A Note on the Wigner Semicircle Law

She Yuting

Supervised by: Prof. Yao Jianfeng (CUHKSZ)

June 2024

1 Introduction

This note intends to give a proof to Wigner's Semicircle Law on the Wigner matrices. We first explore some important properties of the Wigner matrix, then examine the behaviour of the empirical spectral distribution of a sequence of Hermitian matrix ensembles

Definition 1.1.1 Wigner Matrix A random matrix ensemble $M_n = (\mathcal{C}_n)_{1\leq i,j\leq n}$ is called a Hermitian Wigner matrix ensemble, if M_n is Hermitian ($M_n = M_n^*$, where * denotes the transpose conjugate of M_n), in which the upper triangular coefficients $\mathcal{C}_{ij}$ being jointly independent with

(a) The diagonal entries $\mathcal{C}_{i,i}$ being real i.i.d with mean 0 and

(b) The strictly upper triangular entries $\mathcal{C}_{i,j}$, i < j being complex with mean 0 and $\mathbb{E}|\mathcal{C}_{i,j}|^2 = 1$.

One important observation is that all the eigenvalues for a Hermitian matrix are real so it will be sufficient to consider the real eigenvalues of a Wigner matrix M_n. Suppose there are n random eigenvalues which we will denote by

$$\lambda_1(M_n) \geq \lambda_2(M_n) \geq ... \geq \lambda_n(M_n)$$

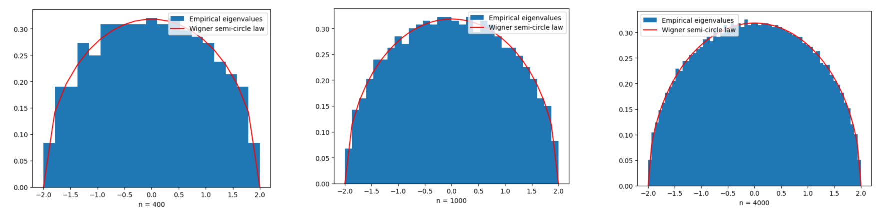
In fact (which will be shown later): these are continuous functions of M_n hence they are random variables themselves. In order to observe the bounded property of the eigenvalues of Wigner matrices, we introduce:

Definition 1.1.2 The Empirical Spectral Distribution Given any $n \times n$ Hermitian matrix M, we can define the empirical Spectral Distribution (ESD) of M as the normalised counting measure of all the eigenvalues.

$$\mu_{M_n} := \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(M_n)}$$

where $\lambda_1(M_n) \geq \lambda_2(M_n) \geq ... \geq \lambda_n(M_n)$, with $\delta_{\lambda_j(X_n)}(x)$ being the indicator function $\mathbf{1}_{\lambda_j(X_n) \leq x}$.

This μ_{M_n} is a random discrete probability measure which puts $\frac{1}{n}$ mass to each (random) eigenvalue. Now consider the behaviour of ESD of a sequence of Hermitian matrix ensembles M_n as $n \to \infty$.
By plotting the histogram of the Hermitian matrix with mean 0 and variance 1, we observe that:



n=400 n=1000 n=4000

Figure 1: Histogram of the eigenvalue distribution of Hermitian matrix with mean 0 and unit variance

We see that as the size of n increases, a semicircle pattern is observed. The semi-circle law is then, to prove that the distribution of ESD sequence, after proper normalization, indeed converges to a deterministic distribution.

2 Convergence of Measures

In order to prove that the distribution of ESD converges, we need to firstly define the convergence concept for random measures.

Definition 2.3.1 If $\{\mu_n\}_{n\geq 1}$ is a sequence of random measures on $(\mathcal{R}, \mathcal{B}(\mathcal{R}))$, we say $\{\mu_n\}_{n\geq 1}$ converges *weakly* to a deterministic measure μ if for every compactly supported (bounded) continuous function f,

$$\lim_{n\to\infty} \int_{\mathcal{R}} f d\mu_n = \int_{\mathcal{R}} f d\mu$$

Since the sequence $\{\mu_n\}_{n\geq 1}$ is itself random, if

$$P(\lim_{n\to\infty} \int_{\mathcal{R}} f d\mu_n = \int_{\mathcal{R}} f d\mu) = 1$$

the sequence is said to converge ***almost surely*** to a probability measure μ.

2.1 Normalization of M_n

Intuitively we can understand the scaling factor $\frac{1}{\sqrt{n}}$ by considering convergence of the moments of the Hermitian matrix M_n.

In general, if the first and the second moment of a random variable converges, then the expectation and the variance of a random variable converges.

In particular, we look at the first and second moment of M_n. The following are some useful results from linear algebra:

Theorem 2.3.2 For a Hermitian matrix A,

$$\text{A} = U^*\Lambda U = \sum_{j=1}^{N} \lambda_j u_j u_j^*$$

where $\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$ and U is a unitary vector with $\{u_i\}$ being the orthonormalised eigenvector for each corresponding engenvalue $\{\lambda_i\}$.

Proof: This result follows immediately from the Principal Axis Theorem. Note that all Hermitian matrices are normal matrices.

Proposition 2.3.3 Let $A = (a_{i,j})_{n\times n}$ and $B = (b_{i,j})_{n\times n}$ be real or complex square matrices. Then

$$\text{Tr(AB)} = \text{Tr(BA)} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}b_{j,i}.$$

This is true because the dimension of A, B are fixed, so the sums are commutative to each other.

It then follows that given matrix $M_n = U^*\Lambda U$, we have

$$\text{Tr}(M_n) = \text{Tr}(U^*\Lambda U) = \text{Tr}(\Lambda U^* U) = \text{Tr}(\Lambda) = \sum_{i=1}^{n} \lambda_i$$

Corollary 2.3.4 Assume that A is diagonalizable. Then for any integer k≥0,

$$\text{Tr}(A^k) = \sum_{i=1}^{n} \lambda^k$$

with a simple induction on Proposition 1.3.5.

The first moment of M_n is thus

$$\frac{1}{n}\sum_{i=1}^{n}\lambda_i = Tr(M_n) = \sum_{i=1}^{n}\mathcal{C}_{i,i}$$

and the second moment of M_n is

$$\frac{1}{n}\sum_{i=1}^{n}\lambda_i^2 = Tr(M_n^2) = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{C}_{i,j}^2 = \frac{1}{n}(\sum_{i=1}^{n}\mathcal{C}_{i,i}^2 + 2\sum_{i<j}\mathcal{C}_{i,j}^2)$$

And it is easy to see that the first and second moments of M_n converges:
It is clear that the first moment converges to 0 by the strong law of large numbers.
For the second moment, since $\frac{1}{n}\sum_{i=1}^{n}\mathcal{C}_{i,i}^2 \to 0$ as $n \to \infty$, by the law of large numbers (as $\mathbb{E}[\mathcal{C}_{i,i}^2] = 1$ by definition), we are then interested in the convergence of the term $\frac{2}{n}\sum_{i<j}\mathcal{C}_{i,j}^2$.

Similarly, by the law of large numbers we know that:

$$\lim_{n\to\infty}\frac{1}{\frac{n(n-1)}{2}}\sum_{i<j}\mathcal{C}_{i,j}^2 = l \text{ for some } l \in \mathbb{R}^+, l < \infty$$

therefore as $n \to \infty$,

$$\lim_{n\to\infty}\frac{2}{n}\sum_{i<j}\mathcal{C}_{i,j}^2 = \lim_{n\to\infty}\frac{2}{n}\frac{n(n-1)}{2}\frac{1}{\frac{n(n-1)}{2}}\sum_{i<j}\mathcal{C}_{i,j}^2 = nl$$

which is of O(n). This suggests that in order to see a meaningful limit, we need to scale the eigenvalues (or the matrix) by $\frac{1}{\sqrt{n}}$.
Although this argument is far from being rigorous, it provides an insight to how we can understand the scaling factor. Since a rigorous proof of this bound involves discussions of the measure concentration, we defer the proof to Annex A after we explore more properties of the measure concentrations (i.e the Talagrand's inequality).

3 Eigenvalue Distribution of Wigner Matrices

3.1 The Semicircle Law

Definition 3.1.1 Let M_n be the top left $n \times n$ minors of an infinite Wigner matrix $(\mathcal{C})_{i,j\geq 1}$. Then the ESDs $\mu_{\frac{1}{\sqrt{n}}M_n}$ converge almost surely (and hence also in probability and expectation) to the Wigner semicircular distribution:

$$\mu_{sc} := \frac{1}{2\pi}\sqrt{4-x^2}$$

There are several ways to establish the semicircular law, including the moment method, Stieltjes transform method, free probability method as well as the Dyson Brownian motion method. Stieltjes transform is considered one of the

most powerful and accurate tools in dealing with the ESD of random Hermitian matrices.
The basic starting point of both moments and the Stieltjes transform method, however, starts from the observation that the moments of the ESD can be written as normalised traces of powers of M_n:

$$\int_{\mathbb{R}} x^k d\mu_{\frac{1}{\sqrt{n}}M_n}(x) = \frac{1}{n}Tr(\frac{1}{\sqrt{n}}M_n)^k.$$

In particular, on taking expectations, we have

$$\int_{\mathbb{R}} x^k d\mathbf{E}[\mu_{\frac{1}{\sqrt{n}}M_n}(x)] = \mathbf{E}\left[\frac{1}{n}Tr(\frac{1}{\sqrt{n}}M_n)^k\right].$$

3.2 The Stieltjes Transform Method

When dealing with random Hermitian matrices, we use Stieltjes transform to characterize the (normalized) ESD proceeding from the identity:

$$\int_{\mathcal{R}} \frac{1}{x-z} d\mu_{\frac{1}{\sqrt{n}}M_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i(\frac{1}{\sqrt{n}}M_n)-z} d\mu_{\frac{1}{\sqrt{n}}M_n}(x) = \frac{1}{n}Tr(\frac{1}{\sqrt{n}}M_n - zI)^{-1}$$

defined for complex number z outside of the support of μ_n, $z \in \mathbb{C}^+$.

We refer to the expression on the left hand side as the **Stieltjes transform** of $\mu_{\frac{1}{\sqrt{n}}M_n}$, and denote it by $s_{\mu_{\frac{1}{\sqrt{n}}M_n}}$ or as s_n for short.

The expression $(\frac{1}{\sqrt{n}}M_n - zI)^{-1}$ is the **normalised resolvent** of M_n, and plays an important role in the spectral theory of that matrix.

Theorem 3.2.1 Stieltjes Transform For any probability measure μ defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the Stieltjes transform of μ is

$$s_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\mu(x)$$

for any $z \notin supp(\mu)$.

Proposition 3.2.2 Let μ be a probability measure on $\mathbb{R}$ and S_μ be its Stieltjes transform, then we have the following:

(i) $S_\mu(z) \in \mathbb{C}^+$, i.e. the Stieltjes transform is well-defined on $\mathbb{C}^+$ and

(ii) S_μ is analytic on $\mathbb{C}^+$.

where $\mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Im(z)} > 0\}$.

Proof. (i) We observe that by writing z := a + ib, $\mathrm{Im}\frac{1}{x-z} = \frac{b}{(x-a)^2+b^2} > 0$ as the reciprocal of a complex number z is given by $\frac{\bar{z}}{|z|^2}$. Therefore we have $S_\mu(z) \in \mathbb{C}^+$.
(ii) A function f(z) is analytic if it has a complex derivative $f'(z)$. Since $\frac{1}{x-z}$ is analytic on $\mathbb{C}^+$, its integral over the real line is also analytic on $\mathbb{C}^+$.

$\square$

The imaginary part of Stieltjes transform is of typical interest. It provides a meaning bound for $s_\mu(z)$. By applying conjugations on the Stieltjes transform $s_\mu(z)$ we obtain the symmetry:

$$\overline{s_\mu(z)} := s_\mu(\bar{z}),$$

From the trivial bound

$$\mid \tfrac{1}{x-z} \mid \leq \tfrac{1}{|Im(z)|}$$

One has the pointwise bound

$$\mid s_\mu(z) \mid \leq \tfrac{1}{|Im(z)|}.$$

Actually, we can recover μ from S_μ using the Stieltjes inversion formula:

Theorem 3.2.3 Stieltjes Inversion Formula For any continuous points $a < b$ of probability measure μ, we have

$$\mu([a,b]) = \lim_{\epsilon \to 0^+} \tfrac{1}{\pi} \int_a^b Im(s_\mu(x+i\epsilon))dx.$$

Proof: We have

$$\begin{aligned}
\frac{1}{\pi}\int_a^b Im(s_\mu(x+i\epsilon))dx &= \frac{1}{\pi}\int_a^b \int_{\mathbb{R}} \frac{\epsilon d_\mu(y)}{(x-y)^2+\epsilon^2} dx \\
&= \frac{1}{\pi}\int_{\mathbb{R}} \int_a^b \frac{\epsilon d_\mu(y)}{(x-y)^2+\epsilon^2} d\mu(y) \text{ (Fubini-Tonelli)} \\
&= \int_{\mathbb{R}} \frac{1}{\pi}\left[\arctan\left(\frac{b-y}{\epsilon}\right) - \arctan\left(\frac{a-y}{\epsilon}\right)\right] d\mu(y).
\end{aligned}$$

Let $\epsilon \to 0^+$, from the dominating convergence theorem, the right-hand side tends to $\mu([a,b])$.

One can also express the imaginary part of the Stieltjes transform as a convolution:

$$\text{Im}(s_\mu(a+ib)) = \pi\mu * P_b(a)$$

where $P_b(a)$ is the Poisson kernel

$$P_b(a) := \tfrac{1}{\pi}\tfrac{b}{x^2+b^2}$$

The kernels form a family of approximations to the identity, and thus $\mu * P_b$ converges in the vague topology to μ. (i.e $\lim\limits_{b \to 0^+}(\mu * P_b) = \mu * \delta_0 = \mu$.)

Thus we see that

$$Im(s_\mu(a+i\epsilon)) \rightharpoonup \pi\mu$$

as $\epsilon \to 0^+$, or equivalently suppose that μ has density ρ at $x \in \mathbb{R}$, we have

$$\rho(x) = \lim_{\epsilon \to 0^+} Im(s_\mu(x + i\epsilon)) = \lim_{\epsilon \to 0^+} \tfrac{s_\mu(x+i\epsilon) - s_\mu(x-i\epsilon)}{2\pi i} = d\mu(x)$$

Thus we see that a probability measure μ can be recovered in terms of limiting behaviour of Stieltjes transform on the real axis.

Theorem 3.2.4 Stieltjes Continuity Theorem, deterministic Let $\{\mu_n\}_{n\geq 1}$ be a sequence of *deterministic* probability measures on the real line, and let μ be a deterministic probability measure. Then μ_n converges in the vague topology to a probability measure μ if and only if

$$\lim_{n\to\infty} s_{\mu_n}(z) = s_\mu(z), \forall z \in \mathbb{C}^+.$$

Theorem 3.2.5 Stieltjes Continuity Theorem, random Let $\{\mu_n\}_{n\geq 1}$ be a sequence of *random* probability measures on the real line, and let μ be a deterministic probability measure. Then μ_n converges in the vague topology to a probability measure if and only if

$$s_{\mu_n}(z) \xrightarrow{a.s.} s_\mu(z), \forall z \in \mathbb{C}^+.$$

We will proceed the Stieltjes transform method in the following 2 steps:

1. For any fixed $z \in \mathbb{C}^+$, $s_n(z) - \mathbb{E}[s_n(z)] \xrightarrow{a.s.} 0$
2. For any fixed z$\in \mathbb{C}^+, \mathbb{E}[s_n(z)] \to s_{sc}(z)$, the Stieltjes transform of the semicircle distribution.

4 Measure Concentration of $s_n(z)$

In this section, we show that for any fixed $z \in \mathbb{C}^+$, $s_n(z) - \mathbb{E}[s_n(z)] \xrightarrow{a.s.} 0$.
The main idea is to compare the transform $s_n(z)$ of the $n \times n$ matrix M_n with the transform $s_{n-1}(z)$ of the top left $n-1 \times n-1$ minor M_{n-1}. It is ideal to have (and will be proven later) that an extra row or column has only a limited amount of influence on the Stieltjes transform $s_n(z)$, then a standard measure concentration result can be applied to control the deviation of $s_n(z)$ from $\mathbb{E}[s_n(z)]$.

We start with an important relationship between the eigenvalues of the two matrices.

4.1 Bounded Total Variation of $s_n(z)$

4.1.1 Cauchy Interlacing Law For any $n \times n$ Hermitian matrix A_n with top left $n-1 \times n-1$ minor A_{n-1}, we have:
$\lambda_1(A_n) \geq \lambda_1(A_{n-1}) \geq \lambda_2(A_n) \geq \lambda_2(A_{n-1})...\lambda_{n-1}(A_n) \geq \lambda_{n-1}(A_{n-1}) \geq \lambda_n(A_n)$.

Proof. WLOG, we can reorder the columns of A_n so that we always have $\lambda_n \leq \lambda_{n-1} \leq ... \leq \lambda_1$. Note that matrix A_n is of the form

$$A_n = \begin{pmatrix} A_{n-1} & y^* \\ y & a_{nn} \end{pmatrix}.$$

where * signifies the conjugate transpose of a matrix. Let D $= diag(\mu_1, \mu_2, ..., \mu_{n-1})$, where $\mu_i := \lambda_i(A_{n-1})$. Then since A_{n-1} is also Hermitian, there exists a unitary matrix U of order $n-1$ such that $U^* A_{n-1} U = D$. Let $U^* y = z = (z_1, z_2, ..., z_{n-1})^T$.
Let

$$\mathrm{V} = \begin{pmatrix} U & 0^* \\ 0 & 1 \end{pmatrix},$$

in which 0 denotes the zero vector. Then V is a unitary matrix and

$$V^* A V = \begin{pmatrix} D & z^* \\ z & a_{nn} \end{pmatrix}.$$

Let $f(x) = det(xI - A) = det(xI - V^* AV)$, where I denotes the identity matrix, expanding $det(xI - V^* AV)$ along the last row, we get:

$$f(x) = (x-a)(x-\mu_1)...(x-\mu_{n-1}) - \sum_{i=1}^{n-1} f_i(x).$$

where $f_i(x) = \left| z_i^2 \right| (x-\mu_1)...(x-\mu_{i-1})(x-\mu_{i+1})...(x-\mu_{n-1})$ for $i = 1, 2, ..., n-1$. Note that $f_i(\mu_j) = 0$ if $j \neq i$ and that

$$f_i(\mu_i) = \begin{cases} > 0 & \text{if i is even} \\ < 0 & \text{if i is odd} \end{cases}$$

Hence that

$$f(\mu_i) = \begin{cases} < 0 & \text{if i is even} \\ > 0 & \text{if i is odd} \end{cases}$$

for $i = 1, 2, ..., n-1$. Since $f(x)$ is a polynomial of degree n with positive leading coefficient, the intermediate value theorem ensures the existence of n roots $\lambda_1, \lambda_2, ..., \lambda_n$ of the equation $f(x) = 0$ such that $\lambda_n \leq \mu_{n-1} \leq \lambda_{n-1} \leq ... \leq \lambda_2 \leq \mu_1 \leq \lambda_1$.

□

Now for $z = a+ib$, $b > 0$, let $f(x) = Im\left(\frac{1}{x-z}\right) = \frac{b}{(x-a)^2+b^2}$, then $0 < f(x) \leq \frac{1}{b}$. The main idea here is to compare the transform $s_n(z)$ of the $n \times n$ matrix M_n and $s_{n-1}(z)$ of the $n-1 \times n-1$ minor M_{n-1}. Together with the fact that f is decreasing, the function has bounded total variation:
for any $a_0 < ... < a_m$,

$$\sum_{k=1}^{m} |f(a_k) - f(a_{k-1})| = |f(a_0) - f(a_n)| \leq \frac{2}{b}$$

Then

$$\begin{aligned}
&\left|\sum_{j=1}^{n} \frac{b}{\left(\frac{\lambda_j(M_n)}{\sqrt{n}} - a\right)^2 + b^2} - \sum_{j=1}^{n-1} \frac{b}{\left(\frac{\lambda_j(M_{n-1})}{\sqrt{n}} - a\right)^2 + b^2}\right| \\
=&\left|\sum_{j=1}^{n} f\left(\frac{\lambda_j(M_n)}{\sqrt{n}}\right) - \sum_{j=1}^{n-1} f\left(\frac{\lambda_j(M_{n-1})}{\sqrt{n}}\right)\right| \\
\leq&\sum_{j=1}^{n-1} \left|f\left(\frac{\lambda_j(M_n)}{\sqrt{n}}\right) - f\left(\frac{\lambda_j(M_{n-1})}{\sqrt{n}}\right)\right| + \left|f\left(\frac{\lambda_n(M_n)}{\sqrt{n}}\right)\right| \\
\leq&\frac{3}{b}
\end{aligned}$$

Hence

$$\left|nIm(s_n(a+ib)) - \sqrt{n(n-1)}Im(s_{n-1}\left(\frac{\sqrt{n}}{\sqrt{n-1}}(a+ib)\right)\right| \leq \frac{3}{b}$$

Then

$$|Im(s_n(a+ib)) - Im(s_{n-1}(a+ib))| = O\left(\frac{1}{n}\right)$$

Similarly, by writing $g(x) = Re\left(\frac{1}{x-z}\right) = \frac{x-a}{(x-a)^2+b^2}$, we have:

$$|Re(s_n(a+ib)) - Re(s_{n-1}(a+ib))| = O\left(\frac{1}{n}\right)$$

Hence

$$|s_n(a+ib) - s_{n-1}(a+ib)| = O\left(\frac{1}{n}\right)$$

Observe that while $s_n(z)$ is dependent on the $n \times n$ matrix M_n, $s_{n-1}(z)$ is only dependent on $n-1 \times n-1$ minor of that matrix. In particular this suggests independence of the n-th row and column of M_n. That suggests the entire row and column have limited influence on the Stielejes transform $s_n(z)$. And by permuting the rows and columns, we know that any row or column of M_n influence $s_n(z)$ at most $O\left(\frac{1}{n}\right)$.

By treating $s_n(z)$ as a n-variate function of the rows of M_n, since all the rows are independent, the following result on measure concentration is applicable.

4.2 Convergence of $s_n(z)$

Before proving the convergence of $s_n(z)$, we prove some useful theorems that are useful in the proof later.

Theorem 4.2.1 Chernoff Bounds If X is a random variable, then for any $a \in \mathbb{R}$, we can write

$$P(X \geq a) = P(e^{sX} \geq e^{sa}), \text{ for } s > 0,$$
$$P(X \leq a) = P(e^{sX} \geq e^{sa}), \text{ for } s < 0.$$

Proof. As $e^{sX} > 0$ for all $s \in \mathbb{R}$, for some $s > 0$,

$$\begin{aligned} P(X \geq a) &= P(e^{sX} \geq e^{sa}) \\ &\leq \frac{E[e^{sX}]}{e^{sa}} \text{ (by Markov's inequality).} \\ &= e^{-sa} M_X(s) \end{aligned}$$

Similarly, for some $s < 0$,

$$\begin{aligned} P(X \leq a) &= P(sX \geq sa) \\ &= P(e^{sX} \geq e^{sa}) \\ &\leq \frac{E[e^{sX}]}{e^{sa}} \\ &= e^{-sa} M_X(s) \end{aligned}$$

□

The Chernoff bound is considered a stronger condition compared to the Markov inequality and/or the Chebyshev inequality.

Theorem 4.2.2 Hoeffding's Inequality Consider the sum $S_n = \sum_{i=1}^{n} X_i$ for independent random variables X_i, then we have for all $s > 0$,

$$\begin{aligned} P(S_n - \mathbb{E}[S_n] \geq t) &= P(exp(s(S_n - \mathbb{E}[S_n])) \geq e^{st}) \\ &\leq e^{-st}\mathbb{E}\left[e^{s(S_n - \mathbb{E}[S_n])}\right] \\ &= e^{-st}\prod_{i=1}^{n}\mathbb{E}[e^{s(X_i - \mathbb{E}[X_i])}] \end{aligned}$$

Which is true by direct application of the Markov's inequality and Chernoff bounds.

Lemma 4.2.3 Hoeffding's Lemma Let X be any real-valued random variable with expected value $\mathbb{E}[X] = 0$ and such that $a \leq X \leq b$ almost surely, for some $a, b \in \mathbb{R}$, $0 \in [a, b]$. Then for all $s \in \mathbb{R}$,

$$\mathbb{E}[e^{sx}] \leq exp\left(\frac{s^2(b-a)^2}{8}\right)$$

Proof. As e^{sx} is convex in x and is this uniformly bounded, i.e.

$$\begin{aligned} e^{sx} &\leq \lambda e^{sb} + (1-\lambda)e^{sa}, \text{ where } \lambda := \frac{x-a}{b-a} \\ &= \frac{x-a}{b-a}e^{sb} + \frac{b-x}{b-a}e^{sa} \end{aligned}$$

Taking expectation on both sides yields

$$\mathbb{E}[e^{sX}] \leq \frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb}$$

Let $\theta := \frac{b}{b-a}$, $u := s(b-a)$, by taking ln on both sides, on the right hand side we have

$$\begin{aligned} \Phi(u) &= ln(\theta e^{sa} + (1-\theta)e^{sb}) \\ &= sa + ln(\theta + (1-\theta)e^{u}) \\ &= (\theta - 1)u + ln(\theta + (1-\theta)e^{u}) \end{aligned}$$

By taking the Taylor's expansion at s = 0, $\exists \xi = \xi(u) \in \mathbb{R}$ s.t.

$$\Phi(u) = \Phi(0) + \Phi'(0) + \tfrac{1}{2}\Phi''(\xi)u^2$$

as $\Phi(0) = 0$ and

$$\Phi'(u) = (\theta - 1) + \tfrac{(1-\theta)e^u}{\theta + (1-\theta)e^u}$$

we have $\Phi'(0) = 0$. Hence $\Phi(u) = \frac{1}{2}\Phi''(\xi)u^2$. Note that $\Phi''(u)$ is bounded above by a constant number:

$$\begin{aligned} \Phi''(u) &= \frac{\theta(1-\theta)e^u}{(\theta + (1-\theta)e^u)^2} \\ &= \left(\frac{\theta}{\theta + (1-\theta)e^u}\right)\left(\frac{(1-\theta)e^u}{\theta + (1-\theta)e^u}\right) \\ &= t(1-t), \text{ where } t := \frac{\theta}{\theta + (1-\theta)e^u}, t \in (0, 1] \\ &\leq \frac{1}{4} \end{aligned}$$

Thus we have $\Phi(u) \leq \frac{u^2}{8}$, which implies $ln\mathbb{E}[e^{sX}] \leq \frac{s^2(b-a)^2}{8}$ and by taking exponential on both sides, we prove the lemma.

□

As M_n is a random matrix with entries of known expectations and variances, we expect the variations between iid copies of M_n to be also small and bounded. It turns out that a particularly useful class of a regularity hypothesis here is a Lipschitz hypothesis – that small variations in $X := (X_1, ..., X_n)$, where X_i

are iid distributed random variables, lead to small variations in $F(X)$, which is some combinations of X. This is described by McDiarmid's inequality.

Theorem 4.2.4 McDiarmid's Inequality Let $X_1, X_2, ..., X_n$ be independent random elements taking values in range $R_1, R_2, ..., R_n$, and let $F : R_1 \times R_2 \times ... \times R_n \to \mathbb{C}$ be a function with the property that if one freezes all but the i^{th} coordinate of $F(x_, ..., x_n)$ for some $\leq i \leq n$, the F only fluctuates by at most $c_i > 0$, thus

$$|F(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n) - F(x_1, ..., x_{i-1}, x_i', x_{i+1}, ..., x_n)| \leq c_i$$

for all $x_i \in X_j, x_j' \in X_i$ for $1 \leq j \leq n$. Then for any $\lambda > 0$, one has

$$\mathbf{P}(|F(X) - \mathbb{E}F(X)| \geq \lambda\sigma) \leq Ce^{-c\lambda^2}$$

for some constants $C, c > 0$, where $\sigma^2 := \sum\limits_{i=1}^{n} c_i^2$.

Proof. Write $\mathbb{E}_i[.]$ to denote expectation conditioned on $X_{1:i} := (X_1, ..., X_i)$. Therefore, $g_i(X_{1:i}) := \mathbb{E}_i[f(X_{1:n})]$ is a function of $X_{1:i}$. Define the random variables:

$$\begin{aligned}
Y_i &:= g_i(X_{1:i}) - g_{i-1}(X_{1:i-1}) \\
A_i &:= \inf_{x_i \in R_i} g_i(X_{1:i-1}, x_i) - g_{i-1}(X_{1:i-1}), \\
B_i &:= \sup_{x_i \in R_i} g_i(X_{1:i-1}, x_i) - g_{i-1}(X_{1:i-1}).
\end{aligned}$$

such that

$$\begin{aligned}
&Y_i \in [A_i, B_i], \\
&\mathbb{E}_{i-1}[Y_i] = 0, \\
&\sum_{i=1}^{n} Y_i = f(X_{1:n}) - \mathbb{E}[f(X_{1:n})].
\end{aligned}$$

Assume that $[A_i, B_i]$ is always an interval of length at most c_i. Let $S_i := \sum\limits_{j=1}^{i} Y_j$, then

$$
\begin{aligned}
\mathbb{E}[e^{\lambda S_n}] &= \mathbb{E}[e^{\lambda(Y_n+S_{n-1})}] \\
&= \mathbb{E}\left[\mathbb{E}_{n-1}\left[e^{\lambda Y_n}\right]e^{\lambda S_{n-1}}\right] \\
&\leq exp(\lambda^2 c_n^2/8)\mathbb{E}[exp(\lambda S_{n-1})] \\
&= exp(\lambda^2 c_n^2/8)\mathbb{E}[exp(\lambda(Y_{n-1}+S_{n-2}))] \\
&= exp(\lambda^2 c_n^2/8)\mathbb{E}[\mathbb{E}_{n-2}[exp(\lambda Y_{n-1})]exp(\lambda S_{n-2})] \\
&\leq exp(\lambda^2 c_n^2/8)exp(\lambda^2 c_{n-1}^2/8)\mathbb{E}[exp(\lambda S_{n-2})] \\
&\vdots \\
&\leq exp(\lambda^2 c_n^2/8)...exp(\lambda^2 c_1^2/8) \\
&= exp\left(\frac{\sum_{i=1}^n c_i^2}{4}.\frac{\lambda^2}{2}\right) \\
&= Ce^{c\lambda^2}.
\end{aligned}
$$

By applying the Hoeffding's lemma inductively, where $C = exp(\frac{\sum_{i=1}^n c_i^2}{4})$, $c = \frac{1}{2}$. Observe that S_n is $\frac{\sum_{i=1}^n c_i^2}{4}$-subgaussian.
Now it remains to prove the assumption that $[A_i, B_i]$ is an interval of length at most c_i. We show that $B_i - A_i \leq c_i$. By independence of $X_1, ..., X_n$, we have

$$
\begin{aligned}
g_i(X_{1:i-1}, x_i) &= \mathbb{E}[f(X_{1:i-1}, x_i, X_{i+1:n})|X_{1:i-1}; X_i = x_i] \\
&= \mathbb{E}[f(X_{1:i-1}, x_i, X'_{i+1:n})|X_{1:i-1}]
\end{aligned}
$$

Where $X'_{i+1:n}$ is an independence copy of $X_{i+1:n}$. Then

$$
\begin{aligned}
B_i - A_i &= \sup_{b\in R_i} g_i(X_{1:i-1}, b) - \inf_{a\in R_i} g_i(X_{1:i-1}, a) \\
&= \sup_{a,b\in R_i}(g_i(X_{1:i-1}, b) - g_i(X_{1:i-1}, a)) \\
&= \sup_{a,b\in R_i}\mathbb{E}[f(X_{1:i-1}, b, X'_{i+1:n}) - f(X_{1:i-1}, a, X'_{i+1:n})|X_{1:i-1}] \\
&\leq \mathbb{E}\left[\sup_{a,b\in R_i}|f(X_{1:i-1}, b, X'_{i+1:n}) - f(X_{1:i-1}, a, X'_{i+1:n})||X_{1:i-1}\right] \\
&\leq c_i \text{ (by the bounded difference property of f).}
\end{aligned}
$$

□

By theorem 4.2.4, we conclude that

$$P(|s_n(a+ib) - \mathbb{E}[s_n(a+ib)]| \geq \lambda) \leq Ce^{-cn\lambda^2}.$$

for some constants $C, c > 0$ and $\sigma^2 := \sum_{i=1}^n c_i^2$. Then from the Borel-Cantelli lemma, for any fixed $z \in \mathbb{C}^+$ we have

$$s_n(z) - \mathbb{E}[s_n(z)] \xrightarrow{a.s.} 0.$$

5 Recursive Equation of $\mathbb{E}[s_n(z)]$

In this section, we show that for any fixed $z \in \mathbb{C}^+$, $\mathbb{E}[s_n(z)] \to s_{sc}(z)$, the Stielejes transform of the semicircular distribution.

5.1 Properties of $\mathbb{E}[s_n(z)]$

Proposition 5.1.1 Determinant of Block Matrices (Schur Complement) Let M be a square matrix of finite order $(a+d) \times (a+d)$ that takes the form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where A $\in \mathbb{C}^{a \times a}$, B $\in \mathbb{C}^{a \times d}$, C $\in \mathbb{C}^{d \times a}$, D$\in \mathbb{C}^{d \times d}$ such that the Schur complement of block A exists in M. Then we have

$$det(M) = det(D - CA^{-1}B)det(A).$$

Proof. Assume that block A is invertible, consider the Schur complement of A:

$$D - CA^{-1}B.$$

Consider the tranformation

$$\begin{pmatrix} I_a & 0 \\ -CA^{-1} & I_d \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} I_a & -A^{-1}B \\ 0 & I_d \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix}$$

Note that det$\begin{pmatrix} I_a & 0 \\ -CA^{-1} & I_d \end{pmatrix}$ and det$\begin{pmatrix} I_a & -A^{-1}B \\ 0 & I_d \end{pmatrix}$ both equal to 1, thus

$$det(M) = det(A(D - CA^{-1}B)) = det(A)det(D - CA^{-1}B)$$

□

Lemma 5.1.2 Let $A_n \in \mathbb{C}^{n \times n}$, and let A_{n-1} be the top left $n-1 \times n-1$ minor. Suppose that

$$A_n = \begin{pmatrix} A_{n-1} & X \\ Y^* & a_{nn} \end{pmatrix}$$

and both A_n, A_{n-1} are invertible. Then

$$[A_n^{-1}]_{nn} = \frac{1}{a_{nn} - Y^* A_{n-1}^{-1} X}$$

Where $[A_n^{-1}]_{nn}$ denotes the (n,n) entry of the matrix A_n^{-1}.

Proof. By the property of the inverse of an invertible matrix

$$A_n^{-1} = \frac{1}{det(A_n)} adj(A_n).$$

by using proposition 5.1.1, we have

$$[A_n^{-1}]_{nn} = \frac{det(A_{n-1})}{det(A_n)} = \frac{det(A_{n-1})}{det(a_{nn}-Y^*A_{n-1}^{-1}X)det(A_{n-1})} = \frac{1}{a_{nn}-Y^*A_{n-1}^{-1}X}.$$

□

Applying Lemma 5.1.2 by letting $A_n = \frac{1}{\sqrt{n}}M_n - zI_n$, we obtain the identity

$$\left(\frac{1}{\sqrt{n}}M_n - zI_n\right)^{-1}_{nn} = \frac{1}{\frac{1}{\sqrt{n}}\mathcal{C}_{nn}-z-\frac{1}{n}X^*(\frac{1}{\sqrt{n}}M_{n-1}-zI_{n-1})^{-1}X}.$$

Since the diagonal entries of $\frac{1}{\sqrt{n}}M_n - zI_n$ have identical distribution, it implies

$$\begin{aligned}
\mathbb{E}[s_n(z)] &= \mathbb{E}\left[\frac{1}{n}Tr\left(\frac{1}{\sqrt{n}}M_n - zI_n\right)^{-1}\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\left(\frac{1}{\sqrt{n}}M_n - zI_n\right)^{-1}_{ii}\right] \\
&= n\cdot\frac{1}{n}\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}M_n - zI_n\right)^{-1}_{nn}\right] \text{ as the diagonal entries of } M_n \text{ are i.i.d distributed} \\
&= \mathbb{E}\left[\frac{1}{\frac{1}{\sqrt{n}}\mathcal{C}_{nn} - z - \frac{1}{n}X^*(\frac{1}{\sqrt{n}}M_{n-1} - zI_{n-1})^{-1}X}\right].
\end{aligned}$$

Denote $R := \left(\frac{1}{\sqrt{n}}M_{n-1} - zI_{n-1}\right)^{-1}$. Consider the stochastic quadratic form $\frac{1}{n}X^*RX$. Here we take vector X that only involves entries that are in M_n but not M_{n-1} and so the random matrix R and the vector X are independent. Recall that z has imaginary part b if $z = a + ib$.

Furthermore, observe that $||X^*RX|| = ||R|| \leq |b|^{-1}$ as the magnitude of the denominator here is bounded below by $|b|$. Then the function $F(X) = X^*RX$ is Lipschitz when X is c-Lipschitz (Here we use the reduction that entries of M_n are bounded):

$$|F(X) - F(Y)| \leq ||F'(u)|| \cdot ||X - Y||$$

where $X, Y \in \Omega^n$ and $||F'(u)|| = ||F'(X)|| = ||2X^*R||$. To show F is Lipschitz, we only need to have $F'(u)$ to be bounded:

$$||F'(u)|| = ||2X^*R|| \leq \sup 2||X|| \cdot ||R|| \leq \tilde{c}\sqrt{n}$$

We see that the upper bound of X^*RX is $O(\sqrt{n})$ or loosely $o(n)$.
Next we wish to show that the expression concentrates around its expectation. We prove this using the Talagrand's inequality.

5.2 The Talagrand's Concentration Inequality

Definition 5.2.1 Hamming Distance Let Ω be a set equipped with the $\sigma-$field $\mathcal{A}$, for each vector w = $(w_1, w_2, ..., w_n) \in \mathbb{R}_+^n$, the weighted Hamming distance between two vectors $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$, $x, y \in \Omega^n$ is defined as

$$d_w(x,y) := \sum_{i\leq n} w_i h_i(x,y) \text{ where } h_i(x,y) = \begin{cases} 1 & \text{if } x_i \neq y_i \\ 0 & \text{otherwise} \end{cases}$$

For a subset A of χ^n, the distances $d_w(x,A)$ and D(x, A) are defined by

$$d_w(x,A) := \inf\{y \in A : d_w(x,y)\}$$
$$D(x,A) := \sup_{w\in W} d_w(X,A)$$

where the supremum is taken over all weights in the set

$$W := \{(w_1,...,w_N) : w_i \geq 0 \text{ for each i and } |w|^2 := \textstyle\sum_{i\leq n} w_i^2 < 1\}$$

Theorem 5.2.2 Talagrand's Inequality For random elements $X = (X_1,...,X_n)$ of Ω^n with independent coodinates and subsets $A \in \mathcal{A}$,

$$\mathbf{P}(A)\mathbf{P}(A_t^c) \leq e^{-\frac{t^2}{4}} \text{ for all } t > 0.$$

where $P(A) := P(x \in A)$ for some $x \in \Omega^n$ and $A_t^c := \{x \in \Omega : D(x,A) \leq t\}$

Proof. Define

$$U_A(x) := \{(s_i)_{i\leq n} \in \{0,1\}^n : \exists y \in A; s_i = 1 \iff x_i \neq y_i\}.$$

Define $V_A(x)$ to be the convex hull of $U(x,A)$. This V_A contains 0 if and only if $x \in A$. The corresponding notion of 'enlargement' of A is then

$$A_t^c = \{x \in \Omega^n : D(x,A) \leq t\}$$

Hence

$$D(x,A) = \sup_{\alpha\in A:|\alpha|_2=1} \inf_{g\in U(x,A)} \alpha.g = \min_{v\in V(x,A)} |v|$$

i.e. the supremum over all directions α of the minimum projection of $U(x,A)$ in that direction. And the maximum of the projection is at most the norm of the projection on the convex combination.
It is then important to establish the inequality (Talagrand 1996),

$$\textstyle\int_\Omega exp(\frac{D(x,A)^2)}{4})dP(x) \leq \frac{1}{\mathbf{P}(A)} \text{ (*)}$$

Assuming (*) is true, then we have for some $t \geq 0$

$$\mathbf{P}\left(D(x,A) \geq t\right) = \mathbf{P}\left(e^{D(x,A)^2/4} \geq e^{t^2/4}\right) \leq \mathbb{E}\left[e^{D(x,A)^2/4}\right] e^{-t^2/4}$$

where the last inequality is by the Markov's inequality. By (*), we see that $\mathbb{E}[e^{D(x,A)^2/4}] \leq \frac{1}{\mathbf{P}(A)}$ and we prove the inequality.

To prove (*), we follow Talagrand to prove by induction on the dimension n.
Base step: when $n = 1$,

$$D(x, A) = \begin{cases} 1 & \text{if } x \notin A \\ 0 & \text{if } x \in A \end{cases}$$

where D(x, A) now is simply denoting whether x is a possible outcome of A. Hence

$$\int_\Omega e^{\frac{D(x,A)^2}{4}} dx \leq \mathbf{P}(A) + (1 - \mathbf{P}(A))e^{1/4}$$

For arbitrary $p \in (0, 1]$, by calculus we have

$$p + (1 - p)e^{1/4} \leq \frac{1}{p}.$$

by setting p = P($x \in A$), n = 1 is true.

Induction step: Assume that the inequality is true when n = k, consider a subset A of Ω^{n+1} and its projection B on Ω^n. For $\omega \in \Omega$, set

$$A(\omega) := \{x \in \Omega^n : (x, \omega) \in A\}$$

To prove this, we introduce an important claim:

for any $\lambda \in [0, 1]$, we have

$$D((x, \omega), A)^2 \leq (1 - \lambda)^2 + \lambda D(x, A(\omega))^2 + (1 - \lambda)D(x, B)^2. \text{ (**)}$$

To prove (**), take $x \in A(\omega)$, $\omega \in \Omega$ and $z = (x, \omega) \in B$ observe that $U_A(z)$ contains both $U_{A(\omega)}(x) \times \{0\}$ and $U_B(x) \times \{1\}$. That is whenever $\exists s, t$ such that

$$\begin{aligned} s \in U_{A(\omega)}(x) &\Rightarrow (s, 0) \in U_A(z) \\ t \in U_B(x) &\Rightarrow (t, 1) \in U_A(z) \end{aligned}$$

Then for the same $s \in V_{A(\omega)}(x)$, $t \in V_B(x)$, $0 \leq \lambda \leq 1$, by convexity we have

$$(\lambda s + (1 - \lambda)t, 1 - \lambda) \in V_A(z).$$

Knowing that $D(z, A)$ is upper bounded by the norm of the convex hull, we have

$$\begin{aligned} D(z, A)^2 &\leq (1 - \lambda)^2 + |(1 - \lambda)s + \lambda t|^2 \\ &\leq (1 - \lambda)^2 + (1 - \lambda)|s|^2 + \lambda|t|^2 \text{ by the Pythagorean theorem} \end{aligned}$$

Let ω be fixed. Substituting from the claim

$$\begin{aligned} \int_x e^{\frac{D(z,A)^2}{4}} dx &\leq e^{\frac{1-\lambda}{4}} \int_x \left(e^{D(x,A(\omega))^2/4}\right)^\lambda \left(e^{D(x,B)^2/4}\right)^{1-\lambda} dx \\ &\leq e^{\frac{(1-\lambda)^2}{4}} \left[\int_x e^{D(x,A(\omega))^2/4}\right]^\lambda \left[\int_x e^{D(x,B)^2/4}\right]^{1-\lambda} \text{ by Holder's inequality} \\ &\leq e^{\frac{(1-\lambda)^2}{4}} \left(\frac{1}{\mathbf{P}(A(\omega))}\right)^\lambda \left(\frac{1}{\mathbf{P}(B)}\right)^{1-\lambda} \text{ by the induction hypothesis} \\ &= \frac{1}{\mathbf{P}(B)} e^{(1-\lambda)^2/4} \left(\frac{\mathbf{P}(A(\omega))}{\mathbf{P}(B)}\right)^{-\lambda} \end{aligned}$$

To proceed, we seek for a bound that is independent of the parameter λ, and intuitively this bound exists as $\lambda \in [0,1]$. By setting $r := \frac{\mathbf{P}(A(\omega))}{\mathbf{P}(B)}$, note that $r \leq 1$ since $A(\omega) \subseteq B$.

Now we introduce one important lemma:
consider $0 \leq r \leq 1$, then

$$\inf_{0\leq\lambda\leq 1} r^{-\lambda} e^{\frac{(1-\lambda)^2}{4}} \leq 2 - r \text{ (***)}$$

To prove (***), set

$$\lambda := \begin{cases} 1 + 2\ln r & , e^{-1/2} \leq r \leq 1 \\ 0 & , 0 < r < e^{-1/2} \end{cases}$$

it suffices to show that $f(r) = \ln(2-r) + \ln(r) + \ln(r)^2 \geq 0$ over $r \in [0,1]$. Since $f(1) = 0$, it suffices to show $f'(r) \leq 0$, which is true by simple calculus.

Thus

$$\int_x e^{\frac{D(z,A)^2}{4}} dx \leq \frac{1}{\mathbf{P}(B)} \left(2 - \frac{\mathbf{P}(A(\omega))}{\mathbf{P}(B)}\right)$$

Integrating over ω gives

$$\int_\omega \int_x e^{\frac{D((x,\omega),A)^2}{4}} dx d\omega \leq \frac{1}{\mathbf{P}(B)} \left(2 - \frac{\mathbf{P}(A)}{\mathbf{P}(B)}\right) = \frac{1}{\mathbf{P}(A)} \frac{\mathbf{P}(A)}{\mathbf{P}(B)} \left(2 - \frac{\mathbf{P}(A)}{\mathbf{P}(B)}\right)$$

As B is the projection of A onto a smaller dimension, we have $\frac{\mathbf{P}(A)}{\mathbf{P}(B)} \in [0,1]$. Since $p(2-p) \leq 1$ for any $p \in [0,1]$, by letting p $= \frac{\mathbf{P}(A)}{\mathbf{P}(B)}$ we have finally

$$\int_\Omega exp(\frac{D(x,A)^2)}{4}) dx \leq \frac{1}{\mathbf{P}(A)}$$

and we prove (*). □

In particular, if $P(X \in A) \geq \frac{1}{2}$, then the probability to not be in within the Hamming distance t decays exponentially. This leads us to propose that the probability actually concentrates around the median of the random variable X, and finally we wish to prove that the expectation of X concentrates around its median.

Corollary 5.2.3 Let $K > 0$, $X = (X_1, ..., X_n)$ be a complex random variable with independent components $|X_i| \leq K$ for all $0 \leq i \leq n$. Let $F : \mathbb{C}^n \to \mathbb{R}$ be a 1-Lipschitz convex function (i.e., $|F(x) - F(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$, where we use Euclidean metric on $\mathbb{R}^n$.). Then for any λ one has

$$\mathbf{P}(|F(X) - \mathbf{M}F(X)| \geq \lambda K) \leq Cexp(-c\lambda^2)$$

and

$$\mathbf{P}(|F(X) - \mathbb{E}[F(X)]| \geq \lambda K) \leq Cexp(-c\lambda^2)$$

for some absolute constants $C, c > 0$, where $\mathbf{M}F(X)$ is a median of $F(X)$.

Proof. Recall that a median $\mathbf{M}(X)$ for a random variable X is a real number satisfying the inequalities $P(X \geq \mathbf{M}(X)) \geq 1/2$ and $P(X \leq \mathbf{M}(X)) \geq 1/2$. Since each component of X has magnitude at most K, we can normalise K = 1. One key observation is that for the special case of A being convex in $[0,1]^n$, the convex distance controls the Euclidean distance.
So we introduce a lemma

Lemma Let A be convex in $[0,1]^n$ and $x \in [0,1]^n$. Then $d_E(x, A) \leq D(x, A)$.
Proof: Suppose $D(x, A) \leq t$, then by definition of convex distance, there exists a convex combination $w := \sum_{i=1}^{n} \lambda_i \vec{s_i}$ of vectors $\vec{s_i} \in U_A(x) \subset \{0,1\}^n, 0 \leq i \leq n$ such that $||w||_E \leq t$. For each $\vec{s_i} \in U_A(x)$, we can find a vector $\vec{z_i} \in A - x$ supported by $\vec{s_i}$. Let $z := \sum_{i=1}^{n} \lambda_i \vec{z_i}$, $z \in A - x$ by convexity. $z_i \in [0,1]^n$ implies each component of z is upper bounded by 1 and since each component of $\vec{z_i}$ is non-zero only when the corresponding $\vec{s_i}$ component is non-zero, we have that each component of $\vec{z_i}$ is upper bounded by the corresponding component of $\vec{s_i}$. Thus we have $d_E(x, A) \leq ||z||_E \leq ||w||_E \leq D(x, A)$.

By the inequality (*) in Theorem 5.2.24, we have

$$\mathbb{E}[e^{d_E(x,A)^2/4}] \leq \frac{1}{\mathbf{P}(A)}$$

Let $a \geq 0$ and take A $= \{F \leq a\}$. Observe that by the Lipschitz property, if $X \in \{F \geq a + \lambda\}$ for some $\lambda \geq 0$, then $d_E(X, A) \geq \lambda$. Then by applying the Talagrand's inequality, one has

$$\mathbf{P}(F(X) \leq a)\mathbf{P}(F(X) \geq a + \lambda) \leq e^{-\lambda^2/4}$$

Then by taking $a = \mathbf{M}F(X)$, we get the upper tail estimate

$$\mathbf{P}(F(X) - \mathbf{M}F(X) \geq \lambda) \leq 2exp(-\lambda^2/4)$$

and by taking $a = \mathbf{M}F(X) - \lambda$, we get the lower tail estimate

$$\mathbf{P}(F(X) - \mathbf{M}F(X) \leq -\lambda) \leq 2exp(-\lambda^2/4)$$

And the desired result follows from the union bound, by taking into consideration of the normalising factor K.
The inequality in turn implies that

$$\mathbb{E}[F(X)] = \mathbf{M}F(X) + O(1),$$

then the second inequality follows immediately.
For another and more elegant proof of the second inequality, we defer the proof to Annex B to avoid too much focus on this section. □

5.3 Convergence of $\mathbb{E}[s_n(z)]$

Recall that in section 5.1, we notice that $F(X) = X^*RX$ is Lipschitz in X of order o(n) when X is bounded. Since X and R are independent, the distribution of X is unchanged conditioning on R. Hence we may apply Corollary 5.2.3 by taking $F(X) = X^*RX$ for some deterministic R:

$$\mathbf{P}(\tfrac{1}{n}|X^*RX - \mathbb{E}[X^*RX]| \geq t|R) \leq Ke^{-knt^2}$$

for any $t > 0$, where k, K are constants independent of R. Note that since entries of M_n are iid with mean 0 and variance 1,

$$\mathbb{E}[X^*RX] = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} r_{ij}\mathbb{E}[\bar{\mathcal{C}_{in}}\mathcal{C}_{jn}] = \sum_{i=1}^{n-1} r_{ii} = Tr(R)$$

where $\mathcal{C}_{in}$ are entries of X and r_{ij} are entries of R. Therefore we have

$$\mathbf{P}(\tfrac{1}{n}|X^*RX - Tr(R)| \geq t|R) \leq Ke^{-knt^2}$$

which implies $\frac{1}{n}X^*RX = \frac{1}{n}Tr(R) + o(1)$.
Note that $\frac{1}{n}Tr(R)$ is essentially $s_{n-1}(z)$ at z. As the normalisation factor is slightly off as we approximate $s_{n-1}(\frac{\sqrt{n}}{\sqrt{n-1}}z)$ by $s_{n-1}(z)$, actually we have

$$Tr(R) = n\tfrac{\sqrt{n}}{\sqrt{n-1}}s_{n-1}\left(\tfrac{\sqrt{n}}{\sqrt{n-1}}z\right)$$

then as $s_n(z) = s_{n-1}(z) + O(\frac{1}{n})$ proven in section 4.1, we conclude that

$$Tr(R) = n(s_n(z) + o(1))$$

Therefore we have

$$\tfrac{1}{n}X^*RX = \tfrac{1}{n}Tr(R) + o(1) \Rightarrow \tfrac{1}{n}X^*RX = \mathbb{E}[s_n(z)] + o(1)$$

by the convergence of $s_n(z)$ towards $\mathbb{E}[s_n(z)]$ proven in section 4.2. Putting this result back to the equation in section 5.1, we have

$$\mathbb{E}[s_n(z)] = -\tfrac{1}{z+\mathbb{E}[s_n(z)]}$$

Then the unique valid solution is

$$s_n(z) = \tfrac{-z+\sqrt{z^2-4}}{2} = s_{sc}(z).$$

As $s_n(z)$ is able to take positive values.

6 Discussion

Note that in Section 5.1, we used the reduction that the entries of M_n are bounded. In this section, we verify the entries of M_n are indeed bounded using the truncation method.

Theorem 6.1.1 Let $M = (\mathcal{C}_{ij})_{1\leq i,j\leq n}$ be a real symmetric matrix whose entries with identical distributions with mean 0 and unit variance, and finite fourth moment $O(1)$ for some $\delta > 0$, then the entries $\mathcal{C}_{ij}$ has magnitude at most $\alpha\sqrt{n}$.

Proof. To prove each independent $\mathcal{C}_{ij}$ is bounded by some function dependent on n denoted by $K(n)$, we can split each $\mathcal{C}_{ij}$ as $\mathcal{C}_{ij,\leq K(n)} + \mathcal{C}_{ij,>K(n)}$. As $||M||_{op}$ has $O(\sqrt{n})$, we may as well compare $K(n)$ with $\alpha\sqrt{n}$ for some constant *alpha* independent of n. That is:

$$\mathcal{C}_{ij} = \mathcal{C}_{ij,\leq\alpha\sqrt{n}} + \mathcal{C}_{ij,>\alpha\sqrt{n}}$$

Since

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{P}(\mathcal{C}_{ij,>\sqrt{n}} \geq \alpha\sqrt{n}) \leq n^2\mathbf{P}(\mathcal{C}_{ij,>\sqrt{n}} \geq \alpha\sqrt{n}) \leq \frac{\mathbb{E}[|\mathcal{C}_{ij,>\alpha\sqrt{n}}|^4]}{\alpha^4}$$

By the Markov's inequality and the fourth moment bound hypothesis, then by Borel-Cantelli lemma we conclude that all entries of M_n are less than $\alpha\sqrt{n}$ almost surely. $\square$

7 Reference

1. Terence Tao, 2011, *Topics in Random Matrix Theory*, American Mathematical Society.

2. Jean-Francois Le Gall, 2022, *Measure Theory, Probability, and Stochastic Process*, Graduate Texts in Mathematics (GTM, volume 295)

3. Talagrand, M., 1994. *Concentration of measure and isoperimetric inequalities in product spaces*, arXiv https://doi.org/10.48550/arXiv.math/9406212.

4. Pollard, D. A 2006. *Note on Talagrand's Convex Hull Concentration Inequality.* arXiv https://doi.org/10.48550/arXiv.math/0611770.