner semicircle distribution μ_{sc} is*

$$\int_{\mathbb{R}} x^k \, d\mu_{\mathrm{sc}}(x) = \begin{cases} 0, & \textit{for } k \textit{ odd} \\ C_{k/2}, & \textit{for } k \textit{ even.} \end{cases}$$

Proof. The case for odd k follows easily from symmetry. For even k, we assume $k = 2m$. Then

$$\begin{aligned} \int_{\mathbb{R}} x^{2m}\, d\mu_{\mathrm{sc}}(x) &= \frac{1}{\pi}\int_0^2 x^{2m}\sqrt{4-x^2}\,dx = \frac{2^{2m+2}}{\pi}\int_0^{\pi/2} \sin^{2m}\theta\cos^2\theta\,d\theta \\ &= \frac{2^{2m+1}}{\pi}\left[\int_0^\pi \sin^{2m}\theta\,d\theta - \int_0^\pi \sin^{2m+2}\theta\,d\theta\right] \end{aligned}$$

We assume $I_{2m} = \int_0^\pi \sin^{2m}\theta \, d\theta$ for $m = 0, 1, 2, \cdots$. Then

$$\begin{aligned} I_{2m+2} &= \int_0^\pi \sin^{2m+2}\theta \, d\theta = \int_0^\pi \sin^{2m}\theta(1-\cos^2\theta)\, d\theta = I_{2m} - \int_0^\pi \sin^{2m}\theta \cos\theta \, d\sin\theta \\ &= I_{2m} + \int_0^\pi \sin\theta \, d(\sin^{2m}\theta\cos\theta) = I_{2m} + \int_0^\pi \sin\theta(2m\sin^{2m-1}\theta\cos^2\theta - \sin^{2m+1}\theta)\, d\theta \\ &= I_{2m} + 2m(I_{2m} - I_{2m+2}) - I_{2m+2} = (2m+1)(I_{2m} - I_{2m+2}). \end{aligned}$$

Then we have

$$I_{2m+2} = \frac{2m+1}{2m+2} I_{2m}.$$

Since $I_0 = \pi$, by induction, we have

$$I_{2m} = \frac{(2m-1)!!}{(2m)!!} I_0 = \frac{(2m)!}{2^{2m}(m!)^2}\pi, \quad m = 0, 1, 2, \cdots.$$

Then

$$\int_{\mathbb{R}} x^{2m} \, d\mu_{\mathrm{sc}}(x) = 2^{2m+1} \frac{(2m)!}{2^{2m}(m!)^2}\left(1 - \frac{(2m+1)(2m+2)}{4(m+1)^2}\right) = \frac{1}{m+1}\frac{(2m)!}{(m!)^2} = C_m.$$

Thus we complete the proof. □

Remark. In fact, we proved that

$$\int_{\mathbb{R}} x^k \, d\mathbb{E}\mu_{\frac{W_n}{\sqrt{n}}}(x) \to \int_{\mathbb{R}} x^k \, d\mu_{\mathrm{sc}}. \tag{1.4}$$

Note that this convergence is deterministic because $\mathbb{E}\mu_{\frac{W_n}{\sqrt{n}}}$ is a deterministic measure.

Lemma 1.9. *Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of $W_n/\sqrt{n}$. For every $k \in \mathbb{N}$, we have*

$$\frac{1}{n}\sum_{j=1}^n |\lambda_j|^k \mathbb{1}_{\{|\lambda_j|>5\}} \to 0 \quad \text{in probability as } n \to \infty.$$

Proof. Let $\epsilon > 0$. For every $m \in \mathbb{N}$, we have

$$\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n |\lambda_j|^m \mathbb{1}_{\{|\lambda_j|>5\}}\right] \le 5^{-m}\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n |\lambda_j|^{2m}\right].$$

By Lemma 1.7 and Markov's inequality,

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^n |\lambda_j|^m \mathbb{1}_{\{|\lambda_j|>5\}} > \epsilon\right) \le \frac{1}{\epsilon 5^m}\lim_{n\to\infty}\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n |\lambda_j|^{2m}\right] = \frac{5^{-m}(2m)!}{\epsilon(1+m)(m!)^2} < \frac{1}{\epsilon}\left(\frac{4}{5}\right)^m.$$

This inequality holds for every $m \in \mathbb{N}$. While the right-hand side decreases to 0 as $m \to \infty$, the left-hand side is increasing in m. Therefore

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^n |\lambda_j|^k \mathbb{1}_{\{|\lambda_j|>5\}} > \epsilon\right) = 0$$

for every $k \in \mathbb{N}$. □

1.2.2 The Variance of Spectral Moments

Lemma 1.10. *For every $k, n \in \mathbb{N}$,*

$$\operatorname{Var}(M_{n,k}) \le \frac{2^{2k+1} k^{2k} R_{2k}}{n^2}, \tag{1.5}$$

where $R_{2k} = \max\left\{\mathbb{E}|\xi_{11}|^{2k}, \mathbb{E}|\xi_{12}|^{2k}\right\}$. Hence for every polynomial $P(x) = c_0 + c_1 x + \cdots + c_m x^m$, we have

$$\int_{\mathbb{R}} P(x)\, d\mu_{\frac{W_n}{\sqrt{n}}}(x) \to \int_{\mathbb{R}} P(x)\, d\mu_{\mathrm{sc}}(x) \quad \textit{in probability.}$$

Proof. For a cycle $\mathbf{i} = (i_1, \cdots, i_k) \in [n]^k$, we write $\xi_{\mathbf{i}} = \xi_{i_1}\xi_{i_2}\cdots\xi_{i_k}$ for simplicity. Then

$$\operatorname{Var}(M_{n,k}) = \frac{1}{n^{2+k}} \sum_{\mathbf{i},\mathbf{i}' \in [n]^k} \left(\mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{i}'}\right] - \mathbb{E}\xi_{\mathbf{i}}\mathbb{E}\xi_{\mathbf{i}'}\right).$$

Each pair $(\mathbf{i}, \mathbf{i}')$ generates a graph with vertices $\mathcal{V}(\mathbf{i}, \mathbf{i}') = \{i_1, \cdots, i_k\} \cup \{i'_1, \cdots, i'_k\}$ and undirected edges $\mathcal{E}(\mathbf{i}, \mathbf{i}') = \{i_1 i_2, \cdots, i_{k-1} i_k, i_k i_1\} \cup \{i'_1 i'_2, \cdots, i'_{k-1} i'_k, i'_k i'_1\}$. The resulting graph has at most two connected components. As before, two pairs $(\mathbf{i}, \mathbf{i}')$ and $(\mathbf{j}, \mathbf{j}')$ are said to be *equivalent* if there is a bijection on $[n]$ mapping corresponding indices to each other.

In order the contribution $\mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{i}'}\right] - \mathbb{E}\xi_{\mathbf{i}}\mathbb{E}\xi_{\mathbf{i}'}$ of $(\mathbf{i}, \mathbf{i}')$ to be nonzero, the following conditions are necessary:

(a) Each edge in $E(\mathbf{i}, \mathbf{i}')$ is traversed at least twice. As a result, there are at most k edges in the graph.

(b) The two graphs generated by cycles $\mathbf{i}$ and $\mathbf{i}'$ have at least one shared edge, otherwise by independence we have $\mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{i}'}\right] - \mathbb{E}\xi_{\mathbf{i}}\mathbb{E}\xi_{\mathbf{i}'} = 0$. As a result, the graph generated by $(\mathbf{i}, \mathbf{i}')$ is connected.

We discuss the contribution of a pair in three cases:

- If $\mathcal{V}(\mathbf{i}, \mathbf{i}')$ has cardinality $h \ge k+2$ and $(\mathbf{i}, \mathbf{i}')$ has nonzero contribution, the resulting graph is connected and should have at least $h - 1 > k$ edges, which contradicts (a). Therefore $(\mathbf{i}, \mathbf{i}')$ has zero contribution.
- If $\mathcal{V}(\mathbf{i}, \mathbf{i}')$ has cardinality $h = k+1$ and $(\mathbf{i}, \mathbf{i}')$ has nonzero contribution, the resulting graph is connected and should have k edges. In this case, there are no cycle in the graph, and each edge would be traversed exactly twice, once in each direction. Since $\mathbf{i}$ begins and ends at i_1, it must traverse each edge an even number of times. The same is true for $\mathbf{i}'$. Thus, each edge in $\mathcal{E}(\mathbf{i}, \mathbf{i}')$ is traversed by either $\mathbf{i}$ or $\mathbf{i}'$, but not both. Then $\mathbf{i}$ and $\mathbf{i}'$ generate distinct edges, a contradiction! Therefore $(\mathbf{i}, \mathbf{i}')$ has zero contribution.
- If $\mathcal{V}(\mathbf{i}, \mathbf{i}')$ has cardinality $h \le k$, there are $n(n-1)\cdots(n-h+1) \le n^k$ equivalent pairs. The contribution of these pairs satisfies
$$\left|\mathbb{E}\left[\xi_{\mathbf{i}}\xi_{\mathbf{i}'}\right] - \mathbb{E}\xi_{\mathbf{i}}\mathbb{E}\xi_{\mathbf{i}'}\right| \le 2R_{2k}.$$

As before, there are no more than $(2k)^{2k}$ distinct equivalent classes of pairs $(\mathbf{i}, \mathbf{i}')$. We summarize the above three cases to obtain

$$\operatorname{Var}(M_{n,k}) = \frac{1}{n^{2+k}} \cdot (2k)^{2k} \cdot n^k \cdot 2R_{2k} = \frac{2^{2k+1} k^{2k} R_{2k}}{n^2}.$$

By Chebyshev's inequality, for any $\epsilon > 0$,

$$\mathbb{P}\left(|M_{n,k} - \mathbb{E}[M_{n,k}]| > \epsilon\right) \le \frac{1}{\epsilon^2}\left(\mathbb{E}[M_{n,k}^2] - (\mathbb{E}[M_{n,k}])^2\right) \le \frac{2^{2k+1} k^{2k} R_{2k}}{\epsilon^2 n^2},$$

which converges to 0 as $n \to \infty$. Combining the above result with Lemma 1.6 and Lemma 1.7, we have

$$M_{n,k} = \int_{\mathbb{R}} x^k\, d\mu_{\frac{W_n}{\sqrt{n}}} \to \int_{\mathbb{R}} x^k\, d\mu_{\mathrm{sc}}(x) \quad \text{in probability.}$$

Since $k \in \mathbb{N}$ is arbitrary, the convergence result holds true for any polynomial. □

Proof of Theorem 1.5 (i). Let $f \in C_c(\mathbb{R})$, and take $N \ge 5$ such that $\operatorname{supp}(f) \subset [-N, N]$. We then apply

Stone-Weierstrass theorem to approximate f on the compact interval $[-N,N]$ by polynomials: for each $\epsilon > 0$, there exists a polynomial P_ϵ on $\mathbb{R}$ such that $|P_\epsilon(x) - f(x)| < \epsilon/4$ for all $x \in [-N,N]$. Note that μ_{sc} is supported in $[-2,2] \subset [-N,N]$. Then

$$\begin{aligned}\left|\int_{\mathbb{R}} f\, d\mu_{\frac{W_n}{\sqrt{n}}} - \int_{\mathbb{R}} f\, d\mu_{\mathrm{sc}}\right| &\le \left|\int_{\mathbb{R}} (f - P_\epsilon)\, d\mu_{\frac{W_n}{\sqrt{n}}} - \int_{\mathbb{R}} (f - P_\epsilon)\, d\mu_{\mathrm{sc}}\right| + \left|\int_{\mathbb{R}} P_\epsilon\, d\mu_{\frac{W_n}{\sqrt{n}}} - \int_{\mathbb{R}} P_\epsilon\, d\mu_{\mathrm{sc}}\right| \\ &\le \left|\int_{\mathbb{R}\setminus[-N,N]} (f - P_\epsilon)\, d\mu_{\frac{W_n}{\sqrt{n}}}\right| + \left|\int_{[-N,N]} (f - P_\epsilon)\, d\mu_{\frac{W_n}{\sqrt{n}}}\right| + \left|\int_{[-N,N]} (f - P_\epsilon)\, d\mu_{\mathrm{sc}}\right| \\ &\quad + \left|\int_{\mathbb{R}} P_\epsilon\, d\mu_{\frac{W_n}{\sqrt{n}}} - \int_{\mathbb{R}} P_\epsilon\, d\mu_{\mathrm{sc}}\right| \\ &\le \left|\int_{\mathbb{R}\setminus[-N,N]} P_\epsilon\, d\mu_{\frac{W_n}{\sqrt{n}}}\right| + \frac{\epsilon}{2} + \left|\int_{\mathbb{R}} P_\epsilon\, d\mu_{\frac{W_n}{\sqrt{n}}} - \int_{\mathbb{R}} P_\epsilon\, d\mu_{\mathrm{sc}}\right| \end{aligned} \tag{1.6}$$

For the first term, note that

$$\left|\int_{\mathbb{R}\setminus[-N,N]} P_\epsilon\, d\mu_{\frac{W_n}{\sqrt{n}}}\right| \le \int_{\mathbb{R}\setminus[-N,N]} |P_\epsilon|\, d\mu_{\frac{W_n}{\sqrt{n}}} \le \int_{\mathbb{R}\setminus[-5,5]} |P_\epsilon|\, d\mu_{\frac{W_n}{\sqrt{n}}} = \frac{1}{n}\sum_{j=1}^{n} |P_\epsilon(\lambda_j)| \mathbb{1}_{\{|\lambda_j| \ge 5\}},$$

which converges to 0 in probability as $n \to \infty$, by Lemma 1.9. Meanwhile, we can also control the last term in (1.6) by Lemma 1.10. Since $\epsilon > 0$ is arbitrary, we have

$$\int_{\mathbb{R}} f\, d\mu_{\frac{W_n}{\sqrt{n}}} \to \int_{\mathbb{R}} f\, d\mu_{\mathrm{sc}} \quad \text{in probability.}$$

Hence $\mu_{\frac{W_n}{\sqrt{n}}} \to \mu_{\mathrm{sc}}$ weakly in probability, and we complete the proof. □

Using the variance bound, we can indeed extend the convergence result to the almost sure case.

Proof of Theorem 1.5 (ii). By Chebyshev's inequality, for every $k \in \mathbb{N}$ and $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(|M_{n,k} - \mathbb{E}[M_{n,k}]| > \epsilon\right) \le \sum_{n=1}^{\infty} \frac{\mathrm{Var}(M_{n,k})}{\epsilon^2} = 2^{2k+1} k^{2k} R_{2k} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

By the Borel-Cantelli lemma, we have

$$\mathbb{P}\left(\limsup_{n\to\infty} |M_{n,k} - \mathbb{E}[M_{n,k}]| > \epsilon\right) = 0.$$

Since $\epsilon > 0$ is arbitrary, we have $M_{n,k} - \mathbb{E}[M_{n,k}] \to 0$ almost surely as $n \to \infty$. Combining this result with (1.4), we obtain that

$$\int_{\mathbb{R}} x^k\, d\mu_{\frac{W_n}{\sqrt{n}}} \to \int_{\mathbb{R}} x^k\, d\mu_{\mathrm{sc}} \quad \text{almost surely.}$$

Using the estimate $C_k \le 4^k$, we have

$$\limsup_{k\to\infty} \frac{1}{2k}\left(\int_{\mathbb{R}} x^{2k}\, d\mu_{\mathrm{sc}}\right)^{1/2k} = \limsup_{k\to\infty} \frac{1}{2k} C_k^{1/2k} \le \limsup_{k\to\infty} \frac{1}{k} = 0 < \infty.$$

By Carleman's continuity theorem, we have $\mu_{\frac{W_n}{\sqrt{n}}} \to \mu_{\mathrm{sc}}$ weakly with probability 1. □

Remark. (i) Following the same approach as the above proof, it is easy conclude that $\mu_{\frac{W_n}{\sqrt{n}}} \to \mu_{\mathrm{sc}}$ weakly in expectation, which is Theorem 1.5 (iii).

(ii) For every $-\infty \le a < b \le \infty$, by the Portmanteau lemma,

$$\frac{1}{n} N_{\frac{W_n}{\sqrt{n}}}([a, b]) \to \int_a^b \rho_{\mathrm{sc}}(x)\, dx \quad \text{almost surely,}$$

where $N_{\frac{W_n}{\sqrt{n}}}([a, b])$ is the number of eigenvalues of $W_n/\sqrt{n}$ lying in $[a, b]$, including repetitions according to algebraic multiplicity.

1.2.3 Remove the Finiteness Assumption for Higher-Order Moments

Now we will remove the assumption (1.1) and extend the convergence results to general Wigner matrices. We will use the Lévy metric between cumulative distribution functions to establish weak convergence.

Lemma 1.11. *Assume that $A, B \in \mathbb{C}^{n\times n}$ are Hermitian matrices. Then*

(i) For every $\alpha > 0$,

$$\rho_L(F_A, F_B)^{1+\alpha} \le \frac{1}{n}\sum_{j=1}^{n} |\lambda_j(A) - \lambda_j(B)|^\alpha, \tag{1.7}$$

where ρ_L is the Lévy metric between two cumulative distribution function (c.d.f.s):

$$\rho_L(F, G) = \inf\{\epsilon > 0 : F(x-\epsilon) - \epsilon \le G(x) \le F(x+\epsilon) + \epsilon \text{ for all } x \in \mathbb{R}\},$$

and F_A, F_B are c.d.f.s of ESDs μ_A and μ_B, respectively.

(ii) In particular,

$$\rho_L(F_A, F_B)^3 \le \frac{1}{n}\|A - B\|_{\mathrm{F}}^2.$$

Remark. Recall that the weak topology on the space of Borel probability measures on $\mathbb{R}$ is metrized by the Lévy metric ρ_L. That is, $F_n \to F$ weakly if and only if $\rho_L(F_n, F) \to 0$.

Proof. (i) Fix $\alpha > 0$. The inequality (1.7) is trivial if $\frac{1}{n}\sum_{j=1}^{n} |\lambda_j(A) - \lambda_j(B)|^\alpha \ge 1$. Then without probability, we can take $\epsilon \in (0, 1)$ such that

$$\frac{1}{n}\sum_{j=1}^{n} |\lambda_j(A) - \lambda_j(B)|^\alpha < \epsilon^{1+\alpha} < 1.$$

Since ϵ is arbitrary, it suffices to prove $\rho_L(F_A, F_B) \le \epsilon$. For each $x \in \mathbb{R}$, let $\mathcal{A}_x = \{j \in [n] : \lambda_j(A) \le x\}$ and $\mathcal{B}_x = \{j \in [n] : \lambda_j(B) \le x + \epsilon\}$. Then for every $j \in \mathcal{A}_x \backslash \mathcal{B}_x$, we have $|\lambda_j(A) - \lambda_j(B)| \ge \epsilon$, and

$$F_A(x) - F_B(x+\epsilon) \le \frac{|\mathcal{A}_x \backslash \mathcal{B}_x|}{n} \le \frac{1}{n\epsilon^\alpha}\sum_{j=1}^{n} |\lambda_j(A) - \lambda_j(B)|^\alpha < \epsilon.$$

Similarly $F_B(x-\epsilon) - F_A(x) \le \epsilon$. Hence $\rho_L(F_A, F_B) \le \epsilon$.

(ii) Let $\alpha = 2$ in (i), and apply Hoffman-Wielandt inequality. □

We next show the stability of ESD under low-rank perturbations.

Lemma 1.12 (Low rank perturbation)**.** *Assume that $A, B \in \mathbb{C}^{n\times n}$ are Hermitian matrices. Then*

$$\rho_L(F_A, F_B) \le \|F_A - F_B\|_\infty \le \frac{\operatorname{rank}(A - B)}{n}, \tag{1.8}$$

where F_A, F_B are cumulative distribution functions of ESDs μ_A and μ_B, respectively.

Proof. Let $r = \operatorname{rank}(A - B)$. Since both sides (1.8) are invariant under a common unitary transformation on A and B, we may transform $A - B$ as $\begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix}$, where $\Sigma \in \mathbb{C}^{r\times r}$ is full-rank. Hence we may assume

$$A = \begin{bmatrix}A_{11} & B_{12}\\ B_{12}^* & B_{12}\end{bmatrix}, \quad B = \begin{bmatrix}B_{11} & B_{12}\\ B_{12}^* & B_{12}\end{bmatrix}.$$

By the Cauchy interlacing theorem,

$$\min\{\lambda_j(A), \lambda_j(B)\} \ge \lambda_j(B_{22}) \ge \max\{\lambda_{j+r}(A), \lambda_{j+r}(B)\}, \quad j = 1, \cdots, n-r.$$

Let $\lambda_0(A) = \infty$ and $\lambda_{n-r+1}(B) = -\infty$. For any $x \in \mathbb{R}$ and choose j with $(\lambda_j(B_{22}), \lambda_{j-1}(B_{22})]$. Then

$$\max\{\lambda_{j+r}(A), \lambda_{j+r}(B)\} \le \lambda_j(B_{22}) \le x \quad \text{and} \quad \min\{\lambda_{j-1}(A), \lambda_{j-1}(B)\} \ge \lambda_{j-1}(B_{22}) > x.$$

Hence

$$1 - \frac{j+r-1}{n} \le \min\{F_A(x), F_B(x)\} \le \max\{F_A(x), F_B(x)\} \le 1 - \frac{j-1}{n},$$

and $|F_A(x) - F_B(x)| \le r/n$ for all $x \in \mathbb{R}$. Finally, if $0 < \epsilon < \rho_L(F_A, F_B)$, we can find $x \in \mathbb{R}$ such that

$$F_A(x-\epsilon) - \epsilon > F_B(x) \quad \text{or} \quad F_A(x+\epsilon) + \epsilon < F_B(x).$$

Then $F_A(x) - F_B(x) \ge F_A(x-\epsilon) - F_B(x) > \epsilon$ for the first case, or $F_B(x) - F_A(x) \ge F_B(x) - F_A(x+\epsilon) > \epsilon$ for the second case. Hence $\epsilon < |F_A(x) - F_B(x)| \le r/n$, and $\rho_L(F_A, F_B) \le r/n$. □

Now we show how to remove the diagonal elements in a Wigner matrix.

Lemma 1.13 (Removing the diagonal)**.** *Let (W_n) be Wigner matrices given in Theorem 1.5. We obtain $\widehat{W}_n$ from W_n by replacing all diagonal entries with 0. Then*

$$\lim_{n\to\infty} \rho\left(F_{\frac{W_n}{\sqrt{n}}}, F_{\frac{\widehat{W}_n}{\sqrt{n}}}\right) = 0.$$

Proof. We truncate the diagonal entries (ξ_{ii}) at $\sqrt{n}$, and let $\Xi_n = \mathrm{diag}\big(\xi_{ii}\mathbb{1}_{\{|\xi_{ii}|\le\sqrt{n}\}}\big)_{i=1}^n$. By Lemma 1.11,

$$\rho_L\left(F_{\frac{W_n-\Xi_n}{\sqrt{n}}}, F_{\frac{\widehat{W}_n}{\sqrt{n}}}\right)^3 \le \frac{1}{n}\left\|\frac{W_n-\Xi_n}{\sqrt{n}} - \frac{\widehat{W}_n}{\sqrt{n}}\right\|_{\mathrm{F}}^2 \le \frac{1}{n^2}\sum_{i=1}^n |\xi_{ii}|^2 \mathbb{1}_{\{|\xi_{ii}|\le\sqrt{n}\}} \le \frac{1}{n}. \tag{1.9}$$

Let $N_n = |\{i \in [n] : \xi_{ii} > \sqrt{n}\}| = \sum_{i=1}^n \mathbb{1}_{\{|\xi_{ii}|>\sqrt{n}\}}$. By Lemma 1.12,

$$\rho_L\left(F_{\frac{W_n}{\sqrt{n}}}, F_{\frac{W_n-\Xi_n}{\sqrt{n}}}\right) \le \frac{\mathrm{rank}(\Xi_n)}{n} = \frac{N_n}{n}. \tag{1.10}$$

Let $p_n = \mathbb{P}(|\xi_{11}| > \sqrt{n}) \to 0$. By Bernstein's inequality, for any $\epsilon > 0$ and sufficiently large n,

$$\mathbb{P}(N_n \ge n\epsilon) = \mathbb{P}\left(\sum_{i=1}^n \left(\mathbb{1}_{\{|\xi_{ii}|>\sqrt{n}\}} - p_n\right) \ge n(\epsilon - p_n)\right) \le \exp\left(-\frac{n^2(\epsilon-p_n)^2/2}{np_n(1-p_n) + \frac{n(\epsilon-p_n)}{3}}\right) \le e^{-\frac{n}{2}(\epsilon-p_n)^2}.$$

By the Borel-Cantelli lemma,

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{N_n}{n} \ge \epsilon\right) = \mathbb{P}\left(\bigcap_{K=1}^\infty\bigcup_{n=K}^\infty\left\{\frac{N_n}{n} \ge \epsilon\right\}\right) \le \lim_{K\to\infty}\sum_{n=K}^\infty e^{-n\epsilon^2/8} = \lim_{K\to\infty}\frac{e^{-K\epsilon^2/8}}{1-e^{-\epsilon^2/8}} = 0.$$

Since $\epsilon > 0$ is arbitrary, $N_n/n \to 0$ almost surely. Combining (1.9) and (1.10), we conclude the proof. □

Finally we present the main result.

Lemma 1.14. *For the Wigner matrices (W_n) in Theorem 1.5, one may assume without loss of generality that the diagonal entries $(\xi_{ii})_{i\ge1}$ are 0 and that the off-diagonal entries $(\xi_{ij})_{1\le i<j}$ are bounded.*

Remark. If we can prove Theorems 1.5 for Wigner matrices with vanished diagonal entries and bounded off-diagonal entries, we may extend the result to a general Wigner matrix W_n using this Lemma.

Proof. We define $(\widetilde{\xi}_{ij})_{1\le i\le j}$ by removing diagonal entries and normalizing truncated off-diagonal entries:

$$\widetilde{\xi}_{ij} = \begin{cases} 0 & \text{if } i=j, \\ \dfrac{\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}} - \mathbb{E}[\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}}]}{\sqrt{\operatorname{Var}(\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}})}} & \text{if } i<j, \end{cases}$$

where N is to be chosen, and let $\widetilde{W}_n = (\widetilde{\xi}_{ij})_{i,j=1}^n/\sqrt{n}$ be the corresponding Wigner matrix for every $n\in\mathbb{N}$. Then

$$\begin{aligned} \frac{1}{n}\operatorname{tr}\left(\widehat{W}_n - \widetilde{W}_n\right)^2 &= \frac{1}{n^2}\sum_{i,j=1}^n |\xi_{ij}-\widetilde{\xi}_{ij}|^2 \\ &\le \frac{2}{n^2}\sum_{1\le i<j\le n}\left|\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|>N\}} - \mathbb{E}[\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|>N\}}]\right|^2 \\ &\quad + \frac{2}{n^2}\sum_{1\le i<j\le n}\left(\frac{1}{\operatorname{Var}(\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}})}-1\right)\left|\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}} - \mathbb{E}[\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}}]\right|^2. \end{aligned} \tag{1.11}$$

Since $(\xi_{ij})_{1\le i<j}$ are i.i.d. and have finite second moments, by the strong law of large numbers,

$$\frac{2}{n(n-1)}\sum_{1\le i<j\le n}\left|\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|>N\}} - \mathbb{E}[\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|>N\}}]\right|^2 \to \operatorname{Var}\left(\xi_{12}\mathbb{1}_{\{|\xi_{12}|>N\}}\right) \quad \text{a.s..}$$

Hence

$$\limsup_{n\to\infty}\frac{2}{n^2}\sum_{1\le i<j\le n}\left|\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|>N\}} - \mathbb{E}[\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|>N\}}]\right|^2 \le \mathbb{E}\left[|\xi_{12}|^2\mathbb{1}_{\{|\xi_{12}|>N\}}\right], \quad \text{a.s..}$$

Similarly

$$\lim_{n\to\infty}\frac{2}{n^2}\sum_{1\le i<j\le n}\left(\frac{1}{\operatorname{Var}(\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}})}-1\right)\left|\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}} - \mathbb{E}[\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\le N\}}]\right|^2 = 1-\operatorname{Var}(\xi_{12}\mathbb{1}_{\{|\xi_{12}|\le N\}}), \quad \text{a.s..}$$

Now given any $\epsilon>0$, we fix $N\in\mathbb{N}$ great enough that

$$\max\left\{\mathbb{E}\left[|\xi_{12}|^2\mathbb{1}_{\{|\xi_{12}|>N\}}\right], 1-\operatorname{Var}(\xi_{12}\mathbb{1}_{\{|\xi_{12}|\le N\}})\right\} \le \frac{\epsilon^3}{2}.$$

By Lemma 1.11,

$$\limsup_{n\to\infty}\rho_L\left(F_{\frac{\widehat{W}_n}{\sqrt{n}}}, F_{\frac{\widetilde{W}_n}{\sqrt{n}}}\right) \le \left[\limsup_{n\to\infty}\frac{1}{n}\operatorname{tr}\left(\widehat{W}_n-\widetilde{W}_n\right)^2\right]^{1/3} < \epsilon, \quad \text{a.s..}$$

Suppose $\mu_{\frac{\widetilde{W}_n}{\sqrt{n}}} \to \mu_{\text{sc}}$ weakly almost surely. By Lemma 1.13,

$$\limsup_{n\to\infty}\rho_L\left(\mu_{\frac{W_n}{\sqrt{n}}},\mu_{\text{sc}}\right) \le \limsup_{n\to\infty}\left[\rho_L\left(F_{\frac{W_n}{\sqrt{n}}}, F_{\frac{\widehat{W}_n}{\sqrt{n}}}\right) + \rho_L\left(F_{\frac{\widehat{W}_n}{\sqrt{n}}}, F_{\frac{\widetilde{W}_n}{\sqrt{n}}}\right) + \rho_L\left(F_{\frac{\widetilde{W}_n}{\sqrt{n}}}, F_{\text{sc}}\right)\right] < \epsilon, \quad \text{a.s..}$$

Since $\epsilon>0$ is arbitrary, we can make $\epsilon_n = n^{-1}\downarrow 0$ and take the intersection of the above events to see

$$\limsup_{n\to\infty}\rho_L\left(F_{\frac{W_n}{\sqrt{n}}}, F_{\text{sc}}\right) = 0, \quad \text{a.s..}$$

Hence $\mu_{\frac{W_n}{\sqrt{n}}} \to \mu_{\text{sc}}$ almost surely. $\square$

1.3 The Resolvent Method

1.3.1 The Stieltjes Transform

Definition 1.15 (Stieltjes transform)**.** Let μ be a Borel measure on the real line $\mathbb{R}$. The *Stieltjes transform* of μ is the function of the complex variable z defined outside the support of μ by the formula

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z}\, d\mu(x), \quad z \in \mathbb{C}\setminus \operatorname{supp}(\mu).$$

In particular, s_μ is well-defined on the upper and lower half-planes in the complex plane $\mathbb{C}$.

Remark. (i) By definition, the imaginary parts of $s_\mu(z)$ and z have the same sign. Since $s_\mu(\overline{z}) = \overline{s_\mu(z)}$ for $z \in \mathbb{C}\backslash\mathbb{R}$, it suffices to study the property of s_μ in the upper half-plane $\mathbb{C}^+ = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$.

(ii) Indeed, the Stieltjes transform $s_\mu : \mathbb{C}^+ \to \mathbb{C}^+$ is a holomorphic function. To see this, we fix $z \in \mathbb{C}\backslash\mathbb{R}$. Then $|x-z|^{-1} \le |\operatorname{Im}(z)|^{-1}$, and $|s_\mu(z)| \le |\operatorname{Im}(z)|^{-1}$. For all $h \in \mathbb{C}$ with $|h| < |\operatorname{Im}(z)|/2$, we have

$$\frac{1}{(x-z)(x-z-h)} \le \frac{2}{|\operatorname{Im}(z)|^2},$$

which is bouded uniformly in $x \in \mathbb{R}$. By the dominated convergenece theorem,

$$\lim_{h\to 0} \frac{s_\mu(z+h) - s_\mu(z)}{h} = \lim_{h\to 0} \int_{\mathbb{R}} \frac{1}{(x-z)(x-z-h)}\, d\mu(x) = \int_{\mathbb{R}} \frac{1}{(x-z)^2}\, d\mu(x).$$

Hence s_μ is complex differentiable in $\mathbb{C}^+$. By holomorphicity, s_μ is infinitely differentiable in $\mathbb{C}^+$, and

$$\frac{d^k}{dz^k} s_\mu(z) = \int_{\mathbb{R}} \frac{1}{(x-z)^{1+k}}\, d\mu(x), \quad z \in \mathbb{C}^+, \quad k \in \mathbb{N}_0. \tag{1.12}$$

Theorem 1.16 (Stieltjes inversion)**.** *For any two points $a < b$ of continuity of F_μ, which is the c.d.f. of μ,*

$$\mu((a,b]) = \lim_{\eta\downarrow 0} \int_a^b \frac{s_\mu(E+i\eta) - s_\mu(E-i\eta)}{2\pi i}\, dE. \tag{1.13}$$

Distinct Borel measures μ on $\mathbb{R}$ have distinct Stieltjes transform s_μ.

Proof. For $\xi \in \mathbb{R}$ and $\eta > 0$, we have

$$\operatorname{Im}(s_\mu(E+i\eta)) = \int_{\mathbb{R}} \operatorname{Im}\left(\frac{1}{x-(E+i\eta)}\right) d\mu(x) = \int_{\mathbb{R}} \frac{\eta}{(x-E)^2+\eta^2}\, d\mu(x).$$

Let $f \in C_b(\mathbb{R})$. By Fubini's theorem and dominated convergence theorem,

$$\begin{aligned}\lim_{\eta\downarrow 0} \frac{1}{\pi} \int_{\mathbb{R}} f(E) \operatorname{Im}(s_\mu(E+i\eta))\, d\xi &= \lim_{\eta\downarrow 0} \int_{\mathbb{R}} \int_{\mathbb{R}} f(E) \frac{\eta}{\pi\left((x-E)^2+\eta^2\right)}\, dE\, d\mu(x) \\ &= \int_{\mathbb{R}} \lim_{\eta\downarrow 0} \int_{\mathbb{R}} f(E) \frac{\eta}{\pi\left((x-E)^2+\eta^2\right)}\, dE\, d\mu(x) = \int_{\mathbb{R}} f(x)\, d\mu(x).\end{aligned}$$

That is, for all bounded continuous function $f : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}} f(x)\, d\mu(x) = \lim_{\eta\downarrow 0} \int_{\mathbb{R}} \frac{s_\mu(E+i\eta) - s_\mu(E-i\eta)}{2\pi i} f(E)\, dE. \tag{1.14}$$

By Riesz representation theorem, μ is uniquely determined by its Stieltjes transform s_μ. The result (1.13) follows from the Portmeanteau lemma. ☐

One can relate weak convergence of measures to pointwise convergence of their Stieltjes transforms.

Theorem 1.17 (Stieltjes continuity theorem)**.** *Let (μ_n) be a sequence of probability measures on $\mathbb{R}$.*

(i) If μ_n converges weakly to a probability measure μ, then $s_{\mu_n}(z) \to s_\mu(z)$ for every $z \in \mathbb{C}\backslash\mathbb{R}$.

(ii) If $s_{\mu_n}(z)$ converges to a limit $s(z)$ for all $z \in \mathbb{R}\backslash\mathbb{C}$, then $S : \mathbb{C}\backslash\mathbb{R} \to \mathbb{C}$ is the Stieltjes transform of a sub-probability measure μ, and $\mu_n \to \mu$ weakly.

Furthermore, assume that (μ_n) are random probability measures and μ is a deterministic probability measure.

(iii) $\mu_n \to \mu$ weakly almost surely if and only if $s_{\mu_n}(z) \to s_\mu(z)$ almost surely for every $z \in \mathbb{C}\backslash\mathbb{R}$.

(iv) $\mu_n \to \mu$ weakly in probability if and only if $s_{\mu_n}(z) \to s_\mu(z)$ in probability for every $z \in \mathbb{C}\backslash\mathbb{R}$.

Proof. (i) For every $z \in \mathbb{C}\backslash\mathbb{R}$, the function $\mathbb{R} \to \mathbb{C} : x \mapsto (x-z)^{-1}$ is bounded and continuous. Hence $\mu_n \to \mu$ weakly implies $s_{\mu_n} \to s_\mu$ pointwise on $\mathbb{C}\backslash\mathbb{R}$.

(ii) By Helly's selection theorem, every subsequence of (μ_n) admits a further subsequence that converges weakly to a sub-probability measure. We let (μ_{n_k}) be a subsequence that converges weakly to a sub-probability measure μ. Then $s_{\mu_{n_k}} \to s_\mu$ by (i), and we have $s_\mu = s$ from the hypothesis. By Theorem 1.16, all weakly convergent subsequences converge to the same μ, and hence $\mu_n \to \mu$.

(iii) is an immediate corollary of (i) and (ii).

(iv) The "only if" part is easy, and we focus on the "if" part. Let $f \in C_c(\mathbb{R})$, and assume f is supported on $[-B/2, B/2]$. We take $f_\eta = f * P_\eta$, where $P_\eta(x) = \frac{\eta}{\pi(x^2+\eta^2)}$ is the Poisson kernel. Then $f * P_\eta \to f$ uniformly on $\mathbb{R}$ as $\eta \downarrow 0$. Given $\epsilon > 0$, we take $\eta > 0$ sufficiently small so that $\|f - f_\eta\|_\infty < \epsilon/5$. Then

$$\begin{aligned}\left|\int_{\mathbb{R}} f\,d\mu_n - \int_{\mathbb{R}} f\,d\mu\right| &\leq \left|\int_{\mathbb{R}} f\,d\mu_n - \int_{\mathbb{R}} f_\eta\,d\mu_n\right| + \left|\int_{\mathbb{R}} f_\eta\,d\mu_n - \int_{\mathbb{R}} f_\eta\,d\mu\right| + \left|\int_{\mathbb{R}} f_\eta\,d\mu - \int_{\mathbb{R}} f\,d\mu\right| \\ &\leq \frac{2\epsilon}{5} + \left|\int_{\mathbb{R}} f_\eta\,d\mu_n - \int_{\mathbb{R}} f_\eta\,d\mu\right|. \end{aligned} \tag{1.15}$$

SImilar to our proof of Theorem 1.16,

$$\int_{\mathbb{R}} f_\eta\,d\mu_n - \int_{\mathbb{R}} f_\eta\,d\mu = \int_{\mathbb{R}} \frac{s_{\mu_n}(E+i\eta) - s_{\mu_n}(E-i\eta)}{2\pi i} f(E)\,dE - \int_{\mathbb{R}} \frac{s_\mu(E+i\eta) - s_\mu(E-i\eta)}{2\pi i} f(E)\,dE.$$

By (1.12), we have $|s'_\mu(z)| \leq |\mathrm{Im}(z)|^{-2}$, and s_μ is Lipschitz. We then divide $\mathrm{supp}(f)$ into 2^k sub-intervals $[E_j^{(k)}, E_{j+1}^{(k)}]$, $j = 1, \cdots, 2^k$ of equal length, and approximate the above integral by a Riemann sum:

$$\left|\int_{\mathbb{R}} f_\eta\,d\mu - \frac{1}{2^k}\sum_{j=1}^{2^k} \frac{s_\mu(E_j^{(k)}+i\eta) - s_\mu(E_j^{(k)}-i\eta)}{2\pi i} f(E_j^{(k)})\right| \leq \frac{B\|f\|_\infty}{2^k\pi\eta^2}.$$

The same result remains true with s_μ replaced by s_{μ_n} for every $n \in \mathbb{N}$. We fix $k \in \mathbb{N}$ great enough such that $B\|f\|_\infty/(2^k\pi\eta^2) < \epsilon/5$. Since for every $z \in \mathbb{R}\backslash\mathbb{C}$, $s_{\mu_n}(z) \to s_\mu(z)$ in probability, we take $N > 1$ sufficiently large so that for every $n \geq N$,

$$\mathbb{P}\left(\left|s_{\mu_n}(E_j^{(k)}+i\eta) - s_\mu(E_j^{(k)}+i\eta)\right| \geq \frac{\epsilon\pi}{5\|f\|_\infty}\right) < \frac{\epsilon}{2^k} \quad \text{for all } j = 1, \cdots, 2^k.$$

Hence with probability at least $1-\epsilon$,

$$\left|\int_{\mathbb{R}} f_\eta\,d\mu_n - \int_{\mathbb{R}} f_\eta\,d\mu\right| \leq \frac{2B\|f\|_\infty}{2^k\pi\eta^2} + \frac{1}{2^k}\sum_{j=1}^{2^k}\frac{1}{2\pi}\cdot\frac{2\epsilon\pi}{5\|f\|_\infty}|f(E_j^{(k)})| < \frac{2\epsilon}{5} + \frac{\epsilon}{5} = \frac{3\epsilon}{5}.$$

Combining with (1.15), we finish the proof. □

1.3.2 The Marcinkiewicz-Zygmund inequality

We introduce a useful inequality for error control. This is a special case of Burkholder-Davis-Gundy inequality.

Lemma 1.18 (Marcinkiewicz-Zygmund inequality). *Let $X_1, \cdots, X_N$ be complex-valued, independent zero-mean random variables. Then for every $p \geq 2$,*

$$\mathbb{E}\left|\sum_{j=1}^{N} X_j\right|^p \leq (Cp)^{p/2}\mathbb{E}\left|\sum_{j=1}^{n} |X_j|^2\right|^{p/2}, \tag{1.16}$$

where C is an absolute constant. Furthermore, if $(a_{ij})_{i,j\in[N]} \in \mathbb{C}^{N\times N}$, we have

$$\left\|\sum_{i\neq j}^{N} a_{ij}\overline{X}_i X_j\right\|_{L^p} \leq 4Cp\left(\sum_{i\neq j}^{N} |a_{ij}|^2\right)^{1/2}\left(\max_{j\in[N]} \mathbb{E}|X_j|^p\right)^{2/p}. \tag{1.17}$$

Proof. Step I. We first assume that $X_1, \cdots, X_N$ are real-valued, and let $\epsilon_1, \cdots, \epsilon_N$ be i.i.d. Rademacher variables independent of $X_1, \cdots, X_N$. We show that

$$\mathbb{E}\left|\sum_{j=1}^{N} X_j\right|^p \leq 2^p\mathbb{E}_X\left[\mathbb{E}_\epsilon\left|\sum_{j=1}^{N}\epsilon_j X_j\right|^p\right], \quad p \geq 2. \tag{1.18}$$

Let Y_j be an independent copy of X_j for $j = 1, \cdots, N$. Then by Jensen's inequality,

$$\mathbb{E}\left|\sum_{j=1}^{N} X_j\right|^p = \mathbb{E}\left|\sum_{j=1}^{N}(X_j - \mathbb{E}Y_j)\right|^p = \mathbb{E}_X\mathbb{E}_Y\left|\sum_{j=1}^{N}(X_j - Y_j)\right|^p.$$

Since $X_j - Y_j$ is symmetric, we have $X_j - Y_j \stackrel{d}{=} \epsilon_j(X_j - Y_j)$. Then

$$\mathbb{E}_X\mathbb{E}_Y\left|\sum_{j=1}^{N}(X_j - Y_j)\right|^p = \mathbb{E}_X\mathbb{E}_Y\mathbb{E}_\epsilon\left|\sum_{j=1}^{N}\epsilon_j(X_j - Y_j)\right|^p \leq \mathbb{E}_X\mathbb{E}_Y\mathbb{E}_\epsilon\left[2^{p-1}\left|\sum_{j=1}^{N}\epsilon_j X_j\right|^p + 2^{p-1}\left|\sum_{j=1}^{N}\epsilon_j Y_j\right|^p\right].$$

Since $(X_1, \cdots, X_N) \stackrel{d}{=} (Y_1, \cdots, Y_N)$, we obtain (1.18).

Step II. By the Khintchine inequality for sub-Gaussian random variables, for all $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$,

$$\left\|\sum_{j=1}^{N} a_j\epsilon_j\right\|_{L^p} \leq 2\sqrt{6p}\,\|a\|_2. \tag{1.19}$$

We use (1.19) conditioning on $X_1, \cdots, X_N$ to get

$$\mathbb{E}_\epsilon\left|\sum_{j=1}^{N}\epsilon_j X_j\right|^p \leq \left(2\sqrt{6p}\right)^p\left(\sum_{j=1}^{n} X_j^2\right)^{p/2}.$$

By (1.18), we obtain

$$\mathbb{E}\left|\sum_{j=1}^{N} X_j\right|^p \leq 2^p\mathbb{E}_X\left[\mathbb{E}_\epsilon\left|\sum_{j=1}^{N}\epsilon_j X_j\right|^p\right] \leq \left(4\sqrt{6p}\right)^p\mathbb{E}\left[\left(\sum_{j=1}^{n} X_j^2\right)^{p/2}\right] \leq (Cp)^{p/2}\mathbb{E}\left[\left(\sum_{j=1}^{n} X_j^2\right)^{p/2}\right],$$

where $C = 96$. Then we finish the proof of (1.16) in the real case.

Step III. If $X_1, \cdots, X_N$ are complex valued, then both $(\operatorname{Re} X_i)_{i=1}^N$ and $(\operatorname{Im} X_i)_{i=1}^N$ are independent real-valued variables. Then we apply (1.16):

$$\mathbb{E}\left|\sum_{j=1}^N X_j\right|^p \leq 2^{p-1}\left(\mathbb{E}\left|\sum_{j=1}^N \mathrm{Re} X_j\right|^p + \mathbb{E}\left|\sum_{j=1}^N \mathrm{Im} X_j\right|^p\right)$$
$$\leq 2^{p-1}(Cp)^{p/2}\left[\mathbb{E}\left|\sum_{j=1}^n |\operatorname{Re} X_j|^2\right|^{p/2} + \mathbb{E}\left|\sum_{j=1}^n |\operatorname{Im} X_j|^2\right|^{p/2}\right] \leq \frac{(2Cp)^{p/2}}{2}\mathbb{E}\left|\sum_{j=1}^n |X_j|^2\right|^{p/2}.$$

Then we finish the proof of (1.16) in the complex case.

Step IV. For every $i, j \in [N]$ with $i \neq j$, we have

$$\frac{1}{2^{N-2}}\sum_{I \sqcup J=[N]} \mathbb{1}_{\{i \in I\}}\mathbb{1}_{\{j \in J\}} = 1,$$

where the sum ranges over all partitions of $[N]$ into two sets I and J. Then

$$\sum_{i \neq j}^N a_{ij}\overline{X}_i X_j = \frac{1}{2^{N-2}}\sum_{i \neq j}^N \sum_{I \sqcup J=[N]} \mathbb{1}_{\{i \in I\}}\mathbb{1}_{\{j \in J\}} a_{ij}\overline{X}_i X_j = \frac{1}{2^{N-2}}\sum_{I \sqcup J=[N]}\sum_{i \in I}\sum_{j \in J} a_{ij}\overline{X}_i X_j.$$

By the triangle inequality, for every $k \in \mathbb{N}$,

$$\left\|\sum_{i \neq j}^N a_{ij}\overline{X}_i X_j\right\|_{L^p} \leq \frac{1}{2^{N-2}}\sum_{I \sqcup J=[N]}\left\|\sum_{i \in I}\sum_{j \in J} a_{ij}\overline{X}_i X_j\right\|_{L^p}.$$

Let $b_j = \sum_{i \in I} a_{ij}\overline{X}_i$. We take the following expectation with respect to $\{X_j : j \in J\}$, denoted by $\mathbb{E}_J$:

$$\mathbb{E}_J\left|\sum_{i \in I}\sum_{j \in J} a_{ij}\overline{X}_i X_j\right|^p = \mathbb{E}_J\left|\sum_{j \in J} b_j X_j\right|^p \leq (Cp)^{p/2}\mathbb{E}_J\left[\left(\sum_{j \in J} b_j^2 X_j^2\right)^{p/2}\right] = (Cp)^{p/2}B^p\mathbb{E}_J\left[\left(\sum_{j \in J}\frac{b_j^2}{B^2}X_j^2\right)^{p/2}\right]$$
$$\leq (Cp)^{p/2}B^p\mathbb{E}_J\left[\sum_{j \in J}\frac{b_j^2}{B^2}X_j^p\right] \leq (Cp)^{p/2}B^p \max_{j \in J}\mathbb{E}|X_j|^p, \tag{1.20}$$

where we take $B^2 = \sum_{j \in J} b_j^2$. Similarly,

$$\mathbb{E}_I|b_j|^p \leq (Cp)^{p/2}\mathbb{E}_I\left[\left(\sum_{i \in I} a_{ij}^2|X_i|^2\right)^{p/2}\right] \leq (Cp)^{p/2}A_j^p \max_{i \in I}\mathbb{E}|X_i|^p, \quad j \in J,$$

where $A_j^2 = \sum_{i \in I} a_{ij}^2$. By Minkowski's inquality,

$$\mathbb{E}_I[B^p] = \|B^2\|_{L^{p/2}}^{p/2} = \left\|\sum_{j \in J} b_j^2\right\|_{L^{p/2}}^{p/2} \leq \left(\sum_{j \in J}\|b_j^2\|_{L^{p/2}}\right)^{p/2} = \left(\sum_{j \in J} CpA_j^2\right)^{p/2}\max_{i \in I}\mathbb{E}|X_i|^p.$$

Plugging into (1.20), we have

$$\mathbb{E}\left|\sum_{i \in I}\sum_{j \in J} a_{ij}\overline{X}_i X_j\right|^p \leq (Cp)^p\left(\sum_{i \in I}\sum_{j \in J} a_{ij}^2\right)^{p/2}\left(\max_{j \in [N]}\mathbb{E}|X_j|^p\right)^2.$$

Note there are in total $2^N - 2$ nontrivial partitions of $[N]$. Then

$$\begin{aligned}\left\|\sum_{i\neq j}^{N} a_{ij}\overline{X}_i X_j\right\|_{L^p} &\leq \frac{1}{2^{N-2}} \sum_{I\sqcup J=[N]} \left\|\sum_{i\in I}\sum_{j\in J} a_{ij}\overline{X}_i X_j\right\|_{L^p} \\ &\leq \frac{2^N-2}{2^{N-2}} Cp \left(\sum_{i\neq j}^{N} a_{ij}^2\right)^{\frac{1}{2}} \left(\max_{j\in[N]} \mathbb{E}|X_j|^p\right)^{\frac{2}{p}} = 4Cp \left(\sum_{i\neq j}^{N} a_{ij}^2\right)^{\frac{1}{2}} \left(\max_{j\in[N]} \mathbb{E}|X_j|^p\right)^{\frac{2}{p}}.\end{aligned}$$

Thus we complete the proof. □

1.3.3 The Stieltjes Transform of the Semicircle Law

For a random Hermitian matrix $A \in \mathbb{C}^{n\times n}$, the Stieljes transform of the ESD of A is related to the trance of its resolvent:

$$s(A,z) := s_{\mu_A}(z) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\lambda_j(A) - z} = \frac{1}{n}\mathrm{tr}\left[(A - z\,\mathrm{Id})^{-1}\right].$$

We consider the Stieltjes transform of normalized Wigner matrices:

$$s_n(z) := s\left(\frac{W_n}{\sqrt{n}}, z\right) = \frac{1}{n}\mathrm{tr}\left[\left(\frac{W_n}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1}\right] = \frac{1}{n}\sum_{j=1}^{n}\left[\left(\frac{W_n}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1}\right]_{jj}.$$

To establish Wigner's semicircle law, if suffices to show that the Stieltjes transform of the ESD of $W_n/\sqrt{n}$ converges pointwise to the Stieltjes transform of the semicircle distribution μ_{sc}. To establish this, we first compute the Siteltjes transform of the semicircle distribution μ_{sc}.

Lemma 1.19. *The Stieltjes transform of the semicircle distribution μ_{sc} is*

$$s_{\mathrm{sc}}(z) = \frac{-z + \sqrt{z^2-4}}{2}, \quad z \in \mathbb{C}^+,$$

where $\mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\}$, and the square root of a complex number z in $\mathbb{C}\backslash\mathbb{R}$ is defined as the branch with the positive imaginary part.

Proof. We let $z \in \mathbb{C}\backslash\mathbb{R}$ and $\mathrm{Im}\, z > 0$. The Stieltjes transform of the semicircle distribution μ_{sc} is

$$s_{\mathrm{sc}}(z) = \frac{1}{2\pi}\int_{-2}^{2} \frac{\sqrt{4-x^2}}{x-z}\,dx.$$

We let $x = 2\cos\theta$. Then

$$\begin{aligned} s_{\mathrm{sc}}(z) &= \frac{2}{\pi}\int_0^{\pi} \frac{\sin^2\theta}{2\cos\theta - z}\,d\theta = \frac{1}{\pi}\int_0^{2\pi} \frac{1}{e^{i\theta}+e^{-i\theta}-z}\left(\frac{e^{i\theta}-e^{i\theta}}{2i}\right)^2 d\theta \\ &= -\frac{1}{4\pi i}\int_{|\zeta|=1} \frac{(\zeta-\zeta^{-1})^2\zeta^{-1}}{\zeta+\zeta^{-1}-z}\,d\zeta = -\frac{1}{4\pi i}\int_{|\zeta|=1} \frac{(\zeta^2-1)^2}{\zeta^2(\zeta^2 - z\zeta + 1)}\,d\zeta \end{aligned}$$

We evaluate the above integral by the residue theorem. The integrand has three poles:

$$\zeta_0 = 0, \quad \zeta_1 = \frac{z+\sqrt{z^2-4}}{2}, \quad \zeta_2 = \frac{z-\sqrt{z^2-4}}{2},$$

where the square root of a complex number in $\mathbb{C}\backslash\mathbb{R}$ is defined as the branch with the positive imaginary part.

By this convention, we have

$$\sqrt{\zeta} = \operatorname{sgn}(\operatorname{Im}\zeta)\sqrt{\frac{|\zeta| + \operatorname{Re}\zeta}{2}} + i\sqrt{\frac{|\zeta| - \operatorname{Re}\zeta}{2}} = \frac{\operatorname{Im}\zeta}{\sqrt{2(|\zeta| - \operatorname{Re}\zeta)}} + \frac{i\,|\operatorname{Im}\zeta|}{\sqrt{2(|\zeta| + \operatorname{Re}\zeta)}}. \tag{1.21}$$

This shows that the real part of $\sqrt{\zeta}$ has the same sign as the $\operatorname{Im}\zeta$. Applying this to ζ_1 and ζ_2, the real part of $\sqrt{z^2-4}$ has the same sign as $\operatorname{Re} z$. Then both real and imaginary parts of ζ_1 are greater than those of ζ_2, and $|\zeta_1| > |\zeta_2|$. Since $\zeta_1\zeta_2 = 1$, we conclude that $|\zeta_1| > 1 > |\zeta_2|$, and the two poles 0 and ζ_2 of the integrad are in the disk $|\zeta| \leq 1$. Note that

$$\begin{aligned}
&\operatorname{Res}\left(\frac{(\zeta^2-1)^2}{\zeta^2(\zeta^2 - z\zeta + 1)}, 0\right) = \lim_{\zeta\to 0}\frac{d}{d\zeta}\left[\frac{(\zeta^2-1)^2}{(\zeta-\zeta_1)(\zeta-\zeta_2)}\right] = \frac{\zeta_1+\zeta_2}{\zeta_1^2\zeta_2^2} = z, \quad \text{and} \\
&\operatorname{Res}\left(\frac{(\zeta^2-1)^2}{\zeta^2(\zeta^2 - z\zeta + 1)}, \zeta_2\right) = \lim_{\zeta\to\zeta_2}\frac{(\zeta-\zeta_2)(\zeta^2-1)^2}{\zeta^2(\zeta^2 - z\zeta + 1)} = \frac{(\zeta_2^2-1)^2}{\zeta_2^2(\zeta_2-\zeta_1)} = \zeta_2 - \zeta_1 = -\sqrt{z^2-4}.
\end{aligned}$$

By Cauchy's residue theorem,

$$s_{\mathrm{sc}}(z) = \frac{-z + \sqrt{z^2-4}}{2}, \quad z \in \mathbb{C}^+.$$

Then we finish the proof. □

Since the expression of $s_{\mathrm{sc}}(z)$ is complicated, directly establish the convergence $s_n(z) \to s_{\mathrm{sc}}(z)$ is difficult. Luckily, we note that $s_{\mathrm{sc}}(z)$ is a fixed point of the function

$$\mathbb{C}\backslash\{-z\} \to \mathbb{C} : s \mapsto -\frac{1}{z+s}.$$

Inspired by this result, we can do the following reduction.

Lemma 1.20. *If $z \in \mathbb{C}^+$ and*

$$s_n(z) + \frac{1}{z + s_n(z)} \to 0 \quad \textit{almost surely},$$

then $s_n(z) \to s_{\mathrm{sc}}(z)$ almost surely.

Proof. We let A be an event of probability 1 on which $s_n(z) + 1/(z + s_n(z)) \to 0$, and fix $\omega \in A$. To show $s_n(z) \to s_{\mathrm{sc}}(z)$ a.s., it suffices to establish the convergence in deterministic case. Since the Siteltjes transforms satisfy $|s_n(z)| \leq |\operatorname{Im} z|^{-1}$, the sequence $(s_n(z))$ is a bounded and has a convergent subsequence $(s_{n_k}(z))$ by Bolzano-Weierstrass theorem. Furthermore, the limit $s = \lim_{k\to\infty} s_{n_k}(z)$ satisfies

$$s + \frac{1}{z+s} = 0, \quad \text{and} \quad s \in \left\{\frac{-z+\sqrt{z^2-4}}{2}, \frac{-z-\sqrt{z^2-4}}{2}\right\}.$$

Now we select the correct branch. Since $\operatorname{Im} s_n(z) > 0$ for all n, we also have $\operatorname{Im} s \geq 0$. Then

$$s = \frac{-z+\sqrt{z^2-4}}{2} = s_{\mathrm{sc}}(z).$$

Essentially, we show that any subsequence of $(s_n(z))$ has a further subsequence converging to $s_{\mathrm{sc}}(z)$. Hence $s_n(z) \to s_{\mathrm{sc}}(z)$, and we finish the proof. □

Now we prove Wigner's semicircle law through the Stieltjes transform. By Lemma 1.14, we may assume that the diagonal entries $(\xi_{ii})_{i\geq 1}$ are zero, and the off-diagonal entries $(\xi_{ii})_{1\leq i<j}$ are bounded by $R < \infty$. We only prove the almost sure convergence result, which is the strongest and implies the other two modes of convergence (in probability and in expectation).

Proof of Theorem 1.5 (ii). Let $W_{n,-j}$ denote the $(n-1)\times(n-1)$ matrix obtained from W_n by removing the j-th row and the j-th column. Then $W_n/\sqrt{n}$ can be written as

$$\frac{W_n}{\sqrt{n}} - z\,\mathrm{Id} = \begin{bmatrix} \frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id} & w_{n,j} \\ w_{n,j}^* & -z \end{bmatrix},$$

where $w_{n,j} = (\xi_{1j}, \cdots, \xi_{j-1,j}, \xi_{j+1,j}, \cdots, \xi_{nj})^\top/\sqrt{n}$. By Schur's complement,

$$\left[\left(\frac{W_n}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1}\right]_{jj} = -\frac{1}{z + w_{n,j}^*\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1} w_{n,j}}.$$

We let

$$\begin{aligned} s_n(z) &= -\frac{1}{n}\sum_{j=1}^{n} \frac{1}{z + w_{n,j}^*\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1} w_{n,j}} \\ &= -\frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{z+s_n(z)} + \frac{s_n(z) - w_{n,j}^*\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1} w_{n,j}}{(z+s_n(z))\left(z + w_{n,j}^*\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1} w_{n,j}\right)}\right] = -\frac{1}{z+s_n(z)} - \delta_n(z), \end{aligned}$$

where

$$\delta_n(z) = \frac{1}{n}\sum_{j=1}^{n} \frac{s_n(z) - w_{n,j}^*\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1} w_{n,j}}{(z+s_n(z))\left(z + w_{n,j}^*\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1} w_{n,j}\right)}.$$

We write

$$\delta_n(z) = \frac{1}{n}\sum_{j=1}^{n}\frac{\upsilon_{n,j}}{(z+s_n(z))(z+s_n(z)-\upsilon_{n,j})}, \qquad \text{where } \upsilon_{n,j} = s_n(z) - w_{n,j}^*\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1} w_{n,j}.$$

Assume $\max_{j\in[n]}|\upsilon_{n,j}| < |\mathrm{Im}\,z|/2$. Since the imaginary parts of $s_n(z)$ and z have the same sign, we have $\mathrm{Im}(z+s_n(z)) \geq \mathrm{Im}z$, and

$$\delta_n(z) \leq \frac{1}{n}\sum_{j=1}^{n}\frac{\upsilon_{n,j}}{|\mathrm{Im}\,z|^2/2} \leq \frac{2}{|\mathrm{Im}\,z|^2}\max_{j\in[n]}|\upsilon_{n,j}|.$$

Therefore, if $\max_{j\in[n]}|\upsilon_{n,j}| \to 0$ almost surely as $n\to\infty$, so does $\delta_n(z)$. Then (1.20) is satisfied, and we can apply Lemma 1.20 to conclude the proof of the semicircle law. We use the following decomposition:

$$\begin{aligned} \upsilon_{n,j} = &\left[s_n(z) - \frac{1}{n}\mathrm{tr}\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1}\right] - \sum_{k=1}^{n-1}\left(|w_{n,j}(k)|^2 - \frac{1}{n}\right)\left[\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1}\right]_{kk} \\ &- \sum_{k\neq k'}^{n-1}\overline{w_{n,j}(k)}w_{n,j}(k')\left[\left(\frac{W_{n,-j}}{\sqrt{n}} - z\,\mathrm{Id}\right)^{-1}\right]_{kk'} =: A_{n,j}(z) + B_{n,j}(z) + C_{n,j}(z). \end{aligned} \tag{1.22}$$

Step I. Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of matrix $W_n/\sqrt{n}$, and $\mu_1, \cdots, \mu_{n-1}$ the eigenvalues of matrix $W_{n,-j}/\sqrt{n}$. Then for $z = E + i\eta$ with $\eta > 0$,

$$A_{n,j}(z) = \frac{1}{n}\left(\sum_{j=1}^{n}\frac{1}{\lambda_j - z} - \sum_{j=1}^{n-1}\frac{1}{\mu_j - z}\right) = \frac{1}{n}\sum_{j=1}^{n-1}\left[\frac{(\lambda_j - E) + i\eta}{(\lambda_j - E)^2 + \eta^2} - \frac{(\mu_j - E) + i\eta}{(\mu_j - E)^2 + \eta^2}\right] + \frac{1}{n}\cdot\frac{1}{\lambda_n - z}. \tag{1.23}$$

By Cauchy's interlacing theorem, $\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \cdots \geq \lambda_{n-1} \geq \mu_{n-1} \geq \lambda_n$. Then the shift of function $\lambda \mapsto \frac{\lambda-E}{(\lambda-E)^2+\eta^2}$ on disjoint intervals $[\mu_1, \lambda_1], \cdots, [\mu_{n-1}, \lambda_{n-1}]$ is bounded by its total variation on $\mathbb{R}$:

$$\frac{1}{n}\sum_{j=1}^{n-1}\left|\frac{\lambda_j - E}{(\lambda_j - E)^2+\eta^2} - \frac{\mu_j - E}{(\mu_j - E)^2+\eta^2}\right| \leq \left\|\frac{\lambda - E}{(\lambda - E)^2+\eta^2}\right\|_{\mathrm{TV}_\lambda} = \frac{2}{\eta}.$$

Similarly,

$$\frac{1}{n}\sum_{j=1}^{n-1}\left|\frac{\eta}{(\lambda_j - E)^2+\eta^2} - \frac{\eta}{(\mu_j - E)^2+\eta^2}\right| \leq \left\|\frac{\eta}{(\lambda - E)^2+\eta^2}\right\|_{\mathrm{TV}_\lambda} = \frac{2}{\eta}.$$

Then from (1.23), we obtain

$$|A_{n,j}(z)| \leq \frac{5}{n\eta}, \quad j = 1, \cdots, n. \tag{1.24}$$

Step II. Let $X_{n,j} = (W_{n,-j}/\sqrt{n} - z\,\mathrm{Id})^{-1}$. Then

$$\frac{1}{n}\|X_{n,j}\|_{\mathrm{F}}^2 = \frac{1}{n}\mathrm{tr}(X_{n,j}^2) = \frac{1}{n}\sum_{j=1}^{n-1}\frac{1}{(\lambda(W_{n,-j}/\sqrt{n}) - z)^2} \leq \frac{1}{\eta^2}.$$

Note that $w_{n,j}$ has independent entries and is independent of $W_{n,-j}$. Then

$$\mathbb{E}\,|B_{n,j}(z)|^6 = \frac{1}{n^6}\sum_{k=1}^{n-1}\mathbb{E}\left[(n|w_{n,j}(k)|^2-1)^6\right]\mathbb{E}\,|X_{n,j}(k,k)|^6 \leq \frac{R^{12}}{n^6}\left(\sum_{k=1}^{n-1}\mathbb{E}\,|X_{n,j}(k,k)|^2\right)^3 \leq \frac{R^{12}}{n^3\eta^6}.$$

Using the case $p = 6$ in Lemma 1.18, there exists an absolute constant $C > 0$ such that

$$\mathbb{E}\,|C_{n,j}(z)|^6 = \mathbb{E}\left|\sum_{\substack{k,k'=1,\\ k\neq k'}}^{n-1}\overline{w_{n,j}(k)}w_{n,j}(k')X_{n,j}(k,k')\right|^6 \leq C\left(\mathbb{E}\|X_{n,j}\|_F^2\right)^3\left(\frac{R}{\sqrt{n}}\right)^{12} \leq \frac{CR^{12}}{n^3\eta^6}.$$

Hence for every $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq j\leq n}|B_{n,j}(z)+C_{n,j}(z)| > \epsilon\right) &\leq \sum_{j=1}^{n}\mathbb{P}\left(|B_{n,j}(z)+C_{n,j}(z)| > \epsilon\right) \leq \frac{1}{\epsilon^6}\sum_{j=1}^{n}\mathbb{E}|B_{n,j}(z)+C_{n,j}(z)|^6\\ &\leq \frac{32}{\epsilon^6}\sum_{j=1}^{n}\left(\mathbb{E}|B_{n,j}(z)|^6+\mathbb{E}|C_{n,j}(z)|^6\right) \leq \frac{32(1+C)R^{12}}{n^2\eta^6}.\end{aligned}$$

Since $\sum_{n=1}^{\infty} n^{-2} < \infty$, by the Borel Cantelli lemma, we have

$$\mathbb{P}\left(\limsup_{n\to\infty}\max_{1\leq j\leq n}|B_{n,j}(z)+C_{n,j}(z)| > \epsilon\right) = 0,$$

and $\max_{1\leq j\leq n}|B_{n,j}(z)+C_{n,j}(z)| \to 0$ almost surely because $\epsilon > 0$ is arbitrary.

Step III. Combining decomposition (1.22), estimate (1.24) and Step II, we have $\max_{1\leq j\leq n}|\upsilon_{n,j}| \to 0$ almost surely, and the proof is completed. □

1.4 Extreme Eigenvalues: Bai-Yin Theorem

The former studies only discuss the limiting spectral distribution of Wigner matrices. In practice, we are also interested in the extreme eigenvalues of random matrices.

Theorem 1.21 (Bai-Yin)**.** *Let W_n be an $n\times n$ complex Hermitian Wigner matrix, i.e. W_n is the topleft $n\times n$ block of the infinite matrix $(\xi_{ij})_{i,j=1}^{\infty}$. Assume that $\mathbb{E}|\xi_{11}^+|^2<\infty$ and $\mathbb{E}|\xi_{12}|^4<\infty$. Then*

$$\lim_{n\to\infty}\frac{\lambda_1(W_n)}{\sqrt{n}}=2,\quad \textit{almost surely.}$$

Remark. In addition, if $\mathbb{E}|\xi_{11}|^2<\infty$ and $\mathbb{E}|\xi_{12}|^4<\infty$, we can apply the above result to both W_n and $-W_n$ to get the asymptotic result of the operator norm $\|W_n\|_2=|\lambda_1(W_n)|\vee|\lambda_n(W_n)|$, which satisfies

$$\lim_{n\to\infty}\frac{\|W_n\|_2}{\sqrt{n}}=2,\quad \text{almost surely.}$$

By Wigner's semicircle law, for any $\epsilon>0$,

$$\frac{1}{n}N_{\frac{W_n}{\sqrt{n}}}([2-\epsilon,2])\to\int_{2-\epsilon}^{2}\rho_{\mathrm{sc}}(x)\,dx>0,\quad \text{almost surely.}$$

As a result, with probability 1, the number of eigenvalues of $W_n/\sqrt{n}$ greater than $2-\epsilon$ goes to ∞ as $n\to\infty$, and the maximum eigenvalue $\lambda_1(W_n/\sqrt{n})$ is greater than $2-\epsilon$. Since $\epsilon>0$ is arbitrary, we have

$$\liminf_{n\to\infty}\frac{\lambda_1(W_n)}{\sqrt{n}}\geq 2,\quad \text{almost surely.}$$

Therefore, to establish Theorem 1.21, it suffices to show that

$$\limsup_{n\to\infty}\frac{\lambda_1(W_n)}{\sqrt{n}}\leq 2,\quad \text{almost surely.}$$

Like the trick we used in previous subsections, we can remove the diagonal entries of Wigner matrices without changing the asymptotics of the largest eigenvalue.

Lemma 1.22. *Without loss of generality, one may assume all diagonal entries $\xi_{ii}=0$ in Theorem 1.21.*

Proof. Use Rayleigh quotient:

$$\begin{aligned}\lambda_1(W_n)&=\sup_{\|u\|=1}\sum_{i,j=1}^{n}z_i\bar{z}_j\xi_{ij}=\sup_{\|u\|=1}\left[\sum_{i\neq j}^{n}z_i\bar{z}_j\xi_{ij}+\sum_{i=1}^{n}\xi_{ii}|z_i|^2\right]\\&\leq\sup_{\|u\|=1}\sum_{i\neq j}^{n}z_i\bar{z}_j\xi_{ij}+\max_{i\in[n]}\xi_{ii}^+\leq\lambda_1(W_n^\circ)+\max_{i\in[n]}\xi_{ii}^+,\end{aligned}$$

where W_n° is obtained from setting diagonal entries of W_n to be 0. To generalize the result, it suffices to show that $\max_{i\in[n]}\xi_{ii}^+/\sqrt{n}\to 0$ almost surely, which implies $\lambda_1(W_n)-\lambda_1(W_n^\circ)=o(1)\sqrt{n}$. We take a dyadic sequence $n_m=2^m$, $m=1,2,\cdots$. By Fubini's theorem,

$$\begin{aligned}\sum_{m=0}^{\infty}\mathbb{P}\left(\max_{i\in[n_m]}\xi_{ii}^+\geq\epsilon\sqrt{n_m}\right)&\leq\sum_{m=0}^{\infty}\sum_{i=1}^{n_m}\mathbb{P}(\xi_{ii}^+\geq\epsilon\sqrt{n_m})\leq\mathbb{E}\left[\sum_{m=0}^{\infty}n_m\mathbb{1}_{\{\epsilon\sqrt{n_m}\leq\xi_{11}^+\}}\right]\\&=\mathbb{E}\left[\sum_{m:n_m\leq(\xi_{11}^+/\epsilon)^2}n_m\right]\leq\frac{2\mathbb{E}|\xi_{11}|^2}{\epsilon^2}<\infty.\end{aligned}$$

By the Borel-Cantelli theorem, with probability 1, we have $\max_{i\in[n_m]} \xi_{ii}^+ \le \epsilon\sqrt{n_m}$ for large enough m. Now for $n_{m-1} \le n \le n_m$,

$$\max_{i\in[n]} \xi_{ii}^+ \le \max_{i\in[n_m]} \xi_{ii}^+ \le \epsilon\sqrt{n_m} \le \epsilon\sqrt{2n}.$$

Therefore, with probability 1, one have

$$\frac{\max_{i\in[n]} \xi_{ii}^+}{\sqrt{n}} \le \sqrt{2}\,\epsilon$$

for large enough n. Since $\epsilon > 0$ is arbitrary, we have $\max_{i\in[n]} \xi_{ii}^+/\sqrt{n} \to 0$ almost surely. □

Lemma 1.23 (Improved moment bound)**.** *Let $W_n = (\xi_{ij})_{n\times n}$ be a random Hermitian matrix in $\mathbb{C}^{n\times n}$. Assume the upper triangular entries $(\xi_{ij})_{1\le i\le j\le n}$ satisfies:*

- *$(\xi_{ij})_{1\le i\le j}$ are jointly independent with mean 0 and variance bounded by 1;*
- *$\sup_{i,j\in[n]} \mathbb{E}|\xi_{ij}|^4 < \infty$; and*
- *$\sup_{i,j\in[n]} |\xi_{ij}| \le O(n^\delta)$ almost surely, where $0 < \delta < 1/2$.*

Let $k \in 2\mathbb{N}$ be a positive even integer of size $O(\log^2 n)$. Then

$$\mathbb{E}\left[\mathrm{tr}\left(W_n^k\right)\right] \le C_{k/2} n^{1+\frac{k}{2}} + O\left(2^k k^{22} n^{2\delta+\frac{k}{2}}\right). \tag{1.25}$$

Proof. By our previous discussion in §1.2.1,

$$\begin{aligned}
\mathbb{E}\left[\mathrm{tr}\left(W_n^k\right)\right] &= \sum_{i_1,\cdots,i_k\in[n]} \mathbb{E}\left[\xi_{i_1i_2}\xi_{i_2i_3}\cdots\xi_{i_{n-1}i_n}\xi_{i_ni_1}\right] \\
&= \sum_{s\in S_k^*} n(n-1)\cdots\left(n-\frac{k}{2}\right)\mathbb{E}\left[\xi_{i_1i_2}\xi_{i_2i_3}\cdots\xi_{i_{n-1}i_n}\xi_{i_ni_1}\right] + \sum_{s\in S_k^\circ}\sum_{(i_1,\cdots,i_k)\in\mathcal{I}_n^s} \mathbb{E}\left[\xi_{i_1i_2}\xi_{i_2i_3}\cdots\xi_{i_{n-1}i_n}\xi_{i_ni_1}\right] \\
&\le C_{k/2} n^{1+\frac{k}{2}} + \underbrace{\sum_{(i_1,\cdots,i_k)\in\mathcal{I}_n^s} \mathbb{E}\left[\xi_{i_1i_2}\xi_{i_2i_3}\cdots\xi_{i_{n-1}i_n}\xi_{i_ni_1}\right]}_{A_{n,k}},
\end{aligned}$$

where S_k^* is the set of shapes with $k/2+1$ vertices and $k/2$ edges, with each edge traversed exactly twice, and S_k° is the set of shapes with at most $k/2$ vertices, with each edge traversed at least twice.

We order the ℓ distinct edges $e_1, \cdots, e_\ell$ by their first appearance in the cycle $(i_1,\cdots,i_k)$, and let $\alpha_1,\cdots,\alpha_\ell$ be the multiplicity of these edges. Then the α_j's are all at least 2 and add up to k. By the moment hypothesis,

$$\mathbb{E}|\xi_{ij}|^\alpha \le \sqrt{\mathbb{E}|\xi_{ij}|^{2(\alpha-2)}\mathbb{E}|\xi_{ij}|^4} = n^{(\alpha-2)\delta}\sqrt{\mathbb{E}|\xi_{ij}|^4}.$$

Since $\alpha_1+\cdots+\alpha_\ell = k$, we have

$$\mathbb{E}\left[\xi_{i_1i_2}\xi_{i_2i_3}\cdots\xi_{i_ki_1}\right] \le O(n^\delta)^{k-2\ell}.$$

Let $N_{\alpha_1,\cdots,\alpha_\ell}$ be the number of cycles $(i_1,\cdots,i_k)$ with edge multiplicities $(\alpha_1,\cdots,\alpha_\ell)$. Then

$$A_{n,k} = \sum_{\ell=1}^{k/2-1} \sum_{\substack{\alpha_1,\cdots,\alpha_\ell\ge 2\\ \alpha_1+\cdots+\alpha_\ell=k}} N_{\alpha_1,\cdots,\alpha_\ell} O(n^\delta)^{k-2\ell}.$$

Given a cycle $(i_1,\cdots,i_k)$, one traverses its k steps one at a time. We use several classifications for the steps:

- *High-multiplicity steps,* which use an edge e_j with multiplicity $\alpha_j \ge 3$.
- *Fresh steps,* which use an edge e_j with multiplicity $\alpha_j = 2$ for the first time. We subdivide them into:
  - *Innovative steps,* which points at a vertex one has not visited before; and
  - *Non-innovative steps,* which points at a vertex one has visited before.

- *Return steps,* which use an edge e_j with multiplicity $\alpha_j = 2$ that is traversed by a previous fresh step. We subdivide them into:
  - *Forced steps,* which start from a vertex v such that, at the time one is performing that step, there is only one available edge from v; and
  - *Unforced steps,* otherwise.

We assume there are h high-multiplicity edges, leading to $\ell - h$ fresh steps and theor $\ell - h$ return step counterparts. Then the number of high multiplicity steps $\sum_{\alpha_j>2} \alpha_j = k - 2(\ell - h) \geq 3h$, and $h \leq k - 2\ell$.

We assume there are m non-innovative steps among the $\ell - h$ fresh steps, leaving $\ell - h - m$ innovative steps. Then we have either $\ell < k/2$ or $m > 0$.

Furthermore, at any given time point in traversing a cycle $(i_1, \cdots, i_k)$, one define an *available edge* to be an edge e_j with $\alpha_j = 2$ such that e_j is already traversed by its fresh step but not by its return step. Then at any given time, there are three cases:

- one travels along a high-multiplicity step;
- one explores a fresh step, thus creating a new available edge;
- one returns along an available edge, thus removing that edge from availability.

We assume there are r unforced return steps among the $\ell - h$ return steps. Let v be a vertex visited by the cycle which is not the initial vertex i_1. Then the very first arrival at v comes from a fresh step, which immediately becomes available. Each departure from v may create another available edge from v, but each subsequent arrival at v will delete an available step from v, unless the arrival is along a non-innovative or high-multiplicity edge. Finally, any return step starting from v will also delete an available edge from v.

This has two consequences. Firstly, if there are no non-innovative or high-multiplicity edges arriving at v, then whenever one arrives at v, there is at most one available edge from v, and so every return step from v is forced. (And there will be only one such return leg.) If instead there are non-innovative or high-multiplicity edges arriving at v, then we see that the number of return steps from v is at most one plus the number of such edges. In both cases, we conclude that the number of unforced return legs from v is bounded by twice the number of non-innovative or high-multiplicity edges arriving at v. Summing over v, we obtain that

$$r \leq 2\left(m + \sum_{\alpha_j>2} \alpha_j\right) = 2(m + k - 2\ell + 2h) \leq 2(m + 3k - 6\ell). \tag{1.26}$$

Now we count $N_{\alpha_1,\cdots,\alpha_\ell}$. We first fix m and r and record the corresponding cycles $(i_1, \cdots, i_k)$.

(i) There are n choices for the initial vertex i_1;

(ii) For each high-multiplicity edge e_j (in increasing order of j), allocate α_j locations in the cycle. There are at most $k^{\sum_{\alpha_j>2}\alpha_j} = k^{k-2(\ell-h)}$ choices.

(iii) Record the destination of (the first occurence of) e_j for each such j, creating n^h choices.

(iv) For each innovative fresh step, record the destination of that step, leading to an additional list of $\ell - h - m$ vertices with at most $n^{\ell-h-m}$ choices.

(v) For each non-innovative step, allocate a position in $\{1, \cdots, k\}$, creating k^m choices.

(vi) For each unforced return step, allocate a position in $\{1, \cdots, k\}$, creating k^r choices.

(vii) Finally, we record a simple random walk of length k, in which we set the difference $+1$ whenever the step is innovative, and -1 otherwise. This creates at most 2^k choices. Note that the positions of innovative steps and forced return steps are determined by (v), (vi) and this walk.

Together with h, m, r, one can reconstruct the original cycle $(i_1, \cdots, i_k)$ from the above data: as one traverses the cycle, the data already tells us which steps are high-multiplicity, which ones are innovative, which ones are non-innovative, and which ones are return steps. In all edges in which one could possibly visit a new vertex, the location of that vertex has been recorded. For all unforced returns, the data tells us which fresh step to backtrack upon to return to. Finally, for forced returns, there is only one available leg to backtrack to, and so one can reconstruct the entire cycle from this data. As a consequence, for fixed h, m and r, and by (1.26), there are at most

$$nk^{k-2(\ell-h)}n^h n^{\ell-h-m}k^m k^r 2^k = n^{1+\ell-m}k^{k+r-2(\ell-h)}2^k \leq n^{1+\ell-m}k^{2m+9(k-2\ell)}2^k$$

contributions to $N_{\alpha_1,\cdots,\alpha_\ell}$. Summing over the possible values of m, r, for $n > 2k^2$, we have

$$N_{\alpha_1,\cdots,\alpha_\ell} \leq 2^k n^{1+\ell}k^{9(k-2\ell)}\sum_{h=0}^{k-2\ell}\sum_{m=0}^{\ell-h}\frac{k^{2m}}{n^m} = 2^{1+k}n^{1+\ell}k^{9(k-2\ell)+1},$$

For $\ell < k/2$, the numbers $\alpha_1 - 1, \cdots, \alpha_\ell - 1$ are positive integers and add up to $k-\ell$. Then there are $\binom{k-\ell}{\ell}$ solutions of $(\alpha_1, \cdots, \alpha_\ell)$ in total, and

$$\begin{aligned}
A_{n,k} &= 2^{1+k}\sum_{\ell=1}^{k/2-1}O(n^\delta)^{k-2\ell}n^{1+\ell}k^{9(k-2\ell)+1}\sum_{\substack{\alpha_1,\cdots,\alpha_\ell\geq 2\\ \alpha_1+\cdots+\alpha_\ell=k}}1 \leq 2^{1+k}\left(\sum_{\ell=1}^{k/2-1}O(n^\delta)^{k-2\ell}n^{1+\ell}k^{10(k-2\ell)+1}\right)\\
&\leq 2^{1+k}k^{1+10k}n^{1+k\delta}O\left(\sum_{\ell=1}^{k/2-1}n^{(1-2\delta)\ell}k^{-20\ell}\right) \leq 2^{1+k}k^{1+10k}n^{1+k\delta}\left(\frac{k}{2}-1\right)O\left(n^{1-2\delta}k^{-20}\right)^{k/2-1}\\
&\leq O\left(2^k k^{22}n^{2\delta+k/2}\right).
\end{aligned}$$

Thus we finish the proof of the improved moment bound (1.25). □

Proof of Theorem 1.21. By Lemma 1.22, we assume that the diagonal entries (ξ_{ii}) are identically zero.

Step I. We pick $\delta = 0.49 \in (0, \frac{1}{2})$ and split each

$$\xi_{ij} = \xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\leq n^\delta\}} + \xi_{ij}\mathbb{1}_{\{|\xi_{ij}|>n^\delta\}} =: \widehat{\xi}_{ij} + \widetilde{\xi}_{ij},$$

and split $W_n = \widehat{W}_n + \widetilde{W}_n$ accordingly. Clearly,

$$\left|\mathbb{E}\widehat{\xi}_{ij}\right| = \left|\mathbb{E}\widetilde{\xi}_{ij}\right| \leq n^{-3\delta}\mathbb{E}\left|\widetilde{\xi}_{ij}\right|^4 \leq n^{-3\delta}\mathbb{E}\left|\xi_{ij}\right|^4.$$

Then

$$\left\|\mathbb{E}\widehat{W}_n u\right\|_2^2 \leq \sum_{i=1}^n\left(\sum_{j=1}^n\left|\mathbb{E}\widehat{\xi}_{ij}\right||u_j|\right)^2 \leq O(n^{-6\delta})\sum_{i=1}^n\left(\sum_{j=1}^n|u_j|\right)^2 = O(n^{2-6\delta}),$$

and consequently $\|\mathbb{E}\widehat{W}_n\|_2 = O(n^{1-3\delta})$, which is smaller than $\epsilon/3$ for large enough n.

Step II. We write $\widehat{W}_n = \overline{W}_n + \mathbb{E}\widehat{W}_n$. For large enough even number $n \in 2\mathbb{N}$ and $k = \lceil\log^2 n\rceil$, by Lemma 1.23,

$$\mathbb{E}\left[\mathrm{tr}(\overline{W}_n^k)\right] \leq C_{k/2}n^{1+\frac{k}{2}} + O\left(2^k k^{22}n^{2\delta+\frac{k}{2}}\right) \leq 2^k n^{1+\frac{k}{2}} + O\left(2^k k^{22}n^{2\delta+\frac{k}{2}}\right) = 2^k O(n^{1+\frac{k}{2}}).$$

Note that $\lambda_1(\overline{W}_n)^k \leq \mathrm{tr}(\overline{W}_n^k) = \lambda_1(\overline{W}_n)^k + \cdots + \lambda_n(\overline{W}_n)^k$. By Markov's inequality,

$$\mathbb{P}\left(\lambda_1(\overline{W}_n) > \left(2+\frac{\epsilon}{3}\right)\sqrt{n}\right) = \left(\frac{2}{2+\epsilon/3}\right)^k O(n) \leq O(n^{1-\log(1+\epsilon/6)\cdot\log n}).$$

Note the series $\sum_{n=1}^\infty n^{1-c\log n}$ converges for each $c > 0$. By the Borel-Cantelli lemma, with probability, for large enough n,

$$\lambda_1(\overline{W}_n) \leq \left(2 + \frac{\epsilon}{3}\right)\sqrt{n}.$$

Step III. To control $\widetilde{W}_n$, we use dyadic sparsification. Take $n_m = 2^m$, $m = 1, 2, \cdots$. We first prove that the entries W_{n_m} is asymptotically almost surely bounded by $O(\sqrt{n_m})$. Note that

$$\mathbb{P}\left(|\xi_{ij}| \geq \frac{\epsilon}{6}\sqrt{n}\right) \leq \frac{1296}{n^2\epsilon^4}\mathbb{E}\left|\xi_{ij}\mathbb{1}_{\{|\xi_{ij}|\geq\epsilon\sqrt{n}/3\}}\right|^4, \quad i < j.$$

By Fubini's theorem,

$$\begin{aligned}\sum_{m=0}^{\infty}\mathbb{P}\left(\max_{i,j\in[n_m]}|\xi_{ij}| \geq \frac{\epsilon}{6}\sqrt{n_m}\right) &\leq \sum_{m=0}^{\infty} n_m^2\mathbb{P}\left(|\xi_{12}| \geq \frac{\epsilon}{6}\sqrt{n_m}\right) \leq \mathbb{E}\left[\sum_{m=0}^{\infty} n_m^2\mathbb{1}_{\{|\xi_{12}|\geq\epsilon\sqrt{n_m}/6\}}\right] \\ &= \mathbb{E}\left[\sum_{m:n_m\leq 36|\xi_{12}|^2/\epsilon^2} n_m^2\right] \leq c_\epsilon \cdot \left(\mathbb{E}|\xi_{12}|^4\right) < \infty,\end{aligned}$$

where the last inequality follows from that for all $q > 1$ and $K \geq 1$,

$$\sum_{m:q^m\leq K} q^m \leq \frac{Kq}{q-1},$$

and $c_\epsilon > 0$ is a constant depending only on ϵ. By the Borel-Cantelli lemma, with probability 1, for large enough m, all entries of $\widetilde{W}_{n_m}$ are bounded by $\epsilon\sqrt{n_m}/3$.

Now we exploit the sparseness of $\widetilde{W}_n$ to control its operator norm. By Markov's inequality and the fourth moment condition, each entry has a probability $O(n^{-4\delta})$ of being zero. Consequently, the probability that at least one column or row of $\widetilde{W}_n$ has two nonzero entries is at most

$$n^2 \cdot O(n^{-4\delta})^2 = O(n^{2-8\delta})$$

Note the geometric series $\sum_{m=0}^{\infty} n_m^{2-8\delta} < \infty$. By the Borel-Cantelli lemma, with probability 1, for large enough m, all columns and rows of the matrix $\widetilde{W}_{n_m}$ has at most one nonzero entry, bounded by $\epsilon\sqrt{n_m}/6$, and

$$\left\|\widetilde{W}_{n_m}\right\|_2 = \sup_{\|u\|_2=1}\left\|\widetilde{W}_{n_m}u\right\|_2 \leq \frac{\epsilon}{6}\sqrt{n_m}.$$

By Cauchy's interlacing theorem,

$$\left\|\widetilde{W}_n\right\|_2 \leq \left\|\widetilde{W}_{n_m}\right\|_2 \leq \frac{\epsilon}{6}\sqrt{n_m} \leq \frac{\epsilon}{3}\sqrt{n}.$$

Step IV. By the last three steps and Weyl's inequality, with probability 1, we have

$$\lambda_1(W_n) \leq \lambda_1(\overline{W}_n) + \left\|\mathbb{E}\widehat{W}_n\right\|_2 + \left\|\widetilde{W}_n\right\|_2 \leq (2+\epsilon)\sqrt{n}$$

for large enough n. Since $\epsilon > 0$ is arbitrary, we take $\epsilon \downarrow 0$ and conclude that

$$\limsup_{n\to\infty}\frac{\lambda_1(W_n)}{\sqrt{n}} \leq 2.$$

Combining this result with Wigner's semicircle law, we finish the proof. □

2 Sample Covariance Matrices and the Marčenko-Pastur Law

General setting. Suppose we observe n independent samples of an m-dimensional feature vector $x^{(j)} = (x_{1j}, \cdots, x_{mj})^\top$, and arrange them as the columns of a data matrix $X_n \in \mathbb{C}^{m\times n}$, i.e. $X_n = [x^{(1)}, \cdots, x^{(n)}]$ with (real or complex) entries (x_{ij}) that are i.i.d., with zero mean, and unit variance. The *sample covariance* is the $m \times m$ Hermitian matrix

$$S_n = \frac{1}{n}\sum_{j=1}^{n} x_j x_j^* = \frac{1}{n} X_n X_n^*.$$

We study the empirical spectral distribution (ESD) of S_n, denoted μ_{S_n}, which places mass $1/m$ at each eigenvalue of S_n. In the high-dimensional regime, we work on a proportional asymptotics model: both the feature dimension $m = m_n$ and the sample size n grow, and their ratio

$$\frac{m_n}{n} \to \alpha \in (0, \infty) \quad \text{as } n \to \infty.$$

where α is called the *aspect ratio.* Note that when $\alpha > 1$, S_n has rank at most $n < m$, so a proportion $1 - 1/\alpha$ of its eigenvalues are exactly zero in the limit.

Theorem 2.1 (Marčenko-Pastur Law)**.** *Suppose that $(x_{ij})_{i,j=1}^{\infty}$ are i.i.d. complex random variables with mean zero and variance 1, and $X_n = (x_{ij})_{i\in[m_n], j\in[n]}$. Also assume that $m_n/n \to \alpha \in (0, \infty)$. Then*

$$\mu_{\frac{1}{n}X_nX_n^*} \to \mu_{\mathrm{MP}_\alpha} \quad \textit{weakly almost surely,}$$

*where μ_{MP_α} is the **Marčenko-Pastur distribution**, which has density function*

$$\rho_{\mathrm{MP}_\alpha}(x) = \frac{\sqrt{(\lambda_{\alpha+} - x)(x - \lambda_{\alpha-})}}{2\pi\alpha x}\mathbb{1}_{[\lambda_{\alpha-},\lambda_{\alpha+}]}(x), \quad x \in \mathbb{R},$$

and has an atom of mass $1 - 1/\alpha$ at the origin if $\alpha > 1$, where $\lambda_{\alpha-} = (1 - \sqrt{\alpha})^2$ and $\lambda_{\alpha+} = (1 + \sqrt{\alpha})^2$. Here, the constant $\alpha \in (0, \infty)$ is the aspect ratio.

Remark. By the Portmanteau lemma, the proportion of eigenvalues of $\frac{1}{n}X_nX_n^*$ in $[a, b]$ is

$$\frac{1}{m_n}N_{[a,b]}\left(\frac{1}{n}X_nX_n^*\right) \to \int_{a\vee\lambda_{\alpha-}}^{b\wedge\lambda_{\alpha+}} \frac{\sqrt{(\lambda_{\alpha+} - x)(x - \lambda_{\alpha-})}}{2\pi\alpha x}\, dx + \left(1 - \frac{1}{\alpha}\right)_+ \mathbb{1}_{\{0\in[a,b]\}}, \quad \text{almost surely.}$$

2.1 The Resolvent Method

2.1.1 Reduction to the Bounded Case

For covariance matrices, we have a rank perturbation result similar to Lemma 1.12.

Lemma 2.2 (Low rank purterbation)**.** *Let $A, B \in \mathbb{C}^{m\times n}$. Then*

$$\|F_{BB^*} - F_{AA^*}\|_\infty \le \frac{\mathrm{rank}(A - B)}{m}, \tag{2.1}$$

where F_{AA^*}, F_{BB^*} are cumulative distribution functions of ESDs of the corresponding covariance matrices.

Proof. We let $D = A - B$, and write $\mathrm{rank}(D) = k$, and by Weyl's inequality,

$$\sigma_{j+k+1}(A) \le \sigma_{j+1}(B), \quad \sigma_{j+k+1}(B) \le \sigma_{j+1}(A), \quad j = 0, \cdots, \min\{m, n\} - k - 1.$$

Then for all $x \in [\sigma_{j+1}(B), \sigma_j(B))$,

$$F_{BB^*}(x) = 1 - \frac{j}{m} = 1 - \frac{j+k}{m} + \frac{k}{m} \le F_{AA^*}(x) + \frac{k}{m}.$$

In fact, this implies

$$F_{BB^*}(x) - F_{AA^*}(x) \le \frac{k}{m}, \quad \text{for all } x \in \mathbb{R}.$$

Similarly, $F_{AA^*}(x) - F_{BB^*}(x) \le k/m$ for all $x \in \mathbb{R}$. This completes the proof of (2.1). □

Here is another estimate similar to Lemma 1.11.

Lemma 2.3. *Let $A, B \in \mathbb{C}^{m \times n}$. Then*

$$\rho_L(F_{AA^*}, F_{BB^*})^4 \le \frac{2\operatorname{tr}(AA^* + BB^*)}{m^2} \|A - B\|_{\mathrm{F}}^2.$$

Proof. By Lemma 1.11, the Cauchy-Schwartz inequality and the Hoffman-Wielandt inequality,

$$\begin{aligned}
\rho_L(F_{AA^*}, F_{BB^*})^2 &\le \frac{1}{m} \sum_{j=1}^m \left|\sigma_j(A)^2 - \sigma_j(B)^2\right| \\
&\le \frac{1}{m} \left(\sum_{j=1}^m |\sigma_j(A) + \sigma_j(B)|^2 \right)^{1/2} \left(\sum_{j=1}^m |\sigma_j(A) - \sigma_j(B)|^2 \right)^{1/2} \\
&\le \frac{1}{m} \left(2\sum_{j=1}^m \sigma_j(A)^2 + 2\sum_{j=1}^m \sigma_j(B)^2 \right)^{1/2} \left(\sum_{j=1}^m |\sigma_j(A) - \sigma_j(B)|^2 \right)^{1/2} \\
&\le \frac{1}{m} \left(2\operatorname{tr}(AA^* + BB^*)\right)^{1/2} \|A - B\|_{\mathrm{F}}.
\end{aligned}$$

Then we finish the proof. □

Lemma 2.4. *In Theorem 2.1, one may assume without loss of generality that the i.i.d. variables $(x_{ij})_{i,j=1}^\infty$ are bounded.*

Proof. We define

$$\overline{x}_{ij} = x_{ij} \mathbb{1}_{\{|x_{ij}| \le N\}}, \quad \widehat{x}_{ij} = \overline{x}_{ij} - \mathbb{E}\left[\overline{x}_{ij}\right], \quad \widetilde{x}_{ij} = \frac{\widehat{x}_{ij}}{\sqrt{\mathbb{E}|\widehat{x}_{ij}|^2}}, \qquad i, j = 1, 2, \cdots.$$

and set $\overline{X}_n = (\overline{x}_{ij})_{i \in [m_n], j \in [n]}$, $\widehat{X}_n = (\widehat{x}_{ij})_{i \in [m_n], j \in [n]}$, and $\widetilde{X}_n = (\widetilde{x}_{ij})_{i \in [m_n], j \in [n]}$. By Lemma 2.3,

$$\begin{aligned}
\rho_L\left(F_{\frac{1}{n}X_nX_n^*}, F_{\frac{1}{n}\overline{X}_n\overline{X}_n^*}\right)^4 &\le \frac{2}{m_n^2 n^2} \left(\|X_n\|_{\mathrm{F}}^2 + \|\overline{X}_n\|_{\mathrm{F}}^2\right) \left\|X_n - \overline{X}_n\right\|_{\mathrm{F}}^2 \\
&= \left[\frac{2}{m_n n} \sum_{i=1}^{m_n} \sum_{j=1}^n \left(|x_{ij}|^2 + |x_{ij}|^2 \mathbb{1}_{\{|x_{ij}| \le N\}}\right) \right] \left[\frac{1}{m_n n} \sum_{i=1}^{m_n} \sum_{j=1}^n |x_{ij}|^2 \mathbb{1}_{\{|x_{ij}| > N\}} \right] \\
&\le \left[\frac{4}{m_n n} \sum_{i=1}^{m_n} \sum_{j=1}^n |x_{ij}|^2 \right] \left[\frac{1}{m_n n} \sum_{i=1}^{m_n} \sum_{j=1}^n |x_{ij}|^2 \mathbb{1}_{\{|x_{ij}| > N\}} \right] \\
&\to 4\mathbb{E}\left[|x_{11}|^2 \mathbb{1}_{\{|x_{11}| > N\}}\right] \quad \text{almost surely.}
\end{aligned}$$

By Lemma 2.2,

$$\rho_L\left(F_{\frac{1}{n}\overline{X}_n^*\overline{X}_n^*}, F_{\frac{1}{n}\widehat{X}_n^*\widehat{X}_n^*}\right) \le \frac{\operatorname{rank}(\mathbb{E}\overline{X}_n)}{m_n} = \frac{1}{m_n},$$

which converges to 0 deterministically. Finally,

$$\begin{aligned}
\rho_L\left(F_{\frac{1}{n}\widehat{X}_n\widehat{X}_n^*}, F_{\frac{1}{n}\widetilde{X}_n\widetilde{X}_n^*}\right)^4 &\le \frac{2}{m_n^2 n^2}\left(\|\widehat{X}_n\|_{\mathrm{F}}^2 + \|\widetilde{X}_n\|_{\mathrm{F}}^2\right)\left\|\widehat{X}_n - \widetilde{X}_n\right\|_{\mathrm{F}}^2 \\
&= 2\left[\frac{1+\mathbb{E}|\widehat{x}_{11}|^2}{m_n n\,\mathbb{E}|\widehat{x}_{11}|^2}\sum_{i=1}^{m_n}\sum_{j=1}^{n}|\widehat{x}_{ij}|^2\right]\left[\frac{\left(1-\sqrt{\mathbb{E}|\widehat{x}_{11}|^2}\right)^2}{m_n n\,\mathbb{E}|\widehat{x}_{11}|^2}\sum_{i=1}^{m_n}\sum_{j=1}^{n}|\widehat{x}_{ij}|^2\right] \\
&\to 2\left(1+\mathrm{Var}(x_{11}\mathbb{1}_{\{|x_{11}|\le N\}})\right)\left(1-\sqrt{\mathrm{Var}(x_{11}\mathbb{1}_{\{|x_{11}|\le N\}})}\right)^2 \quad \text{almost surely.}
\end{aligned}$$

Note that as $N \uparrow \infty$, both $\mathbb{E}\left[|x_{11}|^2\mathbb{1}_{\{|x_{11}|>N\}}\right]$ and $1-\sqrt{\mathrm{Var}\big(x_{11}\mathbb{1}_{\{|x_{11}|\le N\}}\big)}$ converges to 0. Hence given any $\epsilon > 0$, we can find $N > 0$ large enough such that

$$\limsup_{n\to\infty}\rho_L\left(F_{\frac{1}{n}X_nX_n^*}, F_{\frac{1}{n}\widetilde{X}_n\widetilde{X}_n^*}\right) < \epsilon \quad \text{almost surely.}$$

If Theorem 2.1 holds for sample covariance matrices of bounded random variables, then $F_{\frac{1}{n}\widetilde{X}_n\widetilde{X}_n^*} \to F_{\mathrm{MP}_\alpha}$ for all $N > 0$, and

$$\limsup_{n\to\infty}\rho_L\left(F_{\frac{1}{n}X_nX_n^*}, F_{\mathrm{MP}_\alpha}\right) < \epsilon \quad \text{almost surely.}$$

Since $\epsilon > 0$ is arbitrary, we conclude that $F_{\frac{1}{n}X_nX_n^*} \to F_{\mathrm{MP}_\alpha}$ weakly almost surely and complete the proof. □

2.1.2 The Steiltjes transform of the Marčenko-Pastur Law

Now we can derive the limiting distribution of the sample covariance matrix $\frac{1}{n}X_nX_n^*$ by using the Stieltjes transform. To begin with, we compute the Stieltjes transform of the Marčenko-Pastur distribution.

Lemma 2.5. *Let $\alpha > 0$. The Stieltjes transform of the M-P distribution is*

$$s_{\mathrm{MP}_\alpha}(z) = \frac{1-\alpha-z+\sqrt{(1-\alpha-z)^2-4\alpha z}}{2\alpha z}, \quad z \in \mathbb{C}^+.$$

Proof. Fix $z \in \mathbb{C}^+$. If $\alpha < 1$,

$$s_{\mathrm{MP}_\alpha}(z) = \int_{\lambda_{\alpha-}}^{\lambda_{\alpha+}} \frac{\sqrt{(x-\lambda_{\alpha-})(\lambda_{\alpha+}-x)}}{2\pi\alpha(x-z)x}\,dx,$$

where $\lambda_{\alpha-} = (1-\sqrt{\alpha})^2$ and $\lambda_{\alpha+} = (1+\sqrt{\alpha})^2$. We let $x = 1+\alpha+2\sqrt{\alpha}\cos\theta$. Then

$$\begin{aligned}
s_{\mathrm{MP}_\alpha}(z) &= \frac{2}{\pi}\int_0^{\pi}\frac{\sin^2\theta}{(1+\alpha+2\sqrt{\alpha}\cos\theta-z)(1+\alpha+2\sqrt{\alpha}\cos\theta)}\,d\theta \\
&= \frac{1}{\pi}\int_0^{2\pi}\frac{\left(\frac{e^{i\theta}-e^{-i\theta}}{2i}\right)^2}{(1+\alpha+\sqrt{\alpha}(e^{i\theta}+e^{-i\theta})-z)(1+\alpha+\sqrt{\alpha}(e^{i\theta}+e^{-i\theta}))}\,d\theta \\
&= -\frac{1}{4\pi i}\int_{|\zeta|=1}\frac{\left(\zeta-\zeta^{-1}\right)^2}{(1+\alpha+\sqrt{\alpha}(\zeta+\zeta^{-1})-z)(1+\alpha+\sqrt{\alpha}(\zeta+\zeta^{-1}))}\zeta^{-1}\,d\zeta \\
&= -\frac{1}{4\pi i}\int_{|\zeta|=1}\frac{\left(\zeta^2-1\right)^2}{\zeta((1+\alpha-z)\zeta+\sqrt{\alpha}(\zeta^2+1))((1+\alpha)\zeta+\sqrt{\alpha}(\zeta^2+1))}\,d\zeta.
\end{aligned}$$

This integrand has five simple poles at

$$\zeta_0 = 0, \quad \zeta_1 = -\sqrt{\alpha}, \quad \zeta_2 = -\frac{1}{\sqrt{\alpha}},$$
$$\zeta_3 = \frac{-(1+\alpha-z)+\sqrt{(1+\alpha-z)^2-4\alpha}}{2\sqrt{\alpha}}, \quad \zeta_4 = \frac{-(1+\alpha-z)-\sqrt{(1+\alpha-z)^2-4\alpha}}{2\sqrt{\alpha}}.$$

The residuals at these poles are

$$\mathrm{Res}(\zeta_0) = \frac{\left(\zeta_0^2-1\right)^2}{((1+\alpha-z)\zeta_0+\sqrt{\alpha}(\zeta_0^2+1))((1+\alpha)\zeta_0+\sqrt{\alpha}(\zeta_0^2+1))} = \frac{1}{\alpha},$$
$$\mathrm{Res}(\zeta_1) = \frac{\left(\zeta_1^2-1\right)^2}{\alpha\zeta_1(\zeta_1-\zeta_2)(\zeta_1-\zeta_3)(\zeta_1-\zeta_4)} = \frac{\zeta_1(\zeta_1-\zeta_2)}{\alpha(\zeta_1-\zeta_3)(\zeta_1-\zeta_4)} = -\frac{1-\alpha}{\alpha z},$$
$$\mathrm{Res}(\zeta_2) = \frac{\left(\zeta_2^2-1\right)^2}{\alpha\zeta_2(\zeta_2-\zeta_1)(\zeta_2-\zeta_3)(\zeta_2-\zeta_4)} = \frac{\zeta_2(\zeta_2-\zeta_1)}{\alpha(\zeta_2-\zeta_3)(\zeta_2-\zeta_4)} = \frac{1-\alpha}{\alpha z},$$
$$\mathrm{Res}(\zeta_3) = \frac{\left(\zeta_3^2-1\right)^2}{\alpha\zeta_3(\zeta_3-\zeta_1)(\zeta_3-\zeta_2)(\zeta_3-\zeta_4)} = \frac{\zeta_3(\zeta_3-\zeta_4)}{\alpha(\zeta_3-\zeta_1)(\zeta_3-\zeta_2)} = \frac{\sqrt{(1+\alpha-z)^2-4\alpha}}{\alpha z},$$
$$\mathrm{Res}(\zeta_4) = \frac{\left(\zeta_4^2-1\right)^2}{\alpha\zeta_4(\zeta_4-\zeta_1)(\zeta_4-\zeta_2)(\zeta_4-\zeta_3)} = \frac{\zeta_4(\zeta_4-\zeta_3)}{\alpha(\zeta_4-\zeta_1)(\zeta_4-\zeta_2)} = -\frac{\sqrt{(1+\alpha-z)^2-4\alpha}}{\alpha z}.$$

Recalling (1.21), we know that $\mathrm{Re}\sqrt{(1+\alpha-z)^2-4\alpha}$ and $\mathrm{Re}(z-1-\alpha)$ has the same sign, and $|\zeta_3| > |\zeta_4|$. Note that $\zeta_3\zeta_4 = 1$, we have $|\zeta_3| > 1 > |\zeta_4|$. Hence poles ζ_0, ζ_1 and ζ_4 are inside the contour $\{\zeta \in \mathbb{C} : |\zeta| = 1\}$, and their residues should be counted into the integral. By Cauchy's residue theorem,

$$s_{\mathrm{MP}_\alpha}(z) = -\frac{1}{2}\left(\frac{1}{\alpha} - \frac{1-\alpha}{\alpha z} - \frac{\sqrt{(1+\alpha-z)^2-4\alpha}}{\alpha z}\right) = \frac{1-\alpha-z+\sqrt{(1+\alpha-z)^2-4\alpha}}{2\alpha z}.$$

If $\alpha = 1$, the above contour integral becomes

$$s_{\mathrm{MP}_\alpha}(z) = -\frac{1}{4\pi i}\int_{|\zeta|=1} \frac{\left(\zeta^2-1\right)^2}{\zeta((2-z)\zeta+\zeta^2+1)(2\zeta+\zeta^2+1)}\,d\zeta = -\frac{1}{4\pi i}\int_{|\zeta|=1} \frac{\left(\zeta-1\right)^2}{\zeta((2-z)\zeta+\zeta^2+1)}\,d\zeta.$$

The integrand only has three simple poles at ζ_0, ζ_3 and ζ_4. Since the poles ζ_0 and ζ_4 are inside the contour $\{\zeta \in \mathbb{C} : |\zeta| = 1\}$, we have

$$s_{\mathrm{MP}_\alpha}(z) = -\frac{1}{2}\left(1 - \frac{\sqrt{(2-z)^2-4}}{z}\right) = \frac{-z+\sqrt{(2-z)^2-4}}{2z}.$$

Finally, if $\alpha > 1$, the M-P distribution has a point mass $1 - 1/\alpha$ at zero, and

$$s_{\mathrm{MP}_\alpha}(z) = \int_{\lambda_{\alpha-}}^{\lambda_{\alpha+}} \frac{\sqrt{(x-\lambda_{\alpha-})(\lambda_{\alpha+}-x)}}{2\pi\alpha(x-z)x}\,dx - \frac{1}{z}\left(1-\frac{1}{\alpha}\right),$$

We can apply the same contour integral trick as in the case $\alpha < 1$, except in this case the poles ζ_0, ζ_2 and ζ_4 are inside the contour $\{\zeta \in \mathbb{C} : |\zeta| = 1\}$. Then

$$s_{\mathrm{MP}_\alpha}(z) = -\frac{1}{2}\left(\frac{1}{\alpha} + \frac{1-\alpha}{\alpha z} - \frac{\sqrt{(1+\alpha-z)^2-4\alpha}}{\alpha z}\right) - \frac{\alpha-1}{\alpha z} = \frac{1-\alpha-z+\sqrt{(1+\alpha-z)^2-4\alpha}}{2\alpha z}.$$

Note that $\sqrt{(1+\alpha-z)^2-4\alpha} = \sqrt{(1-\alpha-z)^2-4\alpha z}$. Then we complete the proof. $\square$

Like the proof of the semicircle law, we note that $s_{\mathrm{MP}_\alpha}(z)$ is a fixed point of the function

$$\mathbb{C}\backslash\left\{\frac{1-\alpha-z}{\alpha z}\right\} \to \mathbb{C}: \ s \mapsto \frac{1}{1-\alpha-z-\alpha z s}, \tag{2.2}$$

Then we can establish the convergence result through the following lemma.

Lemma 2.6. *Let $\alpha > 0$ and $z \in \mathbb{C}^+$. Then the function (2.2) has two fixed points*

$$s_{+,-} = \frac{1-\alpha-z \pm \sqrt{(1-\alpha-z)^2 - 4\alpha z}}{2\alpha z}.$$

(i) $\operatorname{Im}(1-\alpha-z-\alpha z s_-) \geq -\operatorname{Im}(z/2)$.

(ii) If ν is a probability distribution supported on $[0,\infty)$, then

$$\operatorname{Im}(1-\alpha-z-\alpha z s_\nu(z)) \leq -\operatorname{Im} z. \tag{2.3}$$

(iii) Let μ_n be the ESD of the covariance matrix $\frac{1}{n}X_n X_n^*$, and let S_n be the Stieltjes transform of μ_n. If

$$s_n(z) - \frac{1}{1-\alpha_n - z - \alpha_n z s_n(z)} \to 0 \quad \textit{almost surely,}$$

where $\alpha_n = m_n/n$ for $n \in \mathbb{N}$, then $s_n(z) \to s_+ = s_{\mathrm{MP}_\alpha}(z)$ almost surely.

Proof. (i) Note that the square root always has nonnegative imaginary part. Then

$$\operatorname{Im}(1-\alpha-z-\alpha z s_-) = \operatorname{Im}\left(\frac{1-\alpha-z+\sqrt{(1-\alpha-z)^2-4\alpha z}}{2}\right) \geq -\frac{\operatorname{Im} z}{2}.$$

(ii) We let $z = E + i\eta$, where $\eta > 0$. Then

$$\operatorname{Im}(z s_\nu(z)) = \eta \operatorname{Re} s_\nu(z) + \xi \operatorname{Im} s_\nu(z) = \eta \int_0^\infty \frac{x}{(x-\xi)^2+\eta^2}\, d\nu(x).$$

Then

$$\operatorname{Im}(1-\alpha-z-\alpha z s_\nu(z)) = -\eta\left(1+\alpha\int_0^\infty \frac{x}{(x-\xi)^2+\eta^2}\, d\nu(x)\right) \leq -\eta,$$

which is (2.3). In particulr, if $s_\nu(z)$ is a fixed point of (2.2), it equals $s_+ = s_{\mathrm{MP}_\alpha}(z)$.

(iii) We fix our discussion on an event of probability 1 on which $s_n(z) + 1/(1-\alpha_n - z - \alpha_n z s_n(z)) \to 0$. To show $s_n(z) \to s_{\mathrm{MP}_\alpha}(z)$ a.s., it suffices to establish the convergence in deterministic case. Since $|s_n(z)| \leq |\operatorname{Im} z|^{-1}$, the sequence $(s_n(z))$ is a bounded and has a convergent subsequence $(s_{n_k}(z))$ by Bolzano-Weierstrass theorem. Also, as $n \to \infty$ we have $\alpha_n = m_n/n \to \alpha$, hence the limit $s = \lim_{k\to\infty} s_{n_k}(z)$ satisfies

$$s + \frac{1}{1-\alpha-z-\alpha z s} = 0, \quad \text{and} \quad s \in \{s_+, s_-\}.$$

Now we select the correct branch. Since $X_n X_n^*/n$ is positive-semidefinite, its ESD is supported on $[0,\infty)$, and $\operatorname{Im}(1-\alpha-z-\alpha z s_n(z)) \leq -\operatorname{Im} z$ by Lemma 2.6 (ii). Let $n = n_k \to \infty$, we have $\operatorname{Im}(1-\alpha-z-\alpha z s) \leq -\operatorname{Im} z$, and by (i),

$$s = s_+ = \frac{1-\alpha-z \pm \sqrt{(1-\alpha-z)^2-4\alpha z}}{2\alpha z} = s_{\mathrm{MP}_\alpha}(z).$$

Essentially, we show that every subsequence of $(s_n(z))$ has a further subsequence converging to $s_{\mathrm{MP}_\alpha}(z)$. Therefore $s_n(z) \to s_{\mathrm{MP}_\alpha}(z)$, and we finish the proof. $\square$

2.1.3 Proof of the Marčenko-Pastur Law

Now we use the Siteltjes transform to prove the Marčenko-Pastur law. By Lemma 2.4, we may assume that $|\xi_{11}| \le R$ for some $R \in (0,\infty)$,

Proof of Theorem 2.1. Step I. We denote by $X_{n,-j}$ the $(m_n-1)\times n$ matrix obtained from X_n by removing the j-th row, and denote by $\rho_{n,j} \in \mathbb{R}^n$ the removed row. Then we write

$$\frac{1}{n}X_nX_n^* - z\,\mathrm{Id} = \begin{bmatrix} \frac{1}{n}X_{n,-j}X_{n,-j}^* - z\,\mathrm{Id} & \frac{1}{n}X_{n,-j}\rho_{n,j} \\ \frac{1}{n}\rho_{n,j}^*X_{n,-j}^* & \frac{1}{n}\rho_{n,j}^*\rho_{n,j} - z \end{bmatrix}$$

Using Schur's complement, we have

$$\left[\left(\frac{1}{n}X_nX_n^* - z\,\mathrm{Id}\right)^{-1}\right]_{jj} = \frac{1}{\frac{1}{n}\rho_{n,j}^*\rho_{n,j} - z - \frac{1}{n^2}\rho_{n,j}^*X_{n,-j}^*\left(\frac{1}{n}X_{n,-j}X_{n,-j}^* - z\,\mathrm{Id}\right)^{-1}X_{n,-j}\rho_{n,j}}.$$

We let $\theta_{n,j} = \frac{1}{n}\rho_{n,j}^*\rho_{n,j} - z - \frac{1}{n^2}\rho_{n,j}^*X_{n,-j}^*\left(\frac{1}{n}X_{n,-j}X_{n,-j}^* - z\,\mathrm{Id}\right)^{-1}X_{n,-j}\rho_{n,j}$. Then

$$\begin{aligned} s_n(z) = \frac{1}{m_n}\sum_{j=1}^{m_n}\frac{1}{\theta_{n,j}} &= \frac{1}{m_n}\sum_{j=1}^{m_n}\left[\frac{1}{1-\alpha_n - z - \alpha_n z s_n(z)} - \frac{\theta_{n,j} - 1 + \alpha_n + z + \alpha_n z s_n(z)}{(1-\alpha_n - z - \alpha_n z s_n(z))\,\theta_{n,j}}\right] \\ &= \frac{1}{1-\alpha_n - z - \alpha_n z s_n(z)} - \delta_n(z), \end{aligned}$$

where

$$\delta_n(z) = \frac{1}{m_n}\sum_{j=1}^{m}\frac{\upsilon_{n,j}}{(1-\alpha_n - z - \alpha_n z s_n(z))(1-\alpha_n - z - \alpha_n z s_n(z) + \upsilon_{n,j})}, \quad j = 1,2,\cdots,$$

and $\upsilon_{n,j} = \theta_{n,j} - (1-\alpha_n - z - \alpha_n z s_n(z))$ for $j = 1,2,\cdots$. More specifically,

$$\upsilon_{n,j} = \frac{1}{n}\rho_{n,j}^*\rho_{n,j} - 1 - \frac{1}{n^2}\rho_{n,j}^*X_{n,-j}^*\left(\frac{1}{n}X_{n,-j}X_{n,-j}^* - z\,\mathrm{Id}\right)^{-1}X_{n,-j}\rho_{n,j} + \alpha_n + \alpha_n z s_n(z).$$

We use the following decomposition:

$$\begin{aligned} \upsilon_{n,j} &= \frac{1}{n}\rho_{n,j}^*\rho_{n,j} - 1 \\ &\quad - \frac{1}{n^2}\sum_{k\neq k'}^{n}\overline{\rho_{n,j}(k)}\left[X_{n,-j}^*\left(\frac{1}{n}X_{n,-j}X_{n,-j}^* - z\,\mathrm{Id}\right)^{-1}X_{n,-j}\right]_{kk'}\rho_{n,j}(k') \\ &\quad - \frac{1}{n^2}\sum_{k=1}^{n}\left(|\rho_{n,j}(k)|^2 - 1\right)\left[X_{n,-j}^*\left(\frac{1}{n}X_{n,-j}X_{n,-j}^* - z\,\mathrm{Id}\right)^{-1}X_{n,-j}\right]_{kk} \\ &\quad - \frac{1}{n^2}\operatorname{tr}\left[X_{n,-j}^*\left(\frac{1}{n}X_{n,-j}X_{n,-j}^* - z\,\mathrm{Id}\right)^{-1}X_{n,-j}\right] + \alpha_n + \alpha_n z s_n(z) \\ &=: A_{n,j}(z) + B_{n,j}(z) + C_{n,j}(z) + D_{n,j}(z). \end{aligned} \tag{2.4}$$

Assume $|\upsilon_{n,j}| \le |\mathrm{Im}\,z|/2$. By Lemma 2.6 (ii), $\mathrm{Im}(1-\alpha_n - z - \alpha_n z s_n(z)) < -|\mathrm{Im}\,z|$, and

$$\delta_n(z) = \frac{1}{m_n}\sum_{j=1}^{m_n}\frac{\upsilon_{n,j}}{(1-\alpha_n - z - \alpha_n z s_n(z))(1-\alpha_n - z - \alpha_n z s_n(z) + \upsilon_{n,j})} \le \frac{1}{m_n}\sum_{j=1}^{m_n}\frac{2\upsilon_{n,j}}{|\mathrm{Im}\,z|^2}.$$

Therefore, if we can prove that $\max_{j\in[n]} \upsilon_{n,j} \to 0$ almost surely, then

$$\delta_n = s_n(z) - \frac{1}{1-\alpha_n - z - \alpha_n z s_n(z)} \to 0 \quad \text{almost surely,}$$

and we complete the proof by Lemma 2.6 (iii).

Step II. We first study the term $A_{n,j}(z)$. By (1.16) in Lemma 1.18, we can find an absolute constant $C_0 > 0$ such that

$$\mathbb{E}\,|A_{n,j}(z)|^3 = \frac{1}{n^3}\mathbb{E}\left|\sum_{j=1}^{n}(\rho_{n,j}(k)^2-1)\right|^3 \le \frac{C_0 R^6}{n^3}.$$

Then for any $\epsilon > 0$,

$$\mathbb{P}\left(\max_{j\in[n]}|A_{n,j}(z)| \ge \epsilon\right) \le \sum_{j=1}^{n}\mathbb{P}\left(|A_{n,j}(z)| \ge \epsilon\right) \le \frac{1}{\epsilon^3}\sum_{j=1}^{n}\mathbb{E}\,|A_{n,j}(z)|^3 \le \frac{C_0 R^6}{n^2}.$$

Noticing that $\sum_{n=1}^{\infty} n^{-2} < \infty$, by the Borel-Cantelli lemma,

$$\mathbb{P}\left(\limsup_{n\to\infty}\max_{j\in[n]}|A_{n,j}(z)| \ge \epsilon\right) = 0$$

Since $\epsilon > 0$, we have

$$\max_{j\in[n]}|A_{n,j}(z)| \to 0 \quad \text{almost surely.} \tag{2.5}$$

Step III. Now we deal with the terms $B_{n,j}(z)$ and $C_{n,j}(z)$. We first introduce a technical lemma.

Lemma 2.7. *Let $X \in \mathbb{C}^{m\times n}$ and $z \in \mathbb{C}^+$. Then*

$$\left\|X^*\left(\frac{1}{n}XX^* - z\,\mathrm{Id}\right)^{-1}X\right\|_{\mathrm{F}}^2 \le mn^2\left(1+\frac{|z|}{|\operatorname{Im} z|}\right)^2.$$

Proof. Note that $(A - z\,\mathrm{Id})^{-1}A = \mathrm{Id} + z(A - z\,\mathrm{Id})^{-1}$ for any Hermitian matrix $A \in \mathbb{C}^{n\times n}$. Then

$$\left\|X^*\left(\frac{1}{n}XX^* - z\,\mathrm{Id}\right)X\right\|_{\mathrm{F}}^2 = \left\|\left(\frac{1}{n}XX^* - z\,\mathrm{Id}\right)^{-1}XX^*\right\|_{\mathrm{F}}^2 = n^2\left\|\mathrm{Id} + z\left(\frac{1}{n}XX^* - z\,\mathrm{Id}\right)^{-1}\right\|_{\mathrm{F}}^2.$$

Since all eigenvalues of $\frac{1}{n}XX^* - z\,\mathrm{Id} \in \mathbb{C}^{m\times m}$ have imaginary part $\operatorname{Im} z > 0$, we have

$$\left\|\mathrm{Id} + z\left(\frac{1}{n}XX^* - z\,\mathrm{Id}\right)^{-1}\right\|_{\mathrm{F}}^2 \le \sum_{j=1}^{m}\left(1+\frac{|z|}{|\operatorname{Im} z|}\right)^2 = m\left(1+\frac{|z|}{|\operatorname{Im} z|}\right)^2.$$

Then we finish the proof of (2.7). □

Since $\rho_{n,j}$ has independent entries and is independent of $X_{n,-j}$, by (1.17) in Lemma 1.18 and Lemma 2.7,

$$\mathbb{E}|B_{n,j}(z)|^6 \le \frac{C_1 R^{12}}{n^{12}}\left\|X_{n,-j}^*\left(\frac{1}{n}X_{n,-j}X_{n,-j}^* - z\,\mathrm{Id}\right)^{-1}X_{n,-j}\right\|_{\mathrm{F}}^6 \le \frac{C_1 m_n^3 R^{12}}{n^6}\left(1+\frac{|z|}{|\operatorname{Im} z|}\right)^6,$$

where $C_1 > 0$ is an absolute constant. Note that $\mathbb{E}\left[\rho_{n,j}(k)^2 - 1\right] = 0$ for all $k \in [n]$. Then we apply (1.16) in

Lemma 1.18 and Lemma 2.7 to obtain,

$$\mathbb{E}|C_{n,j}(z)|^6 \le \frac{C_2 R^{12}}{n^{12}} \mathbb{E} \left| \sum_{k=1}^{n} \left| \left[X_{n,-j}^* \left(\frac{1}{n} X_{n,-j} X_{n,-j}^* - z \operatorname{Id} \right)^{-1} X_{n,-j} \right]_{kk} \right|^2 \right|^3 \le \frac{C_2 m_n^3 R^{12}}{n^6} \left(1 + \frac{|z|}{|\operatorname{Im} z|} \right)^6,$$

where $C_2 > 0$ is an absolute constant. Then for any $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\max_{j \in [n]} |B_{n,j}(z) + C_{n,j}(z)| > \epsilon \right) &\le \sum_{j=1}^{n} \mathbb{P}\left(\mathbb{E}|B_{n,j}(z) + C_{n,j}(z)| > \epsilon \right) \le \frac{n|B_{n,j}(z) + C_{n,j}(z)|^6}{\epsilon^6} \\ &\le \frac{32n}{\epsilon^6} \left(\mathbb{E}|B_{n,j}(z)|^6 + \mathbb{E}|C_{n,j}(z)|^6 \right) \le \frac{32(C_1 + C_2) m_n^3 R^{12}}{n^5} \left(1 + \frac{|z|}{|\operatorname{Im} z|} \right)^6. \end{aligned}$$

Since $m_n/n \to \alpha \in (0, \infty)$ as $n \to \infty$, the above probability bound decays in the rate of n^{-2}. Noticing the fact $\sum_{n=1}^{\infty} n^{-2} < \infty$, we can apply Borel-Cantelli lemma to conclude that

$$\max_{j \in [n]} |B_{n,j}(z) + C_{n,j}(z)| \to 0 \quad \text{almost surely.} \tag{2.6}$$

Step IV. Note that $(A - z \operatorname{Id})^{-1} A = \operatorname{Id} + z(A - z \operatorname{Id})^{-1}$ for any Hermitian matrix $A \in \mathbb{C}^{n \times n}$. Then

$$\begin{aligned} \frac{1}{n^2} \operatorname{tr} \left[X_{n,-j}^* \left(\frac{1}{n} X_{n,-j} X_{n,-j}^* - z \operatorname{Id} \right)^{-1} X_{n,-j} \right] &= \frac{1}{n^2} \operatorname{tr} \left[\left(\frac{1}{n} X_{n,-j} X_{n,-j}^* - z \operatorname{Id} \right)^{-1} X_{n,-j} X_{n,-j}^* \right] \\ = -\frac{1}{n} \operatorname{tr} \left[\operatorname{Id} + z \left(\frac{1}{n} X_{n,-j} X_{n,-j}^* - z \operatorname{Id} \right)^{-1} \right] &= -\frac{m_n - 1}{n} - \frac{z}{n} \operatorname{tr} \left[\left(\frac{1}{n} X_{n,-j} X_{n,-j}^* - z \operatorname{Id} \right)^{-1} \right] \end{aligned}$$

Then the term $D_{n,k}(z)$ in (2.4) satisfies

$$\begin{aligned} |D_{n,j}(z)| &= \left| \alpha_n - \frac{m_n - 1}{n} + \frac{\alpha_n z}{m_n} \operatorname{tr} \left[\left(\frac{1}{n} X_n X_n^* - z \operatorname{Id} \right)^{-1} \right] - \frac{z}{n} \operatorname{tr} \left[\left(\frac{1}{n} X_{n,-j} X_{n,-j}^* - z \operatorname{Id} \right)^{-1} \right] \right| \\ &= \frac{1}{n} \left| 1 + z \operatorname{tr} \left[\left(\frac{1}{n} X_n X_n^* - z \operatorname{Id} \right)^{-1} \right] - z \operatorname{tr} \left[\left(\frac{1}{n} X_{n,-j} X_{n,-j}^* - z \operatorname{Id} \right)^{-1} \right] \right|. \end{aligned} \tag{2.7}$$

Lemma 2.8. *Let $A \in \mathbb{C}^{n \times n}$ be an Hermitian matrix, and obtain $A_{-j} \in \mathbb{C}^{(n-1) \times (n-1)}$ from A by removing the j-th row and j-th column. Let $z \in \mathbb{C}^+$. Then*

$$\left| \operatorname{tr}\left[(A - z \operatorname{Id})^{-1} \right] - \operatorname{tr}\left[(A_{-n} - z \operatorname{Id})^{-1} \right] \right| \le \frac{1}{|\operatorname{Im} z|}, \quad j = 1, 2, \cdots, n.$$

Proof. Without loss of generality, we may assume $j = n$ and write

$$A = \begin{bmatrix} A_{-n} & \alpha_n \\ \alpha_n^* & a_{nn} \end{bmatrix}.$$

We let $\beta_n(z) = a_{nn} - z - \alpha_n^* (A_{-n} - z \operatorname{Id})^{-1} \alpha_n$. By Schur's complement,

$$(A - z \operatorname{Id})^{-1} = \frac{1}{\beta_n(z)} \begin{bmatrix} \beta_n(z)(A_{-n} - z \operatorname{Id})^{-1} + (A_{-n} - z \operatorname{Id})^{-1} \alpha_n \alpha_n^* (A_{-n} - z \operatorname{Id})^{-1} & -(A_{-n} - z \operatorname{Id})^{-1} \alpha_n \\ -\alpha_n^* (A_{-n} - z \operatorname{Id}) & 1 \end{bmatrix}.$$

Then

$$\operatorname{tr}\left[(A - z\,\mathrm{Id})^{-1} - (A_{-n} - z\,\mathrm{Id})^{-1}\right] = \frac{1}{\beta_n(z)}\operatorname{tr}\begin{bmatrix}(A_{-n} - z\,\mathrm{Id})^{-1}\alpha_n\alpha_n^*(A_{-n} - z\,\mathrm{Id})^{-1} & 0\\ 0 & 1\end{bmatrix}$$
$$= \frac{1}{\beta_n(z)}\left[1 + \operatorname{tr}\left[(A_{-n} - z\,\mathrm{Id})^{-1}\alpha_n\alpha_n^*(A_{-n} - z\,\mathrm{Id})^{-1}\right]\right]$$
$$= \frac{1}{\beta_n(z)}\left(1 + \alpha_n^*\left(A_{-n} - z\,\mathrm{Id}\right)^{-2}\alpha_n\right).$$

We take the eigendecomposition $A_{-n} = U^*\Lambda U$, where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_{n-1}) \in \mathbb{R}^{n\times n}$ and U is unitary. Then

$$\left|\alpha_n^*\left(A_{-n} - z\,\mathrm{Id}\right)^{-2}\alpha_n\right| = \left|\sum_{j=1}^{n}\frac{|(U\alpha_n)_j|^2}{(\lambda_j - z)^2}\right| \le \sum_{j=1}^{n}\frac{|(U\alpha_n)_j|^2}{(\lambda_j - \xi)^2 + \eta^2}$$
$$= \alpha_n^* U^*\left[(\Lambda - \xi\,\mathrm{Id})^2 + \eta^2\,\mathrm{Id}\right]^{-1} U\alpha_n = \alpha_n^*\left[(A_{-n} - \xi\,\mathrm{Id})^2 + \eta^2\,\mathrm{Id}\right]^{-1}\alpha_n.$$

With a similar trick we also have

$$\alpha_n^*\left(A_{-n} - z\,\mathrm{Id}\right)^{-1}\alpha_n = \sum_{j=1}^{n}\frac{|(U\alpha_n)_j|^2}{\lambda_j - z} = \sum_{j=1}^{n}\frac{(\lambda_j - \xi)|(U\alpha_n)_j|^2}{(\lambda_j - \xi)^2 + \eta^2} + i\eta\sum_{j=1}^{n}\frac{|(U\alpha_n)_j|^2}{(\lambda_j - \xi)^2 + \eta^2},$$

and

$$\operatorname{Im}\beta_n(z) = -\eta - \eta\sum_{j=1}^{n}\frac{|(U\alpha_n)_j|^2}{(\lambda_j - \xi)^2 + \eta^2} = -\eta\left(1 + \alpha_n^*\left[(A_{-n} - \xi\,\mathrm{Id})^2 + \eta^2\,\mathrm{Id}\right]^{-1}\alpha_n\right).$$

Therefore

$$\left|\operatorname{tr}\left[(A - z\,\mathrm{Id})^{-1} - (A_{-n} - z\,\mathrm{Id})^{-1}\right]\right| \le \frac{1 + \alpha_n^*\left[(A_{-n} - \xi\,\mathrm{Id})^2 + \eta^2\,\mathrm{Id}\right]^{-1}\alpha_n}{|\operatorname{Im}\beta_n(z)|} \le \frac{1}{\eta},$$

and we finish the proof. □

Now we let $A = \frac{1}{n}X_nX_n^*$ in Lemma 2.8 and plug-in the result to (2.7) to conclude

$$\max_{j\in[n]}|D_{n,j}(z)| \le \frac{1}{n}\left(1 + \frac{|z|}{|\operatorname{Im} z|}\right), \tag{2.8}$$

which converges to 0 deterministically as $n \to \infty$.

Step V. By (2.4), (2.5), (2.6) and (2.8), we have $\max_{j\in[n]} \upsilon_{n,j} \to 0$ a.s., and the proof is completed. □

2.2 Generalizied Marčenko-Pastur Law

See Silverstein and Bai (1995), Silverstein (1995).

3 Free Probability

3.1 Non-commutative Probability Spaces

Definition 3.1 ($*$-algebra)**.** A *$*$-algebra* is an associative $\mathbb{C}$-algebra $\mathcal{A}$ equipped with a unary operation $* : \mathcal{A} \to \mathcal{A}$ that is an involution and an antiautomorphism, i.e.

(a) $(x+y)^* = x^* + y^*$ for all $x, y \in \mathcal{A}$,

(b) $(xy)^* = y^* x^*$ for all $x, y \in \mathcal{A}$,

(c) $(x^*)^* = x$ for all $x \in \mathcal{A}$, and

(d) $(\alpha x)^* = \overline{\alpha} x^*$ for all $x \in \mathcal{A}$ and all $\alpha \in \mathbb{C}$.

That is, $*$ preserves addition, reverses multiplication, and is antihomogeneous. In particular,

(i) an element $x \in \mathcal{A}$ is said to be *self-adjoint* if it satisfies $x^* = x$, and

(ii) an element $x \in \mathcal{A}$ is said to be *normal* if it satisfies $xx^* = x^* x$.

Definition 3.2 (Non-commutative probability space)**.** A *non-commutative probability space* $(\mathcal{A}, \tau)$ consists of a (potentially non-commutative) $*$-algebra $\mathcal{A}$ with identity $\mathbf{1} \in \mathcal{A}$ and a *trace operator* $\tau : \mathcal{A} \to \mathbb{C}$ which is $*$-linear, maps $\mathbf{1}$ to 1, and is nonnegative, i.e.

(i) $\tau(\alpha X + \beta Y) = \alpha\tau(X) + \beta\tau(Y)$ and $\tau(X^*) = \overline{\tau(X)}$ for all $X, Y \in \mathcal{A}$ and all $\alpha, \beta \in \mathbb{C}$,

(ii) $\tau(\mathbf{1}) = 1$, and

(iii) $\tau(XX^*) \geq 0$ for all $X \in \mathcal{A}$.

Furthermore,

- τ is said to be *faithful* if $\tau(XX^*) = 0$ implies $X = 0$;
- τ is said to be *tracial* if it obeys the *trace axiom*: $\tau(XY) = \tau(YX)$ for every $X, Y \in \mathcal{A}$.

Remark. By definition, the formula $\langle X, Y \rangle_\tau = \tau(XY^*)$ defines a semi-inner product on $\mathcal{A}$. Furthermore, if the trace τ is faithful, then $\langle X, Y \rangle_\tau$ is an inner product.

As a simple generailzation of moments of random variables in classical probability theory, the moments of a random element $X \in \mathcal{A}$ are defined as $\tau(X^k)$, where $k = 1, 2, \cdots$.

Lemma 3.3 (Monotonicity of moments)**.** *Let $(X, \mathcal{A})$ be a non-commutative probability space. For every self-adjoint element $X \in \mathcal{A}$ and $k \in \mathbb{N}$,*

$$\left|\tau(X^{2k-1})\right|^{\frac{1}{2k-1}} \leq \left|\tau(X^{2k})\right|^{\frac{1}{2k}} \leq \left|\tau\left(X^{2k+2}\right)\right|^{\frac{1}{2k+2}}, \tag{3.1}$$

As a consequence, we can define the ***spectral radius*** *$\rho(X)$ of a self-adjoint element X by the formula*

$$\rho(X) = \lim_{k\to\infty} \left|\tau(X^{2k})\right|^{\frac{1}{2k}}. \tag{3.2}$$

Then for all self-adjoint $X \in \mathcal{A}$,

$$\left|\tau(X^k)\right| \leq \rho(X)^k, \quad k = 1, 2, \cdots. \tag{3.3}$$

We say that a self-adjoint element X is ***bounded*** *if its spectral radius $\rho(X) < \infty$.*

Proof. The map $(X, Y) \mapsto \tau(XY^*)$ defines a semi-inner product in $\mathcal{A}$, and the Cauchy-Schwarz inequality $|\tau(XY^*)| = \sqrt{\tau(XX^*)\tau(YY^*)}$ holds. We then proceed the proof by induction. Assume $X = X^*$. For $k = 1$,

$$|\tau(X)| \leq \sqrt{\tau(X^2)\,\tau(1)} = \sqrt{\tau(X^2)} \leq \left|\tau(X^4)\tau(1)\right|^{1/4} = \left|\tau(X^4)\right|^{1/4}$$

By the induction hypothesis, in the k-th step, we have $\left|\tau(X^{2k-2})\right|^{\frac{1}{2k-2}} \le \left|\tau\left(X^{2k}\right)\right|^{\frac{1}{2k}}$. Then

$$\left|\tau(X^{2k-1})\right| \le \left|\tau(X^{2k-2})\,\tau(X^{2k})\right|^{1/2} \le \left|\tau(X^{2k})\right|^{\frac{k-1}{2k}} \left|\tau(X^{2k})\right|^{1/2} = \left|\tau(X^{2k})\right|^{\frac{2k-1}{2k}},$$

and

$$\left|\tau(X^{2k})\right| \le \left|\tau\left(X^{2k-2}\right)\tau\left(X^{2k+2}\right)\right|^{1/2} \le \left|\tau(X^{2k})\right|^{\frac{k-1}{2k}} \left|\tau\left(X^{2k+2}\right)\right|^{1/2}.$$

Combining the above two results, we have

$$\left|\tau(X^{2k-1})\right|^{\frac{1}{2k-1}} \le \left|\tau(X^{2k})\right|^{\frac{1}{2k}} \le \left|\tau\left(X^{2k+2}\right)\right|^{\frac{1}{2k+2}},$$

and the proof is completed. □

We are also interested in the moment of normal elements.

Lemma 3.4. *Let $(X, \mathcal{A})$ be a non-commutative probability space, and $X \in \mathcal{A}$.*
(i) If X is bounded self-adjoint, then for every $R > 0$,

$$\rho(R^2\mathbf{1} + X^2) = R^2 + \rho(X)^2.$$

(ii) If $X \in \mathcal{A}$ is normal,

$$|\tau(X^k)| \le \tau\left((X^*X)^k\right)^{1/2} \le \rho(X^*X)^{k/2}. \tag{3.4}$$

Proof. (i) Let X be bounded self-adjoint and $R \ge 0$. Given any $\epsilon > 0$, by the definition (3.2) of $\rho(X)$, we fix an integer $N \ge 1$ such that

$$|\tau(X^{2j})|^{1/j} \ge \rho(X)^2 - \epsilon \quad \text{for all } j \ge N.$$

Then for every $k > N/2$,

$$\begin{aligned}\tau\left((R^2\mathbf{1} + X^2)^{2k}\right) &= \sum_{j=0}^{2k} \binom{2k}{j} R^{4k-2j} \tau(X^{2j}) \ge \sum_{j=N}^{2k} \binom{2k}{j} R^{4k-2j} \left(\rho(X)^2 - \epsilon\right)^j \\ &= \left(R^2 + \rho(X)^2 - \epsilon\right)^{2k} - \sum_{j=0}^{N-1} \binom{2k}{j} R^{4k-2j} \left(\rho(X)^2 - \epsilon\right)^j.\end{aligned}$$

Note that

$$\left|\sum_{j=0}^{N-1} \binom{2k}{j} R^{4k-2j} \left(\rho(X)^2 - \epsilon\right)^j\right| \le (2k)^N R^{4k} \max\left\{1, |\rho(X)^2 - \epsilon|^N\right\}.$$

Then we have

$$\limsup_{k\to\infty} \left|\sum_{j=0}^{N-1} \binom{2k}{j} R^{4k-2j} \left(\rho(X)^2 - \epsilon\right)^j\right|^{\frac{1}{2k}} \le R^2.$$

Therefore,

$$\begin{aligned}\rho(R^2\mathbf{1} + X^2) = \lim_{n\to\infty} \left|\tau\left((R^2\mathbf{1} + X^2)^{2k}\right)\right|^{\frac{1}{2k}} &\ge \lim_{k\to\infty} \left|\left(R^2 + \rho(X)^2 - \epsilon\right)^{2k} + R^{2k}\right|^{\frac{1}{2k}} \\ &= \max\left\{R^2 + \rho(X)^2 - \epsilon, R^2\right\}.\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we have $\rho(R^2\mathbf{1} + X^2) \geq R^2 + \rho(X)^2$. On the other hand, by (3.3), for all $k \in \mathbb{N}$,

$$\left|\tau\left((R^2\mathbf{1} + X^2)^{2k}\right)\right|^{\frac{1}{2k}} = \left|\sum_{j=0}^{2k}\binom{2k}{j}R^{4k-2j}\tau(X^{2j})\right|^{\frac{1}{2k}} \leq \left|\sum_{j=0}^{2k}\binom{2k}{j}R^{4k-2j}\rho(X)^{2j}\right|^{\frac{1}{2k}} = R^2 + \rho(X)^2$$

Letting $k \to \infty$, it follows that $\rho(R^2\mathbf{1} + X^2) \leq R^2 + \rho(X)^2$.

(ii) is simply a consequence of the Cauchy-Schwarz inequality and (3.3). □

Theorem 3.5. *Let $(\mathcal{A}, \tau)$ be a non-commutative probability space, and let*

$$\mathcal{H}_{\mathcal{A}} = \{X \in \mathcal{A} : X = X^*, \rho(X) < \infty\}$$

be the space of bounded self-adjoint elements.

(i) The spectral radius $\rho : \mathcal{H}_{\mathcal{A}} \to [0, \infty)$ is a seminorm on $\mathcal{H}_{\mathcal{A}}$. If τ is faithful, then ρ is a norm on $\mathcal{H}_{\mathcal{A}}$.

(ii) ρ is submultiplicative. If $X, Y \in \mathcal{H}_{\mathcal{A}}$ are commutative under multiplication, then $\rho(XY) \leq \rho(X)\rho(Y)$.

Proof. (i) It suffices to check the triangle inequality. Let $X, Y \in \mathcal{H}_{\mathcal{A}}$. Then for every $k \in \mathbb{N}$,

$$\begin{aligned}\left|\tau((X+Y)^{2k})\right| &= \left|\sum_{j=0}^{2k}\binom{2k}{j}\tau(X^jY^{2k-j})\right| \leq \sum_{j=0}^{2k}\binom{2k}{j}\sqrt{\tau(X^{2j})\tau(Y^{4k-2j})} \\ &\leq \sum_{j=0}^{2k}\binom{2k}{j}\rho(X)^j\rho(Y)^{2k-j} = |\rho(X)+\rho(Y)|^{2k}.\end{aligned}$$

Raising everything to the $1/(2k)$ power and letting $k \to \infty$, we have $\rho(X+Y) \leq \rho(X) + \rho(Y)$.

(ii) For any $U \in \mathcal{A}$ with $\|U\|_\tau^2 = \tau(UU^*) = 1$, we have

(ii) If $XY = YX$, then for every $k \in \mathbb{N}$,

$$\left|\tau((XY)^{2k})\right| = \left|\tau(X^{2k}Y^{2k})\right| \leq \sqrt{\tau(X^{4k})\tau(Y^{4k})} \leq \rho(X)^{2k}\rho(Y)^{2k}.$$

Raising everything to the $1/(2k)$ power and letting $k \to \infty$, we have $\rho(XY) \leq \rho(X)\rho(Y)$. □

Remark. By this conclusion, if $X \in \mathcal{A}$ is bounded self-adjoint and $P : \mathbb{C} \to \mathbb{C}$ is a polynomial with real coefficients, then $P(X)$ is also bounded self-adjoint.

3.1.1 The Spectral Measures

Theorem 3.6. *Let $X \in \mathcal{A}$ be a bounded self-adjoint element in a noncomutative probability space $(\mathcal{A}, \tau)$. Then there exists a measure μ_X supported on $[-\rho(X), \rho(X)]$, called the* ***spectral measure*** *of X, such that*

$$\tau(P(X)) = \int_{-\rho(X)}^{\rho(X)} P(\lambda)\, d\mu_X(\lambda) \tag{3.5}$$

for all polynomials $P : \mathbb{C} \to \mathbb{C}$ with complex coefficients.

Proof. We write $m_k = \tau(X^k)$ for $k \in \mathbb{N}_0$. Then the Hankel matrix $(m_{j+k})_{j,k\in\mathbb{N}_0}$ is positive semidefinite:

$$\sum_{j,k=0}^{n} c_jc_km_{j+k} = \sum_{j,k=0}^{n} c_jc_k\tau(X^{j+k}) = \tau\left(YY^*\right) \geq 0, \quad \text{where } Y = \sum_{j=0}^{n} c_jX^j.$$

Therefore, the Hamburger's moment problem

$$\int_{\mathbb{R}} x^k \, d\mu(x) = \tau(X^k), \quad k = 0, 1, 2, \cdots \tag{3.6}$$

has a solution. Furthermore, since $\lim_{k\to\infty} \frac{1}{2k} |\tau(X^{2k})|^{\frac{1}{2k}} = 0$, the measure $\mu = \mu_X$ satisfying the equations (3.6) is unique. Since $\int_{\mathbb{R}} d\mu_X = \tau(\mathbf{1}) = 1$, the measure μ_X is a probability measure.

For any $\epsilon > 0$, if $\mu_X \{t \in \mathbb{R} : |t| > \rho(X) + \epsilon\} = \delta > 0$, then we have

$$\rho(X) = \lim_{k\to\infty} |\tau(X^{2k})|^{\frac{1}{2k}} = \lim_{k\to\infty} \left| \int_{\mathbb{R}} x^{2k} \, d\mu_X(x) \right|^{\frac{1}{2k}} \geq \lim_{k\to\infty} \delta^{\frac{1}{2k}} (\rho(X) + \epsilon) = \rho(X) + \epsilon,$$

which is a contradiction. Hence $\mu_X \{t \in \mathbb{R} : |t| > \rho(X) + \epsilon\} = 0$ for every $\epsilon > 0$, and the spectral measure μ_X is supported on $[-\rho(X), \rho(X)]$. Then μ_X is the desired spectral measure satisfying (3.5). □

Remark. By (3.5), we have the bound

$$|\tau(P(X))| \leq \rho(P(X)) \leq \sup_{\lambda \in [-\rho(X), \rho(X)]} |P(\lambda)|. \tag{3.7}$$

By the Stone-Weierstrass theorem, every continuous function $f : [-\rho(X), \rho(X)] \to \mathbb{C}$ can be approximated uniformly by polynomials. Hence we extend the definition

$$\tau(f(X)) = \int_{-\rho(X)}^{\rho(X)} f(\lambda) \, d\lambda, \quad f \in C([-\rho(X), \rho(X)]).$$

Definition 3.7 (Stieltjes transform)**.** Let $X \in \mathcal{A}$ be a bounded self-adjoint element in a noncomutative probability space $(\mathcal{A}, \tau)$. The *Siteltjes transform* of X is a function $S_X : \mathbb{C} \backslash [-\rho(X), \rho(X)]$ defined by

$$S_X(z) = \int_{-\rho(X)}^{\rho(X)} \frac{1}{\lambda - z} \, d\mu_X(\lambda), \quad z \notin [-\rho(X), \rho(X)],$$

where μ_X is the spectral measure of X.

When the spectral measure μ_X is unknown, it is more convenient to write the Stieltjes transform $S_X(z)$ in terms of the moments of X.

Theorem 3.8 (Laurent series representation of the Stieltjes transform)**.** *Let $X \in \mathcal{A}$ be a bounded self-adjoint element in a noncomutative probability space $(\mathcal{A}, \tau)$. Then*

(i) For $z \in \mathbb{C}$ with $|z| > \rho(X)$,

$$S_X(z) = -\sum_{k=0}^{\infty} \frac{\tau(X^k)}{z^{k+1}};$$

(ii) For $z \in \mathbb{C}^+$,

$$S_X(z) = -\sum_{k=0}^{\infty} \frac{\tau((X + iR\mathbf{1})^k)}{(z + iR)^{k+1}}, \quad \textit{where } R > \min\left\{0, \frac{\rho(X)^2 - (\operatorname{Im} z)^2}{2 \operatorname{Im} z}\right\}.$$

In particular, $S_X : \mathbb{C} \backslash [-\rho(X), \rho(X)]$ is an analytical function.

Proof. (i) Fix $|z| > \rho(X)$. For every $\lambda \in [-\rho(X), \rho(X)]$, we have $\frac{1}{|\lambda - z|} \leq \frac{1}{|z| - \rho(X)} < \infty$. By Fubini's theorem,

$$\begin{aligned} S_X(z) &= \int_{-\rho(X)}^{\rho(X)} \frac{1}{\lambda - z}\, d\mu_X(\lambda) = -\frac{1}{z} \int_{-\rho(X)}^{\rho(X)} \sum_{k=0}^{\infty} \frac{\lambda^k}{z^k}\, d\mu_X(\lambda) \\ &= -\sum_{k=0}^{\infty} \int_{-\rho(X)}^{\rho(X)} \frac{\lambda^k}{z^{k+1}}\, d\mu_X(\lambda) = -\sum_{k=0}^{\infty} \frac{\tau(X^k)}{z^{k+1}}. \end{aligned}$$

(ii) For any $R > 0$, by Lemma 3.4, the element $X + iR\mathbf{1}$ satisfies

$$\left|\tau\left((X + iR\mathbf{1})^k\right)\right| \leq \rho\left(R^2\mathbf{1} + X^2\right)^{k/2} = \left(R^2 + \rho(X)^2\right)^{k/2}. \tag{3.8}$$

Fix $z \in \mathbb{C}^+$ and $R > \min\left\{0, \frac{\rho(X)^2 - (\operatorname{Im} z)^2}{2 \operatorname{Im} z}\right\}$. Then

$$|z + iR|^2 \geq (\operatorname{Im} z + R)^2 > \rho(X)^2 + R^2. \tag{3.9}$$

Then we address $(\lambda - z)^{-1}$ by shifting iR and plugging-in the Neumann series:

$$\begin{aligned} S_X(z) &= \int_{-\rho(X)}^{\rho(X)} \frac{1}{(\lambda + iR) - (z + iR)}\, d\mu_X(\lambda) = -\frac{1}{z + iR} \int_{-\rho(X)}^{\rho(X)} \sum_{k=0}^{\infty} \frac{(\lambda + iR)^k}{(z + iR)^k}\, d\mu_X(\lambda) \\ &= -\sum_{k=0}^{\infty} \int_{-\rho(X)}^{\rho(X)} \frac{(\lambda + iR)^k}{(z + iR)^{k+1}}\, d\mu_X(\lambda) = -\sum_{k=0}^{\infty} \frac{\tau((X + iR\mathbf{1})^k)}{(z + iR)^{k+1}}, \end{aligned}$$

where the last series is convergent due to (3.8) and (3.9). ☐

Proposition 3.9. *Let $X \in \mathcal{A}$ be a bounded self-adjoint element of a non-commutative probability space $(\mathcal{A}, \tau)$. Then*

$$\|YX\|_\tau \leq \rho(X)\|Y\|_\tau, \quad \text{for } Y \in \mathcal{A},$$

where $\|\cdot\|_\tau$ is the seminorm $\|Y\|_\tau^2 = \langle Y, Y\rangle_\tau = \tau(YY^*)$.

Proof. Given any $\epsilon > 0$, by the Stone-Weierstrass theorem, there exists a polynomial $P : \mathbb{C} \to \mathbb{C}$ such that

$$\sup_{|x| \leq \rho(X)} |x^2 + P(x)^2 - \rho(X)^2| < \epsilon.$$

We let $E = X^2 + P(X)^2 - \rho(X)^2$. Then

$$\tau(YX^2Y^*) \leq \tau(YX^2Y^*) + \tau(YP(X)^2Y^*) = \rho(X)^2\tau(YY^*) + \tau(YEY^*).$$

By estimate (3.7), we have $\rho(E) \leq \epsilon$, and

$$|\tau(YEY^*)| = |\tau(EY^*Y)| \leq \sqrt{\tau(E^2)\tau((Y^*Y)^2)} \leq \rho(E)\sqrt{\tau((Y^*Y)^2)} \leq \sqrt{\tau((Y^*Y)^2)}\,\epsilon.$$

Note that $\tau((Y^*Y)^2) < \infty$. Combining the above two displays and letting $\epsilon \downarrow 0$, we have

$$\tau(YX^2Y^*) \leq \rho(X)^2\tau(YY^*).$$

This finishes the proof. ☐

3.1.2 Convergence in Moments

Definition 3.10 (Convergence)**.** Let $(\mathcal{A}_n, \tau_n)$ be a sequence of non-commutative probability spaces, and $X_{n,1}, \cdots, X_{n,k} \in \mathcal{A}_n$ random variables in $(\mathcal{A}_n, \tau_n)$ for each n. Let $(\mathcal{A}, \tau)$ be an additional non-commutative probability space, and $X_1, \cdots, X_m \in \mathcal{A}$. We say that the random vector $(X_{n,1}, \cdots, X_{n,m})$ convergence in moments to $(X_1, \cdots, X_m)$, if

$$\tau_n \left(X_{n,i_1} X_{n,i_2} \cdots X_{n,i_k} \right) \to \tau \left(X_{i_1} X_{i_2} \cdots X_{i_k} \right)$$

for every $k \in \mathbb{N}$ and $i_1, i_2, \cdots, i_k \in \{1, \cdots, m\}$. In other words, all mixed moments of $(X_{n,1}, \cdots, X_{n,m})$ converges to those of $(X_1, \cdots, X_m)$.

Theorem 3.11 (Convergence of spectral measures)**.** *Let $(\mathcal{A}_n, \tau_n)$ be a sequence of non-commutative probability spaces, and $X_n \in \mathcal{A}_n$ bounded self-adjoint random variables with $\rho(X_n)$ uniformly bounded. Let $X \in \mathcal{A}$ be another bounded self-adjoint random variable in an additional non-commutative probability space $(\mathcal{A}, \tau)$. Then X_n converges in moments to X if and only if the spectral measures μ_{X_n} converges weakly to μ_X.*

Proof. We take $M > 0$ such that $\sup_{n \in \mathbb{N}} \rho(X_n) \le M$ and $\rho(X) \le M$, so that all μ_{X_n} and μ_X are supported in $[-M, M]$. We choose $\psi \in C_c([-2M, 2M])$ such that $\psi \equiv 1$ on $[-M, M]$. If $\mu_{X_n} \to \mu_X$ weakly, then for every $k \in \mathbb{N}$, the function $\lambda \mapsto x^k \psi(x)$ is bounded continuous on $\mathbb{R}$, and

$$\tau_n(X_n^k) = \int_{\mathbb{R}} \lambda^k \, d\mu_{X_n}(\lambda) = \int_{\mathbb{R}} \lambda^k \psi(\lambda) \, d\mu_{X_n}(\lambda) \to \int_{\mathbb{R}} \lambda^k \psi(\lambda) \, d\mu_X(\lambda) = \int_{\mathbb{R}} \lambda^k \, d\mu_X(\lambda) = \tau(X^k),$$

Conversely, if $\tau_n(X_n^k) \to \tau(X^k)$ for every k, we have

$$\tau(X^k) \le \sup_{n \in \mathbb{N}} \tau_n(X_n^k) \le \sup_{n \in \mathbb{N}} \rho(X_n)^k \le M^k,$$

and

$$\limsup_{k \to \infty} \frac{1}{2k} \left(\int_{\mathbb{R}} \lambda^{2k} \, d\mu_X(\lambda) \right)^{\frac{1}{2k}} \le \limsup_{k \to \infty} \frac{M}{2k} = 0 < \infty.$$

By Carleman's continuity theorem, $\mu_{X_n} \to \mu_X$ weakly, and we finishes the proof. □

3.2 Free Independence

Definition 3.12 (Free independence)**.** Let $(\mathcal{A}, \tau)$ be a non-commutative probability space. Let $(\mathcal{A}_i)_{i\in I}$ be a family of unital sub-algebras of $\mathcal{A}$ over $\mathbb{C}$. Then $(\mathcal{A}_i)_{i\in I}$ are said to be *free* (or *freely independent*) in $(\mathcal{A}, \tau)$, if $\tau(a_1 \cdots a_k) = 0$ whenever

- $k \in \mathbb{N}$;
- $i_1, \cdots, i_k \in J$, and any two adjacent indices are distinct, i.e. $i_1 \neq i_2 \neq i_3 \neq \cdots \neq i_k$, (however e.g. $i_1 = i_3$, and in particular, $i_1 = i_k$ are allowed);
- $a_j \in \mathcal{A}_{i_j}$ and $\tau(a_j) = 0$ for all $j = 1, \cdots, k$.

Note here we do **not** require $\mathcal{A}_i \neq \mathcal{A}_{i'}$ for $i \neq i'$.

Furthermore, let $(X_i)_{i\in I} \subset \mathcal{A}$ be a family of random variables in $(\mathcal{A}, \tau)$. Then $(X_i)_{i\in I}$ are said to be *free* in $(\mathcal{A}, \tau)$, if the unital sub-algebras $\mathcal{A}_i = \mathrm{Alg}(\mathbb{C}\mathbf{1}, X_i)$ they generated are free in $(\mathcal{A}, \tau)$.

Remark. Since the unital sub-algebra over $\mathbb{C}$ generated by $X \in \mathcal{A}$ is $\{P(X) : P \text{ is a polynomial}\}$, the free independence of random variables $(X_i)_{i\in I}$ is equivalent to the condition that one has

$$\tau\left[(P_1(X_{i_1}) - \tau(P_1(X_{i_1}))\mathbf{1})(P_2(X_{i_2}) - \tau(P_2(X_{i_2}))\mathbf{1}) \cdots (P_k(X_{i_k}) - \tau(P_k(X_{i_k}))\mathbf{1})\right] = 0$$

whenever $k \in \mathbb{N}$, $P_1, \cdots, P_k$ are polynomials, and $i_1, \cdots, i_k \in I$ are indices with no two adjacent i_j's equal. Furthermore, the unital sub-algebra generated by a constant variable is the scalar sub-algebra $\mathbb{C}\mathbf{1}$, which is freely independent of any random variable $X \in \mathcal{A}$.

The free independence is closely related to classical independence in the following sense: If a family of random variables is freely independent, then the joint distribution of the family is completely determined by the knowledge of the individual distributions of the variables. A formal statement is:

Theorem 3.13. *Let $(\mathcal{A}, \tau)$ be a non-commutative probability space, and let $(A_i)_{i\in I}$ be freely independent unital sub-algebras of $\mathcal{A}$. Denote by $\mathcal{B}$ the algebra generated by $(A_i)_{i\in I}$, i.e.*

$$\mathcal{B} = \mathrm{Alg}\left(\bigcup_{i\in I} \mathcal{A}_i\right).$$

Then $\tau|_{\mathcal{B}}$ is uniquely determined by $(\tau|_{\mathcal{A}_i})_{i\in I}$. (That is, if $\tilde{\tau}$ is another trace operator such that $(\mathcal{A}, \tilde{\tau})$ is a non-commutative probability space and $\tilde{\tau}|_{\mathcal{A}_i} = \tau|_{\mathcal{A}_i}$ for all $i \in I$, then $\tilde{\tau}|_{\mathcal{B}} = \tau|_{\mathcal{B}}$.)

Proof. By definition, each element of $\mathcal{B}$ is a linear combination of products of the form $a_1 \cdots a_k$, where $k \in \mathbb{N}$, $i_1, \cdots, i_k \in I$ and $a_j \in \mathcal{A}_{i_j}$ for $j = 1, \cdots, k$. We may assume $i_1 \neq i_2 \neq \cdots \neq i_k$, otherwise we just multiply some adjacent factors together to a new element in the same $\mathcal{A}_i$. Then it suffices to prove that $\tau(a_1 \cdots a_k)$ is fully determined by $(\tau|_{\mathcal{A}_i})_{i\in I}$ for all such products $a_1 \cdots a_k \in \mathcal{B}$, which is done by induction over k.

The base case $k = 1$ is clear since $a_1 \in \mathcal{A}_{i_1}$. For the general case $k \geq 2$, we set $\overline{a}_j = a_j - \tau(a_j)\mathbf{1} \in \mathcal{A}_{i_j}$ for $j = 1, \cdots, k$, which satisfies $\tau(\overline{a}_j) = 0$. Then

$$\tau\left(a_1 \cdots a_k\right) = \tau\left((\overline{a}_1 + \tau(a_1)\mathbf{1}) \cdots (\overline{a}_k + \tau(a_k)\mathbf{1})\right) = \tau(\overline{a}_1 \cdots \overline{a}_k) + R,$$

where the remainder

$$R = \sum_{m=0}^{k-1} \sum_{p,q} \tau(\overline{a}_{p_1} \cdots \overline{a}_{p_m})\tau(a_{q_1}) \cdots \tau(a_{q_{k-m}}),$$

and the sum $\sum_{p,q}$ runs over all disjoint decomposition

$$\{p_1, \cdots, p_m\} \cup \{q_1, \cdots, q_{k-m}\} = \{1, \cdots, k\}, \quad p_1 < \cdots < p_m, \quad q_1 < \cdots < q_{k-m}.$$

It is seen that all terms in the remainder R consist of factors of length at most $k-1$, and thus are fully determined by $(\tau|_{\mathcal{A}_i})_{i\in I}$ by the induction hypothesis. On the other hand, by the definition of free independence, $\tau(\overline{a}_1\cdots\overline{a}_k)=0$. Therefore $\tau\left(a_1\cdots a_k\right)$, and the induction step is completed. □

From a combinatorial perspective, free independence is a very special rule for calculating joint moments of freely independent variables out of the moments of the single variables.

Corollary 3.14. *Let $(X_i)_{i\in I}\subset\mathcal{A}$ be a family of freely independent random variables in a non-commutative probability space $(\mathcal{A},\tau)$. Then every joint moment of $(X_i)_{i\in I}$ is a polynomial combination of the individual moments $\tau(X_i^k)$ of the X_i's.*

Following are some concrete examples.

Example 3.15 (Joint moments)**.** Let $X,Y\in\mathcal{A}$ be freely independent, and $p,q,r,s\in\mathbb{N}$. Then

$$\begin{aligned}&\tau(X^pY^q)=\tau(X^p)\tau(Y^q),\quad \tau(X^pY^qX^r)=\tau(X^{p+r})\tau(Y^q),\\&\tau(X^pY^qX^rY^s)=\tau(X^{p+r})\tau(Y^q)\tau(Y^s)+\tau(X^p)\tau(X^r)\tau(Y^{q+s})-\tau(X)^p\tau(Y)^q\tau(X)^r\tau(Y)^s\end{aligned}$$

Proof. (i) By definition of free independence we have

$$0=\tau((X^p-\tau(X^p)\mathbf{1})(Y^q-\tau(Y^q)\mathbf{1}))=\tau(X^pY^q)-\tau(X^p)\tau(Y^q).$$

(ii) By definition of free independence we have

$$\begin{aligned}0&=\tau\left((X^p-\tau(X^p)\mathbf{1})(Y^q-\tau(Y^q)\mathbf{1})(X^r-\tau(X^r)\mathbf{1})\right)\\&=\tau\left(X^p(Y^q-\tau(Y^q)\mathbf{1})(X^r-\tau(X^r)\mathbf{1})\right)\\&=\tau\left(X^pY^q(X^r-\tau(X^r)\mathbf{1})\right)-\tau(Y^q)\tau\left(X^p(X^r-\tau(X^r)\mathbf{1})\right)\\&=\tau(X^pY^qX^r)-\tau(X^{p+r})\tau(Y^q).\end{aligned}$$

(ii) By definition of free independence we have

$$\begin{aligned}0&=\tau\left((X^p-\tau(X^p)\mathbf{1})(Y^q-\tau(Y^q)\mathbf{1})(X^r-\tau(X^r)\mathbf{1})(Y^s-\tau(Y^s)\mathbf{1})\right)\\&=\tau\left(X^p(Y^q-\tau(Y^q)\mathbf{1})(X^r-\tau(X^r)\mathbf{1})(Y^s-\tau(Y^s)\mathbf{1})\right)\\&=\tau\left(X^pY^q(X^r-\tau(X^r)\mathbf{1})(Y^s-\tau(Y^s)\mathbf{1})\right)-\tau(Y^q)\,\tau\left(X^p(X^r-\tau(X^r)\mathbf{1})(Y^s-\tau(Y^s)\mathbf{1})\right)\\&=\tau(X^pY^qX^rY^s)-\tau(X^pY^{q+s})\tau(X^r)-\tau(X^pY^qX^r)\tau(Y^s)+\tau(X^pY^q)\tau(X^r)\tau(Y^s)\\&\quad-\tau(Y^q)\left[\tau(X^{p+r}Y^s)-\tau(X^{p+r})\tau(Y^s)-\tau(X^pY^s)\tau(X^r)+\tau(X^p)\tau(X^r)\tau(Y^s)\right]\\&=\tau(X^pY^qX^rY^s)-\tau(X^p)\tau(X^r)\tau(Y^{q+s})-\tau(X^{p+r})\tau(Y^q)\tau(Y^s)+\tau(X)^p\tau(Y)^q\tau(X)^r\tau(Y)^s.\end{aligned}$$

Then we finish the proof. □

Proposition 3.16. *Let $X,Y\in\mathcal{A}$ be two freely independent self-adjoint random variables in a faithful non-commutative probability space $(\mathcal{A},\tau)$. If X and Y commute with each other, i.e. $XY=YX$, then at least one of them is constant.*

Proof. By (3.15), we have

$$\tau(XYXY)=\tau(X^2)\tau(Y)^2+\tau(X)^2\tau(Y^2)-\tau(X)^2\tau(Y)^2.$$

Since X and Y commute, we also have

$$\tau(XYXY)=\tau(XXYY)=\tau(X^2Y^2)=\tau(X^2)\tau(Y^2).$$

Comparing the two results, we have $\left[\tau(X^2) - \tau(X)^2\right]\left[\tau(Y^2) - \tau(Y)^2\right] = 0$, and at least one of the factors vanishes. Without loss of generality we assume $\tau(X^2) - \tau(X)^2 = 0$. Since X is self-adjoint, we have

$$0 = \tau(X^2) - \tau(X)^2 = \tau\left((X - \tau(X)\mathbf{1})^2\right) = \tau\left((X - \tau(X)\mathbf{1})(X - \tau(X)\mathbf{1})^*\right).$$

Hence $X = \tau(X)\mathbf{1}$, and thus the claim holds. □

3.2.1 Non-Crossing Partitions and Joint Moments

Definition 3.17 (Non-crossing partitions)**.** Let S be a finite, totally ordered set, and write $\Pi(S)$ for the set of all *partitions* of S. That is, for every $\pi = \{V_1, \cdots, V_r\} \in \Pi(S)$, its *blocks* $V_1, \cdots, V_r$ are pairwise disjoint, nonempty subsets of S such that $V_1 \cup \cdots \cup V_r = S$. We write $|\pi| := r$ for the cardinality of π.

- If there exist distinct blocks $V_i, V_j \in \pi$, elements $p_1, q_1 \in V_i$ and $p_2, q_2 \in V_j$ such that $p_1 < p_2 < q_1 < q_2$, then π is said to be a *crossing* partition.
- Otherwise, π is said to be a *non-crossing* partition.
- In addition, if each block of partition π contains exactly two elements of S, then π is said to be a *pair-partition* of S.

Notation. We write $\Pi_{\mathrm{NC}}(S)$ for the set of all non-crossing partitions of S, write $\Pi_2(S)$ for the set of all pair-partitions of S, and $\Pi_{\mathrm{NC}_2}(S) = \Pi_{\mathrm{NC}}(S) \cap \Pi_2(S)$ the set of all non-crossing pair-partitions of S. For the case $S = [n] = \{1, \cdots, n\}$, we simply write $\Pi_{\mathrm{NC}}(n) = \Pi_{\mathrm{NC}}([n])$, $\Pi_2(n) = \Pi_2([n])$ and $\Pi_{\mathrm{NC}_2}(n) = \Pi_{\mathrm{NC}_2}([n])$.

For example, for the set $S = \{1, 2, 3, 4, 5, 6\}$, the partitions $\{\{1, 3, 5\}, \{2\}, \{4, 6\}\}$ and $\{\{1, 6\}, \{2, 4\}, \{3, 5\}\}$ are crossing, and the partitions $\{\{1, 4, 5\}, \{2, 3\}, \{6\}\}$ and $\{\{1, 6\}, \{2, 5\}, \{3, 4\}\}$ are non-crossing. A graphical illustration is given below.

It is seen that non-crossing partitions has a "nested" structure.

Lemma 3.18. *Let $n \in \mathbb{N}$. Both the number of non-crossing partitions of theset $\{1, \cdots, n\}$ and the number of non-crossing pair-partitions of the set $\{1, \cdots, 2n\}$ are given by the Catalan number C_n, i.e.*

$$|\Pi_{\mathrm{NC}}(n)| = |\Pi_{\mathrm{NC}_2}(2n)| = C_n$$

Proof. (i) We let $D_0 = 1$, and write $D_n = |\Pi_{\mathrm{NC}}(n)|$. For $n \geq 1$ and $1 \leq k \leq n$, we write $\Pi_{\mathrm{NC}}^{(k)}(n)$ the set of non-crossing partitions $\pi \in \Pi_{\mathrm{NC}}(n)$ for which the block containing 1 contains k as its largest element. Since every non-crossing partition $\pi \in \Pi_{\mathrm{NC}}^{(k)}(n)$ decomposes canonically into $\pi_1 \cup \pi_2$, where $\pi_1 \in \Pi_{\mathrm{NC}}(k)$ and $\pi_2 \in \Pi_{\mathrm{NC}}(\{k+1, \cdots, n\})$. Hence

$$\Pi_{\mathrm{NC}}^{(k)}(n) \simeq \Pi_{\mathrm{NC}}^{(k)}(k) \times \Pi_{\mathrm{NC}}(n-k).$$

By restricting π_1 to $\{1, \cdots, k-1\}$ and using the non-crossing condition, we can establish a bijection between $\Pi_{\mathrm{NC}}^{(k)}(k)$ and $\Pi_{\mathrm{NC}}(k-1)$. Then

$$\Pi_{\mathrm{NC}}^{(k)}(n) \simeq \Pi_{\mathrm{NC}}(k-1) \times \Pi_{\mathrm{NC}}(n-k).$$

Since $\Pi_{\mathrm{NC}}^{(1)}(n), \cdots, \Pi_{\mathrm{NC}}^{(n)}(n)$ is a partition of $\Pi_{\mathrm{NC}}(n)$, we have

$$|\Pi_{\mathrm{NC}}(n)| = D_n = \sum_{k=1}^{n} D_{k-1} D_{n-k}.$$

This is a recursion characterizes the Catalan numbers (C_n).

(ii) We can establish a bijection between noncrossing partitions $\Pi_{\mathrm{NC}_2}(2n)$ and Dyck paths $\mathcal{D}_{2n}$ as follows. Let $\pi \in \Pi_{\mathrm{NC}_2}(2n)$ and $(x_1, \cdots, x_{2n})$ be a Dyck path. For each $j \in [2n]$, let i be the other element of the pair that contains j (an innovative step). Then $x_j = 1$ if and only if $i > j$, i.e. the block $\{i, j\}$ is never visited before; and $x_j = -1$ if and only if $i < j$, i.e. the pair $\{i, j\}$ is already visited before (a returning step). Hence $|\Pi_{\mathrm{NC}_2}(2n)| = |\mathcal{D}_{2n}| = C_n$, and the proof is complete. □

Definition 3.19 (Kernel)**.** Let I be any nonempty set and $i = (i_1, \cdots, i_k) \in I^k$ a multi-index. We define a equivalence relation $\sim_i$ on $[k]$ by

$$p \sim_i q \quad \text{if and only if} \quad i_p = i_q.$$

Then the *kernel* of i is the partition of $[k]$ into the equivalence classes of $\sim_i$:

$$\ker(i) = [k]/\sim_i = \{B_r : r \in \mathrm{Im}(i)\}, \quad B_j = \{p \in [k] : i_p = r\}.$$

For example, if $p = \{2, 4, 2, 7, 5, 4, 2\}$, then $\pi = \ker(p) = \{\{1, 3, 7\}, \{2, 6\}, \{4\}, \{5\}\}$.

For multi-indices with non-crossing kernels, we have a brief formula for computing joint moments.

Theorem 3.20 (Speicher)**.** *Let $(X_i)_{i \in I} \subset \mathcal{A}$ be a family of freely independent random variables. Let $k \in \mathbb{N}$, and let $i = (i_1, \cdots, i_k) \in I^k$ be a multi-index.*

(a) If $\ker(i)$ is non-crossing. Then

$$\tau\left(X_{i_1} \cdots X_{i_k}\right) = \prod_{V \in \ker(i)} \tau\left(X_{i(V)}^{|V|}\right), \quad \text{where } i(V) \text{ is the index corresponding to block } V.$$

(b) If $\ker(i)$ is crossing and $\tau(X_i) = 0$ for each $i \in I$, then

$$\tau\left(X_{i_1} \cdots X_{i_k}\right) = 0.$$

Proof. We first introducing a lemma which can be seen as a generalization of Example 3.15.

Lemma 3.21. *Let $(\mathcal{A}, \tau)$ be a non-commutative probability space, and let $(A_i)_{i \in I}$ be freely independent unital sub-algebras of $\mathcal{A}$. Let $s, t \in \mathbb{N}$, and $i^* \in I$. Assume that $i_1 \neq i_2 \neq \cdots \neq i_k$, $i_p = i^*$ and $i_j \neq i^*$ for $j = 1, \cdots, p-1, p+1, \cdots, k$. Then for $a_j \in \mathcal{A}_{i_j}$, $j = 1, \cdots, k$, we have*

$$\tau\left(a_1 a_2 \cdots a_k\right) = \tau(a_p)\, \tau(a_1 a_2 \cdots a_{p-1} a_{p+1} \cdots a_k).$$

Proof. We write $\overline{a} = a - \tau(a)\mathbf{1}$ for the centered version of a random variable $a \in \mathcal{A}$, i.e. $\tau(\overline{a}) = 0$. Then it suffices to show that

$$\tau(a_1 a_2 \cdots a_{p-1} \overline{a}_p a_{p+1} \cdots a_k) = 0. \tag{3.10}$$

For every $j \neq p$, we write $a_j = \overline{a}_j + \tau(a_j)\mathbf{1}$ and expand $\tau(a_1 a_2 \cdots a_{p-1} \overline{a}_p a_{p+1} \cdots a_k)$ multilinearly. Then sum consists terms of the form

$$\tau(\overline{a}_{q_1} \cdots \overline{a}_{q_s} \overline{a}_p \overline{a}_{r_1} \cdots \overline{a}_{r_t}) \tau(a_{u_1}) \cdots \tau(a_{u_{p-1-s}}) \tau(a_{v_1}) \cdots \tau(a_{v_{k-p-t}}), \quad 0 \leq s \leq p-1, \ 0 \leq t \leq k-p.$$

Note:

- If $q_1 \neq \cdots \neq q_s \neq p \neq r_1 \neq \cdots r_t$, the term is 0 by the definition of free independence. This is always the case when $0 \leq s, t \leq 1$, i.e. the word $\overline{a}_{q_1} \cdots \overline{a}_{q_s} \overline{a}_p \overline{a}_{r_1} \cdots \overline{a}_{r_t}$ is of length at most 3.
- Otherwise, we merge same-color neighboring factors in $\overline{a}_{q_1} \cdots \overline{a}_{q_s} \overline{a}_p \overline{a}_{r_1} \cdots \overline{a}_{r_t}$ to obtain a word of length at most $s + t$. Repeat this centeralization-expansion procedure to the reduced word. This term will finally vanish after at most $k - 1$ steps.

Therefore $\tau(a_1a_2\cdots a_{p-1}\overline{a}_pa_{p+1}\cdots a_k)=0$, and we finish the proof of (3.10). □

Now we prove Theorem 3.20. If the partition $\ker(i)$ is non-crossing, there exists an "innermost" block $V^*=\{p+1,p+2,\cdots,q\}\in\ker(i)$ in the sense that there is no further block located between elements of V^*, i.e. $i_{p+1}=i_{p+2}=\cdots=i_q=r$, and $i_j\neq r$ for $j\in\{1,\cdots,p\}\cup\{q+1,\cdots,k\}$. By Lemma 3.21,

$$\tau\left(X_{i_1}\cdots X_{i_k}\right)=\tau\left(X_{i_{p+1}}\cdots X_{i_q}\right)\tau\left(X_{i_1}\cdots X_{i_p}X_{i_{q+1}}\cdots X_{i_k}\right)=\tau\left(X_r^{q-p}\right)\tau\left(X_{i_1}\cdots X_{i_p}X_{i_{q+1}}\cdots X_{i_k}\right).$$

Clearly, the kernel of the remaining indices $(i_1,\cdots,i_p,i_{q+1},\cdots,i_k)$ is still non-crossing. We apply the same procedure until blocks of $\ker(i)$ are extracted, which implies

$$\tau\left(X_{i_1}\cdots X_{i_k}\right)=\prod_{V\in\ker(i)}\tau\left(X_{i(V)}^{|V|}\right),\quad \text{where } i(V) \text{ is the index corresponding to block } V.$$

On the other hand, if $\ker(i)$ is crossing, we also multiply out the "innermost" block from $\tau(X_{i_1}\cdots X_{i_k})$ repeatedly until the remaining blocks are pairwise crossing, i.e. for any two blocks V_i, $V_{i'}$, there exist positions $p,q\in V_i$ and $p',q'\in V_{i'}$ such that either $p<p'<q<q'$ or $p'<p<q'<q$. Hence the reduced word has alternating indices and has zero trace by free independence. This finishes the proof. □

Theorem 3.22. *Let $(X_{n,i})_{n\in\mathbb{N},i\in[n]}\subset\mathcal{A}$ be a family of random variables in a nonocommutative probability space $(\mathcal{A},\tau)$, and $(\kappa_p)_{p\in\mathbb{N}}$ is a sequence. Suppose that*

- *for every $n\in\mathbb{N}$, the random variables $X_{n,1},\cdots,X_{n,n}$ are identically distributed;*
- *either for every $n\in\mathbb{N}$, the random variables $X_{n,1},\cdots,X_{n,n}$ are classically independent, or for every $n\in\mathbb{N}$, the random variables $X_{n,1},\cdots,X_{n,n}$ are freely independent; and*
- *for all $n,p\in\mathbb{N}$,*

$$\lim_{n\to\infty}n\cdot\tau(X_{n,i}^p)=\kappa_p,\quad i=1,2,\cdots. \tag{3.11}$$

Then for every $k\in\mathbb{N}$,

$$\lim_{n\to\infty}\tau\left[(X_{n,1}+\cdots+X_{n,n})^k\right]=\begin{cases}\sum_{\pi\in\Pi(k)}\prod_{V\in\pi}\kappa_{|V|},\\ \sum_{\pi\in\Pi_{\mathrm{NC}}(k)}\prod_{V\in\pi}\kappa_{|V|}.\end{cases}$$

Proof. For a partition $\pi\in[k]$ and $n\geq k$, the number of multi-indices $i\in[n]^k$ with $\ker(i)=\pi$ is given by $n(n-1)\cdots(n-|\pi|+1)$. If $(X_{n,i})_{i\in[n]}$ are classically independent,

$$\begin{aligned}\tau\left[(X_{n,1}+\cdots+X_{n,n})^k\right]&=\sum_{i\in[n]^k}\tau(X_{i_1}\cdots X_{i_k})=\sum_{\pi\in\Pi(k)}\sum_{\substack{i\in[n]^k\\ \ker(i)=\pi}}\tau(X_{i_1}\cdots X_{i_k})\\&=\sum_{\pi\in\Pi(k)}\sum_{\substack{i\in[n]^k\\ \ker(i)=\pi}}\prod_{V\in\pi}\tau\left(X_{n,i(V)}^{|V|}\right)\\&=\sum_{\pi\in\Pi(k)}\frac{n(n-1)\cdots(n-|\pi|+1)}{n^{|\pi|}}\prod_{V\in\pi}n\cdot\tau\left(X_{n,i(V)}^{|V|}\right).\end{aligned}$$

Since $|\pi|\leq k$ for all $\pi\in\Pi_{\mathrm{NC}}(k)$, and by (3.11), we have

$$\lim_{n\to\infty}\tau\left[(X_{n,1}+\cdots+X_{n,n})^k\right]=\sum_{\pi\in\Pi(k)}\prod_{V\in\pi}\kappa_{|V|}.$$

If $(X_{n,i})_{i\in[n]}$ are freely independent, by Theorem 3.20,

$$\begin{aligned}
\tau\left[(X_{n,1}+\cdots+X_{n,n})^k\right] &= \sum_{i\in[n]^k}\tau(X_{i_1}\cdots X_{i_k}) = \sum_{\pi\in\Pi_{\mathrm{NC}}(k)}\sum_{\substack{i\in[n]^k\\ \ker(i)=\pi}}\tau(X_{i_1}\cdots X_{i_k}) \\
&= \sum_{\pi\in\Pi_{\mathrm{NC}}(k)}\sum_{\substack{i\in[n]^k\\ \ker(i)=\pi}}\prod_{V\in\pi}\tau\left(X_{n,i(V)}^{|V|}\right) \\
&= \sum_{\pi\in\Pi_{\mathrm{NC}}(k)}\frac{n(n-1)\cdots(n-|\pi|+1)}{n^{|\pi|}}\prod_{V\in\pi}n\cdot\tau\left(X_{n,i(V)}^{|V|}\right).
\end{aligned}$$

Therefore

$$\lim_{n\to\infty}\tau\left[(X_{n,1}+\cdots+X_{n,n})^k\right] = \sum_{\pi\in\Pi_{\mathrm{NC}}(k)}\prod_{V\in\pi}\kappa_{|V|},$$

and we finish the proof. □

3.2.2 Free Central Limit Theorem

3.3 Free Cumulants

Review: the classical cumulants. Let X be a $\mathbb{R}$-valued random variable such that the *moment generating function* $M_X(t) := \mathbb{E}[e^{tX}]$ exists for t in a neighborhood $(-\delta, \delta)$ of the origin. The *cumulant generating function* of X is the function $K_X : (-\delta, \delta) \to \mathbb{R}$ defined as

$$K_X(t) = \log \mathbb{E}\left[e^{tX}\right] = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} t^n,$$

and the *cumulants* $(\kappa_n)_{n \in \mathbb{N}}$ are the coefficients in the Taylor expansion of the cumulant generating function about the origin. Indeed, the n-th cumulant κ_n can be obtained by differentiating the above expansion n times and evaluating the result at zero: $\kappa_n = K^{(n)}(0)$.

Note that

$$M_X(t) = \exp\left(\sum_{n=0}^{\infty} \frac{\kappa_n}{n!} t^n\right) = \sum_{m=0}^{\infty} \frac{1}{m!} \left(\sum_{r=1}^{\infty} \frac{\kappa_r}{r!} t^r\right)^m, \quad t \in (-\delta, \delta).$$

Matching the coefficients of Taylor series, we have

$$\frac{\mathbb{E}[X^n]}{n!} = \sum_{m=1}^{n} \frac{1}{m!} \sum_{\substack{r_1, r_2, \cdots, r_m \geq 1 \\ r_1 + r_2 + \cdots + r_m = n}} \frac{\kappa_{r_1} \kappa_{r_2} \cdots \kappa_{r_m}}{r_1! r_2! \cdots r_m!}. \tag{3.12}$$

We interpret this formula combinatorically. The number of ways to split set $\{1, \cdots, n\}$ into an *ordered* list of m blocks with sizes $(r_1, r_2, \cdots, r_m)$ is

$$\frac{n!}{r_1! r_2! \cdots r_m!},$$

and dividing by $m!$ accounts for neglecting the order of the blocks, turning ordered blocks into set partitions. We multiply both sides of (3.12) to get

$$\mathbb{E}[X^n] = \sum_{m=1}^{n} \frac{1}{m!} \sum_{\substack{r_1, r_2, \cdots, r_m \geq 1 \\ r_1 + r_2 + \cdots + r_m = n}} \frac{n!}{r_1! r_2! \cdots r_m!} \kappa_{r_1} \kappa_{r_2} \cdots \kappa_{r_m} = \sum_{m=1}^{n} \sum_{\pi \in \Pi(n) : |\pi| = m} \prod_{V \in \pi} \kappa_{|V|} = \sum_{\pi \in \Pi(n)} \prod_{V \in \pi} \kappa_{|V|}.$$

Thus we obtain the classical *moment-cumulant* formula:

$$\mathbb{E}[X^n] = \sum_{\pi \in \Pi(n)} \prod_{V \in \pi} \kappa_{|V|}, \quad n = 1, 2, \cdots.$$

Example 3.23. We compute the cumulants of some real-valued random variables.

(i) Let $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. For a Gaussian variable $X \sim \mathcal{N}(\mu, \sigma^2)$, we have

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \quad \Leftrightarrow \quad K_X(t) = \mu t + \frac{\sigma^2 t^2}{2}.$$

Hence $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, and $\kappa_n = 0$ for all $n \geq 3$.

(ii) Let $\lambda > 0$. For a Poisson random variable X with rate λ, i.e. $X \sim \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \delta_n$. Then

$$M_X(t) = \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n e^{-\lambda}}{n!} = e^{\lambda(e^t - 1)} \quad \Leftrightarrow \quad K_X(t) = \lambda(e^t - 1) = \sum_{n=1}^{\infty} \frac{\lambda}{n!} t^n.$$

Hence $\kappa_n = \lambda$ for all $n \in \mathbb{N}$.

In this subsection, we establish an extension of the cumulant to noncommutative probability spaces.

3.3.1 The Möbius Inversion

Definition 3.24. Let P be a finite partially ordered set, and $P^{(2)} = \{(\pi, \sigma) : \pi, \sigma \in P, \pi \leq \sigma\}$.

(i) (Interval). For $(\pi, \sigma) \in P^{(2)}$, define $[\pi, \sigma] = \{\tau \in P : \pi \leq \tau \leq \sigma\}$.

(ii) (Convolution). For every two functions $F, G : P^{(2)} \to \mathbb{C}$, define their convolution $F * G : P^{(2)} \to \mathbb{C}$ as the function

$$(F * G)(\pi, \sigma) = \sum_{\tau \in [\pi, \sigma]} F(\pi, \tau) G(\tau, \sigma), \quad (\pi, \sigma) \in P^{(2)}.$$

If $f : P \to \mathbb{C}$, define $f * G : P^{(2)} \to \mathbb{C}$ as the function

$$(f * G)(\sigma) = \sum_{\tau \in P : \tau \leq \sigma} f(\tau) G(\tau, \sigma), \quad \sigma \in P.$$

(iii) The special functions $\delta, \zeta : P^{(2)} \to \mathbb{C}$ are defined as

$$\delta(\pi, \sigma) = \begin{cases} 1, & \text{if } \pi = \sigma, \\ 0, & \text{if } \pi < \sigma, \end{cases} \quad \text{and} \quad \zeta(\pi, \sigma) = 1, \qquad (\pi, \sigma) \in P^{(2)}.$$

(iv) (Incidence algebra) The set of all functions $F : P^{(2)} \to \mathbb{C}$ equipped with pointwise defined addition and with the convolution $*$ as multiplication is a unital (associative) algebra over $\mathbb{C}$, called the *incidence algebra* of P, with δ as its multiplicative identity.

Remark. By definition, it is clear that δ is the unit of the convolution operation: $\delta * F = F * \delta = F$ for all $F : P^{(2)} \to \mathbb{C}$. Furthermore, for $F, G, H : P^{(2)} \to \mathbb{C}$, note that

$$((F * G) * H)(\pi, \sigma) = (F * (G * H))(\pi, \sigma) = \sum_{\rho, \tau \in P :\, \pi \leq \rho \leq \tau \leq \sigma} F(\pi, \rho) G(\rho, \tau) H(\tau, \sigma)$$

Hence $*$ is associative: $(F * G) * H = (F * G) * H$. Generally, $*$ is not commutative.

Theorem 3.25 (Möbius inversion)**.** *Let P be a finite partially ordered set. The zeta function ζ is invertible in the incidence algebra of P, i.e. there exists a function $\mu : P^{(2)} \to \mathbb{C}$, called the **Möbius function**, such that*

$$\mu * \zeta = \zeta * \mu = \delta.$$

Proof. Recursively define

$$\mu(\pi, \pi) = 1, \quad \mu(\pi, \sigma) = - \sum_{\tau \in P : \pi \leq \tau < \sigma} \mu(\pi, \tau).$$

Then we have $\mu * \zeta = \delta$:

$$(\mu * \zeta)(\pi, \sigma) = \sum_{\tau \in [\pi, \sigma]} \mu(\pi, \tau) \zeta(\tau, \sigma) = \sum_{\tau \in [\pi, \sigma]} \mu(\pi, \tau) = \begin{cases} 1, & \mu = \sigma, \\ \mu(\pi, \sigma) - \sum_{\tau \in P : \pi \leq \tau < \sigma} \mu(\pi, \tau) = 0, & \mu < \sigma. \end{cases}$$

We let $\mu(\pi, \sigma) = \zeta(\pi, \sigma) = \delta(\pi, \sigma) = 0$ for $\pi \not\leq \sigma$. Then we have

$$\sum_{\tau \in P} \mu(\pi, \tau) \zeta(\tau, \sigma) = \delta(\tau, \sigma).$$

That is, the matrices $M = (\mu(\pi, \sigma))_{\pi, \sigma \in P}$ and $Z = (\zeta(\pi, \sigma))_{\pi, \sigma \in P}$ satisfies $M \cdot Z = \text{Id}$, and by linear algebra we have $Z \cdot M = \text{Id}$. Hence $\mu * \zeta = \delta$, which finishes the proof. $\square$

Remark. By the above proof, the Möbius function μ on P is recursively defined by

$$\mu(\pi,\pi)=1,\quad \mu(\pi,\sigma)=-\sum_{\tau\in P:\pi\le\tau<\sigma}\mu(\pi,\tau)=-\sum_{\tau\in P:\pi<\tau\le\sigma}\mu(\tau,\sigma),\quad \text{for } \pi<\sigma \text{ in } P.$$

The value of Möbius function μ at (π,σ) depends on the interval $[\pi,\sigma]$.

Corollary 3.26. *Let $f,g:P\to\mathbb{C}$. Then the following statements are equivalent:*

(i) $f=g*\zeta$, meaning

$$f(\sigma)=\sum_{\tau\in P:\pi\le\sigma}g(\pi)\quad \textit{for all } \sigma\in P.$$

(ii) $g=f*\mu$, meaning

$$g(\sigma)=\sum_{\tau\in P:\pi\le\sigma}f(\pi)\mu(\pi,\sigma)\quad \textit{for all } \sigma\in P.$$

Finally, we introduce the invariance of Möbius functions under isomorphisms and Möbius functions on product spaces.

Proposition 3.27. *(i) Let P,Q be finite partial ordered sets, and let $\Phi:P\to Q$ be a order embedding, i.e. $\Phi(\pi)\le\Phi(\sigma)$ in Q if and only if $\pi\le\sigma$ in P. Also assume that $[\Phi(\pi),\Phi(\sigma)]\subset\Phi(P)$ for all $\pi\le\sigma$ in P. Then*

$$\mu_P(\pi,\sigma)=\mu_Q(\Phi(\pi),\Phi(\sigma)),\quad \textit{for all } \pi\le\sigma \textit{ in } P,$$

where μ_P and μ_Q are the Möbius functions on P and Q, respectively.

(ii) Let $P_1,P_2,\cdots,P_k$ be finite partial ordered sets, and consider their direct product

$$P=P_1\times P_2\times\cdots\times P_k,\quad (\pi_1,\cdots,\pi_k)\le(\sigma_1,\cdots,\sigma_k)\ \Leftrightarrow\ \pi_j\le\sigma_j \textit{ for all } j\in[k].$$

Let μ_j be the Möbius function on P_j for $j\in[k]$, and μ the Möbius function on P. Then for $\pi_1\le\sigma_1$ in P_1, $\pi_2\le\sigma_2$ in P_2, $\cdots$, $\pi_k\le\sigma_k$ in P_k, we have

$$\mu\left((\pi_1,\cdots,\pi_k),(\sigma_1,\cdots,\sigma_k)\right)=\mu_1(\pi_1,\sigma_1)\cdots\mu_n(\pi_k,\sigma_k). \tag{3.13}$$

Proof. (i) If $\Phi(\sigma)=\Phi(\pi)$, we have both $\sigma\le\pi$ and $\sigma\ge\pi$ on P, which implies that $\Phi:P\to Q$ is injective. We let $\nu(\widehat{\pi},\widehat{\sigma})=\mu_P(\Phi^{-1}(\widehat{\pi}),\Phi^{-1}(\widehat{\sigma}))$ for all $\widehat{\pi}\le\widehat{\sigma}$ in $\Phi(P)$. Then

$$\begin{aligned}(\nu*\zeta)(\widehat{\pi},\widehat{\sigma})&=\sum_{\widehat{\tau}\in[\widehat{\pi},\widehat{\sigma}]}\nu(\widehat{\pi},\widehat{\tau})=\sum_{\widehat{\tau}\in[\widehat{\pi},\widehat{\sigma}]}\mu_P(\Phi^{-1}(\widehat{\pi}),\Phi^{-1}(\widehat{\tau}))\\&=\sum_{\tau\in[\Phi^{-1}(\widehat{\pi}),\Phi^{-1}(\widehat{\sigma})]}\mu_P(\Phi^{-1}(\widehat{\pi}),\tau)=\delta_P(\Phi^{-1}(\widehat{\pi}),\Phi^{-1}(\widehat{\sigma}))=\delta_Q(\widehat{\pi},\widehat{\sigma}).\end{aligned}$$

Hence $\nu*\zeta=\delta_Q|_{\Phi(P)}$. Simiarly we can prove $\zeta*\nu=\delta_Q|_{\Phi(P)}$, and thus ν is the Möbius function on $\Phi(P)$.

(ii) We let $\widetilde{\mu}$ be the right-hand side of (3.13). For $\pi=(\pi_1,\cdots,\pi_k)\le\sigma=(\sigma_1,\cdots,\sigma_k)$,

$$\begin{aligned}(\widetilde{\mu}*\zeta)(\pi,\sigma)&=\sum_{\tau\in[\pi,\sigma]}\widetilde{\mu}(\pi,\tau)=\sum_{\tau_1\in[\pi_1,\sigma_1]}\cdots\sum_{\tau_k\in[\pi_k,\sigma_k]}\mu_1(\pi_1,\tau_1)\cdots\mu_k(\pi_k,\tau_k)\\&=\prod_{j=1}^{k}\sum_{\tau_j\in[\pi_j,\sigma_j]}\mu_j(\pi_j,\tau_j)=\prod_{j=1}^{k}\delta_j(\pi_j,\sigma_j)=\delta(\pi,\sigma).\end{aligned}$$

Simiarly we can prove $\zeta*\widetilde{\mu}=\delta$. Hence $\widetilde{\mu}$ is the Möbius function on P. □

3.3.2 Free Cumulants

In the following discussion, we assume $P = \Pi_{\text{NC}}(n)$, where the partial order $\leq$ is defined as follows:

$$\pi \leq \sigma \quad \Leftrightarrow \quad \text{every block of } \pi \text{ is contained in a block of } \sigma.$$

For example, $\{\{1,7\},\{2,5\},\{3,4\},\{6\}\} \leq \{\{1,6,7\},\{2,5\},\{3,4\}\} \leq \{\{1,6,7\},\{2,3,4,5\}\}$. We denote the minimal element by $0_n = \{\{1\},\{2\},\cdots,\{n\}\}$, and the maximal element by $1_n = \{\{1,2,\cdots,n\}\}$.

We first show that $\Pi_{\text{NC}}(n)$ is a lattice under the partial order defined above.

Theorem 3.28 (Non-crossing lattice)**.** *For each $n \in \mathbb{N}$, the partially ordered set $\Pi_{\text{NC}}(n)$ is a lattice: for every $\pi, \sigma \in \Pi_{\text{NC}}(n)$,*

- *there exists a unique smallest $\upsilon \in \Pi_{\text{NC}}(n)$ with the properties $\upsilon \geq \pi$ and $\upsilon \geq \sigma$, which is written $\pi \vee \sigma$ and called the **join** of π and σ; and*
- *there exists a unique largest $\lambda \in \Pi_{\text{NC}}(n)$ with the properties $\lambda \leq \pi$ and $\lambda \leq \sigma$, which is written $\pi \wedge \sigma$ and called the **meet** of π and σ.*

Proof. For $\pi, \sigma \in \Pi_{\text{NC}}(n)$, we simply define

$$\lambda = \pi \wedge \sigma = \{V_i \cap W_j : V_i\}, \quad \upsilon = \pi \vee \sigma = \bigwedge \{\rho \in \Pi_{\text{NC}}(n) : \rho \geq \pi, \rho \geq \sigma\}.$$

Then $\lambda = \pi \wedge \sigma$ is a finer parition than π and σ and is non-crossing, and λ is maximal since $p \sim_\pi q$ and $p \sim_\sigma q$ implies $p \sim_\lambda q$. Also, $\upsilon = \pi \vee \sigma$ is coarser. By induction, υ is also non-crossing, which finishes the proof. $\square$

Now we see how to define a multiplicative family on these lattices.

Definition 3.29 (Multiplicative family)**.** Let $\mathcal{A}$ be a unital associative algebra over $\mathbb{C}$, and $\varphi_n : \mathcal{A}^n \to \mathbb{C}$ a family of multilinear functionals. We extend $(\varphi_n)_{n\in\mathbb{N}}$ to a family $(\varphi_\pi)_{\pi\in\Pi_{\text{NC}}}$ of multilinear functionals on

$$\Pi_{\text{NC}} = \bigcup_{n=1}^{\infty} \Pi_{\text{NC}}(n),$$

by defining, for each $n \in \mathbb{N}$, $\pi \in \Pi_{\text{NC}}(n)$ and $a_1, \cdots, a_n \in \mathcal{A}$,

$$\varphi_\pi(a_1, \cdots, a_n) = \prod_{V\in\pi} \varphi_{|V|}(a_1, \cdots, a_n | V),$$

where for $V = \{i_1, \cdots, i_s\}$ with $1 \leq i_1 < \cdots < i_s \leq n$, we have $\varphi_s(a_1, \cdots, a_n|V) = \varphi_s(a_{i_1}, \cdots, a_{i_s})$. Then $(\varphi_\pi)_{\pi\in\Pi_{\text{NC}}}$ is called the *multiplicative family of functionals* on Π_{NC} determined by $(\varphi_n)_{n\in\mathbb{N}}$.

Definition 3.30 (Free cumulants)**.** Let $(\mathcal{A}, \tau)$ be a non-commutative probability space. We define, for every $n \in \mathbb{N}$, the multilinear functional

$$\varphi_n(a_1, \cdots, a_n) = \tau(a_1 \cdots a_n), \quad a_1, \cdots, a_n \in \mathcal{A},$$

and extend $(\varphi_n)_{n\in\mathbb{N}}$ to a multiplicative familt of functionals $(\varphi_\pi)_{\pi\in\Pi_{\text{NC}}}$ by defining

$$\varphi_\pi(a_1, \cdots, a_n) = \prod_{V\in\pi} \varphi_{|V|}(a_1, \cdots, a_n|V).$$

The corresponding *free cumulants* $(\kappa_\pi)_{\pi\in\Pi_{\text{NC}}}$ are the multilinear functionals defined by $\kappa = \varphi * \mu$, i.e.

$$\kappa_\sigma(a_1, \cdots, a_n) = \sum_{\pi\in\Pi_{\text{NC}}(n):\pi\leq\sigma} \varphi_\pi(a_1, \cdots, a_n)\mu(\pi, \sigma), \quad a_1, \cdots, a_n \in \mathcal{A}.$$

Proposition 3.31. *Let $(\mathcal{A},\tau)$ be a non-commutative probability space with free cumulants $(\kappa_\pi)_{\pi\in\Pi_{\mathrm{NC}}}$. Then $(\kappa_\pi)_{\pi\in\Pi_{\mathrm{NC}}}$ is the multiplicative family on Π_{NC} defined by $(\kappa_n)_{n\in\mathbb{N}}$, where $\kappa_n=\kappa_{1_n}$ for every $n\in\mathbb{N}$.*

Proof. Let $n\in\mathbb{N}$, $\sigma=\{V_1,\cdots,V_r\}\in\Pi_{\mathrm{NC}}(n)$ and $a_1,\cdots,a_n\in\mathcal{A}$. If $\pi\in\Pi_{\mathrm{NC}}(n)$ and $\pi\le\sigma$, we decompose $\pi=\pi_1\cup\pi_2\cup\cdots\cup\pi_r$, where $\pi_j\in\Pi_{\mathrm{NC}}(V_j)$ for every $j\in[r]$. Then the interval $[\pi,\sigma]$ decomposes accordingly:

$$[\pi,\sigma]\simeq\left[\pi_1,1_{|V_1|}\right]\times\cdots\times\left[\pi_r,1_{|V_r|}\right]\subset\Pi_{\mathrm{NC}}(V_1)\times\cdots\times\Pi_{\mathrm{NC}}(V_r).$$

Since $\mu(\pi,\sigma)$ depends only on the interval $[\pi,\sigma]$, and by Proposition 3.27,

$$\mu(\pi,\sigma)=\mu\left(\pi_1,1_{|V_1|}\right)\cdots\mu\left(\pi_r,1_{|V_r|}\right),$$

and thus

$$\begin{aligned}\kappa_\sigma(a_1,\cdots,a_n)&=\sum_{\pi\in\Pi_{\mathrm{NC}}(n):\pi\le\sigma}\varphi_\pi(a_1,\cdots,a_n)\mu(\pi,\sigma)\\&=\sum_{\pi_1\in\Pi_{\mathrm{NC}}(V_1)}\cdots\sum_{\pi_r\in\Pi_{\mathrm{NC}}(V_r)}\prod_{j=1}^r\varphi_{\pi_j}(a_1,\cdots,a_n|V_j)\mu\left(\pi_j,1_{|V_j|}\right)\\&=\prod_{j=1}^r\underbrace{\sum_{\pi_j\in\Pi_{\mathrm{NC}}(V_j)}\varphi_{\pi_j}(a_1,\cdots,a_n|V_j)\mu\left(\pi_j,1_{|V_j|}\right)}_{\kappa_{|V_j|}(a_1,\cdots,a_n|V_j)}=\prod_{V\in\sigma}\kappa_{|V|}(a_1,\cdots,a_n|V).\end{aligned}$$

Then we finish the proof. ☐

Remark. By the Möbius inversion, $\kappa=\varphi*\mu$ implies $\varphi=\kappa*\zeta$. Hence for any $\pi\le\sigma$ in $\Pi_{\mathrm{NC}}(n)$,

$$\kappa_\sigma(a_1,\cdots,a_n)=\sum_{\pi\in\Pi_{\mathrm{NC}}(n):\pi\le\sigma}\varphi_\pi(a_1,\cdots,a_n)\mu(\pi,\sigma),\quad a_1,\cdots,a_n\in\mathcal{A},$$

and

$$\varphi_\sigma(a_1,\cdots,a_n)=\sum_{\pi\in\Pi_{\mathrm{NC}}(n):\pi\le\sigma}\kappa_\pi(a_1,\cdots,a_n),\quad a_1,\cdots,a_n\in\mathcal{A}.$$

These are known as the *free moment-cumulant formula.*

Example 3.32. Let $(\mathcal{A},\tau)$ be a non-commutative probability space. By the free moment-cumulant formula,

$$\begin{aligned}\tau(a_1)&=\kappa_1(a_1),\\\tau(a_1a_2)&=\kappa_2(a_1a_2)+\kappa_1(a_1)\kappa_1(a_2),\\\tau(a_1a_2a_3)&=\kappa_3(a_1a_2a_3)+\kappa_2(a_1a_2)\kappa_1(a_3)+\kappa_2(a_1a_3)\kappa_1(a_2)+\kappa_1(a_1)\kappa_2(a_2a_3)+\kappa_1(a_1)\kappa_1(a_2)\kappa_1(a_3),\\\tau(a_1a_2a_3a_4)&=\kappa_4(a_1a_2a_3a_4)+\kappa_2(a_1a_2)\kappa_2(a_3a_4)+\kappa_2(a_1a_4)\kappa_2(a_2a_3)+\kappa_1(a_1)\kappa_1(a_2)\kappa_1(a_3)\kappa_1(a_4)\\&\quad+\kappa_1(a_1)\kappa_3(a_2a_3a_4)+\kappa_3(a_1a_3a_4)\kappa_1(a_2)+\kappa_3(a_1a_2a_4)\kappa_1(a_3)+\kappa_3(a_1a_2a_3)\kappa_1(a_4)\\&\quad+\kappa_1(a_1)\kappa_1(a_2)\kappa_2(a_3a_4)+\kappa_1(a_1)\kappa_1(a_3)\kappa_2(a_2a_4)+\kappa_1(a_1)\kappa_1(a_4)\kappa_2(a_2a_3)\\&\quad+\kappa_2(a_1a_4)\kappa_1(a_2)\kappa_1(a_3)+\kappa_2(a_1a_3)\kappa_1(a_2)\kappa_1(a_4)+\kappa_2(a_1a_2)\kappa_1(a_3)\kappa_1(a_4),\quad\cdots.\end{aligned}$$

Therefore

$$\begin{aligned}\kappa_1(a_1)&=\tau(a_1),\\\kappa_2(a_1a_2)&=\tau(a_1a_2)-\tau(a_1)\tau(a_2),\\\kappa_3(a_1a_2a_3)&=\tau(a_1a_2a_3)-\tau(a_1)\tau(a_2a_3)-\tau(a_2)\tau(a_1a_3)-\tau(a_3)\tau(a_1a_2)+2\tau(a_1)\tau(a_2)\tau(a_3),\quad\cdots.\end{aligned}$$

Now we fix positive integers $m \le n$ and indices $0 = i_0 < i_1 < i_2 < \cdots < i_{m-1} < i_m = n$. We write

$$\upsilon = \{\{1, \cdots, i_1\}, \{i_1 + 1, \cdots, i_2\}, \cdots, \{i_{m-1} + 1, \cdots, n\}\} = \{V_1, \cdots, V_m\}$$

For each non-crossing partition $\pi \in \Pi_{\mathrm{NC}}(m)$, we let

$$\widehat{\pi} = \left\{ \bigcup_{j \in U} V_j : U \in \pi \right\} \in [\upsilon, 1_n] \subset \Pi_{\mathrm{NC}}(n). \tag{3.14}$$

It is easy to verify that $\pi \mapsto \widehat{\pi}$ is an order embedding from $\Pi_{\mathrm{NC}}(m)$ onto $[\upsilon, 1_n] \subset \Pi_{\mathrm{NC}}(n)$. Furthermore, by Proposition 3.27, $\mu(\pi, \sigma) = \mu(\widehat{\pi}, \widehat{\sigma})$ for all $\pi \le \sigma$ in $\Pi_{\mathrm{NC}}(m)$.

Proposition 3.33. *Let $(\mathcal{A}, \tau)$ be a non-commutative probability space with free cumulants $(\kappa_n)_{n \in \mathbb{N}}$. Fix positive integers $m \le n$ and indices $0 = i_0 < i_1 < i_2 < \cdots < i_{m-1} < i_m = n$, and write*

$$\upsilon = \{\{1, \cdots, i_1\}, \{i_1 + 1, \cdots, i_2\}, \cdots, \{i_{m-1} + 1, \cdots, n\}\} = \{V_1, \cdots, V_m\}.$$

For $a_1, \cdots, a_n \in \mathcal{A}$, define $A_j = a_{i_{j-1}+1} \cdots a_{i_j}$ for $j \in [m]$. Then for all $\sigma \in \Pi_{\mathrm{NC}}(m)$,

$$\kappa_\sigma(A_1, \cdots, A_m) = \sum_{\substack{\pi \in \Pi_{\mathrm{NC}}(n) \\ \upsilon \vee \pi = \widehat{\sigma}}} \kappa_\pi(a_1, \cdots, a_n).$$

In particular,

$$\kappa_m(A_1, \cdots, A_m) = \sum_{\substack{\pi \in \Pi_{\mathrm{NC}}(n) \\ \upsilon \vee \pi = 1_n}} \kappa_\pi(a_1, \cdots, a_n).$$

Proof. For every $\sigma \in \Pi_{\mathrm{NC}}(m)$, we have

$$\begin{aligned}
\kappa_\sigma(A_1, \cdots, A_m) &= \sum_{\rho \in \Pi_{\mathrm{NC}}(m) : \rho \le \sigma} \varphi_\rho(A_1, \cdots, A_m) \mu(\rho, \sigma) \\
&= \sum_{\omega \in [\upsilon, \widehat{\sigma}]} \varphi_\omega(a_1, \cdots, a_n) \mu(\omega, \widehat{\sigma}) \qquad \text{(change the variable } \omega = \widehat{\rho}\text{)} \\
&= \sum_{\omega \in [\upsilon, \widehat{\sigma}]} \sum_{\pi \in [0_n, \omega]} \kappa_\pi(a_1, \cdots, a_n) \mu(\omega, \widehat{\sigma}) \\
&= \sum_{\pi \in [0_n, \widehat{\sigma}]} \sum_{\omega \in [\upsilon \vee \pi, \widehat{\sigma}]} \kappa_\pi(a_1, \cdots, a_n) \mu(\omega, \widehat{\sigma}) \\
&= \sum_{\pi \in \Pi_{\mathrm{NC}}(n)} \kappa_\pi(a_1, \cdots, a_n) \sum_{\omega \in [\upsilon \vee \pi, \widehat{\sigma}]} \mu(\omega, \widehat{\sigma}) = \sum_{\substack{\pi \in \Pi_{\mathrm{NC}}(n) \\ \upsilon \vee \pi = \widehat{\sigma}}} \kappa_\pi(a_1, \cdots, a_n),
\end{aligned}$$

where the last equality follows from Möbius inversion $\zeta * \mu = \delta$. This finishes the proof. □

Following is a useful corollary of the above Proposition.

Proposition 3.34. *Let $(\mathcal{A}, \tau)$ be a non-commutative probability space with free cumulants $(\kappa_n)_{n \in \mathbb{N}}$. Let $n \in \mathbb{N}$ and $n \ge 2$, and $a_1, \cdots, a_n \in \mathcal{A}$. If there exists at least one $i \in [n]$ such that $a_i = \mathbf{1}$, then $\kappa_n(a_1, \cdots, a_n) = 0$.*

Proof. Since $\mathbf{1}$ commutes with all elemnts in $\mathcal{A}$, we may assume $a_n = \mathbf{1}$, and proceed by induction. For the base case $n = 2$, we have $\kappa_2(a_1, \mathbf{1}) = \tau(a_1 \cdot \mathbf{1}) - \tau(a_1)\tau(\mathbf{1}) = 0$. Now we assume $\kappa_r(a_1, \cdots, a_{r-1}, \mathbf{1}) = 0$ for $r = 1, \cdots, n$. Let $\upsilon = \{\{1\}, \{2\}, \cdots, \{n-1\}, \{n, n+1\}\}$. Then

$$\kappa_n(a_1, \cdots, a_n \cdot \mathbf{1}) = \sum_{\substack{\pi \in \Pi_{\mathrm{NC}}(n+1) \\ \upsilon \vee \pi = 1_{n+1}}} \kappa_\pi(a_1, \cdots, a_n, 1).$$

If $\pi \in \Pi_{\rm NC}(n+1)$ and $\upsilon \vee \pi = 1_{n+1}$, either of the following cases holds:

- $\pi = 1_n$, and $\kappa_\pi(a_1, \cdots, a_n, 1) = \kappa_{n+1}(a_1, \cdots, a_n, 1)$; or
- there exists $r \in \mathbb{N}_0$ with $r < n$ such that $\pi = \{\{1, 2, \cdots, r, n+1\}, \{r+1, \cdots, n\}\}$,

$$\kappa_\pi(a_1, \cdots, a_n, 1) = \kappa_{r+1}(a_1, \cdots, a_r, \mathbf{1})\kappa_{n-r}(a_{r+1}, \cdots, a_n) = \begin{cases} \kappa_1(\mathbf{1})\kappa_n(a_1, \cdots, a_n), & \text{if } r = 0, \\ 0, & \text{if } r > 0, \end{cases}$$

where the case $r > 0$ follows by induction hypothesis.

To summarize,

$$\kappa_n(a_1, \cdots, a_n \cdot \mathbf{1}) = \kappa_{n+1}(a_1, \cdots, a_n, 1) + \kappa_1(\mathbf{1})\kappa_n(a_1, \cdots, a_n).$$

Since $\kappa_1(\mathbf{1}) = \tau(\mathbf{1}) = 1$, we have $\kappa_{n+1}(a_1, \cdots, a_n, 1) = 0$, which finishes the induction step. □

Now we can establish of the equivalence between free independence and vanishing of mixed cumulants.

Theorem 3.35 (Speicher). *Let $(\mathcal{A}, \tau)$ be a non-commutative probability space with free cumulants $(\kappa_n)_{n\in\mathbb{N}}$, and let $(\mathcal{A}_i)_{i\in I}$ be a family of unital subalgebras of $\mathcal{A}$ over $\mathbb{C}$. Then, the following statements are equivalent:*

(i) The sub-algebras $(\mathcal{A}_i)_{i\in I}$ are freely independent in $(\mathcal{A}, \tau)$;

(ii) Mixed cumulants in the sub-algebras $(\mathcal{A}_i)_{i\in I}$ vanish, i.e. for all $n \in \mathbb{N}$ with $n \geq 2$, all $i \in I^n$ and all $a_1 \in \mathcal{A}_{i_1}$, $\cdots$, $a_n \in \mathcal{A}_{i_n}$, we have $\kappa_n(a_1, \cdots, a_n) = 0$ whenever there exists $j, k \in [n]$ such that $i_j \neq i_k$.

Proof. (ii) ⇒ (i). Fix $n \in \mathbb{N}$, $i \in I^n$ such that $i_1 \neq i_2 \neq \cdots \neq i_{n-1} \neq i_n$, and $a_1, \cdots, a_n \in \mathcal{A}$ such that $a_j \in \mathcal{A}_{i_j}$ and $\tau(a_j) = 0$ for all $j \in [n]$. It suffices to show that $\tau(a_1 \cdots a_n) = 0$. By the free moment-cumulant formula,

$$\tau(a_1 \cdots a_n) = \sum_{\pi\in\mathrm{NC}(n)} \kappa_\pi(a_1, \cdots, a_n) = \sum_{\pi\in\mathrm{NC}(n)} \prod_{V\in\pi} \kappa_{|V|}(a_1, \cdots, a_n|V).$$

For every $\pi \in \Pi_{\rm NC}(n)$, we can take an innermost block V^* which contains either only one number $j \in [n]$ or two consecutive numbers $\{j, j+1\} \subset [n]$, with $i_j \neq i_{j+1}$. In either case we have $\kappa_{|V^*|}(a_1, \cdots, a_n|V^*) = 0$. Hence the product $\prod_{V\in\pi} \kappa_{|V|}(a_1, \cdots, a_n|V)$ vanishes, and $\tau(a_1 \cdots a_n) = 0$.

(i) ⇒ (ii). Let $n \in \mathbb{N}$, $n \in I^n$ and $a_1, \cdots, a_n \in \mathcal{A}$ such that $a_j \in \mathcal{A}_{i_j}$ for all $j \in [n]$. We first assume that $a_1, \cdots, a_n$ are centered and alternating, i.e. $i_1 \neq i_2 \neq \cdots \neq i_{n-1} \neq i_n$. Similar to our reasoning in (ii) ⇒ (i), for every $\pi \in \Pi_{\rm NC}(n)$, we have $\varphi_{|V^*|}(a_1, \cdots, a_n|V^*) = 0$ for an innermost block $V^* \in \Pi_{\rm NC}(n)$, and

$$\kappa_n(a_1, \cdots, a_n) = \sum_{\pi\in\Pi_{\rm NC}(n)} \prod_{V\in\pi} \varphi_{|V|}(a_1, \cdots, a_n|V)\, \mu(\pi, 1_n) = 0.$$

By Proposition 3.34, we can drop the assumption $\tau(a_j) = 0$ for all $j \in [n]$, since

$$\kappa_n(a_1, \cdots, a_n) = \kappa_n(a_1 - \tau(a_1)\mathbf{1}, \cdots, a_n - \tau(a_n)\mathbf{1}).$$

Then it remains to show vanishing of the cumulant if arguments are only mixed, not necessarily alternating, i.e. there exists $j, k \in [n]$ such that $i_j \neq i_k$, but not necessarily $i_1 \neq i_2 \neq \cdots \neq i_{n-1} \neq i_n$.

We prove this by induction. For the base case $n = 2$, variables a_1, a_2 are mixed means they are free, and $\kappa_2(a_1, a_2) = \tau(a_1 a_2) - \tau(a_1)\tau(a_2) = 0$. For $n \geq 3$, we multiply together neighbors of the same color to get an alternating representation $A_1 \cdots A_m = a_1 \cdots a_n$, where $2 \leq m \leq n$ because $a_1, \cdots, a_n$ are mixed, and $A_j \in \mathcal{A}_{i'_j}$ with $i'_1 \neq i'_2 \neq \cdots \neq i'_{m-1} \neq i'_m$. We may assume $m < n$, otherwise the case is already handled in the alternating case. By the above conclusion and Proposition 3.33 (we keep the notation υ),

$$0 = \kappa_m(A_1, \cdots, A_m) = \kappa_n(a_1, \cdots, a_n) + \sum_{\substack{\pi\in\Pi_{\rm NC}(n)\\ \pi\vee\upsilon=1_n, \pi<1_n}} \kappa_\pi(a_1, \cdots, a_n)$$

By induction hypothesis, any $\pi \in \Pi_{\text{NC}}(n)$ can yield a potentially nonzero cumulant $\kappa_\pi(a_1, \cdots, a_n)$ only if each block of π connects exclusively elements from the same subalgebra, i.e. $\pi \le \ker(i)$. Note that $\upsilon \le \ker(i)$ also. Hence $\pi \vee \upsilon = 1_n$ only if $\ker(i) = 1_n$, i.e. all (a_j) are from the same sub-algebra. But this would contradict the fact $m \ge 2$. Hence there are no $\pi \in \Pi_{\text{NC}}(n)$ preceding 1_n yielding nonzero cumulants, and

$$\kappa_n(a_1, \cdots, a_n) = \kappa_m(A_1, \cdots, A_m) = 0.$$

Thus we finish the proof. □

We can refine this criterion to a similar characterization of free independence for random variables.

Theorem 3.36 (Speicher). *Let $(\mathcal{A}, \tau)$ be a non-commutative probability space with free cumulants $(\kappa_n)_{n\in\mathbb{N}}$, and let $(X_i)_{i\in I} \subset \mathcal{A}$ be a family of random variables. Then, the following statements are equivalent:*

(i) The random variables $(X_i)_{i\in I}$ are freely independent in $(\mathcal{A}, \tau)$;

(ii) Mixed cumulants in the random variables $(X_i)_{i\in I}$ vanish, i.e. for all $n \in \mathbb{N}$ with $n \ge 2$ and all $i \in I^n$, we have $\kappa_n(X_{i_1}, \cdots, X_{i_n}) = 0$ whenever there exists $j, k \in [n]$ such that $i_j \ne i_k$.

Proof. The direction (i) to (ii) is just a special case of 3.35. To prove the direction (ii) to (i), we can show that the mixed cumulants of unital sub-algebras $\mathcal{A}_i = \text{Alg}(X_i, \mathbf{1})$ vanish, which is also clear by multilinearity of cumulants and the condition (ii). □

Following are some immediate corollaries of the vanishing of mixed cumulants.

Corollary 3.37. *Let $(\mathcal{A}_i)_{i\in I}$ be a freely independent family of unital sub-algebras of a non-commutative probability space $(\mathcal{A}, \tau)$, and $I_1, \cdots, I_m \subset I$ are pairwise disjoint subsets of I. Then the family of sub-algebras*

$$\mathcal{B}_j = \text{Alg}\left(\bigcup_{i\in I_j} \mathcal{A}_i\right), \quad j = 1, 2, \cdots, m$$

is also freely independent.

Corollary 3.38. *Let $X_1, X_2, \cdots, X_k$ be freely independent random variables in $(\mathcal{A}, \tau)$. Then*

$$\kappa_n(X_1 + X_2 + \cdots + X_k) = \kappa_n(X_1) + \kappa_n(X_2) + \cdots + \kappa_n(X_k)$$

for all $n \in \mathbb{N}$, where $\kappa_n(X)$ is short for $\kappa_n(\underbrace{X, \cdots, X}_{n})$.

3.3.3 Free Cumulant-Generating Functions

To end this part, we study the cumulants of a single random variable X in a noncommutative probability space $(\mathcal{A}, \tau)$ with cumulants $(\kappa_n)_{n\in\mathbb{N}}$. The moment and cumulant sequences of X, denoted by $(m_n)_{n\in\mathbb{N}}$ and $(\kappa_n)_{n\in\mathbb{N}}$, respectively, are

$$m_n = \tau(X^n) = \varphi_n(X, \cdots, X), \quad \kappa_n = \kappa_n(X, \cdots, X), \quad n = 1, 2, \cdots.$$

We can extend (m_n) and (κ_n) to multiplicative functions $m : \Pi_{\text{NC}} \to \mathbb{C}$ and $\kappa : \Pi_{\text{NC}} \to \mathbb{C}$ via

$$m(\pi) = m_\pi := \prod_{V\in\pi} m_{|V|}, \quad \kappa(\pi) = \kappa_\pi := \prod_{V\in\pi} \kappa_{|V|}, \quad \pi \in \Pi_{\text{NC}}. \tag{3.15}$$

Then m and κ satisfies $\kappa = m * \mu$ and $m = \kappa * \zeta$. These combinatorial relations are nice but not convenient for concrete calculations. We introduce an analytic reformulation.

Theorem 3.39. *Let $(m_n)_{n\in\mathbb{N}}$ and $(\kappa_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{C}$, and let $\kappa, \pi : \Pi_{\mathrm{NC}} \to \mathbb{C}$ be the multiplicative functions defined in* (3.15)*. Consider the corresponding formal power series in $\mathbb{C}[\![z]\!]$:*

$$M(z) = 1 + \sum_{n=1}^{\infty} m_n z^n, \quad C(z) = 1 + \sum_{n=1}^{\infty} \kappa_n z^n.$$

Then the following statements are equivalent:

(i) $m = \kappa * \zeta$, i.e. $m_n = \sum_{\pi\in\Pi_{\mathrm{NC}}(n)} \kappa_\pi$ for all $n \in \mathbb{N}$;

(ii) For all $n \in \mathbb{N}$,

$$m_n = \sum_{r=1}^{n} \sum_{\substack{i_1,\cdots,i_r\geq 0\\ i_1+\cdots+i_r+r=n}} \kappa_r m_{i_1}\cdots m_{i_r}.$$

(iii) We have as functional relation in $\mathbb{C}[\![z]\!]$ that

$$C(z\cdot M(z)) = M(z).$$

Proof. (i) $\Rightarrow$ (ii). Fix $n \in \mathbb{N}$, $\pi \in \Pi_{\mathrm{NC}}(\pi)$, and let $V \in \pi$ be the block containing 1. If $|V| = r$, we write

$$\pi = \{V\} \cup \pi_1 \cup \cdots \cup \pi_r, \quad V = \{1, i_1+2, i_1+i_2+3, \cdots, i_1+\cdots+i_{r-1}+r\},$$

where $\pi_j \in \Pi_{\mathrm{NC}}(i_j)$ is the sub-partition between the j-th and $(j+1)$-th elements of V for $j = 1, \cdots, r-1$, and $\pi_r \in \Pi_{\mathrm{NC}}(i_r)$ is the sub-partition on the right-side of i_r. Note it could be the case $i_j = 0$ and $\pi_j = \emptyset$. Using this decomposition, we have

$$\begin{aligned} m_n = \sum_{\pi\in\Pi_{\mathrm{NC}}(n)} \kappa_\pi &= \sum_{r=1}^{n} \sum_{\substack{i_1,\cdots,i_r\geq 0\\ i_1+\cdots+i_r+r=n}} \sum_{\pi_1\in\Pi_{\mathrm{NC}}(i_1)} \cdots \sum_{\pi_r\in\Pi_{\mathrm{NC}}(i_r)} \kappa_r\kappa_{\pi_1}\cdots\kappa_{\pi_r} \\ &= \sum_{r=1}^{n} \sum_{\substack{i_1,\cdots,i_r\geq 0\\ i_1+\cdots+i_r+r=n}} \kappa_r \left(\sum_{\pi_1\in\Pi_{\mathrm{NC}}(i_1)} \kappa_{\pi_1}\right)\cdots\left(\sum_{\pi_r\in\Pi_{\mathrm{NC}}(i_r)} \kappa_{\pi_r}\right) \\ &= \sum_{r=1}^{n} \sum_{\substack{i_1,\cdots,i_r\geq 0\\ i_1+\cdots+i_r+r=n}} \kappa_r m_{i_1}\cdots m_{i_r}. \end{aligned}$$

(ii) $\Rightarrow$ (iii). We plug in the expression of m_n in (ii) to the expansion of $M(z)$ to obtain

$$\begin{aligned} M(z) = 1 + \sum_{n=1}^{\infty} m_n z^n &= 1 + \sum_{n=1}^{\infty}\sum_{r=1}^{n} \sum_{\substack{i_1,\cdots,i_r\geq 0\\ i_1+\cdots+i_r+r=n}} \kappa_s m_{i_1}\cdots m_{i_r} z^n \\ &= 1 + \sum_{n=1}^{\infty}\sum_{r=1}^{n} \sum_{\substack{i_1,\cdots,i_r\geq 0\\ i_1+\cdots+i_r+r=n}} (\kappa_r z^r)(m_{i_1} z^{i_1})\cdots(m_{i_r} z^{i_r}) \\ &= 1 + \sum_{r=1}^{\infty} \sum_{i_1,\cdots,i_r=0}^{\infty} (\kappa_r z^r)(m_{i_1} z^{i_1})\cdots(m_{i_r} z^{i_r}) \\ &= 1 + \sum_{r=1}^{\infty} \kappa_r z^r \left(\sum_{i_1=0}^{\infty} m_{i_1} z^{i_1}\right)^r = 1 + \sum_{r=1}^{\infty} \kappa_r z^r M(z)^r = C(z\cdot M(z)). \end{aligned}$$

(iii) $\Rightarrow$ (i). Since both (i) and (iii) determine a unique relation between sequences $(m_n)_{n\in\mathbb{N}}$ and $(\kappa_n)_{n\in\mathbb{N}}$, the implication (i) $\Rightarrow$ (iii) also gives (iii) $\Rightarrow$ (i). □

Proposition 3.40. *Assume $m = \kappa * \zeta$. Then $(\kappa_n)_{n\in\mathbb{N}}$ is exponentially bounded if and only if $(m_n)_{n\in\mathbb{N}}$ is exponentially bounded.*

Proof. Step I. Note that the sequences $\delta_n = \delta(0_n, 1_n)$ and $\mu_n = \mu(0_n, 1_n)$ satisfies $\delta = \mu * \zeta$. We consider the analytic function C defined by

$$\begin{aligned} C(z) = \frac{1+\sqrt{1+4z}}{2} &= 1 + \frac{1}{2}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{(2n-1)!!}{2^n n!}(4z)^n \\ &= 1 + \frac{1}{2}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{(2n-2)!}{2^{2n-1}n!(n-1)!}(4z)^n = 1 + \sum_{n=1}^{\infty}(-1)^{n-1}C_{n-1}(4z)^n. \end{aligned}$$

Then we have $C(z+z^2) = 1+z$. Note that $M(z) = 1+z$ is the generating function for sequences (δ_n). Since $\delta = \mu * \zeta$, and $C(z \cdot M(z)) = 1+z$, by Theorem 3.39, we have $C(z) = 1 + \sum_{n=1}^{\infty} \mu_n z^n$, and

$$\mu(0_n, 1_n) = (-1)^{n-1}C_{n-1}, \quad n = 1, 2, \cdots.$$

Step II. Now For any $0_n \le \pi = \{V_1, \cdots, V_r\} < 1_n$, we already show in Lemma 3.31 that

$$[0_n, \pi] \simeq \Pi_{\mathrm{NC}}(|V_1|) \times \cdots \Pi_{\mathrm{NC}}(|V_1|) = [0_{|V_1|}, 1_{|V_1|}] \times \cdots \times [0_{|V_r|}, 1_{|V_r|}].$$

Then by Proposition 3.27,

$$\mu(0_n, \pi) = \prod_{V\in\pi}\mu(0_{|V|}, 1_{|V|}) = \prod_{V\in\pi}(-1)^{|V|-1}C_{|V|-1}, \quad 0_n \le \pi < 1_n.$$

Step III. We define the *Kreweras complement* $K : \Pi_{\mathrm{NC}}(n) \to \Pi_{\mathrm{NC}}(n)$ as follows: consider additional numbers $\overline{1}, \cdots, \overline{n}$ and interlace them with $1, \cdots, n$ in the alternating way: $(1, \overline{1}, 2, \overline{2}, \cdots, n, \overline{n})$. Let π be a non-crossing partition of $\{1, \cdots, n\}$ Then its Kreweras complement $K(\pi) \in \Pi_{\mathrm{NC}}(n)$ is defined as

$$K(\pi) = \sup_{\rho\in\mathcal{C}(\pi)} \rho = \sup\left\{\rho \in \Pi_{\mathrm{NC}}(n) : \pi \cup \overline{\rho} \in \Pi_{\mathrm{NC}}(1, \overline{1}, \cdots, n, \overline{n})\right\},$$

where $\overline{\rho}$ is the same partition as ρ, but on barred labels $\{\overline{1}, \cdots, \overline{n}\}$. Then one can verify that

- K is a bijection on $\Pi_{\mathrm{NC}}(\pi)$, and
- for any $\pi \le \sigma$, one have $\mathcal{C}(\pi) \supset \mathcal{C}(\sigma)$, and $K(\pi) \ge K(\sigma)$.

To summarize, K is an order anti-homomorphism on $\Pi_{\mathrm{NC}}(n)$, and

$$\mu(\pi, 1_n) = \mu(0_n, K(\pi)) = \prod_{V\in K(\pi)}(-1)^{|V|-1}C_{|V|-1}, \quad 0_n < \pi \le 1_n.$$

Step IV. By Steps I and III and the bound $C_n \le 4^n$, we have

$$|\mu(\pi, 1_n)| \le 4^n, \quad \pi \in \Pi_{\mathrm{NC}}(n).$$

If $|m_n|^{1/n} \le \rho$ for all $n \in \mathbb{N}$ and some $\rho > 0$, we have

$$|\kappa_n| \le \sum_{\pi\in\Pi_{\mathrm{NC}}(n)}\prod_{V\in\pi}|m_{|V|}| \cdot |\mu(\pi, 1_n)| \le \sum_{\pi\in\Pi_{\mathrm{NC}}(n)} 4^n\rho^n \le (16\rho)^n.$$

The other direction follows easily from the moment-cumulant formula. ∎

3.4 Additive Free Convolution

In this subsection, we discuss the sum of freely independent random variables in non-commutative probabiity spaces. Let μ, ν be two compactly supported probability measures on $\mathbb{R}$. Assume X and Y are two freely independent random variables in a non-commutative probability space $(\mathcal{A}, \tau)$ with spectral measures μ and ν, respectively. The spectral measure of $X+Y$ is called the *additive free convolution* of μ and ν, written

$$X + Y \sim \mu \boxplus \nu.$$

We should note that the convolution $\mu \boxplus \nu$ does not depend on the specific choice of probability space $(\mathcal{A}, \tau)$, because the moments (and hence the law) of $X+Y$ are determined by the moments of $\{\tau(X^k)\}_{k\in\mathbb{N}}$ and $\{\tau(Y^k)\}_{k\in\mathbb{N}}$ and the free independence, by Theorem 3.13.

Cauchy transform. For convenience, we often use a variant of the Stieltjes transform in free probability. Given a probability measure μ on $\mathbb{R}$, define the *Cauchy transform* of μ as the function G_μ on $\mathbb{C}\backslash \operatorname{supp}(\mu)$:

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-x}\, d\mu(x) = -s_\mu(z), \quad z \in \mathbb{C}\backslash \operatorname{supp}(\mu).$$

By Theorem 3.8, for the moment sequence $m_n = \int_{\mathbb{R}} x^k \, d\mu(x),\ n = 0, 1, 2, \cdots$ of μ, we have

$$G_\mu(z) = \sum_{k=0}^{\infty} \frac{m_n}{z^{k+1}} = \frac{1}{z} M\left(\frac{1}{z}\right).$$

Now we derive a formal inverse of the Cauchy transform.

Theorem 3.41 (Voiculescu)**.** *For compactly supported probability measures one has the following analytic properties of the Cauchy transform and the R-transform.*

(i) Let μ be a probability measure on $\mathbb{R}$ with compact support, contained in an interval $[-\rho, \rho]$. Consider its Cauchy transform G_μ as an analytic function in $U := \{z \in \mathbb{C} : |z| > 4\rho\}$. Then G_μ is injective on U, and

$$V := \left\{z \in \mathbb{C} : |z| < \frac{1}{6\rho}\right\} \subset G_\mu(U) \subset \left\{z \in \mathbb{C} : |z| < \frac{1}{3\rho}\right\}.$$

Hence G_μ has an analytic inverse $K_\mu = G_\mu^{-1} : V \to U$, which satisfies

$$G_\mu(K_\mu(z)) = z \quad \text{for } |z| < \frac{1}{6\rho}, \quad \text{and} \quad K_\mu(G_\mu(z)) = z \quad \text{for } |z| < \frac{1}{7\rho}.$$

(ii) The function K_μ has on V has the Laurent series expansion

$$K_\mu(z) = \frac{1}{z} + R_\mu(z), \quad \text{where } R_\mu(z) = \sum_{n=1}^{\infty} \kappa_n z^{n-1}, \quad z \in V.$$

*The power series function R_μ is called the **R-transform** of μ.*

Proof. (i) For all $z \in U$, define $f(z) := G_\mu(1/z)$. Then f has a power series expansion

$$f(z) = \sum_{n=0}^{\infty} m_n z^{n+1}, \quad |z| < \frac{1}{\rho},$$

and

$$|f(z)| \le \sum_{n=0}^{\infty} |m_n|\, |z|^{n+1} < \sum_{n=0}^{\infty} \rho^n \left(\frac{1}{4\rho}\right)^{n+1} = \frac{1}{3\rho} \quad \text{for all } |z| < \frac{1}{4\rho}.$$

Now we consider $z_1, z_2 \in \mathbb{C}$ with $|z_1|, |z_2| < \frac{1}{4\rho}$. If $z_1 \neq z_2$, by the mean value theorem,

$$\left|\frac{f(z_2) - f(z_1)}{z_2 - z_1}\right| \geq \operatorname{Re}\left(\frac{f(z_2) - f(z_1)}{z_2 - z_1}\right) = \int_0^1 \operatorname{Re}\left[f'(z_1 + t(z_2 - z_1))\right] dt.$$

Note that

$$\operatorname{Re} f'(z) = \operatorname{Re}\left(1 + \sum_{n=1}^{\infty}(n+1)m_n z^n\right) \geq 1 - \sum_{n=1}^{\infty}(n+1)m_n|z|^n \geq 2 - \sum_{n=0}^{\infty}\frac{n+1}{4^n} = \frac{2}{9}, \quad |z| < \frac{1}{4\rho}.$$

Combining the last two results, we have

$$|f(z_2) - f(z_1)| \geq \frac{2}{9}|z_2 - z_1|, \quad \text{for all } z_1, z_2 \in \left\{z \in \mathbb{C} : |z| < \frac{1}{4\rho}\right\}.$$

Therefore f is injective on $B(0, \frac{1}{4\rho})$, and G_μ is injective on $U = \{z \in \mathbb{C} : |z| > 4\rho\}$. Furthermore, if $|w| < \frac{1}{6\rho}$, we consider the function $h(z) = f(z) - w$.

$$|h(z) - (z - w)| = |f(z) - z| = \left|\sum_{n=1}^{\infty} m_n z^{n+1}\right| = \sum_{n=1}^{\infty}\rho^n\left(\frac{1}{4\rho}\right)^{n+1} = \frac{1}{12\rho} < |z - w|, \quad \text{for all } |z| = \frac{1}{4\rho}.$$

By Rouché's theorem, the analytic functions h and $z \mapsto z - w$ have the same number of zeros inside $B(0, \frac{1}{4\rho})$. Therefore $h(z) = f(z) - w$ has a simple zero in $B(0, \frac{1}{4\rho})$, and $w \in f(B(0, \frac{1}{4\rho}))$. Consequently, $G_\mu(U) \supset V$, and f has an analytic inverse $f^{-1} : V \to B(0, \frac{1}{4\rho})$. The inverse of G_μ is given by $K_\mu = 1/f^{-1}$.

Since f^{-1} has a simple zero at 0 and has no other zeroes, the function K_μ has simple pole at 0, and has the representation

$$K(z) = \frac{c}{z} + R(z),$$

where $c \in \mathbb{C}$ and $R : V \to U$ is some analytic function. Furthermore,

$$z = f(f^{-1}(z)) = G_\mu\left(\frac{1}{f^{-1}(z)}\right) = G_\mu(K_\mu(z)) = G_\mu\left(\frac{c}{z} + R(z)\right), \quad z \in V.$$

For $z \in \mathbb{C}$ with $|z| > 7\rho$, it suffices to show that $G_\mu(z) \in V$, i.e. $|G_\mu(z)| < \frac{1}{6\rho}$. After that, we have $K_\mu(G_\mu(z))$ by construction. Note that

$$|f(z)| \leq \sum_{n=0}^{\infty} m_n|z|^{n+1} < \sum_{n=0}^{\infty}\rho^n\left(\frac{1}{7\rho}\right)^{n+1} = \frac{1}{6\rho}, \quad |z| < \frac{1}{7\rho}.$$

Hence $|G_\mu(z)| < \frac{1}{6\rho}$ for $|z| > 7\rho$, and we finish the proof.

(ii) By Theorem 3.39, for some sufficiently small $\delta > 0$, we have

$$C(G_\mu(z)) = C\left(\frac{1}{z}M\left(\frac{1}{z}\right)\right) = M\left(\frac{1}{z}\right) = z \cdot G_\mu(z), \quad \text{for } z \in \mathbb{C} \text{ such that } |G_\mu(z)| < \delta.$$

Hence $z = C(G_\mu(z))/G_\mu(z)$. We define $\widetilde{R}(z) = \sum_{n=1}^{\infty}\kappa_n z^{n-1}$ and $\widetilde{K}(z) = \frac{1}{z} + \widetilde{R}(z) = C(z)/z$. Then

$$\widetilde{K}(G_\mu(z)) = \frac{1}{G_\mu(z)} + \widetilde{R}(G_\mu(z)) = \frac{C(G_\mu(z))}{G_\mu(z)} = z, \quad \text{for } z \in \mathbb{C} \text{ such that } |G_\mu(z)| < \delta.$$

Since G_μ is injective on U, we know that at least K_μ and $\widetilde{K}$ agree on a small neighborhood of 0. By uniqueness of power series representation, we have $R_\mu(z) = \sum_{n=1}^{\infty}\kappa_n z^{n-1}$ and $K_\mu(z) = \frac{1}{z} + R_\mu(z)$ on V. □

According to Theorem 3.40, for a bounded random variable X, its cumulants satisfy $\kappa_n(X) \leq (16\rho(X))^n$ for all $n \in \mathbb{N}$. As a consequence, its R-transform $R_X(z) = \sum_{n=0}^{\infty} \kappa_n(X) z^{n-1}$ is well-defined in a neighborhood of 0. The following theorem states that the R-trasform linearizes additive free convolution.

Theorem 3.42 (Voiculescu)**.** *The R-transform linearizes additive free convolution, i.e. for compactly supported probability measures μ, ν on $\mathbb{R}$,*

$$R_{\mu \boxplus \nu} = R_\mu + R_\nu.$$

Proof. Let $X \sim \mu$ and $Y \sim \nu$ be two random variables in some non-commutative probability space $(\mathcal{A}, \tau)$ with cumulants $(\kappa_n)_{n \in \mathbb{N}}$ such that X and Y are freely independent. Then

$$\kappa_n(X+Y) = \kappa_n(X) + \kappa_n(Y), \quad n = 1, 2, \cdots .$$

Then $X + Y \sim \mu \boxplus \nu$, and

$$R_{\mu \boxplus \nu}(z) = \sum_{n=0}^{\infty} \kappa_n(X+Y) z^{n-1} = \sum_{n=0}^{\infty} \kappa_n(X) z^{n-1} + \sum_{n=0}^{\infty} \kappa_n(Y) z^{n-1} = R_\mu(z) + R_\nu(z).$$

Thus we finish the proof. □

It is easy to verify the following properties of additive free convolution.

Proposition 3.43. *Let μ, ν, λ be compactly supported probability measures on $\mathbb{R}$.*

(i) (Commutativity). $\mu \boxplus \nu = \nu \boxplus \mu$.

(ii) (Associativity). $(\mu \boxplus \nu) \boxplus \lambda = \mu \boxplus (\nu \boxplus \lambda)$.

(iii) (Neutral element). $\delta_0 \boxplus \mu = \mu$.

(iv) (Translation). $\delta_t \boxplus \mu = \mu_{(t)}$ *for every* $t \in \mathbb{R}$, *where* $\mu_{(t)}(B) = \mu\{x - t : x \in B\}$ *for all Borel sets* $B \subset \mathbb{R}$.

Now we hanble a special example of additive free convolution.

Example 3.44. We consider the discrete measure

$$\mu = \frac{1}{2}(\delta_1 + \delta_{-1}).$$

The Cauchy transform of μ is

$$G_\mu(z) = \frac{1}{2(z-1)} + \frac{1}{2(z+1)} = \frac{z}{z^2 - 1},$$

and we find the inverse of G_μ by solving $G_\mu(K_\mu(z)) = z$, which has two solutions

$$K_\mu(z) = \frac{1 \pm \sqrt{1 + 4z^2}}{2z},$$

and

$$R_\mu(z) = K_\mu(z) - \frac{1}{z} = \frac{1 \pm \sqrt{1 + 4z^2}}{2z} - \frac{1}{z} = \frac{-1 \pm \sqrt{1 + 4z^2}}{2z}.$$

Note that $R_\mu(0) = 0$. Then we choose the branch

$$R_\mu(z) = \frac{-1 + \sqrt{1 + 4z^2}}{2z}.$$

Then

$$R_{\mu \boxplus \mu}(z) = \frac{-1 + \sqrt{1 + 4z^2}}{z}, \quad \text{and} \quad K_{\mu \boxplus \mu}(z) = \frac{\sqrt{1 + 4z^2}}{z}.$$

Consequently,

$$G_{\mu\boxplus\mu}(z) = \frac{1}{\sqrt{z^2-4}},$$

and by the Stieltjes inversion formula,

$$\frac{d(\mu\boxplus\mu)(t)}{dt} = -\frac{1}{\pi}\lim_{\eta\downarrow 0}\operatorname{Im}\frac{1}{\sqrt{(t+i\eta)^2-4}} = -\frac{1}{\pi}\operatorname{Im}\frac{1}{\sqrt{t^2-4}} = \frac{\mathbb{1}_{[-2,2]}(t)}{\pi\sqrt{4-t^2}}.$$

Hence $\mu\boxplus\mu$ is the arcsine distribution, which is continuous.

Example 3.45. The R-transform of the Marčenko-Pastur distribution is

$$R_\mu(z) = \sum_{n=0}^{\infty}\alpha z^{n-1} = \frac{\alpha}{1-z}.$$

Then

$$K_\mu(z) = \frac{1}{z} + \frac{\alpha}{1-z},$$

and

$$\frac{1}{G_\mu(z)} + \frac{\alpha}{1-G_\mu(z)} = z \quad\Rightarrow\quad G_\mu(z) \in \frac{1-\alpha+z\pm\sqrt{(1-\alpha+z)^2-4z}}{2z}$$

Since $G_\mu(i\eta)\to 0$ as $\eta\to\infty$, we select the branch

$$G_\mu(z) = \frac{1-\alpha+z-\sqrt{(1-\alpha+z)^2-4z}}{2z} = \frac{1-\alpha+z-\sqrt{(z-1-\alpha)^2-4\alpha}}{2z}.$$

Then for $t\neq 0$,

$$\begin{aligned}\frac{d\mu(t)}{dt} &= -\frac{1}{\pi}\lim_{\eta\downarrow 0}\operatorname{Im}\frac{1-\alpha+t+i\eta-\sqrt{(t+i\eta-1-\alpha)^2-4\alpha}}{2(t+i\eta)}\\ &= -\frac{1}{\pi}\operatorname{Im}\frac{1-\alpha+t-\sqrt{(t-1-\alpha)^2-4\alpha}}{2t}\\ &= \frac{\sqrt{(t-\lambda_{\alpha-})(\lambda_{\alpha+}-t)}}{2\pi t}\mathbb{1}_{[\lambda_{\alpha-},\lambda_{\alpha+}]}(t),\end{aligned}$$

and

$$\mu(\{0\}) = 1-\int_{\lambda_{\alpha-}}^{\lambda_{\alpha+}}\frac{\sqrt{(t-\lambda_{\alpha-})(\lambda_{\alpha+}-t)}}{2\pi t}\,dt = \left(1-\frac{1}{\alpha}\right)_+.$$

4 Gaussian Ensembles

In this section, we study two special kinds of of Wigner matrices $(\xi_{ij})_{i,j\in\mathbb{N}}$ introduced in Example 1.4:

- *Gaussian Orthogonal Ensemble (GOE).* The diagonal entries $(\xi_{ii})_{i=1}^{\infty}$ are i.i.d. $\mathcal{N}_{\mathbb{R}}(0,2)$ variables, and the off-diagonal entries $(\xi_{ij})_{1\le i<j}$ are i.i.d. $\mathcal{N}_{\mathbb{R}}(0,1)$ variables. In this case, In this case, the density function of W_n is

$$\begin{aligned}\rho_{n,\mathrm{GOE}}(X) &= \prod_{i=1}^{n}\frac{1}{2\sqrt{\pi}}e^{-\frac{x_{ii}^2}{4}}\prod_{1\le i<j\le n}\frac{1}{\sqrt{2\pi}}e^{-x_{ij}^2/2}\\ &= \frac{1}{2^{n(n+3)/4}\pi^{n(n+1)/4}}\exp\left(-\frac{1}{4}\sum_{i=1}^{n}x_{ii}^2-\frac{1}{2}\sum_{1\le i<j\le n}x_{ij}^2\right)\\ &= \frac{1}{2^{n(n+3)/4}\pi^{n(n+1)/4}}\exp\left(-\frac{1}{4}\operatorname{tr}(X^2)\right),\qquad X=(x_{ij})_{n\times n}\in\mathbb{R}^{n\times n}\text{ is symmetric.}\end{aligned}$$

  The GOE distribution is invariant under orthogonal similarity transformation, i.e. $W_n \overset{d}{=} QW_nQ^*$ for any orthogonal matrix $Q\in\mathbb{R}^{n\times n}$.

- *Gaussian Unitary Ensemble (GUE).* The diagonal entries $(\xi_{ii})_{i=1}^{\infty}$ are i.i.d. $\mathcal{N}_{\mathbb{R}}(0,1)$ variables, and the off-diagonal entries $(\xi_{ij})_{1\le i<j}$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ variables, i.e.

$$\mathbb{P}(\xi_{12}\in A)=\int_A\frac{1}{\pi}e^{-|z|^2}\,dz,\quad A\subset\mathbb{C}\text{ is Borel.}$$

  In this case, In this case, the density function of W_n is

$$\begin{aligned}\rho_{n,\mathrm{GUE}}(X) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-x_{ii}^2/2}\prod_{1\le i<j\le n}\frac{1}{\pi}e^{-|x_{ij}|^2}\\ &= \frac{1}{2^{n/2}\pi^{n^2/2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}x_{ii}^2-\frac{1}{2}\sum_{1\le i<j\le n}|x_{ij}|^2\right)\\ &= \frac{1}{2^{n/2}\pi^{n^2/2}}\exp\left(-\frac{1}{2}\operatorname{tr}(X^2)\right),\qquad X=(x_{ij})_{n\times n}\in\mathbb{C}^{n\times n}\text{ is Hermitian.}\end{aligned}$$

  The GUE distribution is invariant under unitary similarity transformation, i.e. $W_n \overset{d}{=} UW_nU^*$ for any unitary matrix $U\in\mathbb{C}^{n\times n}$.

To summarize, the density function of GOE/GUE is given by

$$\rho_{n,\beta}(X)=\frac{1}{Z_{n,\beta}}\exp\left(-\frac{\beta}{4}\operatorname{tr}(X^2)\right),\quad\text{where the Dyson index }\beta=\begin{cases}1 & \text{for GOE,}\\ 2 & \text{for GUE,}\end{cases}$$

and $Z_{n,\beta}>0$ is a normalizing constant.

The space of $n\times n$ Hermitian matrices as $\mathbb{R}^{n^2}$. An Hermitian matrix $H\in\mathbb{C}^{n\times n}$ is fully determined by its upper triangular entries $(H_{ij})_{1\le i\le j\le n}$, with the diagonal entries $H_{11},\cdots,H_{nn}\in\mathbb{R}$ and the off diagonal entries $H_{12},H_{13},H_{23},\cdots,H_{1n},\cdots,H_{n-1,n}\in\mathbb{C}\cong\mathbb{R}^2$. Hence the total real dimension is $n+2\times\frac{n(n-1)}{2}=n^2$. We define a linear isomorphism ψ from the vector space of Hermitian matrices $\mathcal{H}_n$ to $\mathbb{R}^{n^2}$:

$$\psi(H)=(H_{11},H_{22},\cdots,H_{nn},\mathrm{Re}(H_{12}),\cdots,\mathrm{Re}(H_{n-1,n}),\mathrm{Im}(H_{12}),\cdots,\mathrm{Im}(H_{n-1,n})).$$

This identifies $\mathcal{H}_n\cong\mathbb{R}^{n^2}$ as real vector spaces. Similarly $\mathcal{S}_n\cong\mathbb{R}^{n(n+1)/2}$.

4.1 Joint Distribution of Eigenvalues

In this subsection, we derive a closed-form density function for the spectral distribution of GOEs and GUEs.

Theorem 4.1 (Joint spectral density of Gaussian ensembles)**.** *Let $W_n = (\xi_{ij})_{1\le i,j\le n}$ be a GOE/GUE, and let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ be the eigenvalues of W_n, including repititions according to algebraic multiplicity. Then the density of the joint distribution of $(\lambda_1, \lambda_2, \cdots, \lambda_n)$ is given by*

$$\rho_{n,\beta}(\lambda_1, \cdots, \lambda_n) = \frac{\mathbb{1}_{\{\lambda_1 > \cdots > \lambda_n\}}}{Z_{n,\beta}} |\Delta_n(\lambda_1, \cdots, \lambda_n)|^\beta e^{-\frac{\beta}{4}(\lambda_1^2 + \cdots + \lambda_n^2)}, \quad \text{where } \beta = \begin{cases} 1 & \text{for GOE,} \\ 2 & \text{for GUE,} \end{cases}$$

where Δ_n is the $n \times n$ ***Vandemonde determinant***

$$\Delta_n(\lambda_1, \cdots, \lambda_n) = \det \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{pmatrix} = \prod_{1 \le i < j \le n} (\lambda_j - \lambda_i),$$

and $Z_{n,\beta} > 0$ is a normalizing constant.

4.1.1 Analysis of Spectra and Eigenvectors

We first prove that the spectrum of a GOE/GUE is almost surely simple.

Lemma 4.2 (Sylvester resultant)**.** *Consider two polynomials*

$$f(z) = a_0 + a_1 z + \cdots + a_n z^n, \quad g(z) = b_0 + b_1 z + \cdots + b_m z^m,$$

where $a_n \ne 0$ and $b_m \ne 0$. Define the ***Sylvester matrix***

$$S_{f,g} = \left(\begin{array}{ccccccc} a_n & a_{n-1} & \cdots & a_0 & 0 & \cdots & 0 \\ 0 & a_n & a_{n-1} & \cdots & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & a_n & a_{n-1} & \cdots & a_0 \\ \hline b_m & b_{m-1} & \cdots & b_0 & 0 & \cdots & 0 \\ 0 & b_m & b_{m-1} & \cdots & b_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & b_m & b_{m-1} & \cdots & b_0 \end{array}\right) \in \mathbb{C}^{(m+n)\times(m+n)},$$

where the upper block has m rows and the lower block has n rows. Let $\zeta_1, \cdots, \zeta_n \in \mathbb{C}$ be the zeros of f, and $\eta_1, \cdots, \eta_m \in \mathbb{C}$ the zeros of g, including the repetition according to multiplicity. Then

$$\det(S_{f,g}) = (-1)^m b_m^n \prod_{j=1}^m f(\eta_j) = a_n^m \prod_{k=1}^n g(\zeta_k) = a_n^m b_m^n \prod_{j=1}^m \prod_{k=1}^n (\zeta_k - \eta_j).$$

In particular, f and g have a common zero in $\mathbb{C}$ if and only if $\det(S_{f,g}) = 0$.

Remark. The determinant of the Sylvester matrix $S_{f,g}$ is also called the *Sylvester resultant.*

Proof. Step I. We first assume $f \cdot g$ has no repeated zeros, i.e. $\zeta_1, \cdots, \zeta_n, \eta_1, \cdots, \eta_n$ are mutually distinct. Given two polynomials $u(z) = u_0 + u_1 z + \cdots + u_{m-1} z^{m-1}$, $v(z) = v_0 + v_1 z + \cdots + v_{n-1} z^{n-1}$, define the linear mapping

$$\Phi : \mathbb{C}[\![z]\!]_{<m} \oplus \mathbb{C}[\![z]\!]_{<n} \to \mathbb{C}[\![z]\!]_{<m+n}, \quad \Phi(u,v)(z) = u(z) \cdot f(z) + v(z) \cdot g(z).$$

Then $S_{f,g}^{\top}$ is the coefficient matrix of Φ in the monomial bases, i.e. the coefficients of the polynomial $\Phi(u,v)$ (in decreasing order) is given by $S_{f,g}^{\top}(u_{m-1}, \cdots, u_1, u_0, v_{n-1}, \cdots, v_1, v_0)^{\top}$. Now we write

$$f(z) = a_n \prod_{j=1}^{n} (z - \zeta_j), \quad g(z) = b_m \prod_{j=1}^{m} (z - \eta_j),$$

and evaluate $h = \Phi(f,g)$ at the zeros of f and g:

- At $z = \zeta_j$, $j = 1, \cdots, n$, one have $h(\zeta_j) = v(\zeta_j) \cdot g(\zeta_j)$, since $f(\zeta_j)$ vanishes;
- At $z = \eta_j$, $j = 1, \cdots, m$, one have $h(\eta_j) = u(\eta_j) \cdot f(\eta_j)$, since $g(\eta_j)$ vanishes.

Inspired by this, we switch the evaluation coordinates:

- On the domain, use the transformation

  $$E : \mathbb{C}[\![z]\!]_{<m} \oplus \mathbb{C}[\![z]\!]_{<n} \to \mathbb{C}_m \oplus \mathbb{C}_n, \quad (u,v) \mapsto (u(\eta_1), \cdots, u(\eta_m), v(\zeta_1), \cdots, v(\zeta_n))\,;$$

  In the monomial basis, the coefficient matrix of E is a block diagonal matrix, whose upper-left and lower-right blocks are the Vandermonde matrices $(\eta_j^{k-1})_{j,k\in[m]}$ and $(\zeta_j^{k-1})_{j,k\in[n]}$, respectively. Then

  $$\det(E) = \Delta_m(\eta_1, \cdots, \eta_m) \cdot \Delta_n(\zeta_1, \cdots, \zeta_n).$$

- On the codomain, use the transformation

  $$F : \mathbb{C}[\![z]\!]_{<m+n} \to \mathbb{C}_{m+n}, \quad h \mapsto (h(\eta_1), \cdots, h(\eta_m), h(\zeta_1), \cdots, h(\zeta_n))\,.$$

  Likewise, we have

  $$\det(F) = \Delta_{m+n}(\eta_1, \cdots, \eta_m, \zeta_1, \cdots, \zeta_n) = \Delta_m(\eta_1, \cdots, \eta_m) \cdot \Delta_n(\zeta_1, \cdots, \zeta_n) \cdot \prod_{j=1}^{m} \prod_{k=1}^{n} (\eta_j - \zeta_k).$$

Clearly, in the evaluation coordinates, the coefficient matrix of Φ is given by the diagonal matrix

$$D = \operatorname{diag}(g(\zeta_1), \cdots, g(\zeta_n), f(\eta_1), \cdots, f(\eta_m)).$$

By the change of basis formula, we have $S_{f,g}^{\top} = F^{-1}DE$, and

$$\det(S_{f,g}) = \frac{\prod_{j=1}^{m} f(\eta_j) \cdot \prod_{k=1}^{n} g(\zeta_k) \cdot \det(E)}{\det(F)} = \frac{\prod_{j=1}^{m} f(\eta_j) \cdot \prod_{k=1}^{n} g(\zeta_k)}{\prod_{j=1}^{m} \prod_{k=1}^{n} (\eta_j - \zeta_k)}.$$

Note that $f(\eta_j) = a_n \prod_{k=1}^{n} (\eta_j - \zeta_k)$ and $g(\zeta_k) = (-1)^m b_m \prod_{j=1}^{m} (\eta_j - \zeta_k)$. Then the desired result follows.

Step II. If $f \cdot g$ has repeated zeros, we use perturbation. Take a small number $\epsilon > 0$ such that ϵ is smaller than the distance of any two distinct elements of $\{\zeta_1, \cdots, \zeta_n, \eta_1, \cdots, \eta_m\}$, and let

$$f^{\epsilon}(z) = a_n \prod_{k=1}^{n} \left(z - \zeta_k - 2^{-k}\epsilon\right) =: a_0^{\epsilon} + a_1^{\epsilon} z + \cdots + a_{n-1}^{\epsilon} z^{n-1} + a_n^{\epsilon} z^n,$$

$$g^{\epsilon}(z) = b_m \prod_{j=1}^{m} \left(z - \eta_j - i2^{-j}\epsilon\right) =: b_0^{\epsilon} + b_1^{\epsilon} z + \cdots + b_{m-1}^{\epsilon} z^{n-1} + b_m^{\epsilon} z^m.$$

Then their zeros are mutually distinct, and

$$\det(S_{f^\epsilon,g^\epsilon}) = \prod_{j=1}^{m}\prod_{k=1}^{n}\left(\zeta_k - \eta_j - (2^{-k} + i2^{-j})\epsilon\right).$$

As ϵ goes to 0, we have $\max_{k\in[n]} |a_k^\epsilon - a_k| \to 0$ and $\max_{j\in[m]} |b_j^\epsilon - b_j| \to 0$, and hence $\det(S_{f^\epsilon,g^\epsilon}) \to \det(S_{f,g})$. Taking $\epsilon \downarrow 0$ in the last display, we obtain the desired result for the general case. $\square$

Lemma 4.3. *We consider the Lebesgue measure.*

(i) Let $f : \mathbb{R}^N \to \mathbb{C}$ be a nonzero polynomial with complex coefficients. Then the zero set $\mathcal{Z}_f = \{f = 0\}$ of f is of zero Lebesgue measure on $\mathbb{R}^N$.

(ii) The set of symmetric matrices $A = (A_{ij})_{n\times n} \in \mathbb{C}^{n\times n}$ with repeated eigenvalues is of zero Lebesgue measure on $\mathbb{R}^{n(n+1)/2}$.

(iii) The set of Hermitian matrices $H = (H_{ij})_{n\times n} \in \mathbb{C}^{n\times n}$ with repeated eigenvalues is of zero Lebesgue measure on $\mathbb{R}^{n^2}$.

Remark. Since the density of GUE W_n is absolutely continuous with respect to the Lebesgue measure, the spectrum of W_n is almost surely simple, i.e. every eigenvalue of W_n is of algebraic multiplicity 1.

Proof. (i) We prove the result by induction. For the base case $N = 1$, by the fundamental theorem of algebra, any nonzero polynomial f of degree m has no more than m zeros. Therefore $\mathcal{Z}_f$ is a finite set and has zero Lebesgue measure. For the induction step, we write

$$f(x_1, \cdots, x_N) = \sum_{k=1}^{m} p_k(x_1, \cdots, x_{N-1})\, x_N^m,$$

where $p_1, \cdots, p_k : \mathbb{C}^{N-1} \to \mathbb{C}$ are polynomials. Then if $x \in \mathcal{Z}_f$, there are two possibilities:

- either $p_1 = \cdots = p_k = 0$, or
- x_N is a root of the nontrivial univariate polynomial $g(t) = \sum_{k=1}^{m} p_k(x_1, \cdots, x_{N-1})\, t^k$.

Let A, B be the subsets of $\mathbb{C}^n$ where these respective conditions hold, so that $\mathcal{Z}_f = A \cup B$.

- Using the inductive hypothesis, the Lebesgue measure of A is zero.
- Using the fundamental theorem of algebra, for each $(x_1, \cdots, x_{N-1}) \in \mathbb{C}^{N-1}$, there are finitely many t such that $(x_1, \cdots, x_{N-1}, t) \in \mathcal{Z}_f$. By Fubini's theorem, B also has zero Lebesgue measure.

Since (ii) and (iii) are similar, we only prove (iii).

(iii) Consider the characteristic polynomial $f(\lambda) = \det(H - \lambda\,\mathrm{Id}) = a_0 + a_1\lambda + \cdots + a_n\lambda^n$, where coefficients $a_0, a_1, \cdots, a_n$ are homogeneous polynomials of the entries of H. Then H has no repeated eigenvalues if and only if $f(\lambda)$ and $f'(\lambda) = b_0 + b_1\lambda + \cdots + b_{n-1}\lambda^{n-1}$, where $b_k = (k+1)a_{k+1}$, have no common zeros.

We take the *Sylvester matrix* $S_{f,f'}$. By Lemma 4.2, Hermitian matrix H has no repeated eigenvalues if and only if $\det(S_{f,f'}) = 0$. Since $S_{f,f'}$ is a polynomial of the entries of H, the result follows from (i). $\square$

Next we study the property of eigenvectors.

Lemma 4.4. *Fix $n \in \mathbb{N}$. We write*

- $\mathrm{U}(n)$ *for the set of $n \times n$ unitary matrices,*
- $\mathrm{U}^+(n)$ *for the set of $n \times n$ unitary matrices U such that every entry u_{ij} is nonzero, and*
- $\mathrm{U}^{++}(n)$ *for the set of matrices $U \in \mathrm{U}^+(n)$ such that every diagonal entry u_{ii} is strictly positive real.*

We also write $\mathbb{R}^n_> = \{\Lambda \in \mathbb{R}^{n\times n} :\ \Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\},$ with $\lambda_1 > \lambda_2 > \cdots > \lambda_n\}$.

(i) Define

$$\mathcal{H}_n^+ = \{H \in \mathcal{H}_n : \textit{there exists } U \in \mathrm{U}^+(n) \textit{ and } \Lambda \in \mathbb{R}^n_> \textit{ such that } H = U\Lambda U^*\}$$

Then $\mathcal{H}_n^+$ is of full Lebesgue measure on $\mathbb{R}^{n^2}$.

(ii) the map $(\mathbb{R}^n_>, \mathrm{U}^{++}(n)) \to \mathcal{H}_n^+$ given by $(\Lambda, U) \mapsto U\Lambda U^*$ is a bijection.

Proof. To prove (i), we write $H = U\Lambda U^*$ for the eigendecomposition of an Hermitian matrix $H \in \mathcal{H}_n$. By Lemma 4.3 (ii), we may assume $\Lambda \in \mathbb{R}^n_>$. The column of U consists of eigenvectors of H.

For an eigenvalue λ of H, let $G = H - \lambda\,\mathrm{Id}_n$. Then $GG^\sharp = \det(G)\,\mathrm{Id}_n = 0$, where $G^\sharp$ is the adjugate of G, i.e. $G^\sharp_{ij} = (-1)^{i+j}\det(G_{-j,-i})$, where $G_{-j,-i}$ is the $(n-1)\times(n-1)$ matrix obtained from G by removing j-th row and i-th column. Since the spectrum of H is simple, the null space of G has (complex) dimension 1, and all columns of $G^\sharp$ is a multiple of some eigenvector $u_\lambda \in \mathbb{C}^n$, which is a column of U. If $G^\sharp_{ii} = \det(H_{-i,-i} - \lambda\,\mathrm{Id}_{n-1}) = 0$, then the characteristic polynomials of H and $H_{-i,-i}$ have a common zero, and the corresponding Sylvester resultant, which is a nonzero polynomials about entries of H, vanishes. By Lemma 4.3 (i), this happens only for a set of matrices H of zero Lebesgue measure in $\mathcal{H}_n \cong \mathbb{R}^{n^2}$. Outside this set, we have $G^\sharp_{ii} \neq 0$, and $v_\lambda(i) \neq 0$. This is true for all eigenvalues λ and all indices $i \in [n]$. Consequently, all entries of U are nonzero, and $\mathcal{H}_n^+$ is of full Lebesgue measure.

The second part of the lemma is immediate, since the eigenspace corresponding to each eigenvalue is of dimension 1, the eigenvectors are fixed by the forcing $u_{ii} > 0$ for every $i \in [n]$, and the multiplicity arises from the possible permutations of the order of the eigenvalues. ☐

4.1.2 Change-of-Variable Technique

In this part, we use the change-of-variable formula to derive the joint density function of eigenvalues.

Lemma 4.5. *(Orthogonal and unitary groups).*

*(i) Let $\mathfrak{o}(n)$ be the space of $n\times n$ skew-symmetric real matrices, and $\mathrm{O}(n)$ the group of $n\times n$ orthogonal matrices. The exponential map $\exp : \mathfrak{o}(n) \to \mathrm{O}(n)$ is a surjective, locally one-to-one mapping. Thus $\mathfrak{o}(n)$ is the **Lie algebra** of $\mathrm{O}(n)$.*

*(ii) Let $\mathfrak{u}(n)$ be the space of $n\times n$ skew-Hermitian matrices, and $\mathrm{U}(n)$ the group of $n\times n$ unitary matrices. The exponential map $\exp : \mathfrak{u}(n) \to \mathrm{U}(n)$ is a surjective, locally one-to-one mapping. Thus $\mathfrak{u}(n)$ is the **Lie algebra** of $\mathrm{U}(n)$.*

(iii) Let $\mathrm{U}^\star(n)$ be the set of $n\times n$ unitary matrices U such that each diagonal entry $u_{ii} \in \mathbb{R}$, which is a submanifold of $\mathrm{U}(n)$. Via the exponential map, $\mathrm{U}^\star(n)$ is locally parameterized by an (n^2-n)-dimensional real vector space

$$\mathfrak{u}^\star(n) = \{S \in \mathfrak{u}(n) : S_{ii} = 0,\ i = 1, \cdots, n\}.$$

Proof. We first prove (ii). It is trivial to verify that e^S unitary is for any skew-Hermitian matrix $S = -S^*$. To check surjectivity, we fix a unitary matrix $U \in \mathrm{U}(n)$. Since U is normal, we consider its eigendecomposition $U = VDV^*$, where $D = (e^{\mathrm{i}\theta_1}, \cdots, e^{\mathrm{i}\theta_n})$ and $V \in \mathrm{U}(n)$. Then the matrix

$$S := V\operatorname{diag}\{\mathrm{i}\theta_1, \cdots, \mathrm{i}\theta_n\}V^*$$

is skew-Hermitian, i.e. $S = -S^*$, and satisfies $U = e^S$, $U^* = e^{-S}$.

To show that the exponential mapping is locally one-to-one, it suffices to show that it is one-to-one on some neighborhood of the zero matrix in $\mathfrak{u}(n)$ by group invariance. Since a skew-Hermitian matrix S satisfies $S = -S^*$, the space $\mathfrak{u}(n)$ is parameterized by the upper-triangular entries

$$(s_{ij})_{1\le i\le j\le n} = ((s_{ii})_{1\le i\le n}, (\operatorname{Re} s_{ij}, \operatorname{Im} s_{ij})_{1\le i<j\le n}),$$

which has real dimension n^2. Set $U = e^S$ and consider the above upper-triangular entries $(x_{ij})_{1\le i\le j\le n}$ as a function of $(s_{ij})_{1\le i\le j\le n}$. Then $e^{tS} = 1 + tS + O(t^2)$, and the partial derivatives at $S = 0$ are

$$\frac{\partial u_{ii}}{\partial s_{i'i'}} = \delta_{ii'}, \quad \frac{\partial \operatorname{Re} u_{ij}}{\partial \operatorname{Re} s_{i'j'}} = \frac{\partial \operatorname{Im} u_{ij}}{\partial \operatorname{Im} s_{i'j'}} = \delta_{ii'}\delta_{jj'}.$$

Hence the Jacobian matriix $DU(S)$ is an identity, which is invertible. By the inverse function theorem, there exist open neighborhoods V of zero matrix in $\mathfrak{u}(n)$ and W of Id in $\mathrm{U}(n)$ such that $\exp|_V : V \to W$ is a diffeomorphism. By group invariance, $\exp : \mathfrak{u}(n) \to \mathrm{U}(n)$ is locally one-to-one.

To prove (iii), we let $\{U(t), 0 \le t \le T\}$ be a smooth curve in $\mathrm{U}^\star(n)$ with $U(0) = \mathrm{Id}_n$. Then the skew-Hermitian matrix $S = U'(0)$ satisfies $s_{ii} = \frac{d}{dt}U_{ii}|_{t=0}$. For each $i \in [n]$, note that the diagonal entry s_{ii} of a skew-Hermitian matrix is purely imaginary, and $U_{ii}(t)$ is real. Hence $s_{ii} = 0$, and the tangent space of $\mathrm{U}^\star(n)$ at Id_n consists of skew-Hermitian matrices with all diagonal entries 0.

Finally, for the statement (i), we apply the following decomposition for real normal matrix U:

$$U = Q \operatorname{diag}\{1, \cdots, 1, -1, \cdots, -1, \Theta_1, \cdots, \Theta_r\}\, Q^\top,$$

where $\Theta_i = \begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix}$, $i = 1, \cdots, n$ are blocks of 2×2 rotation matrices, and $U \in \mathrm{O}(n)$. The remaining part of the proof for $\exp : \mathfrak{o}(n) \to \mathrm{O}(n)$ is similar to the unitary case. □

Lemma 4.6. *Consider the bijective map $(\mathbb{R}^n_>, \mathrm{U}^{++}(n)) \to \mathcal{H}_n^+ : (\Lambda, U) \mapsto X = U\Lambda U^*$. Then the Jacobian determinant of X with respect to (Λ, U) has the form*

$$|\det DX(\Lambda, U)| = |\Delta_n(\lambda_1, \cdots, \lambda_n)|^2 f(U),$$

where $f : \mathrm{U}^{++}(n) \to \mathbb{R}$ is a function of entries of U.

Proof. We note that $\mathrm{U}^+(n)$ is an open subset of $\mathrm{U}^\star(n)$. We view $X = U\Lambda U^* = e^S \Lambda e^{-S}$ as a function of Λ and S, where $S \in \mathfrak{u}^\star(n)$ is skew-Hermitian with all diagonal entries 0, and $U = e^S \in \mathrm{U}^+(n)$. Note Λ_n has degree of freedom n and S has degree of freedom $n^2 - n$, which is compatible with the real dimension of X.

Given $1 \le i < j \le n$, let E_{ij} be the skew matrix whose (i,j)-entry is 1 and (j,i)-entry is -1, with all other entries 0. Then $(E_{ij})_{1\le i<j\le n}$ forms a basis for $n \times n$ skew-Hermitian matrices. Furthermore, for $\epsilon \to 0$ in $\mathbb{C}$,

$$\begin{aligned} e^{E_{ij}}\Lambda e^{-E_{ij}} &= (1 + \epsilon E_{ij} + O(\epsilon^2))\,\Lambda\,(1 - \epsilon E_{ij} + O(\epsilon^2)) = \Lambda + \epsilon(E_{ij}\Lambda - \Lambda E_{ij}) + O(\epsilon^2) \\ &= \Lambda + \epsilon(\lambda_j - \lambda_i)E_{ij} + O(\epsilon^2). \end{aligned}$$

Hence for all indices $1 \le i < j \le n$ and $1 \le i' < j' \le n$,

$$\left.\frac{\partial x_{ii}}{\partial \lambda_{i'}}\right|_{S=0} = \delta_{ii'}, \quad \left.\frac{\partial x_{ij}}{\partial s_{i'j'}}\right|_{S=0} = (\lambda_j - \lambda_i)\delta_{ii'}\delta_{jj'}.$$

This can be summarized as $dX|_{S=0} = d\Lambda + (dS)\Lambda - \Lambda(dS)$. For the general case, the differential form of $X = U\Lambda U^* = e^S\Lambda e^{-S}$ is given by the unitary transformation

$$dX = U\,[d\Lambda + (dS)\Lambda - \Lambda(dS)]\,U^*, \quad \text{where } U = e^S.$$

Hence

$$dx_{ij} = \sum_{k=1}^n u_{ik}\overline{u}_{jk}\, d\lambda_k + \sum_{k\ne\ell}^n u_{ik}\overline{u}_{j\ell}(\lambda_\ell - \lambda_k)\, ds_{k\ell}.$$

For any $1 \le k < \ell \le n$, all entries of the two columns in $DX(\Lambda, S)$ corresponding to the derivatives with respect to $\operatorname{Re} s_{k\ell}$ and $\operatorname{Im} s_{k\ell}$ have a common factor $(\lambda_\ell - \lambda_k)$, and hence $(\lambda_\ell - \lambda_k)^2$ is a factor of the Jacobian. Hence $|\Delta_n(\lambda_1, \cdots, \lambda_n)|^2$ is a factor of $DX(\Lambda, S)$. Note that $DX(\Lambda, S)$ should be a homogeneous polynomial on $(\lambda_1, \cdots, \lambda_n)$ of order at most $n(n-1)$, whose coefficients are functions of entries of S. Then the Jacobian of X with respect to Λ, S has the form

$$|\det DX(\Lambda, S)| = |\Delta_n(\lambda_1, \cdots, \lambda_n)|^2\, g(S),$$

where $g : \mathfrak{u}^\star(n) \to \mathbb{R}$ is a function of entries of S. Since $S \mapsto U = e^S$ is a local diffeomorphism between $\mathfrak{u}^\star(n)$ and $\mathrm{U}^{++}(n)$, and its Jacobian depends only on $U = e^S$, the result (4.6) follows from change-of-variables. □

Proof of Theorem 4.1. We consider the GUE case, where $\beta = 2$. *Step I.* We first consider the mapping $\Phi : \mathbb{R}^n_{\ge} \times \mathrm{U}(n) \to \mathcal{H}_n, (\Lambda, U) \to U\Lambda U^*$, where $\mathbb{R}^n_{\ge}$ is the space of real diagonal matrices with non-increasing entries $\lambda_1 \ge \cdots \ge \lambda_n$. The pullback of the GUE distribution $\rho_{n,\mathrm{GUE}}(X)\, dX$ under Φ is denoted by $P(d\Lambda, dU) = \nu(\Lambda, dU)\mu(d\Lambda)$, where $\nu(\Lambda, \cdot)$ is the regular conditional distribution of U given Λ. Since the GUE distribution is invariant under unitary transformation $X \mapsto VXV^*$, where $V \in \mathrm{U}(n)$, the conditional distribution $\nu(\Lambda, dU)$ is invariant under left-multiplication $U \mapsto VU$. By uniqueness of the Haar measure, $\nu(\Lambda, dU)$ is the normalized left Haar measure on $\mathrm{U}(n)$, which does not depend on Λ. Hence Λ and U are independent, and $P(d\Lambda, dU) = \mu(d\Lambda)\nu(dU)$.

By Lemma 4.4, ν is concentrated on the set $\mathrm{U}^+(n)$, and we write $\pi : \mathrm{U}^+(n) \to \mathrm{U}^{++}(n)$ for the projection onto $\mathrm{U}^{++}(n)$, i.e. $\pi(U)_{ij} = u_{ij}\overline{u}_{jj}/|u_{jj}|$ for all $i, j \in [n]$, where $U = (u_{ij})_{i,j\in[n]}$. We denote by $\widetilde{\nu} = \nu \circ \pi^{-1}$ the pushforward of left-Haar measure ν on $\mathrm{U}^+(n)$ under π. Then $\mu(d\Lambda)\,\widetilde{\nu}(dU)$ is the pullback of $\rho_{n,\mathrm{GUE}}(X)\, dX$ under the bijection $\Psi : \mathbb{R}^n_{>} \times U^{++}(n) \to \mathcal{H}_n^+, (\Lambda, U) \to U\Lambda U^*$ in Lemma 4.4.

Step II. Now we focus on the measure $\mu(d\Lambda)$. For $X \in \mathcal{H}_n^+$, writing $X = \Psi(\Lambda, U) = U\Lambda U^*$. Then for any continuous function $\varphi : \mathbb{R}^n_{>} \to \mathbb{C}$, we have

$$\begin{aligned}
\int_{\mathbb{R}^n_{>}} \varphi(\Lambda)\, \mu(d\Lambda) &= \int_{\mathbb{R}^n_{>}} \int_{U^{++}(n)} \varphi(\Lambda)\, \widetilde{\nu}(dU)\, \mu(d\Lambda) = \int_{\mathcal{H}_n^+} (\varphi \circ \Psi^{-1})(X) \rho_{n,\mathrm{GUE}}(X)\, dX \\
&= \int_{\mathbb{R}^n_{>}} \int_{U^{++}(n)} \varphi(\Lambda) \rho_{n,\mathrm{GUE}}(U\Lambda U^*)\, |\det D\Psi(\Lambda, U)|\, dU\, d\Lambda \\
&= \int_{\mathbb{R}^n_{>}} \int_{U^{++}(n)} e^{-\frac{1}{2}\operatorname{tr}(\Lambda^2)}\, |\Delta_n(\lambda_1, \cdots, \lambda_n)|^2\, f(U)\, dU\, d\Lambda \\
&= c_n \int_{\mathbb{R}^n_{>}} e^{-\frac{1}{2}\operatorname{tr}(\Lambda^2)}\, |\Delta_n(\lambda_1, \cdots, \lambda_n)|^2\, d\Lambda,
\end{aligned}$$

where c_n is a constant depending only on n. Hence the density function of eigenvalues $\Lambda = \operatorname{diag}\{\lambda_1, \cdots, \lambda_n\}$ is given by

$$\rho_{n,\mathrm{GUE}}(\lambda_1, \cdots, \lambda_n) \propto \mathbb{1}_{\{\lambda_1 > \cdots > \lambda_n\}} e^{-\frac{1}{2}(\lambda_1^2 + \cdots + \lambda_n^2)}\, |\Delta_n(\lambda_1, \cdots, \lambda_n)|^2 .$$

The case for GOE distribution is similar. □

4.2 Determinantal Laws in the GUE

In this subsection, we study how Hermite polynomials and wave functions arise naturally from the spectral density of GUE:

$$\rho_{\mathrm{GUE},n}(x_1,\cdots,x_n) \propto e^{-\frac{1}{2}(x_1^2+\cdots+x_n^2)}\,|\Delta_n(x_1,\cdots,x_n)|^2\,\mathbb{1}_{\{x_1>\cdots>x_n\}}.$$

We let $(P_k)_{k=0}^{n-1}$ be a (univariate) polynomial family such that for every $k\in\mathbb{N}$, P_k is a monic polynomial of degree k, i.e. the leading coefficient of P_k is 1. Since determinant is invariant when adding a scalar multiple of one column to another column, for every $x_1,\cdots,x_n\in\mathbb{R}$, we have

$$\Delta_n(x_1,\cdots,x_n) = \det\begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix} = \det\begin{bmatrix} 1 & P_1(x_1) & \cdots & P_{n-1}(x_1) \\ 1 & P_1(x_2) & \cdots & P_{n-1}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & P_1(x_n) & \cdots & P_{n-1}(x_n) \end{bmatrix}$$

Then

$$|\Delta_n(x_1,\cdots,x_n)|^2 = \det\left[\sum_{k=0}^{n-1} x_i^k x_j^k\right]_{i,j=1}^n = \det\left[\sum_{k=0}^{n-1} P_k(x_i)P_k(x_j)\right]_{i,j=1}^n.$$

As a result, the spectral density of $n\times n$ GUE satisfies

$$\rho_{\mathrm{GUE},n}(\lambda_1,\cdots,\lambda_n) \propto \det\left[\sum_{k=0}^{n-1} e^{-x_i^2/4}P_k(x_i)e^{-x_j^2/4}P_k(x_j)\right]_{i,j=1}^n, \quad \lambda_1\ge\lambda_2\ge\cdots\ge\lambda_n. \tag{4.1}$$

A nice choice of $(P_k)_{k=0}^{n-1}$ is the Hermite polynomial family.

4.2.1 Hermite Polynomials

The Hermite polynomials are a family of orthogonal polynomials under the Gaussian measure.

Theorem 4.7 (Hermite polynomials)**.** *Consider the family of* ***Hermite polynomials****:*

$$\mathfrak{H}_n(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}, \quad n=0,1,2,\cdots. \tag{4.2}$$

(i) *$\mathfrak{H}_n(x)$ is a monic polynomial in x of degree n, i.e. the leading coefficient is* 1.

(ii) *(Rodrigues' formula).*

$$\mathfrak{H}_n = (-1)^n\left(\frac{d}{dx}-x\right)^n 1, \quad n=0,1,2,\cdots.$$

(iii) *(Derivatives).*

$$\mathfrak{H}_n^{(m)}(x) = \frac{n!}{(n-m)!}\mathfrak{H}_{n-m}(x), \quad 0\le m\le n.$$

(iv) *(Hermite differential equation).* $\mathfrak{H}_n'' - x\mathfrak{H}_n' + n\mathfrak{H}_n = 0$, $n=1,2,\cdots$.

(v) *(Orthogonality). The Hermitian polynomials are orthogonal under the Gaussian inner product, i.e. for $n,m\in\mathbb{N}$,*

$$\int_{-\infty}^{\infty}\mathfrak{H}_n(x)\mathfrak{H}_m(x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = n!\,\delta_{nm}.$$

(vi) *(Christoffel-Darboux). For $x\ne y$,*

$$\sum_{k=0}^{n-1}\frac{\mathfrak{H}_k(x)\mathfrak{H}_k(y)}{k!} = \frac{\mathfrak{H}_n(x)\mathfrak{H}_{n-1}(y)-\mathfrak{H}_{n-1}(x)\mathfrak{H}_n(y)}{(n-1)!(x-y)}, \quad n=1,2,\cdots.$$

Proof. (i) follows easily from $\mathfrak{H}_0 = 1$ and induction.

(ii) Note that for any differentiable function $f : \mathbb{R} \to \mathbb{C}$,

$$e^{x^2/2}\frac{d}{dx}e^{-x^2/2}f(x) = \left(\frac{d}{dx} - x\right) f(x).$$

By definition, we have

$$\mathfrak{H}_n = (-1)^n \cdot e^{x^2/2}\frac{d}{dx}e^{-x^2/2}(-1)^{n-1}\mathfrak{H}_{n-1} = -\left(\frac{d}{dx} - x\right)\mathfrak{H}_{n-1} = \cdots = (-1)^n\left(\frac{d}{dx} - x\right)^n 1.$$

(iii) We first claim that

$$\frac{d}{dx}\left(\frac{d}{dx} - x\right)^n f = \left(\frac{d}{dx} - x\right)^n \frac{d}{dx}f - n\left(\frac{d}{dx} - x\right)^{n-1} f, \quad n \in \mathbb{N}.$$

This can be proved by induction. For the base case $n = 1$,

$$\frac{d}{dx}\left(\frac{d}{dx} - x\right) f - \left(\frac{d}{dx} - x\right)\frac{d}{dx}f = \left(\frac{d^2}{dx^2}f - x\frac{d}{dx}f - f\right) - \left(\frac{d^2}{dx^2}f - x\frac{d}{dx}f\right) = -f.$$

By induction hypothesis,

$$\begin{aligned}
\frac{d}{dx}\left(\frac{d}{dx} - x\right)^n f &= \left(\frac{d}{dx} - x\right)\frac{d}{dx}\left(\frac{d}{dx} - x\right)^{n-1} f - \left(\frac{d}{dx} - x\right)^{n-1} f \\
&= \left(\frac{d}{dx} - x\right)\left[\left(\frac{d}{dx} - x\right)^{n-1}\frac{d}{dx} - (n-1)\left(\frac{d}{dx} - 1\right)^{n-2}\right] f - \left(\frac{d}{dx} - x\right)^{n-1} f \\
&= \frac{d}{dx}\left(\frac{d}{dx} - x\right)^n f - n\left(\frac{d}{dx} - x\right)^{n-1} f.
\end{aligned}$$

Using this conclusion, we have

$$\begin{aligned}
\mathfrak{H}_n' &= (-1)^n\frac{d}{dx}\left(\frac{d}{dx} - x\right)^n 1 = (-1)^n\left[\left(\frac{d}{dx} - x\right)^n\frac{d}{dx} - n\left(\frac{d}{dx} - x\right)^{n-1}\right] 1 \\
&= (-1)^{n-1} n\left(\frac{d}{dx} - x\right)^{n-1} 1 = n\mathfrak{H}_{n-1}.
\end{aligned}$$

The general case $\mathfrak{H}_n^{(m)} = \frac{n!}{(n-m)!}\mathfrak{H}_{n-m}$ follows from induction.

(iv) We multiply (4.2) by $e^{-x^2/2}$ and differentiate with respect to x on both sides:

$$(e^{-x^2/2}\mathfrak{H}_n)'(x) = (-1)^n\frac{d^{n+1}}{dx^{n+1}}e^{-x^2/2}.$$

Using the product rule and dividing by $e^{-x^2/2}$, we have

$$\mathfrak{H}_n'(x) - x\mathfrak{H}_n(x) = -\mathfrak{H}_{n+1}(x).$$

Again, we apply differentiation on both sides and use (iii) with $m = 1$:

$$\mathfrak{H}_n''(x) - x\mathfrak{H}_n'(x) - \mathfrak{H}_n(x) = -\mathfrak{H}_{n+1}'(x) = -(n+1)\mathfrak{H}_n(x).$$

This is the Hermite differential equation: $\mathfrak{H}_n''(x) - x\mathfrak{H}_n'(x) + n\mathfrak{H}_n(x) = 0$.

(v) We write $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ for the standard Gaussian density. By definition, $\mathfrak{H}_n(x)\phi(x) = (-1)^n\phi^{(n)}(x)$. Then the inner product

$$\langle \mathfrak{H}_n, \mathfrak{H}_m\rangle_\phi = \int_{-\infty}^{\infty} \mathfrak{H}_n(x)\mathfrak{H}_m(x)\phi(x)\,dx = (-1)^n \int_{-\infty}^{\infty} \mathfrak{H}_n(x)\phi^{(m)}(x)\,dx = \int_{-\infty}^{\infty} \mathfrak{H}_n^{(m)}(x)\phi(x)\,dx,$$

where the last equality follows from integration by parts and the fact that ϕ and its derivatives vanish at $\pm\infty$. By (iii),

$$\langle \mathfrak{H}_n, \mathfrak{H}_m\rangle_\phi = \int_{-\infty}^{\infty} \mathfrak{H}_m^{(n)}(x)\phi(x)\,dx = \begin{cases} \frac{n!}{(n-m)!}\int_{-\infty}^{\infty} \mathfrak{H}_{n-m}(x)\phi(x)\,dx, & n \geq m, \\ 0, & n < m. \end{cases}$$

In particular, $\langle \mathfrak{H}_n, 1\rangle_\phi = 0$ for $n \geq 1$. Hence $\langle \mathfrak{H}_n, \mathfrak{H}_m\rangle_\phi = n!\delta_{nm}$.

(vi) is simply the Christoffel-Darboux formula for orthogonal polynomials. □

4.2.2 Determinantal Laws

By (4.1), the spectral density of GUE can be written as the determinantal form:

$$\rho_{\mathrm{GUE},n}(x_1, \cdots, x_n) \propto \det\left[\sum_{k=1}^{n} e^{-x_i^2/4}\mathfrak{H}_n(x_i)\cdot e^{-x_j^2/4}\mathfrak{H}_n(x_j)\right]_{i,j=1}^{n}, \quad x_1 > x_2 > \cdots > x_n.$$

For simplicity, we often use the following oscillator wave functions obtained from Hermite polynomials. These functions form an orthonormal basis in $L^2(\mathbb{R})$ under the Lebesgue measure.

Proposition 4.8. *Consider the family of* ***normalized oscillator wave functions***

$$\psi_n(x) = \frac{1}{(2\pi)^{1/4}\sqrt{n!}}e^{-x^2/4}\mathfrak{H}_n(x), \quad n = 0, 1, 2, \cdots. \tag{4.3}$$

(i) (Orthogonality). The normalized oscillator wave functions are orthonormal in $L^2(\mathbb{R})$ under the Lebesgue measure:

$$\langle \psi_n, \psi_m\rangle_{L^2(\mathbb{R})} = \int_{\mathbb{R}} \psi_n(x)\psi_n(x)\,dx = \delta_{nm}, \quad n, m = 0, 1, 2, \cdots.$$

(ii) (Derivative).

$$\psi_n' = \sqrt{n}\psi_{n-1} - \frac{x}{2}\psi_n, \quad n = 0, 1, 2, \cdots.$$

(iii) (Harmonic oscillator).

$$\frac{x^2}{4}\psi_n - \psi_n'' = \left(n + \frac{1}{2}\right)\psi_n, \quad n = 0, 1, 2, \cdots. \tag{4.4}$$

(iv) (Christoffel-Darboux). For $x \neq y$,

$$\sum_{k=0}^{n-1}\psi_k(x)\psi_k(y) = \sqrt{n}\,\frac{\psi_n(x)\psi_{n-1}(y) - \psi_{n-1}(x)\psi_n(y)}{x - y}.$$

If $x = y$, taking the limit gives

$$\sum_{k=0}^{n-1}\psi_k(x)\psi_k(x) = \sqrt{n}\left[\psi_{n-1}(x)\psi_n'(x) - \psi_n(x)\psi_{n-1}'(x)\right].$$

Proof. The statements (ii) and (iv) follow easily from definition. For (ii), we differentiate twice on both sides of (4.3) and apply Theorem 4.7 (iii) with $m = 1$ to obtain

$$\psi_n'(x) = \frac{e^{-x^2/4}}{(2\pi)^{1/4}\sqrt{n!}} \left(\mathfrak{H}_n'(x) - \frac{x}{2}\mathfrak{H}_n(x)\right) = \frac{e^{-x^2/4}}{(2\pi)^{1/4}\sqrt{n!}} \left(n\mathfrak{H}_{n-1}(x) - \frac{x}{2}\mathfrak{H}_n(x)\right) = \sqrt{n}\psi_{n-1} - \frac{x}{2}\psi_n.$$

Finally, we differentiate twice on both sides of (4.3) and apply Theorem 4.7 (iv) to obtain

$$\begin{aligned}\psi_n''(x) &= \frac{1}{(2\pi)^{1/4}\sqrt{n!}} e^{-x^2/4} \left(\mathfrak{H}_n''(x) - x\mathfrak{H}_n'(x) + \left(\frac{x^2}{4} - \frac{1}{2}\right)\mathfrak{H}_n(x)\right) \\ &= \frac{1}{(2\pi)^{1/4}\sqrt{n!}} e^{-x^2/4} \left(-n - \frac{1}{2} + \frac{x^2}{4}\right)\mathfrak{H}_n(x) = \left(\frac{x^2}{4} - n - \frac{1}{2}\right)\psi_n(x).\end{aligned}$$

This is the Haarmonic oscillator equation (iii). □

Since $(\psi_k)_{k=1}^{\infty}$ is an orthonormal basis of $L^2(\mathbb{R})$, the kernel for the orthogonal projection operator Π_{V_n} onto the subspace $V_n = \operatorname{span}\{\psi_0, \psi_1, \cdots, \psi_{n-1}\}$ is given by

$$K_n(x, y) = \sum_{k=0}^{n-1} \psi_k(x)\psi_k(y). \tag{4.5}$$

That is,

$$(\Pi_{V_n} f)(x) = \int_{\mathbb{R}} K_n(x, y) f(y)\, dy = \sum_{k=0}^{n-1} \langle f, \psi_k\rangle_{L^2(\mathbb{R})} \psi_k(x), \quad f \in L^2(\mathbb{R}).$$

Using this notation, one can write the spectral density of GUE as

$$\rho_{\mathrm{GUE},n}(x_1, \cdots, x_n) \propto \det \left[\sum_{k=0}^{n-1} \psi_k(x_j)\psi_k(x_j)\right]_{i,j=1}^{n} = \det \left[K(x_i, x_j)\right]_{i,j=1}^{n}, \quad x_1 > x_2 > \cdots > x_n. \tag{4.6}$$

Following are some useful identities of K_n which can be easily obtained from orthonormality.

Proposition 4.9. *Let $n \in \mathbb{N}$ and $K_n : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ the kernel* (4.5)*.*

(i) (Trace). $\int_{\mathbb{R}} K_n(x, x)\, dx = n$.

(ii) (Reproducing kernel). $\int_{\mathbb{R}} K_n(x, y)K_n(y, z)\, dy = K_n(x, z)$.

Lemma 4.10 (Determinantal integration)**.** *Let $n \in \mathbb{N}$, $0 \le k \le n$, and $K_n : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ the kernel* (4.5)*. Then for any $x_1, \cdots, x_k \in \mathbb{R}$,*

$$\int_{\mathbb{R}} \det \left[K_n(x_i, x_j)\right]_{i,j=1}^{k+1} dx_{k+1} = (n - k) \det \left[K_n(x_i, x_j)\right]_{i,j=1}^{k}. \tag{4.7}$$

In particular,

(i) the case $k = 0$ corresponds to the trace identity $\int_{\mathbb{R}} K_n(x, x)\, dx = n$, and

(ii) for $k = n$, we have

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det \left[K_n(x_i, x_j)\right]_{i,j=1}^{n} dx_1 \cdots dx_n = n!.$$

Proof. We let $a_{ij} = K_n(x_i, x_j)$ and $A_{p,q}$ be the upper-left $p \times q$ block of $(a_{ij})_{i,j=1}^{k+1} \in \mathbb{R}^{(k+1)\times(k+1)}$. We let $A_{p,q}^{(\ell)}$ be the matrix obtained from $A_{p,q}^{(r)}$ by removing the r-th column $(a_{1r}, \cdots, a_{pr})$. Using the cofactor expansion,

$$\det A_{k+1,k+1} = \sum_{r=1}^{k} (-1)^{k+1+r} a_{k+1,r} \det A_{k,k+1}^{(r)} + a_{k+1,k+1} \det A_{k,k}. \tag{4.8}$$

We can easily find the integral of the second term:

$$\int_{\mathbb{R}} a_{k+1,k+1} \det A_{k,k}\, dx_{k+1} = \det A_{k,k} \int_{\mathbb{R}} K_n(x_{k+1}, x_{k+1})\, dx_{k+1} = n \det A_{k,k}. \tag{4.9}$$

By scaling the last column of $A^{(r)}_{k,k+1}$ by $a_{k+1,r}$, we have

$$a_{k+1,r} \det A^{(r)}_{k,k+1} = \det \left[A^{(r)}_{k,k} \,\middle|\, (a_{k+1,r} a_{i,k+1})_{i=1}^n \right],$$

with all dependence on x_{k+1} in the last column. For the i-th entry, using the reproducing kernel property:

$$\int_{\mathbb{R}} a_{k+1,r} a_{i,k+1}\, dx_{k+1} = \int_{\mathbb{R}} K_n(x_{k+1}, x_r) K_n(x_i, x_{k+1})\, dx_{k+1} = K(x_i, x_r) = a_{ir}.$$

Combining the last two identities, we have

$$\int_{\mathbb{R}} a_{k+1,r} \det A^{(r)}_{k,k+1}\, dx_{k+1} = \det \left[A^{(r)}_{k,k} \,\middle|\, (a_{i,r})_{i=1}^n \right] = (-1)^{k-r} \det A_{k,k}. \tag{4.10}$$

Pluggin-in (4.9) and (4.10) to (4.8), we obtain

$$\int_{\mathbb{R}} \det A_{k+1,k+1}\, dx_{k+1} = (n-k) \det A_{k,k}.$$

Then we complete the proof of (4.7). The statement (ii) follows by recursion. □

We denote by $\mathrm{S}(n)$ the group of permutation of $\{1, \cdots, n\}$, i.e.

$$\mathrm{S}(n) = \{\sigma : \{1, \cdots, n\} \to \{1, \cdots, n\} \mid \sigma \text{ is a bijection}\}.$$

Since the determinant could only change sign under permutation, we have

$$\det [K_n(x_i, x_j)]_{i,j=1}^n = \det \left[K_n(x_{\sigma(i)}, x_{\sigma(j)})\right]_{i,j=1}^n, \quad \sigma \in \mathrm{S}(n).$$

By extending $\mathbb{R}^n_> = \{(x_1, \cdots, x_n) \in \mathbb{R}^n : x_1 > x_2 > \cdots > x_n\}$ to the entire space $\mathbb{R}^n$, we have

$$\begin{aligned} n! = \int_{\mathbb{R}^n} \det [K_n(x_i, x_j)]_{i,j=1}^n\, dx &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det [K_n(x_i, x_j)]_{i,j=1}^n \sum_{\sigma \in \mathrm{S}(n)} \mathbb{1}_{\{x_{\sigma(1)} > \cdots > x_{\sigma(n)}\}}\, dx \\ &= n! \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det [K_n(x_i, x_j)]_{i,j=1}^n \mathbb{1}_{\{x_1 > \cdots > x_n\}}\, dx. \end{aligned}$$

Hence $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det [K_n(x_i, x_j)]_{i,j=1}^n \mathbb{1}_{\{x_1 > \cdots > x_n\}}\, dx = 1$, and the normalizing constant in (4.6) is 1. This is also known as the Gaudin-Mehta formula for GUE spectral density.

Theorem 4.11 (Gaudin-Mehta)**.** *Let $W_n = (\xi_{ij})_{1 \le i,j \le n}$ be a GUE, and K_n the kernel defined in* (4.5)*. Then the spectral density of W_n is given by*

$$\rho_{\mathrm{GUE},n}(x_1, \cdots, x_n) = \det [K_n(x_i, x_j)]_{i,j=1}^n \mathbb{1}_{\{x_1 > \cdots > x_n\}}.$$

For the marginal densities for GUE eigenvalues, we have the following conclusion.

Proposition 4.12 (Correlation)**.** *Let $1 \le k \le n$. Then the k-point correlation function of $\{\lambda_1, \cdots, \lambda_n\}$ is*

$$\rho_{n,k}(x) = \frac{(n-k)!}{n!} \det [K_n(x_i, x_j)]_{i,j=1}^k, \quad x_1, \cdots, x_k \in \mathbb{R}.$$

That is, for any measurable function $f : \mathbb{R} \to [0, \infty)$,

$$\mathbb{E}\left[\frac{1}{\binom{n}{k}} \sum_{1 \le i_1 < \cdots < i_k \le n} f(\lambda_{i_1}(W_n), \cdots, \lambda_{i_k}(W_n))\right] = \int_{\mathbb{R}^k} f(x)\rho_{n,k}(x)\, dx$$
$$= \frac{(n-k)!}{n!} \int_{\mathbb{R}^k} f(x) \det\left[K_n(x_i, x_j)\right]_{i,j=1}^k \, dx.$$

In particular,

$$\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n f(\lambda_j(W_n))\right] = \frac{1}{n}\int_{\mathbb{R}} f(x) K_n(x,x)\, dx.$$

Proof. By permutation invariance and Lemma 4.10,

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \sum_{1 \le i_1 < \cdots < i_k \le n} f(x_{i_1}, \cdots, x_{i_k}) \det\left[K_n(x_i, x_j)\right]_{i,j=1}^n \mathbb{1}_{\{x_1 > \cdots > x_n\}}\, dx_1 \cdots dx_n$$
$$= \frac{1}{n!}\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \sum_{1 \le i_1 < \cdots < i_k \le n} f(x_{i_1}, \cdots, x_{i_k}) \det\left[K_n(x_i, x_j)\right]_{i,j=1}^n \, dx_1 \cdots dx_n$$
$$= \frac{1}{n!}\binom{n}{k} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \cdots, x_k) \det\left[K_n(x_i, x_j)\right]_{i,j=1}^n \, dx_1 \cdots dx_n$$
$$= \frac{1}{n!}\binom{n}{k} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \cdots, x_k) \left[\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det\left[K_n(x_i, x_j)\right]_{i,j=1}^n \, dx_{k+1} \cdots dx_n\right] dx_1 \cdots dx_k$$
$$= \frac{1}{n!}\binom{n}{k} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \cdots, x_k) \cdot (n-k)! \det\left[K_n(x_i, x_j)\right]_{i,j=1}^k \, dx_1 \cdots dx_k$$
$$= \frac{1}{k!} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \cdots, x_k) \det\left[K_n(x_i, x_j)\right]_{i,j=1}^k \, dx_1 \cdots dx_k.$$

The result follows from dividing both sides by $\binom{n}{k}$. ◻

5 Circular Law

5.1 A Brief Journey

The *circular distribution* is the uniform probability measure on the unit disk $B_{\mathbb{C}}(0,1) = \{\lambda \in \mathbb{C} : |\lambda| \le 1\}$:

$$\mu_{\mathrm{circ}}(A) = \frac{1}{\pi}\int_{|z|\le 1} \mathbb{1}_A(z)\,dx\,dy, \quad A \subset \mathbb{C} \text{ is Borel.}$$

For general non-Hermitian matrices, the relationship between eigenvalues and singular values are captured by a set of inequalities due to Weyl.

Theorem 5.1 (Weyl)**.** *Let $A \in \mathbb{C}^{n\times n}$ be a non-Hermitian matrix. Let $\lambda_1(A), \cdots, \lambda_n(A)$ be the eigenvalues of A ordered in decreasing modulus, i.e. $|\lambda_1(A)| \ge \cdots \ge |\lambda_n(A)|$, with growing phases, and let $\sigma_1(A) \ge \cdots \ge \sigma_n(A)$ be singular values of A. Then*

$$\prod_{j=1}^k |\lambda_j(A)| \le \prod_{j=1}^k \sigma_j(A), \quad k = 1, \cdots, n.$$

Proof. Using Schur's unitary triangularization theorem, there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ and an upper triangular matrix $T \in \mathbb{C}^{n\times n}$, with diagonal entries given by $T_{jj} = \lambda_j(A)$, $j = 1, \cdots, n$, such that $A = UTU^*$. Since singular values are invariant under unitary transformation, we have

$$\sigma_j(A) = \sigma_j(T), \quad j = 1, \cdots, n.$$

Now let $M \in \mathbb{C}^{n\times n}$ be a complex matrix with singular values $\sigma_1(B) \ge \cdots \ge \sigma_n(B)$. We fix $k \in [n]$, and prove that for any $k \times k$ submatrix B of M,

$$|\det(B)| \le \prod_{j=1}^k \sigma_j(M). \tag{5.1}$$

Assume B is obtained by selecting rows $i_1 < \cdots < i_k$ and columns $j_1 < \cdots < j_k$ of M. Take orthogonal matrices $R = [e_{i_1}, \cdots, e_{i_k}]^\top \in \mathbb{R}^{k\times n}$ and $C = [e_{j_1}, \cdots, e_{j_k}] \in \mathbb{R}^{n\times k}$, so that $B = RMC$. By Courant-Fisher max-min principle, for every $j \in [k]$, we have

$$\sigma_j(MC) = \max_{\dim V = j} \min_{\substack{v\in V\\ \|v\|_2\le 1}} \|MCv\|_2 \le \max_{\dim U = j} \min_{\substack{u\in U\\ \|u\|_2\le 1}} \|Mu\|_2 = \sigma_j(M),$$

and

$$\sigma_j(RMC) = \max_{\dim V = j} \min_{\substack{v\in V\\ \|v\|_2\le 1}} \|RMCv\|_2 \le \max_{\dim V = j} \min_{\substack{v\in V\\ \|v\|_2\le 1}} \|R\|_2\,\|MCv\|_2 = \|R\|_2\,\sigma_j(MC).$$

Hence

$$\sigma_j(B) = \sigma_j(RMC) \le \|R\|_2\,\sigma_j(MC) \le \|R\|_2\,\|C\|_2\,\sigma_j(M) = \sigma_j(M),$$

and the result (5.1) follows. Furthermore, if we let $M = T$ and B be the upper left $k \times k$ minor of T, we have

$$\prod_{j=1}^k |\lambda_j(A)| \le \prod_{j=1}^k \sigma_j(T) = \prod_{j=1}^k \sigma_j(A),$$

which is exactly Weyl's inequality. □

Theorem 5.2 (Circular law)**.** *Let $(X_{ij})_{i,j\ge 1}$ be an array of i.i.d. random variables with zero mean and unit variance. Let $X_n = (X_{ij})_{1\le i,j,\le n}$ be the upper left $n \times n$ block of the infinite array. Then almost surely, as $n \to \infty$,*

$$\mu_{n^{-1/2}X_n} \to \mu_{\mathrm{circ}} \quad \textit{weakly.}$$

5.2 Main Tools

5.2.1 Logarithmic Potential

Definition 5.3 (Logarithmic potential)**.** Let $\mathcal{P}_\infty(\mathbb{C})$ be the set of probability mesures on $\mathbb{C}$ which integrate $\log|\cdot|$ in a neighborhood of infinity, i.e. for $\mu \in \mathcal{P}_\infty(\mathbb{C})$, there exists $R > 1$ such that

$$\int_{\mathbb{C}} \log^+ |\lambda|\, d\mu(\lambda) < \infty, \quad \text{where } \log^+ r = \max\{\log r, 0\}.$$

The *logarithm potential* of $\mu \in \mathcal{P}_\infty(\mathbb{C})$ is the function $U_\mu : \mathbb{C} \to (-\infty, \infty]$ defined by

$$U_\mu(z) = -\int_{\mathbb{C}} \log|z-\lambda|\, d\mu(\lambda) = -(\log|\cdot| * \mu)(z), \quad z \in \mathbb{C}.$$

Remark. By definition, for every $z \in \mathbb{C}$, the function $\lambda \mapsto -\log|z-\lambda|$ is quasi-integrable with respect to the measure $\mu \in \mathcal{P}_\infty(\mathbb{C})$. We note that $\lambda \mapsto -\log|z-\lambda|$ is bounded, and hence integrable on the compact set $\{\lambda \in \mathbb{C} : |\lambda - z| \geq 1, |\lambda| \leq |z|+1\}$. In the neighborhood of z, we have

$$-\int_{|\lambda-z|<1} \log|z-\lambda|\, d\mu(\lambda) \in [0,\infty],$$

and in the neighborhood of infinity, we have

$$\int_{|\lambda|>|z|+1} \log|z-\lambda|\, d\mu(\lambda) \leq \int_{|\lambda|>|z|+1} \log(2|\lambda|)\, d\mu(\lambda) \leq \log 2 + \int_{|\lambda|>1} \log|\lambda|\, d\mu(\lambda) < \infty.$$

Hence $U_\lambda(z) \in (-\infty, \infty]$.

Example 5.4. The logarithmic potential of the circular distribution μ_{circ} is given by

$$U_{\mu_{\text{circ}}}(z) = \begin{cases} \dfrac{1-|z|^2}{2}, & |z| \leq 1, \\ -\log|z|, & |z| > 1. \end{cases}$$

Proof. Since μ_{circ} is the uniform probability measure on the unit disk, $U_{\mu_{\text{circ}}}$ is a radial function. For $|z| > 1$, since $\lambda \mapsto \log|z-\lambda|$ is harmonic in a neighborhood of the unit disk, by the mean-value property,

$$U_{\mu_{\text{circ}}}(z) = -\frac{1}{\pi}\int_{|\lambda|\leq 1} \log|z-\lambda|\, dS(\lambda) = -\log|z|.$$

For $0 \leq |z| \leq 1$, we let $r = |z|$. Then

$$U_{\mu_{\text{circ}}}(z) = U_{\mu_{\text{circ}}}(r) = -\frac{1}{\pi}\int_{|\lambda|\leq 1} \log|r-\lambda|\, dS(\lambda) = -\frac{1}{\pi}\int_0^1 \int_0^{2\pi} \log\left|r - \rho e^{i\theta}\right| \rho\, d\theta\, d\rho.$$

We first compute the inner integral. Note that

$$\int_0^{2\pi} \log\left|r-\rho e^{i\theta}\right| d\theta = 2\pi \log r + \int_0^{2\pi} \log\left|1 - \frac{\rho}{r} e^{i\theta}\right| d\theta.$$

Note that for $0 < \beta < 1$, the function $z \mapsto \log|1-\beta z|$ is harmonic in a neighborhood of the unit disc. By the mean-value property, if $0 < \rho < r$, we have

$$\frac{1}{2\pi}\int_0^{2\pi}\log\left|1-\frac{\rho}{r}e^{i\theta}\right|d\theta = \frac{1}{2\pi}\int_{|z|=1}\log\left|1-\frac{\rho}{r}z\right|dS = \log|1| = 0.$$

For $0 < r < \rho$, since

$$|r-\rho e^{i\theta}| = \sqrt{r^2+\rho^2-2r\rho\cos\theta} = |\rho - re^{i\theta}|,$$

we exchange the roles of ρ and r to obtain $\int_0^{2\pi}\log\left|r-\rho e^{i\theta}\right|d\theta = 2\pi\log\rho$. Hence

$$\int_0^{2\pi}\log\left|r-\rho e^{i\theta}\right|d\theta = 2\pi\log\max\{\rho, r\}.$$

Now compute the outer integral:

$$U_{\mu_{\mathrm{circ}}}(z) = -2\left(\int_0^r \rho\log r\,d\rho + \int_r^1 \rho\log\rho\,d\rho\right) = \frac{1-r^2}{2}.$$

Then we finish the proof. □

Proposition 5.5. *Let $\mu\in\mathcal{P}_\infty(\mathbb{C})$. Then $U_\mu\in L^1_{\mathrm{loc}}(\mathbb{C})$.*

Proof. Let $K\subset\mathbb{C}$ be a compact set. By Tonelli-Fubini theorem,

$$\int_K |U_\mu(z)|\,dx\,dy = \int_{\mathbb{C}}\left(\int_K |\log|z-\lambda||\,dx\,dy\right)d\mu(\lambda)$$

Since K is compact, we take $N > 1$ such that $K\subset B(0,N)$.

$$f(\lambda) = \int_K |\log|z-\lambda||\,dx\,dy$$

- If $|\lambda|\le 1+2N$, we have

$$f(\lambda) = -\int_{K\cap B(\lambda,1)}\log|z-\lambda|dx\,dy + \int_{K\setminus B(\lambda,1)}\log|z-\lambda|dx\,dy.$$

  For the first part, change the variable $w = z-\lambda$ to get

$$-\int_{K\cap B(\lambda,1)}\log|z-\lambda|dx\,dy \le -\int_{B(\lambda,1)}\log|z-\lambda|dx\,dy = -\int_{B(0,1)}\log|w|\,dx\,dy = \frac{\pi}{2}.$$

  For the second part, note that $|z-\lambda|\le|z|+|\lambda|\le 1+3N$ for $z\in K$. Then

$$\int_{K\setminus B(\lambda,1)}\log|z-\lambda|\,dx\,dy \le \int_K \log(1+3N)\,dx\,dy \le \pi N^2\log(1+3N).$$

  To summarize,

$$\sup_{|\lambda|\le 1+2N} f(\lambda) \le \frac{\pi}{2} + \pi N^2\log(1+3N).$$

- If $|\lambda|\ge 2N$, we have $|\lambda|/2\le|\lambda|-N\le|z-\lambda|\le|\lambda|+N\le 3|\lambda|/2$ for $z\in K$. This bound implies

$$|\log|\lambda-z||\le\max\left\{\log\frac{|\lambda|}{2},\log\frac{3|\lambda|}{2}\right\}\le 1+\log|\lambda|,\quad z\in K.$$

Then

$$f(\lambda) = \int_K |\log|\lambda - z||\, dx\, dy \le \int_K (1 + \log|\lambda|)\, dx\, dy = \pi N^2 (1 + \log|\lambda|), \quad \lambda > 2N.$$

Combining the two cases, for some constant $C_N > 0$ depending on N only, we have

$$f(\lambda) \le C_N(1 + \log^+|\lambda|) \quad \text{for all } \lambda \in \mathbb{C}.$$

Then

$$\int_K |U_\mu(z)|\, dx\, dy \le \int_{\mathbb{C}} f(\lambda)\, d\lambda \le C_N \int_{\mathbb{C}} \left(1 + \log^+|\lambda|\right) d\mu(\lambda) < \infty,$$

where the last inequality follows because $\mu \in \mathcal{P}_\infty(\mathbb{C})$. □

Distribution theory review. Since every $\mu \in \mathcal{P}_\infty(\mathbb{C})$ is a Radon measure on $\mathbb{C}$, we view it as a Schwartz-Sobolev distribution, i.e. $\mu \in \mathcal{D}'(\mathbb{C})$, which is a linear functional on the space $C_c^\infty(\mathbb{C})$ of test function:

$$\langle \mu, \phi \rangle = \int_{\mathbb{C}} \phi(\lambda)\, d\mu(\lambda), \quad \phi \in \mathbb{C}_c^\infty(\mathbb{C}).$$

Also, by Proposition 5.5, $U_\mu \in L^1_{\text{loc}}(\mathbb{C})$ is a distribution.

Next, we define the first-order differential operators in $\mathcal{D}'(\mathbb{C})$ as $\partial = \frac{1}{2}(\partial_x - i\partial_y)$ and $\overline{\partial} = \frac{1}{2}(\partial_x + i\partial y)$, and define the *Laplace operator* $\Delta = 4\partial\overline{\partial} = 4\overline{\partial}\partial = \partial_x^2 + \partial_y^2$. Note that $\log|\cdot|$ is harmonic in $\mathbb{C}\backslash\{0\}$ and ϕ is compactly supported. By Green's second identity,

$$\begin{aligned} -\int_{\mathbb{C}} \log|z|\, \Delta\phi(z)\, dx\, dy &= -\lim_{\epsilon\downarrow 0} \int_{|z|\ge\epsilon} \log|z|\, \Delta\phi(z)\, dx\, dy \\ &= \lim_{\epsilon\downarrow 0} \left[\int_{|z|=\epsilon} \phi(z)\nabla\log|z| \cdot \mathbf{n}(z)\, ds - \int_{|z|=\epsilon} \log|z|\, \nabla\phi(z)\cdot\mathbf{n}(z)\, ds \right]. \end{aligned}$$

Note that $\nabla\log|z| = z/|z|^2$, and the outer unit normal $\mathbf{n}(z) = -z/|z|$. Then the second term

$$\left| \int_{|z|=\epsilon} \log|z|\, \nabla\phi(z)\cdot\mathbf{n}(z)\, dS \right| \le \log\epsilon \cdot \int_{|z|=\epsilon} |\nabla\phi(z)|\, dS \le 2\pi\epsilon\log\epsilon \cdot \sup_{|z|\le 1} |\nabla\phi(z)|,$$

which vanishes as $\epsilon \downarrow 0$, and the first term

$$\int_{|z|=\epsilon} \phi(z)\nabla\log|z|\cdot\mathbf{n}(z)\, dS = -\frac{1}{\epsilon}\int_{|z|=\epsilon} \phi(z)\, dS,$$

which converges to $-2\pi\phi(0)$ as $\epsilon \downarrow 0$. Therefore

$$\langle \Delta\log|\cdot|, \phi\rangle = \langle \log|\cdot|, \Delta\phi\rangle = \int_{\mathbb{C}} \Delta\phi(z)\log|z|\, dx\, dy = 2\pi\phi(0) = \langle 2\pi\delta_0, \phi\rangle.$$

Hence $\Delta\log|\cdot| = 2\pi\delta_0$. In fact, $\frac{1}{2\pi}\log|\cdot|$ is the Green's function for Poisson's equation $\Delta u = f$ in $\mathbb{R}^2$, i.e.

$$u(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \log|x-y|\, f(y)\, dy \quad \text{satisfies } \Delta u = f.$$

Also, by Tonelli-Fubini theorem, for any probability measure $\mu \in \mathcal{P}_\infty(\mathbb{C})$ and test function $\phi \in C_c^\infty(\mathbb{C})$,

$$\langle \Delta U_\mu, \phi\rangle = \langle U_\mu, \Delta\phi\rangle = -\int_{\mathbb{C}} \left(\int_{\mathbb{C}} \log|z-\lambda|\, \Delta\phi(z)\, dx\, dy \right) d\mu(\lambda) = -2\pi \int_{\mathbb{C}} \phi(\lambda)\, d\mu(\lambda) = -\langle 2\pi\mu, \phi\rangle.$$

In other words,

$$\Delta U_\mu = -2\pi\mu \quad \text{in } \mathcal{D}'(\mathbb{C}). \tag{5.2}$$

Theorem 5.6 (Unicity)**.** *Let $\mu, \nu \in \mathcal{P}_\infty(\mathbb{C})$. Then $U_\mu = U_\nu$ a.e. if and only if $\mu = \nu$.*

Proof. Clearly $\mu = \nu$ implies $U_\mu = U_\nu$. Now if $U_\mu = U_\nu$ a.e., we have $\Delta U_\mu = \Delta U_\nu$ in $\mathcal{D}'(\mathbb{C})$, and (5.2) implies $\mu = \nu$ in $\mathcal{D}'(\mathbb{C})$. Since μ and ν are both Radon measures on $\mathbb{C}$, we have $\mu = \nu$. □

Theorem 5.7 (Convergence in potentials and weak convergence)**.** *Let (μ_n) be a sequence in $\mathcal{P}_\infty(\mathbb{C})$. Assume that $\log(1+|\cdot|)$ is uniformly integrable for $(\mu_n)_{n\in\mathbb{N}}$. Then the following two statements are equivalent:*

(i) There exists a function $U : \mathbb{C} \to (-\infty, \infty]$ such that $U_{\mu_n}(z) \to U(z)$ for a.e. $z \in \mathbb{C}$.

(ii) There exists $\mu \in \mathcal{P}_\infty(\mathbb{C})$ such that $\mu_n \to \mu$ weakly.

Furthermore, if function U satisfies (i) and μ satisfies (ii), then $U_\mu = U$ a.e., and $\mu = -\frac{1}{2\pi}\Delta U$ in $\mathcal{D}'(\mathbb{C})$.

Proof. (i) ⇒ (ii). For every $N > 1$, by de la Vallée Poussin criterion for uniform integrability, there exists a non-decreasing, convex function $\varphi : [0,\infty) \to [0,\infty)$, which may depend on N, such that $\varphi(t)/t \to \infty$ as $t \uparrow \infty$, $\varphi(t) \le 1 + t^2$ for all $t \ge 0$, and

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{C}} \varphi\left(\log(N + |\lambda|)\right) d\mu_n(\lambda) < \infty.$$

We take a compact set $K \subset \mathbb{C}$, and fix $N > 1$ such that $B(0,N) \supset K$. By the non-decreasing property of φ, Jensen's inequality and Tonelli-Fubini theorem,

$$\int_K \varphi(|U_n(z)|)\, dx\, dy \le \int_{\mathbb{C}} \int_K \varphi\left(|\log|z-\lambda||\right) dx\, dy\, d\mu_n(\lambda).$$

Note that for every $z \in K$,

$$\varphi\left(|\log|z-\lambda||\right) \le \left(1 + |\log|z-\lambda||^2\right) \mathbb{1}_{\{|\lambda|\le N\}} + \varphi\left(\log(N+|\lambda|)\right) \mathbb{1}_{\{|\lambda|>N\}}.$$

To control the second term, we split and use local integratability of $1 + (\log|\cdot|)^2$ on $\mathbb{C}$:

$$\begin{aligned}
&\int_{|\lambda|\le N} \int_K \left(1 + |\log|z-\lambda||^2\right) dx\, dy\, d\mu_n(\lambda) \\
&\le \int_{|\lambda|\le N} \int_{K\cap B(\lambda,1)} \left(1 + |\log|z-\lambda||^2\right) dx\, dy\, d\mu_n(\lambda) + \int_{|\lambda|\le N} \int_{K\setminus B(\lambda,1)} \left(1 + |\log|z-\lambda||^2\right) dx\, dy\, d\mu_n(\lambda) \\
&\le \int_{|\lambda|\le N} \int_{B(0,1)} \left(1 + |\log|z||^2\right) dx\, dy\, d\mu_n(\lambda) + \int_{|\lambda|\le N} \int_{K\setminus B(\lambda,1)} \left(1 + |\log(2N)|^2\right) dx\, dy\, d\mu_n(\lambda) \\
&\le \int_{B(0,1)} \left(1 + |\log|z||^2\right) dx\, dy + \pi N^2 \left(1 + |\log(2N)|^2\right) := C_N,
\end{aligned}$$

where $C_N \in (0,\infty)$ is a constant depending on N only. To control the second term, note that

$$\int_{|\lambda|>N} \int_K \varphi\left(\log(N+|\lambda|)\right) dx\, dy\, d\mu_n(\lambda) \le \pi N^2 \int_{\mathbb{C}} \varphi\left(\log(N+|\lambda|)\right) d\mu_n(\lambda).$$

Hence

$$\sup_{n\in\mathbb{N}} \int_K \varphi(|U_n(z)|)\, dx\, dy \le C_N + \pi N^2 \sup_{n\in\mathbb{N}} \int_{\mathbb{C}} \varphi\left(\log(N+|\lambda|)\right) d\mu_n(\lambda) < \infty.$$

Again by de la Vallée Poussin criterion, and since K is arbitrary, $(U_{\mu_n})_{n\in\mathbb{N}}$ is locally uniformly Lebesgue integrable on $\mathbb{C}$. Then by assumption (i), U is locally Lebesgue integrable on $\mathbb{C}$, and $U_{\mu_n} \to U$ in $L^1_{\text{loc}}(\mathbb{C})$.

By continuity of Laplace operator Δ in $\mathcal{D}'(\mathbb{C})$, endowed with the weak-* topology, $\Delta U_{\mu_n} \to \Delta U$ in $\mathcal{D}'(\mathbb{C})$. Note that for sequence of Radon measures, convergence in $\mathcal{D}'(\mathbb{C})$ implies weak convergence. By (5.2), we have

$\mu_n \to \mu = -\frac{1}{2\pi}\Delta U$ weakly, and μ is a probability measure in $\mathcal{P}_\infty(\mathbb{C})$, since

$$\int_{\mathbb{C}} \log^+ |\lambda| \, d\mu(\lambda) \le \int_{\mathbb{C}} \log(1+|\lambda|) \, d\mu(\lambda) \le \lim_{n\to\infty} \int_{\mathbb{C}} \log(1+|\lambda|) \, d\mu_n(\lambda) \le \sup_{n\in\mathbb{N}} \int_{\mathbb{C}} \log(1+|\lambda|) \, d\mu_n(\lambda) < \infty.$$

Finally it remains to check $U_\mu = U$ a.e., which automatically follow from the following result.

(ii) $\Rightarrow$ (i). Note that for any $\phi \in C_c^\infty(\mathbb{C})$,

$$\langle U_{\mu_n}, \phi \rangle = -\int_{\mathbb{C}} \left(\int_{\mathbb{C}} \phi(z) \log|z-\lambda| \, dx \, dy \right) d\mu_n(\lambda) = -\int_{\mathbb{C}} (\phi * \log|\cdot|)(\lambda) \, d\mu_n(\lambda).$$

Since $\phi \in C_c^\infty(\mathbb{C})$ and $\log|\cdot|$ is locally integrable, if $\lambda_j \to \lambda$ in $\mathbb{C}$, we take a common compact support K of $(\phi(\lambda_j - \cdot))_{j\in\mathbb{N}}$ and restrict $\log|\cdot|$ on K. Then some multiple of $|\log|\cdot||\mathbb{1}_K$ is a common L^1-majorant for functions $\phi(\lambda_j - \cdot)\log|\cdot|$, and by dominated convergence theorem,

$$\int_{\mathbb{C}} \phi(\lambda_j - z) \log|z| \, dx \, dy \to \int_{\mathbb{C}} \phi(\lambda - z) \log|z| \, dx \, dy \quad \text{as } j \to \infty.$$

Hence

$$\phi * \log|\cdot| : \lambda \mapsto \int_{\mathbb{C}} \phi(z) \log|\lambda - z| \, dx \, dy = \int_{\mathbb{C}} \phi(\lambda - z) \log|z| \, dx \, dy$$

is a continuous function. Using the same approach as in Proposition 5.5, we have $|\phi * \log|\cdot|| \le C_\phi(1+\log^+|\cdot|)$ for some constant $C_\phi > 0$ depending only on ϕ. Hence $\phi * \log|\cdot|$ is also uniformly integrable for $(\mu_n)_{n\in\mathbb{N}}$, and

$$\langle U_{\mu_n}, \phi \rangle = -\int_{\mathbb{C}} (\phi * \log|\cdot|)(\lambda) \, d\mu_n(\lambda) \to -\int_{\mathbb{C}} (\phi * \log|\cdot|)(\lambda) \, d\mu(\lambda) = \langle U_\mu, \phi \rangle.$$

Therefore $U_{\mu_n} \to U_\mu$ in $\mathcal{D}'(\mathbb{C})$. If $U_{\mu_n} \to U$ also, since both U_μ and U are in $L^1_{\mathrm{loc}}(\mathbb{C})$, they must agree a.e.. $\square$

5.2.2 Hermitization

Spectral logarithm potential. Let $A \in \mathbb{C}^{n\times n}$ be a non-Hermitian matrix, and let $P_A(z) = \det(A - z\,\mathrm{Id})$ be its characteristic polynomial. Then for every $z \in \mathbb{C}\backslash\{\lambda_1(A), \cdots, \lambda_n(A)\}$,

$$U_{\mu_A}(z) = -\int_{\mathbb{C}} \log|z-\lambda| \, d\mu(\lambda) = -\frac{1}{n}\sum_{j=1}^n \log|z - \lambda_j(A)| = -\frac{1}{n}\log|\det(A - zI)| = -\frac{1}{n}\log|P_A(z)|.$$

We also have the determinantal Hermitization form:

$$U_{\mu_A}(z) = -\frac{1}{n}\log\det\left(\sqrt{(A-zI)(A-zI)^*}\right) = -\int_0^\infty \log t \, d\nu_{A-zI}(t).$$

Therefore, the knowledge of ν_{A-zI} for a.e. $z \in \mathbb{C}$ suffices to determine μ_A. Furthermore, by (5.2),

$$2\pi \int_{\mathbb{C}} \phi \, d\mu_A = \frac{1}{n}\int_{\mathbb{C}} \Delta\phi(z) \log|P_A(z)| \, dx \, dy.$$

For our later discussion, we also use uniform integrability. A Borel function f is said to be *uniformly integrable* for a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ on E, if

$$\lim_{N\uparrow\infty} \sup_{n\in\mathbb{N}} \int_{\{|f|>N\}} |f| \, d\mu_n = 0.$$

Lemma 5.8 (Logarithmic majorization and uniform integrability)**.** *Let $(\alpha_{n,k})_{1\le k\le n}$ and $(\beta_{n,k})_{1\le k\le n}$ be two* *triangular arrays in* $\mathbb{R}_+$. *Define discrete measures*

$$\mu_n = \frac{1}{n}\sum_{k=1}^{\infty}\delta_{\alpha_{n,k}}, \quad \textit{and} \quad \nu_n = \frac{1}{n}\sum_{k=1}^{\infty}\delta_{\beta_{n,k}}, \quad n = 1, 2, \cdots .$$

Assume the following properties hold:

(i) $\alpha_{n,1} \geq \alpha_{n,2} \geq \cdots \geq \alpha_{n,n}$ *and* $\beta_{n,1} \geq \beta_{n,2} \geq \cdots \geq \beta_{n,n}$ *for large enough* n,

(ii) $\prod_{k=1}^n \alpha_{n,k} = \prod_{k=1}^n \beta_{n,k}$ *for large enough* n,

(iii) $\prod_{j=1}^k \alpha_{n,j} \leq \prod_{j=1}^k \beta_{n,j}$ *for every* $1 \leq k \leq n$ *for large enough* n,

(iv) $\nu_n \to \nu$ *weakly as* $n \to \infty$ *for some probability measure* ν, *and*

(v) log *is uniformly integrable for* $(\nu_n)_{n\in\mathbb{N}}$.

Then

(a) $(\mu_n)_{n\in\mathbb{N}}$ *is a tight sequence of probability measures,*

(b) the function log *is uniformly integrable for* $(\mu_n)_{n\in\mathbb{N}}$,

(c) as $n \to \infty$,

$$\lim_{n\to\infty}\int_0^{\infty}\log t\, d\mu_n(t) = \lim_{n\to\infty}\int_0^{\infty}\log t\, d\nu_n(t) = \int_0^{\infty}\log t\, d\nu(t), \tag{5.3}$$

and in particular, for every accumulate point μ *of* $(\mu_n)_{n\in\mathbb{N}}$,

$$\int_0^{\infty}\log t\, d\mu(t) = \int_0^{\infty}\log t\, d\nu(t).$$

Proof. Using the de la Vallée Poussin theorem, the property (v) implies the existence of some non-decreasing, convex function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi(t)/t \to \infty$ as $t \uparrow \infty$, and

$$\sup_{n\in\mathbb{N}}\int_0^{\infty}\varphi(|\log t|)\, d\nu_n(t) = \sup_{n\in\mathbb{N}}\frac{1}{n}\left(\sum_{k=1}^{n}\varphi(|\log \beta_{n,k}|)\right) < \infty.$$

We let $a_j = \log \alpha_{n,j}$ and $b_j = \log \beta_{n,j}$ for $j \in [n]$. By properties (i) and (iii), $a = (a_{n,j})_{j\in[n]}$ is strongly majorized by $b = (b_{n,j})_{j\in[n]}$, i.e.

$$a_1 \geq \cdots \geq a_n, \quad b_1 \geq \cdots \geq b_n, \quad \sum_{j=1}^n a_j = \sum_{j=1}^n b_j, \quad \text{and} \quad \sum_{j=1}^k a_j \leq \sum_{j=1}^k b_j, \quad \text{for all } k = 1, \cdots, n.$$

We then use a proof from Hardy-Littlewood-Pólya to show the existence of a bistochastic matrix $S = (s_{ij})_{i,j\in[n]}$ such that $a = Sb$, i.e. $a_i = \sum_{j=1}^n s_{ij} b_j$ for all $i \in [n]$.

- If $a_1 = b_1$, then we leave b_1 as it is and let $T^{(1)} = I$.
- If $a_1 < b_1$, there must exist $k > 1$ such that $a_k > b_k$. We pick smallest such k and $0 < \theta < 1$ such that $\theta b_1 + (1-\theta) b_k = a_1$. We let matrix $T^{(1)} \in \mathbb{R}^{n\times n}$ satisfies $T^{(1)}_{11} = T^{(1)}_{kk} = \theta$ and $T^{(1)}_{1k} = T^{(1)}_{k1} = 1-\theta$, with all other diagonal entries 1 and off-diagonal entries 0. Then $b^{(1)} = T^{(1)} b$ satisfies $b^{(1)}_1 = a_1$, and $b^{(1)}_j = b_j$ for all $j \in [n]\backslash\{1, k\}$. Furthermore, the new vector $b^{(1)}$ still strongly majorizes a.
- We inductively repeat the above steps for vectors restricted to coordinates $j, j+1, \cdots, n$ to adjust coordinate j, where $j = 2, 3, \cdots$. This yields a sequence of transforms $T^{(2)}, \cdots, T^{(n-1)}$ until all coordinates of a and b matches. Then $a = T^{(n-1)} \cdots T^{(2)} T^{(1)} b$, and $S = T^{(n-1)} \cdots T^{(2)} T^{(1)}$ is bistochastic.

Then for any convex function ψ on $\mathbb{R}$,

$$\sum_{i=1}^n \psi(a_{n,i}) \leq \sum_{i=1}^n \psi\left(\sum_{j=1}^n s_{ij} b_{n,j}\right) \leq \sum_{i=1}^n\sum_{j=1}^n s_{ij}\psi(b_{n,j}) = \sum_{j=1}^n\left(\sum_{i=1}^n s_{ij}\right)\psi(b_{n,j}) = \sum_{j=1}^n \psi(b_{n,j}).$$

We choose the function $\psi(x) = \varphi(|x|)$, which is convex. Then

$$\sup_{n\in\mathbb{N}} \int_0^\infty \varphi(|\log t|)\, d\mu_n(t) = \sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{k=1}^n \varphi\left(|\log \alpha_{n,k}|\right) \le \sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{k=1}^n \varphi\left(|\log \beta_{n,k}|\right) < \infty.$$

Again by de la Vallée Poussin theorem, log is uniformly integrable for $(\mu_n)_{n\in\mathbb{N}}$. This also implies the fact that $(\mu_n)_{n\in\mathbb{N}}$ is tight, since

$$\sup_{n\in\mathbb{N}} \mu_n([N,\infty)) \le \sup_{n\in\mathbb{N}} \int_{|\log\lambda|>\log N} d\mu_n(\lambda) \le \sup_{n\in\mathbb{N}} \int_{|\log\lambda|>\log N} |\log\lambda|\, d\mu_n(\lambda) \to 0 \quad \text{as } e < N \uparrow \infty.$$

Finally, the (5.3) follows from the property (ii), and we finish the proof. □

Following is the main theorem we will make use of in the proof of the circular law.

Theorem 5.9 (Girko Hermitization). *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of complex random matrices where A_n is of size $n \times n$. Suppose there exists a family of (non-random) probability measures $(\nu_z)_{z\in\mathbb{C}}$ on $\mathbb{R}_+ = [0,\infty)$ such that for almost every $z \in \mathbb{C}$, almost surely,*

(i) $\nu_{A_n - zI} \to \nu_z$ weakly as $n \to \infty$, and

(ii) log *is uniformly integrable for $(\nu_{A_n-zI})_{n\in\mathbb{N}}$.*

Then there exists a probability measure $\mu \in \mathcal{P}_\infty(\mathbb{C})$ such that

(a) almost surely, $\mu_{A_n} \to \mu$ weakly as $n \to \infty$, and

(b) for almost every $z \in \mathbb{C}$,

$$U_\mu(z) = -\int_0^\infty \log s\, d\nu_z(s). \tag{5.4}$$

Proof. First, we consider the product measure $\mathbb{P} \otimes m$ on $\Omega \times \mathbb{C}$, where $(\Omega, \mathbb{P})$ is the underlying probability space and m is the Lebesgue measure on $\mathbb{C}$. By Tonelli-Fubini theorem, the quantifiers "for a.e. $z \in \mathbb{C}$" and "for a.s. $\omega \in \Omega$" can be swapped.

Next, we condition on an event E of probability 1 such that properties (i)-(ii) holds for a.e. $z \in \mathbb{C}$ on E, and fix a realization $\omega \in E$. Then we can focus on the determinisitc case. Also we fix $N_\omega \subset \mathbb{C}$ of Lebesgue measure zero such that properties (i)-(ii) holds for all $z \notin N_\omega$.

For every $z \notin N_\omega$, we set $\nu = \nu_z$ and define triangular arrays $(\alpha_{n,k})_{1\le k\le n}$ and $(\beta_{n,k})_{1\le k\le n}$ by

$$\alpha_{n,k} = \left|\lambda_k(A_n(\omega) - zI)\right|, \quad \beta_{n,k} = \sigma_k(A_n(\omega) - zI), \quad 1 \le k \le n.$$

By Theorem 5.1, the properties (i)-(iii) in Lemma 5.8 are satisfied. Also properties (iv)-(v) in Lemma 5.8 is satisfied by assumptions (i)-(ii). Note that $\mu_{A_n(\omega)-zI} = \mu_{A_n(\omega)} * \delta_{-z}$ for all $z \in \mathbb{C}$, which is a traslated version of $\mu_{A_n(\omega)}$. Then we apply Lemma 5.8 implies that

- $(\mu_{A_n(\omega)})_{n\in\mathbb{N}}$ is tight, and
- for a.e. $z \in \mathbb{C}$, the function $\lambda \mapsto \log|z - \lambda|$ is uniformly integrable for $(\mu_{A_n(\omega)})_{n\in\mathbb{N}}$, and that

$$\lim_{n\to\infty} U_{\mu_{A_n(\omega)}}(z) = -\lim_{n\to\infty} \int_0^\infty \log s\, d\nu_{A_n-zI}(s) = -\int_0^\infty \log s\, d\nu_z(s) =: U(z).$$

By Prokhorov's theorem, every subsequence of the tight sequence $(\mu_{A_n(\omega)})_{n\in\mathbb{N}}$ admits a further subsequence that converges weakly. Then by the subsequnece criterion, it suffices to show that $(\mu_{A_n(\omega)})_{n\in\mathbb{N}}$ has only one accumulate point of weak convergence. Assume that μ_ω and μ'_ω are both accumulate points of $(\mu_{A_n(\omega)})_{n\in\mathbb{N}}$. By the uniform integrability of $\log|\cdot|$, we have $\mu_\omega, \mu'_\omega \in \mathcal{P}_\infty(\mathbb{C})$ and $U_{\mu_\omega} = U_{\mu'_\omega} = U$ a.e.. By Theorem 5.6, $\mu_\omega = \mu'_\omega$, and hence $\mu_{A_n(\omega)} \to \mu_\omega$. Since the logarithm potential U is deterministic, it follows that $\omega \to \mu_\omega$ is deterministic by Theorem 5.6 again. Therefore $\mu_{A_n} \to \mu$ on E, and $U_\mu(z) = U(z) = -\int_0^\infty \log s\, d\mu_z(s)$. □

Finally, we weaken the uniform integrability condition (ii) in Theorem 5.9 to simplify subsequent analysis.

Lemma 5.10 (Weakening uniform integrability). *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of complex random matrices where A_n is of size $n \times n$. Assume that for almost every $z \in \mathbb{C}$, there exists $p > 0$ such that almost surely,*

$$\limsup_{n\to\infty} \int_0^\infty s^{-p} d\nu_{A_n - zI}(s) < \infty, \quad \textit{and} \quad \limsup_{n\to\infty} \int_0^\infty s^{p} d\nu_{A_n - zI}(s) < \infty. \tag{5.5}$$

Then for almost every $z \in \mathbb{C}$, almost surely, the function log *is uniformly integrable for $(\nu_{A_n - zI})_{n\in\mathbb{N}}$.*

Proof. Let $(\Omega, \mathbb{P})$ be the underlying probability space, and m the Lebesgue measure on $\mathbb{C}$. By Tonelli-Fubini,

$$(\mathbb{P}\otimes m)\left\{(\omega, z) : z \text{ is an eigenvalue of } A_n(\omega) \text{ for some } n \in \mathbb{N}\right\} = \int_\Omega \int_{\mathbb{C}} \mathbb{1}_{\bigcup_{n=1}^\infty \{\det(A_n(\omega) - zI) = 0\}}\, dm(z)\, d\mathbb{P}(\omega)$$
$$\leq \int_\Omega \left[\sum_{n=1}^\infty \int_{\mathbb{C}} \mathbb{1}_{\{\det(A_n(\omega)-zI)=0\}}\, dm(z)\right] d\mathbb{P}(\omega) = 0,$$

where the last equality follows because $A_n(\omega)$ has at most n eigenvalues in $\mathbb{C}$, and spectrum of $A_n(\omega)$ is of Lebesgue measure 0. Hence for a.e. $z \in \mathbb{C}$, almost surely, z is not an eigenvalue of A_n for any $n \in \mathbb{N}$. This implies that for a.e. $z \in \mathbb{C}$,

$$\int_0^\infty |\log s|\, d\nu_{A_n - zI}(s) < \infty \quad \text{a.s. for all } n \in \mathbb{N}.$$

Therefore, to show uniform integrability, we may replace the sup in definition by lim sup and prove

$$\lim_{N\uparrow\infty} \limsup_{n\to\infty} \int_{|\log s| \geq N} |\log s|\, d\nu_{A_n - zI}(s) = 0 \quad \text{a.s.}. \tag{5.6}$$

We fix a small $\delta > 0$. Then $|\log s|^\delta / N^\delta \geq 1$ on $\{|\log s| \geq N\}$. By Markov's inequality,

$$\int_{|\log s|\geq N} |\log s|\, d\nu_{A_n - zI}(s) \leq \frac{1}{N^\delta} \int_0^\infty |\log s|^{1+\delta}\, d\nu_{A_n - zI}(s)$$

Note that for any $q > 0$,

$$|\log s| \leq \frac{s^{-q}}{q} \mathbb{1}_{\{0\leq s\leq 1\}} + \frac{s^q}{q} \mathbb{1}_{\{s\geq 1\}}.$$

Choose $q = \frac{p}{1+\delta}$. Then

$$|\log s|^{1+\delta} \leq \left(\frac{1+\delta}{p}\right)^{1+\delta} \left(s^{-p}\mathbb{1}_{\{0\leq s\leq 1\}} + s^p \mathbb{1}_{\{s\geq 1\}}\right),$$

and

$$\int_{|\log s|\geq N} |\log s|\, d\nu_{A_n - zI}(s) \leq \frac{1}{N^\delta}\left(\frac{1+\delta}{p}\right)^{1+\delta} \left[\int_0^\infty s^{-p} d\nu_{A_n-zI}(s) + \int_0^\infty s^p d\nu_{A_n - zI}(s)\right]$$

By (5.5), as $N \uparrow \infty$,

$$\limsup_{n\to\infty} \int_{|\log s|\geq N} |\log s|\, d\nu_{A_n - zI}(s) \to 0 \quad \text{a.s.},$$

which is exactly (5.6). Thus we finish the proof. $\square$

5.3 Proof of the Circular Law

5.3.1 Convergence of Singular Values Measure

To verify that for each $z \in \mathbb{C}$, there exists a probability measure ν_z on $\mathbb{R}_+$ such that $\nu_{n^{-1/2}X_n - zI} \to \nu_z$ weakly a.s., we need to study the spectral measure of the Hermitian matrices

$$\left(n^{-1/2}X_n - zI\right)\left(n^{-1/2}X_n - zI\right)^*, \quad n = 1, 2, \cdots.$$

Theorem 5.11 (Dozier-Silverstein)**.** *Let $z \in \mathbb{C}$. Almost surely, the empirical spectral measure of*

$$\left(n^{-1/2}X_n - zI\right)\left(n^{-1/2}X_n - zI\right)^* \in \mathbb{C}^{n\times n}$$

converges weakly to a probability measure μ_z depending on z only. Furthermore, μ_z is uniquely defined by its Stieltjes transform $s : \mathbb{C}^+ \to \mathbb{C}^+$, which satisfies the fixed point equation

$$s(\zeta) = \frac{1}{\frac{|z|^2}{1+s(\zeta)} - \zeta(1 + s(\zeta))}, \quad \zeta \in \mathbb{C}^+. \tag{5.7}$$

Lemma 5.12 (Reduction)**.** *In Theorem 5.11, one may assume that for every $n \in \mathbb{N}$, the matrix X_n has i.i.d. entries $(x_{ij})_{i,j\in[n]}$ bounded by $\log n$.*

Proof. We fix a sequence (κ_n) in $\mathbb{R}_+$ that grows to ∞, and define

$$\overline{x}_{ij} = x_{ij}\mathbb{1}_{\{|x_{ij}|\le\kappa_n\}}, \quad \widehat{x}_{ij} = \overline{x}_{ij} - \mathbb{E}\left[\overline{x}_{ij}\right], \quad \widetilde{x}_{ij} = \frac{\widehat{x}_{ij}}{\sqrt{\mathbb{E}|\widehat{x}_{ij}|^2}}, \qquad i, j = 1, 2, \cdots.$$

and set $\overline{X}_n = (\overline{x}_{ij})_{i,j\in[n]}$, $\widehat{X}_n = (\widehat{x}_{ij})_{i,j\in[n]}$, and $\widetilde{X}_n = (\widetilde{x}_{ij})_{i,j\in[n]}$. We fix $N > 0$. By Lemma 2.3, for large enough n, we have $\kappa_n > N$, and

$$\begin{aligned}
&\rho_L\left(F_{(n^{-1/2}X_n-zI)(n^{-1/2}X_n-zI)^*}, F_{(n^{-1/2}\overline{X}_n-zI)(n^{-1/2}\overline{X}_n-zI)^*}\right)^4 \\
&\le \frac{2}{n^2}\left(\left\|n^{-1/2}X_n - zI\right\|_{\mathrm{F}}^2 + \left\|n^{-1/2}\overline{X}_n - zI\right\|_{\mathrm{F}}^2\right)\left\|n^{-1/2}X_n - n^{-1/2}\overline{X}_n\right\|_{\mathrm{F}}^2 \\
&\le \frac{2}{n^2}\left[4\|zI\|_{\mathrm{F}}^2 + \frac{2}{n}\sum_{i,j=1}^{n}\left(|x_{ij}|^2 + |x_{ij}|^2\mathbb{1}_{\{|x_{ij}|\le\kappa_n\}}\right)\right]\left[\frac{1}{n}\sum_{i,j=1}^{n}|x_{ij}|^2\mathbb{1}_{\{|x_{ij}|>\kappa_n\}}\right] \\
&\le \left[8|z|^2 + \frac{8}{n^2}\sum_{i,j=1}^{n}|x_{ij}|^2\right]\left[\frac{1}{n^2}\sum_{i,j=1}^{n}|x_{ij}|^2\mathbb{1}_{\{|x_{ij}|>N\}}\right] \\
&\to 8(1+|z|^2)\,\mathbb{E}\left[|x_{11}|^2\mathbb{1}_{\{|x_{11}|>N\}}\right] \quad \text{almost surely.}
\end{aligned} \tag{5.8}$$

Next, by Lemma 2.2,

$$\rho_L\left(F_{(n^{-1/2}\overline{X}_n-zI)(n^{-1/2}\overline{X}_n-zI)^*}, F_{(n^{-1/2}\widehat{X}_n-zI)(n^{-1/2}\widehat{X}_n-zI)^*}\right) \le \frac{\operatorname{rank}(\mathbb{E}\overline{X}_n)}{n} = \frac{1}{n}, \tag{5.9}$$

which converges to 0 deterministically. Finally, for large enough n, we have $\kappa_n > N$, and

$$
\begin{aligned}
&\rho_L\left(F_{(n^{-1/2}\widehat{X}_n - zI)(n^{-1/2}\widehat{X}_n - zI)^*}, F_{(n^{-1/2}\widetilde{X}_n - zI)(n^{-1/2}\widetilde{X}_n - zI)^*}\right)^4 \\
&\le \frac{2}{n^2}\left(\left\|n^{-1/2}\widehat{X}_n - zI\right\|_{\mathrm{F}}^2 + \left\|n^{-1/2}\widetilde{X}_n - zI\right\|_{\mathrm{F}}^2\right)\left\|n^{-1/2}\widehat{X}_n - n^{-1/2}\widetilde{X}_n\right\|_{\mathrm{F}}^2 \\
&\le 4\left[|z|^2 + \frac{1+\mathbb{E}|\widehat{x}_{11}|^2}{n^2\,\mathbb{E}|\widehat{x}_{11}|^2}\sum_{i,j=1}^n |\widehat{x}_{ij}|^2\right]\left[\frac{\left(1-\sqrt{\mathbb{E}|\widehat{x}_{11}|^2}\right)^2}{n^2\mathbb{E}|\widehat{x}_{11}|^2}\sum_{i,j=1}^n |\widehat{x}_{ij}|^2\right] \qquad (5.10)\\
&\le 4\left[|z|^2 + \frac{C}{n^2}\sum_{i,j=1}^n |x_{ij}|^2\right]\left[\frac{C}{n^2}\sum_{i,j=1}^n |x_{ij}|^2 \mathbb{1}_{\{|x_{ij}|\ge N\}}\right] \\
&\to 4\left(C+|z|^2\right)\left(1-\sqrt{\mathrm{Var}(x_{11}\mathbb{1}_{\{|x_{11}|\le N\}})}\right)^2 \quad \text{almost surely,}
\end{aligned}
$$

where C is some constant not depending on n. As $N \uparrow \infty$, the a.s. bounds (5.8) and (5.10) converge to 0. Hence it suffices to show the weak convergence of the ESD of matrices $\widetilde{X}_n$, which have i.i.d. entries with mean 0, variance 1 and amplitude $O(\kappa_n)$. Choosing $\kappa_n = O(\log n)$ concludes the proof. □

Lemma 5.13. *Let $A = (a_{ij})_{i,j\in[n]}$ be an $n \times n$ complex matrix with $\|C\|_2 \le 1$, and $Y = (Y_1, \cdots, Y_n)$, where $Y_1, \cdots, Y_n$ are i.i.d. random variables with $\mathbb{E}Y_1 = 0$, $\mathbb{E}|Y_1|^2 = 1$ and $|Y_1| \le \log n$ a.s.. Then*

$$\mathbb{E}\left|Y^*AY - \operatorname{tr} A\right|^6 \le Kn^3(\log n)^{12}.$$

Proof. Since $\sqrt{\lambda_1(AA^*)} = \|A\|_2 \le 1$, it follows $|a_{ii}| \le 1$ for each $i \in [n]$. Note that

$$\mathbb{E}\left|Y^*AY - \operatorname{tr} A\right|^p \le 2^{p-1}\left[\mathbb{E}\left|\sum_{i=1}^n a_{ii}(|Y_i|^2-1)\right|^p + \mathbb{E}\left|\sum_{i\ne j} a_{ij}Y_iY_j\right|^p\right].$$

By Lemma 1.18, there exists a constant $K_p > 0$ depending on p only, such that

$$\mathbb{E}\left|\sum_{i=1}^n a_{ii}(|Y_i|^2-1)\right|^p \le K_p\,\mathbb{E}\left|\sum_{i=1}^n |a_{ii}|^2\left||Y_i|^2-1\right|^2\right|^{p/2} \le K_p n^{p/2}(\log n)^{2p},$$

and

$$\mathbb{E}\left|\sum_{i\ne j} a_{ij}Y_iY_j\right|^p \le K_p\left(\sum_{i\ne j}|a_{ij}|^2\right)^{p/2}\left(\max_{j\in[n]}\mathbb{E}\,|Y_j|^p\right)^2$$

Since $\sum_{i,j=1}^n |a_{ij}|^2 = \operatorname{tr}(AA^*) \le n\lambda_1(AA^*) \le n$, the bound is at most $K_p n^{p/2}(\log n)^{2p}$. In particular, there exists $K > 0$ such that a constant $K = 64K_6 > 0$ such that

$$\mathbb{E}\left|Y^*AY - \operatorname{tr} A\right|^6 \le Kn^3(\log n)^{12}.$$

The we finishes the proof. □

Proof of Theorem 5.11. We fix $z \in \mathbb{C}$ with $r = |z|$, and write

$$C_n = \left(\frac{X_n}{\sqrt{n}} - zI\right)\left(\frac{X_n}{\sqrt{n}} - zI\right)^* = \sum_{j=1}^n y_jy_j^*, \qquad \text{where } y_j = \frac{x_j}{\sqrt{n}} - ze_j, \quad j = 1, \cdots, n.$$

Step I. Fix $\zeta = E + i\eta \in \mathbb{C}^+$, with $\eta > 0$, and let

$$\beta_n = \frac{r^2}{1+s_n} - \zeta(1+s_n), \quad \text{where } s_n = s_{C_n}(\zeta) \text{ is the Stieltjes transform of ESD of } C_n.$$

Let $D_n = C_n - \zeta I$ and $D_{n,-j} = D_n - y_j y_j^*$. Then $s_n = n^{-1} \operatorname{tr} D_n^{-1}$. By Sherman-Morrison formula,

$$D_n^{-1} = D_{n,-j}^{-1} - \frac{D_{n,-j}^{-1} y_j y_j^* D_{n,-j}^{-1}}{1 + y_j^* D_{n,-j}^{-1} y_j}, \quad j = 1, \cdots, n.$$

Then

$$I + \zeta D_n^{-1} = D_n^{-1}(D_n + \zeta I) = \sum_{j=1}^n D_n^{-1} y_j y_j^* = \sum_{j=1}^n \left(D_{n,-j}^{-1} - \frac{D_{n,-j}^{-1} y_j y_j^* D_{n,-j}^{-1}}{1 + y_j^* D_{n,-j}^{-1} y_j} \right) y_j y_j^*$$

Taking the trace on both sides and dividing by n, we have

$$1 + \zeta s_n = \frac{1}{n} \sum_{j=1}^n y_j^* D_{n,-j}^{-1} y_j \left(1 - \frac{y_j^* D_{n,-j}^{-1} y_j}{1 + y_j^* D_{n,-j}^{-1} y_j} \right) = 1 - \frac{1}{n} \sum_{j=1}^n \frac{1}{1 + y_j^* D_{n,-j}^{-1} y_j}. \tag{5.11}$$

Then

$$\begin{aligned} \beta_n^{-1} I - D_n^{-1} &= \beta_n (D_n - \beta_n I) D_n^{-1} = \beta_n^{-1} \left(\sum_{j=1}^n y_j y_j^* - \frac{r^2 I}{1+s_n} + \zeta s_n I \right) D_n^{-1} \\ &= \beta_n^{-1} \left[\sum_{j=1}^n \left(y_j y_j^* - \frac{1}{n\left(1 + y_j^* D_{n,-j}^{-1} y_j\right)} \right) - \frac{r^2 I}{1+s_n} \right] D_n^{-1}. \end{aligned} \tag{5.12}$$

For notation simplicity, for each $j \in [n]$, we write

$$\omega_{n,j} = \frac{1}{n} x_j^* D_{n,-j} x_j, \quad \theta_{n,j} = \frac{1}{\sqrt{n}} \overline{z} e_j^* D_{n,-j}^{-1} x_j, \quad \vartheta_{n,j} = \frac{1}{\sqrt{n}} z x_j^* D_{n,-j}^{-1} e_j, \quad \tau_{n,j} = r^2 e_j^* D_{n,-j}^{-1} e_j.$$

Then $y_j^* D_{n,-j}^{-1} y_j = \omega_{n,j} + \theta_{n,j} + \vartheta_{n,j} + \tau_{n,j}$. Again, we take the trace on both sides of (5.12) and divide by n to obtain

$$\begin{aligned} \beta_n^{-1} - s_n &= \frac{1}{n\beta_n} \sum_{j=1}^n \left(y_j^* D_n^{-1} y_j - \frac{\operatorname{tr} D_n^{-1}}{n\left(1 + y_j^* D_{n,-j}^{-1} y_j\right)} - \frac{r^2 e_j^* D_n^{-1} e_j}{1+s_n} \right) \\ &= \frac{1}{n\beta_n} \sum_{j=1}^n \left(y_j^* D_n^{-1} y_j - \frac{s_n}{1 + y_j^* D_{n,-j}^{-1} y_j} - \frac{r^2 e_j^* D_{n,-j}^{-1} e_j}{1+s_n} + \frac{r^2 e_j^* D_{n,-j}^{-1} y_j y_j^* D_{n,-j}^{-1} e_j}{(1+s_n)(1 + y_j^* D_{n,-j}^{-1} y_j)} \right) \\ &= \frac{1}{n\beta_n} \sum_{j=1}^n \left(\frac{y_j^* D_{n,-j}^{-1} y_j - s_n}{1 + y_j^* D_{n,-j}^{-1} y_j} - \frac{r^2 e_j^* D_{n,-j}^{-1} e_j}{1+s_n} + \frac{r^2 e_j^* D_{n,-j}^{-1} y_j y_j^* D_{n,-j}^{-1} e_j}{(1+s_n)(1 + y_j^* D_{n,-j}^{-1} y_j)} \right) \\ &= \frac{1}{n\beta_n} \sum_{j=1}^n \left(\frac{\omega_{n,j} + \theta_{n,j} + \vartheta_{n,j} + \tau_{n,j} - s_n}{1 + y_j^* D_{n,-j}^{-1} y_j} - \frac{\tau_{n,j}}{1+s_n} + \frac{(\theta_{n,j} + \tau_{n,j})(\vartheta_{n,j} + \tau_{n,j})}{(1+s_n)(1 + y_j^* D_{n,-j}^{-1} y_j)} \right) \\ &= \frac{1}{n\beta_n} \sum_{j=1}^n \left(\frac{(\omega_{n,j} - s_j) + \theta_{n,j} + \vartheta_{n,j}}{1 + y_j^* D_{n,-j}^{-1} y_j} + \frac{\tau_{n,j}(s_n - \omega_{n,j}) + \theta_{n,j}\vartheta_{n,j}}{(1+s_n)(1 + y_j^* D_{n,-j}^{-1} y_j)} \right). \end{aligned} \tag{5.13}$$

Note that

$$\operatorname{Im}(\zeta s_n) = \operatorname{Im} \left(\frac{1}{n} \sum_{j=1}^n \frac{\zeta}{\lambda_j(C_n) - \zeta} \right) = \frac{1}{n} \sum_{j=1}^n \frac{\lambda_j(C_n)\, \eta}{|\lambda_j(C_n) - \zeta|^2} \geq 0.$$

Then

$$\frac{1}{|1+s_n|} = \frac{|\zeta|}{|\zeta+\zeta s_n|} \le \frac{|\zeta|}{\operatorname{Im}\zeta + \operatorname{Im}(\zeta s_n)} \le \frac{|\zeta|}{\eta}, \tag{5.14}$$

and

$$|\beta_n| \ge |\operatorname{Im}\beta_n| = \left|\frac{r^2 \operatorname{Im} s_n}{|1+s_n|^2} + \operatorname{Im}(\zeta s_n) + \operatorname{Im}(\zeta)\right| \ge \frac{1}{\eta}. \tag{5.15}$$

Since all eigenvalues of D_n and $D_{n,-j}$ have imaginary part η, we have $\|D_n^{-1}\|_2 \le 1/\eta$ and $\|D_{n,-j}^{-1}\|_2 \le 1/\eta$. Consequently

$$|\tau_{n,j}| \le r^2 \|D_{n,-j}^{-1}\| \le \frac{1}{\eta}. \tag{5.16}$$

Combining (5.13), (5.14), (5.15) and (5.16), we have

$$|\beta_n^{-1} - s_n| \le \frac{\eta}{n}\sum_{j=1}^{n}\left(\left(1+\frac{|\zeta|}{\eta^2}\right)|\omega_{n,j}-s_n| + |\theta_{n,j}| + |\vartheta_{n,j}| + \frac{|\zeta|}{\eta}|\theta_{n,j}\vartheta_{n,j}|\right). \tag{5.17}$$

Step II. Now we handle the first term in (5.17). By Lemma 5.13,

$$\begin{aligned}
\mathbb{E}|\omega_{n,j}-s_n|^6 &= \frac{1}{n^6}\mathbb{E}\left|x_j^* D_{n,-j}^{-1} x_j - \operatorname{tr} D_n^{-1}\right|^6 \\
&\le \frac{32}{n^6}\left(\mathbb{E}\left|x_j^* D_{n,-j}^{-1} x_j - \operatorname{tr} D_{n-j}^{-1}\right|^6 + \mathbb{E}\left|\operatorname{tr}(D_{n,-j}^{-1} - D_n^{-1})\right|^6\right) \\
&\le \frac{32}{n^6}\left(\mathbb{E}\left|x_j^* D_{n,-j}^{-1} x_j - \operatorname{tr} D_{n-j}^{-1}\right|^6 + \mathbb{E}\left|\frac{y_j^* D_{n,-j}^{-2} y_j}{1+y_j^* D_{n,-j}^{-1} y_j}\right|^6\right) \\
&\le \frac{32}{n^6}\left(\frac{Kn^3(\log n)^{12}}{\eta^6} + \frac{1}{\eta^{12}}\mathbb{E}[y_j^* y_j]\right) \le \frac{32}{n^3}\left(\frac{K(\log n)^{12}}{\eta^6} + \frac{1+r^2}{n^3\eta^{12}}\right).
\end{aligned}$$

Then for any $\epsilon > 0$, by Markov's inequality,

$$\mathbb{P}\left(\max_{1\le j\le n}|\omega_{n,j}-s_n| > \epsilon\right) \le \sum_{j=1}^{n}\mathbb{P}\left(|\omega_{n,j}-s_n| > \epsilon\right) \le \sum_{j=1}^{n}\frac{\mathbb{E}|\omega_{n,j}-s_n|^6}{\epsilon^6} \le \frac{32}{n^2}\left(\frac{K(\log n)^{12}}{\eta^6} + \frac{1+r^2}{n^3\eta^{12}}\right).$$

Since the dominating term $\sum_{n=1}^{\infty}(\log n)^{12}/n^2 < \infty$, by the Borel Cantelli lemma,

$$\limsup_{n\to\infty}\max_{1\le j\le n}|\omega_{n,j}-s_n| < \epsilon, \quad \text{a.s..}$$

And since $\epsilon > 0$ is arbitrary, we have $\max_{1\le j\le n}|\omega_{n,j}-s_n| \to 0$ a.s..

Step III. Next we bound the remaining terms in (5.17). By Lemma 5.13,

$$\begin{aligned}
\mathbb{E}|\vartheta_{n,j}|^{12} &= \mathbb{E}\left|\frac{z x_j^* D_{n,-j}^{-1} e_j}{\sqrt{n}}\right|^{12} \le \frac{r^{12}}{n^6}\mathbb{E}\left|x_j^* D_{n,-j}^{-1} e_j e_j^* D_{n,-j}^{-1} x_j\right|^6 \\
&\le \frac{32 r^{12}}{n^6}\left(\mathbb{E}\left|x_j^* D_{n,-j}^{-1} e_j e_j^* D_{n,-j}^{-1} x_j - \operatorname{tr}\left(D_{n,-j}^{-1} e_j e_j^* D_{n,-j}^{-1}\right)\right|^6 + \mathbb{E}\left|\operatorname{tr}\left(D_{n,-j}^{-1} e_j e_j^* D_{n,-j}^{-1}\right)\right|^6\right) \\
&\le \frac{32 r^{12}}{n^6}\left(\frac{Kn^3(\log n)^{12}}{\eta^{12}} + \frac{1}{\eta^{12}}\right) = \frac{32 r^{12}}{n^3\eta^{12}}\left(K(\log n)^{12} + \frac{1}{n^3}\right).
\end{aligned}$$

Similarly,

$$\mathbb{E}|\theta_{n,j}|^{12}, \mathbb{E}|\theta_{n,j}\vartheta_{n,j}|^6 \le \frac{32 r^{12}}{n^3\eta^{12}}\left(K(\log n)^{12} + \frac{1}{n^3}\right).$$

Similar to Step II, we can use Borel-Cantellu lemma to deduce that

$$\max_{1\le j\le n}\{|\theta_{n,j}|,|\theta_{n,j}|,|\vartheta_{n,j}\vartheta_{n,j}|\}\to 0 \quad \text{a.s..}$$

Therefore $|\beta_n^{-1}-s_n|\to 0$ a.s.. More specifically,

$$\frac{1}{\frac{r^2}{1+s_n}-\zeta(1+s_n)}-s_n\to 0 \quad \text{a.s..} \tag{5.18}$$

Step IV. We fix $\omega\in\Omega$ such that (5.18) holds. By (5.11), (s_n) is a bounded sequence:

$$|s_n|\le\frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\zeta(1+y_j^*D_{n,-j}^{-1}y_j)}\right|\le\frac{1}{\eta}.$$

By Bolzano-Weierstrass theorem, it suffices to show that (s_n) has only one limit point, which must satisfy the fixed point equation (5.7). Once we show this, the convergence result follows from Stieltjes continuity theorem. Note that $s_n\in\mathbb{C}^+$ and $\operatorname{Im}(\zeta s_n)\ge 0$. Therefore we finish our proof by the following uniqueness lemma. $\square$

Lemma 5.14 (Uniqueness)**.** *Let $z\in\mathbb{C}$, and $\zeta,s,t\in\mathbb{C}^+$ with $\operatorname{Im}(\zeta s)\ge 0$ and $\operatorname{Im}(\zeta t)\ge 0$. If both s and t satisfies the fixed point equation* (5.7)*, then $s=t$.*

Proof. Write $r=|z|\ge 0$. By (5.7), we have

$$s-t=\frac{1}{\frac{r^2}{1+s}-(1+s)\zeta}-\frac{1}{\frac{r^2}{1+t}-(1+t)\zeta}=\frac{\frac{r^2}{(1+s)(1+t)}+\zeta}{\left(\frac{r^2}{1+s}-(1+s)\zeta\right)\left(\frac{r^2}{1+t}-(1+t)\zeta\right)}(s-t):=\alpha(s-t).$$

We define

$$G(u)=\frac{1}{\frac{r^2}{1+u}-(1+u)\zeta},\quad u\in\mathbb{C}^+.$$

Then $G(s)=s$, $G(t)=t$, and

$$\alpha=\frac{rG(s)}{1+s}\cdot\frac{rG(t)}{1+t}+\zeta G(s)G(t).$$

Since $s=G(s)$, we have $\operatorname{Re}s=\operatorname{Re}(G(s))$ and $\operatorname{Im}s=\operatorname{Im}(G(s))$. More specifically,

$$\operatorname{Re}s=\left[\frac{r^2(1+\operatorname{Re}s)}{|1+s|^2}-\operatorname{Re}\zeta-\operatorname{Re}(\zeta s)\right]|G(s)|^2,\quad\text{and}\quad \operatorname{Im}s=\left[\frac{r^2\operatorname{Im}s}{|1+s|^2}+\operatorname{Im}\zeta+\operatorname{Im}(s\zeta)\right]|G(s)|^2. \tag{5.19}$$

The first part of (5.19) implies

$$\left[1-\frac{r^2|G(s)|^2}{|1+s|^2}+\operatorname{Re}\zeta|G(s)|^2\right](1+\operatorname{Re}s)=1+\operatorname{Im}\zeta\operatorname{Im}s|G(s)|^2, \tag{5.20}$$

and the second part implies

$$\left[1-\frac{r^2|G(s)|^2}{|1+s|^2}-\operatorname{Re}\zeta|G(s)|^2\right]\operatorname{Im}s=(1+\operatorname{Re}s)\operatorname{Im}\zeta|G(s)|^2. \tag{5.21}$$

We plug-in (5.21) to (5.20) and rearrange to obtain

$$\left[\left(1-\frac{r^2|G(s)|^2}{|1+s|^2}\right)^2-(\operatorname{Re}\zeta)^2|G(s)|^4-(\operatorname{Im}\zeta)^2|G(s)|^4\right]\frac{\operatorname{Im}s}{\operatorname{Im}\zeta|G(s)|^2}=1,$$

which also writes

$$\left(1 - \frac{r^2|G(s)|^2}{|1+s|^2}\right)^2 - |\zeta|^2|G(s)|^4 = \frac{\operatorname{Im}\zeta|G(s)|^2}{\operatorname{Im} s} > 0.$$

By (5.19), we have $\left(1 - \frac{r^2|G(s)|^2}{|1+s|^2}\right) = (\operatorname{Im}\zeta + \operatorname{Im}(s\zeta))|G(s)|^2 > 0$. Hence

$$1 - \frac{r^2|G(s)|^2}{|1+s|^2} > |\zeta||G(s)|^2.$$

A similar inequality also holds for t. Using the inequality $\sqrt{1-x}\sqrt{1-y} \le 1 - \sqrt{xy}$ for $x, y \in [0,1]$, we have

$$\begin{aligned} |\alpha| &\le \left|\frac{rG(s)}{1+s}\right| \left|\frac{rG(t)}{1+t}\right| + \zeta\, |G(s)|\, |G(t)| \\ &< \sqrt{1 - |\zeta||G(s)|^2}\sqrt{1 - |\zeta||G(t)|^2} + |\zeta|\, |G(s)|\, |G(t)| \\ &\le 1 - |\zeta|\, |G(s)|\, |G(t)| + |\zeta|\, |G(s)|\, |G(t)| = 1. \end{aligned}$$

Hence $|\alpha| < 1$, and $s = t$. □

5.3.2 Count of Small Singular Values

Lemma 5.15 (Tao-Vu)**.** *Let $1 \le m \le n$, and let $A \in \mathbb{C}^{n\times m}$ be a matrix of full rank, with columns $A_1, \cdots, A_m \in \mathbb{C}^n$, and $V_k = \operatorname{span}\{A_j : j \in [m], j \ne k\}$ for every $k \in [m]$. Then*

$$\sum_{j=1}^m \sigma_j(A)^{-2} = \sum_{j=1}^m \operatorname{dist}(A_j, V_j)^{-2},$$

where $\operatorname{dist}(x, V) := \inf_{y\in V} \|x - y\|_2$ is the induced Euclidean distance between a vector and a set.

Proof. Let $A_{-j} \in \mathbb{R}^{n\times(m-1)}$ be the matrix obtained from A by removing the j-th row. Then the orthogonal projection of A_j onto $V_j = \mathfrak{R}(A_{-j})$ is given by $A_{-j}(A^*_{-j}A_{-j})^{-1}A^*_{-j}A_j$. By the Pythagorean theorem,

$$\|A_j\|_2^2 - \operatorname{dist}(A_j, V_{-j})^2 = \|A_{-j}(A^*_{-j}A_{-j})^{-1}A^*_{-j}A_j\|_2^2 = A_j^* A_{-j}(A^*_{-j}A_{-j})^{-1}A^*_{-j}A_j.$$

On the other hand, by Schur's complement, for any invertible matrix $B \in \mathbb{C}^{m\times m}$ and partition $[n] = I \cup I^c$,

$$(B^{-1})_{I,I} = \left(B_{I,I} - B_{I,I^c}B^{-1}_{I^c,I^c}B_{I^c,I}\right)^{-1}.$$

We let $B = A^*A$ and $I = \{j\}$ for $j = 1, \cdots, m$ to obtain

$$((A^*A)^{-1})_{jj} = \left(A_j^*A_j - (A^*_{-j}A_j)^*(A_{-j}A^*_{-j})^{-1}(A^*_{-j}A_j)\right)^{-1} = \operatorname{dist}(A_j, V_{-j})^{-2}, \quad j = 1, \cdots, m.$$

The desired result then follows by taking the sum of the above over $j = 1, \cdots, m$. □

Lemma 5.16 (Tao-Vu)**.** *There exist $\gamma > 0$ and $\delta > 0$ such that for large enough $n \in \mathbb{N}$, any $1 \le j \le n$, any deterministic vector $v \in \mathbb{C}^n$ and any subspace H of $\mathbb{C}^n$ with $1 \le \dim H \le n - n^{1-\gamma}$, we have, denoting $Y := (X_{1j}, \cdots, X_{nj}) + v$,*

$$\mathbb{P}\left(\operatorname{dist}(Y, H) \le \frac{1}{2}\sqrt{n - \dim(H)}\right) \le e^{-n^\delta}.$$

Proof. Step I. We denote by H' the subspace of $\mathbb{C}^n$ spanned by H and vector v. Then $\dim H' \le \dim H + 1$, and $\operatorname{dist}(Y, H) \ge \operatorname{dist}(Y, H')$. Since the original dimension assumption is $\dim H \le n^{1-\gamma}$, adding 1 does not change the asymptotic "smallness" of the subspace. We may thus directly suppose without loss of generality that $v = 0$. Let $U \in \mathbb{C}^{n\times \dim(H)}$ be a matrix whose columns form an orthonormal basis of H. Then

$$\mathbb{E}\left[\mathrm{dist}(Y,H)^2\right] = \mathbb{E}\left[\|Y\|_2^2 - \|UU^*Y\|_2^2\right] = \mathbb{E}\left[YY^* - \mathrm{tr}(UU^*YY^*)\right] = n - \mathrm{tr}(UU^*) = n - \dim H. \tag{5.22}$$

Let $0 < \epsilon < 1/3$. By Markov's inequality we have $\mathbb{P}(|X_{kj}| \geq n^{\epsilon}) \leq n^{-2\epsilon}$. Using Hoeffding's inequality,

$$\begin{aligned}\mathbb{P}\left(\sum_{k=1}^{n} \mathbb{1}_{\{|X_{kj}|\leq n^{\epsilon}\}} < n - n^{1-\epsilon}\right) &\leq \exp\left(-\frac{2(\sum_{k=1}^{n}\mathbb{P}(|X_{kj}| \leq n^{\epsilon}) - (n-n^{1-\epsilon}))^2}{n}\right) \\ &\leq \exp\left(-2n^{1-2\epsilon}(1-n^{-\epsilon})^2\right) \leq \exp(-n^{1-2\epsilon}), \quad \text{for } n \gg 1. \end{aligned}\tag{5.23}$$

Step II. By the above result, there are at least $n - n^{1-\epsilon}$ entries in $(X_{kj})_{k=1}^{n}$ bounded by n^{ϵ} with high probability. By premutation invariance, we define event

$$E_m := \bigcap_{k=1}^{m} \{|X_{kj}| \leq n^{\epsilon}\}, \quad \text{with } m = \lceil n - n^{1-\epsilon} \rceil.$$

Since the bad event (5.23) has probability less than $O(\exp(-n^{1-2\epsilon}))$, it suffices to condition on E_m. Let $\mathscr{F}_m$ be the σ-algebra generated by $(X_{m+1,j}, \cdots, X_{nj})$. We let $\mathbb{E}_m$ be the expectation conditional on E_m and $\mathscr{F}_m$, i.e.

$$\mathbb{E}_m[\xi] = \frac{\mathbb{E}\left[\xi \mathbb{1}_{E_m}|\mathscr{F}_m\right]}{\mathbb{E}[\mathbb{1}_{E_m}|\mathscr{F}_m]}.$$

Let W be the subspace spanned by

$$H,\ u = (0, \cdots, 0, X_{m+1,j}, \cdots, X_{nj}),\ w = (\mathbb{E}_m[X_{1j}], \cdots, \mathbb{E}_m[X_{mj}], 0, \cdots, 0),$$

and let $Z = (X_{1j} - \lambda, \cdots, X_{1m} - \lambda, 0, \cdots, 0) = Y - u - w$, where $\lambda = \mathbb{E}[X_{1j}]$. Then $\dim W \leq \dim H + 2$, and $\mathrm{dist}(Y,H) \geq \mathrm{dist}(Y,W) = \mathrm{dist}(Z,W)$. Similar to our deduction in Step I, it suffices prove the result for $\mathrm{dist}(Z,W)$. We note that

$$\sigma^2 := \mathbb{E}_m\left[Z_1^2\right] = \frac{1}{\mathbb{E}\mathbb{1}_{\{|X_{ij}|\leq n^{\epsilon}\}}}\mathbb{E}\left[\left(X_{1j} - \frac{\mathbb{E}\left[X_{ij}\mathbb{1}_{\{|X_{ij}|\leq n^{\epsilon}\}}\right]}{\mathbb{E}\mathbb{1}_{\{|X_{ij}|\leq n^{\epsilon}\}}}\right)^2 \mathbb{1}_{\{|X_{ij}|\leq n^{\epsilon}\}}\right] = 1 - o(1).$$

Step III. We define $f : x \in D_{\epsilon}^{m} \mapsto \mathrm{dist}((x, 0, \cdots, 0), W)$, where $D_{\epsilon} = \{z \in \mathbb{C} : |z| \leq n^{\epsilon}\}$. Then f is a 1-Lipschitz function, and by Talagrand's concentration inequality,

$$\mathbb{P}_m\left(|\mathrm{dist}(Z,W) - M_m| \geq t\right) \leq 4\exp\left(-\frac{t^2}{32n^{2\epsilon}}\right), \tag{5.24}$$

where M_m is the median of $\mathrm{dist}(Y,W)$ under $\mathbb{E}_m$. By Fubini's theorem,

$$\mathbb{E}_m\left|\mathrm{dist}(Z,W) - M_m\right|^2 \leq 4\int_0^{\infty} 2t\exp\left(-\frac{t^2}{32n^{2\epsilon}}\right)dt = 128n^{2\epsilon}.$$

By the triangle inequality,

$$\sqrt{\mathbb{E}_m\left|\mathrm{dist}(Z,W)\right|^2} \leq \sqrt{\mathbb{E}_m\left|\mathrm{dist}(Z,W) - M_m\right|^2} + M_m \leq 8\sqrt{2}n^{\epsilon} + M_m.$$

On the other hand, similar to our calculation in (5.22),

$$\mathbb{E}_m |\mathrm{dist}(Z,W)|^2 \geq \sigma^2(m-\dim W) \geq \sigma^2(n-n^{1-\epsilon}-\dim H-2)$$

Therefore

$$M_m \geq \sqrt{\sigma^2\left(n-n^{1-\epsilon}-\dim H-2\right)} - 8\sqrt{2}n^{\epsilon}$$

We select $\gamma \in (0,\epsilon)$, with $0 \leq \dim H \leq n-n^{1-\gamma}$. Then $n-\dim H \gg n^{1-\epsilon}$ as $n\to\infty$, and there exists a constant $\frac{1}{2} < c < 1$ such that $M_m \geq c\sqrt{n-\dim H}$ for $n \gg 1$. We take $t = (c-\frac{1}{2})\sqrt{n-\dim H}$ in (5.24) to obtain

$$\mathbb{P}_m\left(\mathrm{dist}(Z,W) \leq \frac{1}{2}\sqrt{n-\dim H}\right) \leq 4\exp\left(-\frac{(c-\frac{1}{2})^2(n-\dim H)}{32n^{2\epsilon}}\right).$$

The exponent behaves asymptotically like $O(n^{1-\gamma-2\epsilon})$, with $1-\gamma-2\epsilon>0$ since we choose $0<\gamma<\epsilon<1/3$. Therefore, there exists $\delta>0$ such that the probability is bounded by $\exp(n^{-\delta})$. □

Lemma 5.17 (Count of small singular values)**.** *There exist absolute constants $c_0>0$ and $0<\gamma<1$ such that for any fixed sequence $M_n \in \mathbb{C}^{n\times n}$, almost surely, for large enough n and all indices $n^{1-\gamma} \leq j \leq n-1$,*

$$\sigma_{n-j}\left(n^{-1/2}X_n + M_n\right) \geq c_0\frac{j}{n}.$$

Proof. For simplicity we write $\sigma_{n-j} = \sigma_{n-j}\left(n^{-1/2}X_n+M_n\right)$. Up to increasing γ, it is enough to prove the statement for all $2n^{1-\gamma} \leq j \leq n-1$ for some $\gamma\in(0,1)$ to be chosen later.

We fix $2n^{1-\gamma} \leq j \leq n-1$, and let Y_n be the matrix formed by the first $m := n-\lceil j/2\rceil$ columns of $X_n + \sqrt{n}M_n$. Let $\tau_1 \geq \cdots \geq \tau_m$ be the singular values of Y_n. By Courant-Fisher max-min principle,

$$n^{-1/2}\tau_k = n^{-1/2}\max_{\dim V=k}\min_{u\in V\cap S^{n-1}}\|u^\top Y_n\|_2 \leq \sigma_k, \quad k=1,\cdots,m.$$

By Lemma 5.15, if Y_n is of full rank, then

$$\sum_{n=N}^{\infty}\tau_1^{-2}+\cdots+\tau_{n-\lceil j/2\rceil}^{-2} = \mathrm{dist}(Y_{n,1},H_{n,1})^{-2}+\cdots+\mathrm{dist}(Y_{n,n-\lceil j/2\rceil},H_{n,n-\lceil j/2\rceil})^{-2},$$

where $Y_{n,j}$ is the j-th column of Y and $H_{n,j}$ is the subspace spanned by the remaining columns of Y_n. In particular,

$$\frac{j}{n}\sigma_{n-j}^{-2} \leq j\tau_{n-j}^{-2} \leq \sum_{k=n-j}^{n-\lceil j/2\rceil}\tau_k^{-2} \leq \sum_{k=1}^{n-\lceil j/2\rceil}\mathrm{dist}(Y_{n,k},H_{n,k})^{-2}. \tag{5.25}$$

Since H_k is independent of Y_k and $\dim H_k \leq n-j/2 \leq n-n^{1-\gamma}$, for the choice of $\gamma,\delta>0$ given in Lemma 5.16, there exists some large enough $N\in\mathbb{N}$ such that

$$\sum_{n=N}^{\infty}\mathbb{P}\left(\bigcup_{j=2n^{1-\gamma}}^{n-1}\bigcup_{k=1}^{n-\lceil j/2\rceil}\left\{\mathrm{dist}(Y_{n,k},H_{n,k}) \leq \frac{\sqrt{j}}{2\sqrt{2}}\right\}\right) \leq \sum_{n=N}^{\infty}n^2\exp(-n^{\delta}) < \infty.$$

Consequently, by the Borel-Cantelli lemma, almost surely, for large enough n, all $2n^{1-\gamma} \leq j \leq n-1$ and all $1 \leq k \leq n-\lceil j/2\rceil$,

$$\mathrm{dist}(Y_{n,k},H_{n,k}) \geq \frac{\sqrt{j}}{2\sqrt{2}}.$$

Consequently, Y_n is a.s. of full column rank, and by (5.25), we have $\sigma_{n-j}^2 \geq j^2/(8n^2)$. Putting all together, we obtain the desired result with $c_0 = 2\sqrt{2}$. □

5.3.3 The Smallest Singular Value

Lemma 5.18 (Rudelson-Vershynin)**.** *Let $A \in \mathbb{C}^{n\times n}$ with columns $A_1, \cdots, A_n \in \mathbb{C}^n$, and define subspaces $V_k = \operatorname{span}\{A_j : j \in [n], j \neq k\}$ for every $k \in [n]$. Then*

$$\frac{1}{\sqrt{n}} \min_{1\le j\le n} \operatorname{dist}(A_j, V_j) \le \sigma_n(A) \le \min_{1\le j\le n} \operatorname{dist}(A_j, V_j).$$

Proof. For every $x \in \mathbb{C}^n$, we have $Ax = x_1A_1 + \cdots + x_nA_n$, and by the triangle inequality,

$$\|Ax\|_2 \ge \operatorname{dist}(Ax, V_j) = \min_{v\in V_j} \|Ax - v\|_2 = \min_{v\in V_j} \|x_jA_j - v\|_2 = |x_j| \operatorname{dist}(A_j, V_j).$$

If $\|x\|_2 = 1$, there exists $j \in [n]$ such that $|x_j| \ge n^{-1/2}$, and

$$\sigma_n(A) = \min_{1\le j\le n} \|Ax\|_2 \ge \min_{1\le j\le n} |x_j| \operatorname{dist}(A_j, V_j) \ge \frac{1}{\sqrt{n}} \min_{1\le j\le n} \operatorname{dist}(A_j, V_j).$$

On the other hand, for every $j \in [n]$, by Gram-Schmidt, there exists $y \in \mathbb{C}^n$ with $y_j = 1$ such that

$$\operatorname{dist}(A_j, V_j) = \|y_1A_1 + \cdots + y_nA_n\| = \|Ay\|_2 \ge \sigma_n(A)\,\|y\|_2 \ge \sigma_n(A).$$

Then we finish the proof. □

Lemma 5.19 (Tao-Vu)**.** *For any $a, q > 0$, there exists a constant $b > 0$ depending on a and q such that for all large enough $n \in \mathbb{N}$ and deterministic $M \in \mathbb{C}^{n\times n}$ with $\sigma_1(M) \le n^q$,*

$$\mathbb{P}\left(\sigma_n(X_n + M) \le n^{-b}\right) \le n^{-a}.$$

In particular, there exists $b > 0$ depending on q only such that a.s. for large enough n,

$$\sigma_n(X_n + M) \ge n^{-b}.$$

Proof with bounded density assumption. Let $A_1, \cdots, A_n$ be the rows of $X_n + M$, and $V_k = \operatorname{span}\{A_j : j \in [n], j \neq k\}$ for $k \in [n]$. Then

$$\min_{1\le j\le n} \operatorname{dist}(A_j, V_j) \le \sqrt{n}\sigma_n(X_n + M).$$

Using a union bound, we have

$$\mathbb{P}\left(\sqrt{n}\sigma_n(X_n + M) \le t\right) \le n \max_{1\le j\le n} \mathbb{P}\left(\operatorname{dist}(A_j, V_j) \le t\right), \quad t \ge 0.$$

Now we fix $j \in [n]$, and let Y_j be a unit vector orthogonal to V_j. We may fix our choice of Y_j by normalizing the leftmost nonzero column of the projection matrix $I - A_{-j}A_{-j}^\dagger$ onto the subspace V_j, hence Y_j depends only on the columns A_{-j} and is independent of A_j. Furthermore, by Cauchy-Schwarz inequality,

$$\operatorname{dist}(A_j, V_j) = \left\|(I - A_{-j}A_{-j}^\dagger)A_j\right\|_2 \cdot \|Y_i\|_2 \ge |A_j \cdot Y_j|\,.$$

Let ν_j be the distribution of V_j on the sphere S^{n-1} of $\mathbb{C}^n$. Then

$$\mathbb{P}\left(\operatorname{dist}(A_j, V_j) \le t\right) \le \mathbb{P}\left(|A_j \cdot Y_j| \le t\right) = \int_{S^{n-1}} \mathbb{P}\left(|A_j \cdot y| \le t\right)\, d\nu_j(y).$$

We assume X_{11} has a bounded density φ on $\mathbb{C}$. For any $y \in S^{n-1}$, since $\|y\|_2 = 1$, there exists $k \in [n]$ such that $|y_k| \geq n^{-1/2}$, hence the density of $\overline{y}_k A_{kj}$ is bounded by $\sqrt{n}\|\varphi\|_\infty$. Since $A_j \cdot y = \overline{y}_1 A_{1j} + \cdots + \overline{y}_n A_{nj}$ is a sum of independent random variables containing $\overline{y}_j A_{kj}$, by a basic property of convolutions of probability measures, $A_j \cdot y$ also has a density bounded by $\sqrt{n}\|\varphi\|_\infty$. Hence

$$\mathbb{P}\left(|A_j \cdot y| \leq t\right) = \int_{\mathbb{C}} \mathbb{1}_{\{|s| \leq t\}} \varphi_j(s)\, ds \leq \sqrt{n}\pi t^2 \|\varphi\|_\infty.$$

Therefore, for every $r > 0$, we choose $t = n^{-r}$ to obtain

$$\mathbb{P}\left(\sigma_n(X_n + M) \leq n^{-r-\frac{1}{2}}\right) \leq n^{\frac{3}{2}-2r}\pi\|\varphi\|_\infty.$$

Then we may choose r with $\frac{3}{2} - 2r < -a$ and set $b = r + \frac{1}{2}$. For the second statement, we take $a > 1$ and apply Borel-Cantelli lemma to complete the proof. □

5.4 Prove the Circular Law

Proof of Theorem 5.2. We prove the circular law by verifying the two conditions in Theorem 5.9.

Step I. We fix $z \in \mathbb{C}$. For $p < 2$, by Hölder's inequality,

$$\left|\int_0^\infty s^p \, d\nu_{n^{-1/2}X_n - zI}(s)\right|^{\frac{1}{p}} \leq \left|\int_0^\infty s^2 \, d\nu_{n^{-1/2}X_n - zI}(s)\right|^{\frac{1}{2}}.$$

By Weyl's inequality and the strong law of large numbers,

$$\begin{aligned}\int_0^\infty s^2 \, d\nu_{n^{-1/2}X_n - zI}(s) &= \frac{1}{n}\sum_{j=1}^n \sigma_j\left(\frac{X_n}{\sqrt{n}} - zI\right)^2 \leq \frac{2}{n}\sum_{j=1}^n \sigma_j\left(\frac{X_n}{\sqrt{n}}\right)^2 + 2|z|^2 \\ &= \frac{2}{n^2}\operatorname{tr}(X_n X_n^*) + 2|z|^2 = \frac{2}{n^2}\sum_{i,j=1}^n |\xi_{ij}|^2 + 2|z|^2 \to 2\left(1 + |z|^2\right), \quad \text{almost surely.}\end{aligned}$$

Combining the above two displays, we have

$$\limsup_{n\to\infty} \int_0^\infty s^p \, d\nu_{n^{-1/2}X_n - zI}(s) < \infty \;\text{ a.s.,} \quad p < 2.$$

Step II. For notation simplicity we write $\sigma_j = \sigma_j(n^{-1/2}X_n - zI)$. We take the constant $c_0 > 0$ and fix $M_n = -zI_n$ in Lemma 5.17, choose $M = z\sqrt{n}I$, $q > 1/2$ and take the constant $b > 0$ in Lemma 5.19. Then almost surely, for large enough n, we have

$$\frac{1}{n}\sum_{j=1}^n \sigma_j^{-p} \leq \frac{1}{n}\sum_{j=1}^{n-\lfloor n^{1-\gamma}\rfloor} \sigma_j^{-p} + \sum_{j=n-\lfloor n^{1-\gamma}\rfloor+1}^n \sigma_n^{-p} \leq \frac{1}{n}\sum_{j=1}^n \left(\frac{c_0 j}{n}\right)^{-p} + n^{-\gamma+bp} \leq \int_0^1 s^{-p}\, ds + n^{-\gamma+bp}.$$

Note that $\int_0^1 s^{-p}\, ds$ converges when $0 < p < 1$. Then we choose $0 < p < \min\{\gamma/b, 1\}$, which satisfies

$$\limsup_{n\to\infty} \int_0^\infty s^{-p} \, d\nu_{n^{-1/2}X_n - zI}(s) < \infty.$$

Step III. By Theorem 5.11, $\nu_{n^{-1/2}X_n - zI} \to \nu_z$ a.s. for all $z \in \mathbb{C}$, where ν_z is the pushforward of μ_z under the square root $\mathbb{R}_+ \to \mathbb{R}_+ : s \mapsto \sqrt{s}$.

By Theorem 5.9, the empirical spectral distribution $\mu_{n^{-1/2}X_n - zI}$ converges a.s. to a probability measure

$\mu \in \mathcal{P}_\infty(\mathbb{C})$, with logarithm potential

$$U_\mu(z) = -\int_0^\infty \log s \, d\nu_z(s), \quad z \in \mathbb{C}.$$

Furthermore, since ν_z depends only on z, the measure μ does not depend on the distribution of X_{11}. □

Note that

$$\mathbb{E}\left[\int_{\mathbb{C}} f(z)\, d\mu_{n^{-1/2}X_n}(z)\right] = \mathbb{E}\left[\sum_{j=1}^{n} f\left(\frac{\lambda_j(X_n)}{\sqrt{n}}\right)\right]$$

The 1-point correlation function is given by

$$\varphi_{n,1}(z) = \frac{1}{n\pi} e^{-|z|^2} \sum_{k=0}^{n-1} \frac{|z|^{2k}}{k!},$$

and

$$n\varphi_{n,1}(\sqrt{n}z) = \frac{1}{\pi} e^{-n|z|^2} \sum_{k=0}^{n-1} \frac{(n|z|^2)^k}{k!} = \frac{1}{\pi} \mathbb{P}\left(S_n \le n-1\right),$$

where S_n is a Poisson random variable of rate $n|z|^2$. By Poisson's central limit theorem,

$$\frac{S_n - n|z|^2}{\sqrt{n}|z|} \to \mathcal{N}_{\mathbb{R}}(0,1) \quad \text{weakly}.$$

Then

$$\mathbb{P}\left(S_n \le n-1\right) = \mathbb{P}\left(\frac{S_n - n|z|^2}{\sqrt{n}|z|} \le \frac{n(1-|z|^2)-1}{\sqrt{n}|z|}\right) \to \begin{cases} \mathbb{P}(Z \le \infty) = 1, & |z| < 1, \\ \mathbb{P}(Z \le 0) = 1/2, & |z| = 1, \\ \mathbb{P}(Z \le -\infty) = 0, & |z| > 1. \end{cases}$$

Hence $n\varphi_{n,1}(\sqrt{n}z) \to \pi^{-1}\mathbb{1}_{|z|\le 1}$ almost everywhere on $\mathbb{C}$.