Wigner Semicircle Law

November 27, 2020

Abstract

The Wigner Semicircle Law claims that the empirical density of eigenvalues of a random matrix is given by the universal semicircle distribution. In this exposition, we begin with definitions of Auxiliary and Empirical Distribution functions that we will use to prove the Wigner Semicircle Law, in addition to a few results that will push us towards proving the Semicircle Law. Specifically, we will devote considerable attention towards the Resolvent, and the Stieljes Transform, in our proof of the Main Theorem. After our proof, we will give a few examples to illustrate how the theorem applies to more concrete examples that I have chosen.

Contents

1 Introduction

1.1 Random Matrices

Let

$$\mathcal{A} = \begin{bmatrix} \xi_{11} & \cdots \\ \xi_{21} & \cdots \\ \xi_{31} & \cdots \\ \vdots & \cdots \end{bmatrix}$$

be a symmetric ∞ by ∞, *Hermitian*, matrix $\mathcal{A} = \{\xi_{ij}\}_{1 \le i,j \le \infty}$. The random entries $\{\xi_{ij}\}$ must satisfy $\mathbf{E}\xi_{ij} = 0$, $\mathbf{E}(\xi_{ij})^2 = 1$ for $i \neq j$, $\mathbf{E}(\xi_{ij})^2 = \sigma^2$ for $i = j$ (mean 0 and variance σ), and $\xi_{ij} = \bar{\xi}_{ji}$. With the Hermitian constraint (last condition), we can obtain $\mathcal{A}$ by filling the upper triangular region with random entries and then reflecting the entries to fill the lower triangular region of $\mathcal{A}$. With $\mathcal{A}$, we can obtain a sequence

$$\mathcal{A}_n = \begin{bmatrix} \xi_{11} & \cdots \\ \vdots & \cdots \\ \xi_{n1} & \cdots \end{bmatrix}$$

of Hermitian matrices of increasing order n, where for $n \geq 2$

$$\mathcal{A}_n = \begin{bmatrix} \mathcal{A}_{n-1} & v_n \\ v_n^{\mathbf{T}} & \xi_{nn} \end{bmatrix},$$

where $\mathbf{T}$ denotes the conjugate transpose of the column vector v_n. We will denote the eigenvalues of $\mathcal{A}_n$ with $\lambda_k^{(n)}$, and $\{u_k^{(n)}\}$ the corresponding orthonormalized column eigenvector, for $1 \le k \le n$. We will arrange the eigenvalues of $\mathcal{A}_n$ in increasing order with $\lambda_1^{(n)} < \lambda_2^{(n)} < \cdots < \lambda_n^{(n)}$. The corresponding (column) eigenvectors $u_k^{(n)}$, for $k = 1, \cdots, n$, can be chosen to form a unitary matrix $\mathcal{U}_n = (u_1^{(n)}, \cdots, u_k^{(n)})$. This gives $\mathcal{U}_n^{\mathbf{T}} \mathcal{A}_n \mathcal{U}_n = \text{diag}(\lambda_k^{(n)})$.

Remark: One ensemble of random matrices that converges to the Wigner Semicircle Distribution is the collection of Wigner matrices. From the conditions on the random entries of $\mathcal{A}$ that I have listed, we can obtain a **Wigner** Matrix by reflecting all of the complex entries from the upper triangular portion of the matrix to the lower triangular portion with the **Hermitian constraint**.

From the n by n submatrix $\mathcal{A}_n$, we can form another random matrix [1] $\mathcal{A}_n/\sqrt{n} = (1/\sqrt{n})\{\xi_{ij}\}$. This matrix will emerge in several proofs in order to calculate the trace of the resolvent. Besides this, we will begin our discussion of the Semicircle Law by defining the auxiliary and distribution functions $X_n(t)$ and $F_n(x)$. We will also give definitions of the Resolvent and Stieljes Transform.

1.2 Definitions

From the submatrix $\mathcal{A}_n$, we will choose 2 vectors at random to compute the magnitude of their dot product. We will let $v_n^{\mathbf{T}} = (\xi_{n1}, \xi_{n2}, \cdots, \xi_{n,n-1})$. The **Auxiliary Function** $X_n(t)$ will bookkeep the length of projections of the new row $v_n^{\mathbf{T}}$ of $\mathcal{A}_n$ onto the eigenvectors of the preceding matrix $\mathcal{A}_{n-1}$. If $v_k^{\mathbf{T}}$ and $u_k^{(n-1)}$ are the vectors chosen at random, we can show that the normalization $1/\sqrt{(n(n-1))}$ from our definition of $X_n(t)$ is appropriate, where the auxiliary function $X_n(t)$ is

Auxiliary Function (1)

[1] This matrix is Wigner.

$$X_n(t) = (n(n-1))^{-1/2} \sum_{k=1}^{\lfloor (n-1)t \rfloor} |v_n^{\mathbf{T}} u_k^{(n-1)}|^2$$

with $t \in [0,1]$ [2].

Besides the **Auxiliary Function**, we are also interested in the asymptotic shape of the **Empirical Distribution** $F_n(x)$. From the eigenvalues $\lambda_k^{(n)}$ of $\mathcal{A}_n/\sqrt{n}$, we will study the distribution $F_n(x)$

Empirical Distribution of Eigenvalues (2)

$$F_n(x) = \frac{1}{n}\{\#\lambda_k^{(n)} \leq x\sqrt{n}, 1 \leq k \leq n\}$$

This distribution $F_n(x)$ is a step function function with points of discontinuity at the eigenvalues of $\mathcal{A}_n/\sqrt{n}$, and with $x \in \mathbb{R}$.

We will make use of the **Resolvent** in the proofs of **Lemmas R1, R2**.

Resolvent (3)

The *Resolvent* is defined over a unital algebra, A, which is defined a field of the complex numbers $\mathbb{C}$. By definition the Spectrum $\mathrm{Sp}(x)$ is the space

$$\mathrm{Sp}(x) = \{t \in \mathbb{C} | x - t * 1 \text{ not invertible in A}\} \ .$$

More formally, the *Resolvent* can be defined with the map $\phi : \mathbb{C} - \mathrm{Sp}(x) \longrightarrow A$ [3] In ϕ, $\mathrm{Sp}(x)$ is the spectrum that we have defined. For the computations that will soon follow in **Lemma R2**, we will make use of $\mathbf{Tr}R(z, \mathcal{A}_n)$ with the definition that $R(z, \mathcal{A}_n) = (\mathcal{A}_n - zI_n)^{-1}$. The *resolvent* is an operator that we will manipulate.

We will also define and state properties of the **Stieljes Transform**.

Stieljes Transform (4)

Let μ be a finite positive measure on the Borel sets of the real line $\mathbb{R}$, and $F(x) = \mu((-\infty, x))$ its distribution function. Then

$$\hat{\mu}(z) = \hat{F}(z) = \int_{x=-\infty}^{\infty} (x-z)^{-1} \mathrm{d}\mu = \int_{x=-\infty}^{\infty} (x-z)^{-1} \mathrm{d}F(x)$$

exists for all complex z with $\mathrm{Im}(z) \neq 0$, and it called the *Stieljes Transform* of μ or F, respectively. We will confine ourself to the upper half plane $\mathrm{Im}(z) > 0$. **Alternatively**, some people would define the *Stieljes Transform* $\mathcal{S}_\rho(z)$ of a density $\rho(z)$ with

$$\mathcal{S}_\rho(z) = \int_I \frac{\rho(z)}{z-t} dt \text{ , where } x \in I \subset \mathbb{R} \text{ and } z \in \mathbb{C} \backslash \mathbb{R} \ .$$

Remark: Our proof of **Theorem 1** on Page 8 demonstrates why it is necessary to include the normalization factor of $(n(n-1))^{-1/2}$ in front of the summation, with computations from the Resolvent equations.

[2]Similar to the first definition that *Arnold* gives on page 185. A small difference is that we decided to take the ceiling instead of the floor.

[3]For more motivation, I have included another *remark* with more specific definitions of the **classical location** of the eigenvalues that we have obtained from $\mathcal{A}_n$. This remark makes use of the previous one and more clearly illustrates how the *resolvent* is used.

2 Properties and Lemmas

In the second section, we will establish properties of **Definitions 2-4** that will be useful for our proof of the **Main Theorem**. We would first like to determine why the factor of $\sqrt{n}$ in our definition of $F_n(x)$ is appropriate.[4] Later, we will make use of the *Stieljes Transform* to obtain the transform of a distribution function $F_n(x)$, which we will denote as $\hat{F}_n(x)$.

2.1 Properties of the Empirical Distribution

2.1.1 Lemma ED1

Let A be an n by n symmetric $(0,1)$ matrix, in which the density of ones is d (i.e. the number of ones is n^2d). Then the largest eigenvalues λ_1 of A satisfies the inequality

$$nd \le \lambda_1 \le n\sqrt{d}\ .$$

Proof of ED1: Let $e = (1, \cdots, 1)$ be the n-dimensional all-one vector. Then $\lambda_1 \ge \frac{(e,Ae)}{(e,e)} = \frac{n^2d}{n}$. If λ_i are the eigenvalues of the matrix, then $\sum_{i=1}^{n} \lambda_i^2 = n^2d$. Hence $\lambda_1 \le n\sqrt{d}$. This completes the proof.

Remark: The factor $\sqrt{n}$ is important for later discussion. From the **Main Theorem** that will be stated and proved, this factor allows us to pick convergent subsequences.

From $\mathcal{A}_n$ that we defined on **Page 1**, we see that **Proposition 1** justifies why $\sqrt{n}$ is necessary for the Distribution $F_n(x)$. We can also apply the *Stieljes Transform* to obtain the transform $\hat{F}_n(x)$ of the distribution $F_n(x)$.

Remark: For the next section, we will denote

$$||\mathcal{A}|| = \sup_{x\in\mathbb{C}^n:|x|=1}|\mathcal{A}x|\ .$$

Remark: We will prove **Lemma ED2** with the statements that I have given. These statement, and proof, follow directly from the last section of slides.

2.1.2 Lemma ED2, Factor of $\sqrt{n}$ in the Empirical Distribution of Eigenvalues

Suppose that the entries ξ_{ij} of $\mathcal{A}$ are independent, with mean 0 and uniformly bounded in magnitude by 1. Then there exists absolute constants $C, c > 0$ such that

$$\mathbf{P}(||\mathcal{A}|| < \mathcal{A}\sqrt{n}) \le C\exp(-cAn)\ ,$$

for all $\mathcal{A} \ge C$. In particular, we have that $||\mathcal{A}|| = \mathrm{O}(\sqrt{n})$ with overwhelming probability.

Before proving **Lemma ED2**, we will state **Lemma ED2***

2.1.3 Lemma ED2*

Suppose that the coefficients ξ_{ij} of $\mathcal{A}$ satisfy the same conditions in the statement of **Lemma ED2**. Let x be a unit vector in $\mathbb{C}^n$. Then for sufficiently large A, larger than some absolute constant, one has

$$\mathbf{P}(|\mathcal{A}x| \ge A\sqrt{n}) \le C\exp(-cAn)$$

[4] More motivation is given towards this construction in a further comment in the **Supplementary Section**. In the second remark, I will distinguish between different orders on $[-2, 2]$. This interval is important because it is the one on which the *Wigner Semicircle* distribution is defined.

for some absolute constants $C, c > 0$.

Proof: Let $X_1, \cdots X_n$ be the n rows of $\mathcal{A}$, then the column vector $\mathcal{A}x$ has coefficients $X_i x$ for $i = 1, \ldots, n$. If we let $x_1, \ldots, x_n$ be the coefficients of x, so that $\sum_{j=1}^{n} |x_j|^2 = 1$, then $X_i x$ is just $\sum_{j=1}^{n} \xi_{ij} x_j$. Applying standard concentration of measure results, we see that each $X_i x$ uniformly subgaussian, thus

$$\mathbf{P}(|X_i x| \geq \lambda) \leq C\exp(-c\lambda^2)$$

for some absolute constants $C, c > 0$. In particular, we have

$$\mathbf{E}\exp(c|X_i x|^2) \leq C \ ,$$

for some slightly different absolute constants $C, c > 0$. Multiplying these inequalities gives

$$\mathbf{E}\exp(c|\mathcal{A}x|^2) \leq C^n \ .$$

The claim follows from Markov's inequality.

We have to use another result, which we shall label as **Lemma ED2****

2.1.4 Lemma ED2**

Let $\mathcal{F}$ be a maximal 1/2-*net* of the sphere $\mathcal{S}$, namely a set of points in $\mathcal{S}$ that are separated from each other by distance of at least 1/2, and which is maximal with respect to set inclusion. Then for any n by n matrix $\mathcal{A}$ with complex entries, and any $\lambda > 0$

$$\mathbf{P}(||\mathcal{A}|| > \lambda) \leq \mathbf{P}(\bigvee_{y \in \mathcal{F}} |\mathcal{A}y| > \lambda/2) \ .$$

Proof of ED2**: From the definition of $||\mathcal{A}||$, and compactness, we can determine an $x \in \mathcal{S}$ so that

$$|\mathcal{A}x| = ||\mathcal{A}|| \ .$$

The point x that we are choosing does not have to lie in $\mathcal{F}$. However, as $\mathcal{F}$ is a maximal 1/2-net of $\mathcal{S}$, there exists some x that lie within 1/2 of some point y in $\mathcal{F}$.
If x did not exist, it would contradict the maximality of $\mathcal{F}$. Since $|x - y| \leq 1/2$,

$$|\mathcal{A}(x-y)| \leq ||\mathcal{A}||/2 \ .$$

By the *triangle inequality*, we have that

$$|\mathcal{A}y| \geq ||\mathcal{A}||/2$$

In particular, if $||\mathcal{A}|| > \lambda$, then $|\mathcal{A}y| > \lambda/2$ for some $y \in \mathcal{F}$, and the claim follows. This finishes the proof of **ED2****.

We will now complete the proof of **ED2**.

Proof of ED2 We can bound each of the probabilities $\mathbf{P}(|\mathcal{A}y| > \mathcal{A}\sqrt{n}/2)$ with $C\exp(-c\mathcal{A}n)$, for $\mathcal{A}$ sufficiently large. From another result that we gave, we also know that the maximal 1/2 - net of S has cardinality $\mathrm{O}(1/n)$. As a result, for $\mathcal{A}$ sufficiently large, the entropy loss of $\mathrm{O}(1/n)$ can be absorbed into the exponential gain of $\exp(-c\mathcal{A}n)$ by perturbing c slightly, which establishes **Lemma ED2**. This concludes the proof.

2.1.5 Lemma ED3

Let $A_n = (a_{ij})$ be an n by n symmetric matrix where $a_{ii} = 0$, a_{ij} for $i > j$ are independent random variables. Suppose that $\mathbf{P}(a_{ij} = 1) = p$, $\mathbf{P}(a_{ij} = 0) = q = 1-p$. If $\lambda_1 = \lambda_1(n)$ is the largest eigenvalue of the matrix A_n, then $\lim_{n\longrightarrow\infty}\frac{\lambda_1}{n} = p$ in probability.

Proof of ED3: We know:

$$\min_{1\leq i\leq n}\sum_{j=1}^{n} a_{ij} \leq \lambda_1 \leq \max_{1\leq i\leq n}\sum_{j=1}^{n} a_{ij}$$

From the *Perron-Frobenius Theorem for Regular Matrices.* I have included a statement of it in **III** of the **Supplementary Section**. Since for a given value of i the probability $\mathbf{P}(|\frac{1}{n}\sum_{j=1}^{n} a_{ij} - p| > \delta)$ is exponentially small,

$$\lim_{n\longrightarrow\infty}\mathbf{P}\left(\max_{1\leq i\leq n}\left|\frac{1}{n}\sum_{j=1}^{n} a_{ij} - p\right| > \delta\right) = 0$$

This completes the proof of **2**.

2.2 Properties of Stieljes Transform

There are also properties of the Stieljes Transform that are useful for our discussion:

Property 1: $\hat{F}$ is analytic, $\mathrm{Im}(\hat{F}(z)) \geq 0$ and for all $k \geq 0$

$$|\hat{F}^k(z)| \leq \mathrm{Im}(z)^{-k-1}F(\infty) \ .$$

Property 2, *Inversion Formula*: F is uniquely determined by $\hat{F}$. In particular, for point $x < y$ of continuity of F,

$$F(y) - F(x) = \pi^{-1}\lim_{v\longrightarrow 0}\int_{u=x}^{y}\mathrm{Im}(\hat{F}(u+\mathrm{i}v)du \ .$$

Property 3, *Fundamental Theorem of Grommer and Hamburger*: Let $\{\mu_n\}$ converge vaguely to μ, i.e., let $\int f d\mu + \int f d\mu_n$ for every continuous function f with compact support, and let $\sup_n \mu_n(R) < \infty$. Then $\bar{\mu}_n(z) \longrightarrow \bar{\mu}(z)$ uniformly in compact sets of the UHP. Conversely, for a sequence $\{\mu_n\}$ of measure with $\sup_n \mu_n(R) < \infty$ let $\{\bar{\mu}_n(z)\}$ converge for a z set with limit point inside the UHP. Then the convergence holds uniformly in the compact z sets, the limit is the *Stieljes Transform* $\bar{\mu}$ of a finite measure μ, and $\mu_n \longrightarrow \mu$.

- Note for myself

Motivate **2** and **3** the most, they will be used later.

2.3 Properties of Resolvent

Finally, the *resolvent* is helpful in our discussion of the Semicircle Law because we can multiply the identity matrix by a complex variable z, and we can invert the operator $A - zI$ for $z \notin \sigma(L)$. Clearly, the inverse is $(A - zI)^{-1}$. More generally, the matrix $z - A$ and $(z - A)^{-1}$ are diagnolizable and commute, and A, $z - A$ and $(z - A)^{-1}$ have the same eigenspaces.

2.3.1 Lemma R1 (First Resolvent Equation)

If $L \in \mathcal{L}(V)$ and $z_1 \neq z_2 \notin \sigma(L)$, then $R(z_1) \circ R(z_2) = \dfrac{R(z_1) - R(z_2)}{z_1 - z_2}$.

Proof of Lemma 1: The equality follows directly, in which $R(z_1) - R(z_2) = R(z_1)(L - z_2 I)R(z_2) - R(z_1)(L - z_1 I)R(z_2) = (z_1 - z_2)R(z_1)R(z_2)$. Finally, dividing by $z_1 - z_2$ on each sides yields the given equality.

With the *First Resolvent Equation*, we will now make use of a similar result in the *Second Resolvent Equation* to compute the trace of the resolvent in the following **Lemma**:

2.3.2 Lemma R2 (Trace of Resolvent)

Let $R(z, A_n)$ be the resolvent of A_n, with $\mathrm{Im}(z) \neq 0$. [5] Then

$$\mathbf{Tr}R(z, \mathcal{A}_k) = \sum_{k=1}^{n} \frac{1 + v_k^{\mathbf{T}} R(z, \mathcal{A}_{k-1})^2 v_k}{\xi_{kk} - z - v_k^{\mathbf{T}} R(z, \mathcal{A}_{k-1}) v_k} \ .$$

Remark: If we take $n = 1$, we can calculate the trace of the resolvent for a submatrix of $\mathcal{A}$ consisting of the single entry ξ_{11}. In the denominator, the term ξ_{kk} for $n = 1$ is ξ_{11}. Even more so, the resolvent terms $R(z, \mathcal{A}_{k-1})$ and $R(z, \mathcal{A}_{k-1})^2$ in the denominator and numerator, respectively, vanish, which gives that the terms $v_k^{\mathbf{T}} R(z, \mathcal{A}_{k-1}) v_k$ and $v_k^{\mathbf{T}} R(z, \mathcal{A}_{k-1})^2 v_k$ also vanish. Altogether, this gives that the trace of the resolvent for $n = 1$ is equal to $(\xi_{11} - z)^{-1}$, which coincides with the definition of the resolvent if we take $I \equiv 1$ and $A \equiv \xi_{11}$ from the previous page.

Proof of Lemma 2: From our remark, it suffices to prove that the equality holds for $n \geq 2$. If we start from the second resolvent equation, we define submatrices of $\mathcal{A}$, $\mathcal{A}^*$ and $\mathcal{B}$, as follows:

$$\mathcal{A}^* = \begin{bmatrix} \mathcal{A}_{n-1} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathcal{B}^* = \begin{bmatrix} 0 & v_n \\ v_n & \xi_{nn} \end{bmatrix}$$

From our choice of $\mathcal{A}^*$ and $\mathcal{B}^*$, it is clear that $\mathcal{A}^* + \mathcal{B} = \mathcal{A}_n$. We can obtain the desired quantity from our previous discussion. From the *Second Resolvent Equation* [6], we know that $R(z, \mathcal{A}^* + \mathcal{B}^*) = R(z, \mathcal{A}^*)[I + \mathcal{B}^* R(z, \mathcal{A})]^{-1}$. With the following computations, we can obtain $\mathbf{Tr}R(z, \mathcal{A}_k)$:

$$R(z, \mathcal{A}_n) = (\mathcal{A}_n - zI_n)^{-1}$$

We will give **Proposition 2**, and its proof here.

2.3.3 Proposition

Let A be an n by n Hermitian matrix, $\mathcal{E}$ the n by n matrix of all elements equal to 1, $\mathcal{D} = \mathrm{diag}(d_1, \ldots, d_n)$ a real diagonal matrix. Denote by F, F_1 and F_2 the empirical distribution functions of the eignevalues of $\mathcal{A}$, $\mathcal{A} + a\mathcal{E}$ (a real) and $\mathcal{A} + \mathcal{D}$, respectively. Then

$$|\hat{F}(z) - \hat{F}_1(z)| \leq (n\mathrm{Im}(z))^{-1} \ ,$$

[5] **Lemma 2** refers to *Arnold*'s **Lemma 1** on page 188. First, I verified the *second resolvent equation* by expanding the terms of each expression and formally verifying that the series are equal. After these steps, I then substituted the expression β_n repeatedly to get the correct expression within the summation.

[6] Originally thought to use a geometric series expansion for the resolvent. I couldn't get the terms out correctly.

and

$$|\hat{F}(z)-\hat{F}_2(z)| \leq (\mathrm{Im}(z))^{-2}\mathrm{max}_{1\leq i\leq n}|d_i| \; .$$

Proof of Proposition:

We have that $\mathrm{Im}(z)>0$ for $\hat{F}(z)=n^{-1}\mathbf{Tr}R(z,A)$ and $\hat{F}_1(z)=n^{-1}\mathbf{Tr}R(z,A+a\mathcal{E})$, where $R(z,A)=(A-zI)^{-1}$ denotes the resolvent and $\mathbf{Tr}A$ the trace of A. Inserting

$$(I_n+a\mathcal{E}R(z,A))^{-1}=I_n-\left(\frac{a}{ae'R(z,A)e}\right)\mathcal{E}R(z,\mathcal{A}) \; ,$$

where

$$e'=(1,1,\ldots,1) \text{ n times.}$$

into the **second resolvent equation**

$$R(z,A+a\mathcal{E})=R(z,A)(I-a\mathcal{E}R(z,A))^{-1}$$

and passing to traces gives [7]

$$\hat{F}(z)-\hat{F}_1(z)=n^{-1}\frac{a+e'R(z,A)^2e}{1+ae'R(z,A)e} \; .$$

With a slight change of notation, this argument resembles terms in the formula for the Trace of Resolvent.

If $f(z)=\int(x-z)^{-1}d\mu(x)$, where $\mu(x)=e'\mathcal{S}(x)e$, $\mathcal{S}$ being the spectral matrix of A, we have $e'R(z,A)e=f(z)$ and $e'R(z,A)^2e=f'(z)$. This gives

$$\hat{F}(z)-\hat{F}_1(z)=n^{-1}\frac{af'(z)}{1+af(z)} \; .$$

We obtain

$$|\hat{F}(z)-\hat{F}_1(z)|\leq(n\mathrm{Im}(z))^{-1}$$

from the fact that

$$|1+af'(z)|\geq|a|\mathrm{Im}(f(z))=|a|\mathrm{Im}(z)\int|x-z|^{-2}d\mu(x)$$

and

[7] This follows a computation from Arnold's *On Wigner's Semicircle Law for the Eigenvalues of Random Matrices* on page 197.

$$|af(z)| \leq |a| \int |x-z|^{-2} d\mu(x) \ .$$

Putting these estimates together shows that **(i)** holds. The second statement also directly follows once we apply the definition of the *Second Resolvent Equation*,

$$R(z, A+\mathcal{D}) - R(z, A) = R(z, A+\mathcal{D})\mathcal{D}R(z, A)$$

therefore

$$\begin{aligned} |\hat{F}(z) - \hat{F}_2(z)| &\leq n^{-1} |\mathbf{Tr} R(z, A+\mathcal{D})\mathcal{D}R(z, A)| \\ &\leq ||R(z, A+\mathcal{D})\mathcal{D}R(z, A)|| \\ &\leq (\mathrm{Im}(z))^{-2} \mathrm{max}_{1 \leq i \leq n} |d_i| \ . \end{aligned}$$

The **second** condition holds once we apply the second Resolvent equation. It is on **Page 197** of (5) [8].

Remark: From these manipulations, we can continue using higher order resolvent equations. From our verification of the *Second Resolvent Equation*, we will compute $R(z, \mathcal{A}_n)$ directly. The next following steps in **Proof of Lemma 2** give us the direct computation to get the expression for $\mathbf{Tr}R(z, \mathcal{A}_n)$. [9] The following computations have to be repeatedly substituted in to get the expression for $\mathbf{Tr}R(z, \mathcal{A}_n)$.

We have

$$R(z, \mathcal{A}_n) = \begin{bmatrix} R(z, \mathcal{A}_{n-1}) - \beta_n \frac{R(z, \mathcal{A}_{n-1}) v_n v_n^{\mathbf{T}} R(z, \mathcal{A}_{n-1})}{z} & \beta_n \frac{R(z, \mathcal{A}_{n-1}) v_n}{z} \\ \beta_n \frac{v_n^{\mathbf{T}} R(z, \mathcal{A}_{n-1})}{z} & -\beta_n \frac{1}{z} \end{bmatrix} .$$

For convenience, I have separated the quantity β_n whenever it appears in the matrix above. From the resolvent equations, we conclude that $\beta_n = [1 - \xi_{nn}/z + v_n^{\mathbf{T}} R(z, \mathcal{A}_{n-1}) v_n / z]^{-1}$ resembles the resolvent of smaller matrices. From the matrix for $R(z, \mathcal{A}_n)$, we can obtain $\mathbf{Tr}R(z, \mathcal{A}_n)$ by summing over the diagonal elements of $R(z, \mathcal{A}_n)$. We get

$$\mathbf{Tr}R(z, \mathcal{A}_n) = \sum_{i=1}^{n} \xi_{ii} = \mathbf{Tr}R(z, \mathcal{A}_{n-1}) - \beta_n \frac{v_n^{\mathbf{T}} R(z, \mathcal{A}_{n-1})^2 v_n}{z} - \beta_n \frac{1}{z} \ .$$

Factoring common terms from the last two expressions gives

$$\mathbf{Tr}R(z, \mathcal{A}_{n-1}) - \frac{\beta_n}{z} \left(v_n^{\mathbf{T}} R(z, \mathcal{A}_{n-1})^2 v_n + 1 \right) \ .$$

With this expression for $\mathbf{Tr}R(z, \mathcal{A}_n)$, we can now obtain the expression given in the statement of **Lemma 2** by substituting in the value of β_n and simplifying

[8] See References list on the last page.

[9] This follows the beginning of *Arnold*'s computations from *Deterministic Version of Wigner's Semicircle Law for the Distribution of Matrix Eigenvalues* on page 189.

$$
\begin{aligned}
&\Rightarrow \mathbf{Tr}R(z,\mathcal{A}_{n-1}) - \frac{[1-\frac{\xi_{nn}}{z}+v_n^{\mathbf{T}}R(z,\mathcal{A}_{k-1})\frac{v_n}{z}]^{-1}}{z}\left(v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})^2v_n+1\right)\\
&= \mathbf{Tr}R(z,\mathcal{A}_{n-1}) - \frac{1}{z[1-\frac{\xi_{nn}}{z}+v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})\frac{v_n}{z}]}\left(v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})^2v_n+1\right)\\
&= \mathbf{Tr}R(z,\mathcal{A}_{n-1}) - \frac{1}{z-\xi_{nn}+v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})v_n}\left(v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})^2v_n+1\right).
\end{aligned}
$$

Although combining terms with $(z-\xi_{kk}+v_k^{\mathbf{T}}R(z,\mathcal{A}_{k-1})v_k)$ as a common denominator may appear to be the next correct step, we would not be able to simplify terms to get the desired expression for $\mathbf{Tr}R(z,\mathcal{A}_n)$. Instead, we will apply the same argument that we gave at the beginning of the page; namely, we have

$$
R(z,\mathcal{A}_{n-1}) = \begin{bmatrix} R(z,\mathcal{A}_{n-2}) - \beta_{n-1}\frac{R(z,\mathcal{A}_{n-2})v_{n-1}v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})}{z} & \beta_{n-1}\frac{R(z,\mathcal{A}_{n-2})v_{n-1}}{z} \\ \beta_{n-1}\frac{v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})}{z} & -\beta_{n-1}\frac{1}{z} \end{bmatrix},
$$

where

$$
\beta_{n-1} = \left(1-\frac{\xi_{n-1,n-1}}{z}+v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})\frac{v_{n-1}}{z}\right)^{-1}.
$$

With this, we can continue repeating the steps that I have given **above**. Evaluating $R(z,\mathcal{A}_{n-1})$ gives additional terms in our expression

$$
\mathbf{Tr}R(z,\mathcal{A}_{n-1}) - \frac{v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})^2v_{n-1}+1}{z-\xi_{n-1,n-1}+v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})v_{n-1}} - \frac{v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})^2v_n+1}{z-\xi_{nn}+v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})v_n}.
$$

As we continue iterating this procedure to evaluate each **new** *Resolvent* term, we will accumulate more terms

$$
= \left(\underbrace{\cdots}_{n-2\text{ terms for }n-2\text{ iterations.}} - \frac{v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})^2v_{n-1}+1}{z-\xi_{n-1,n-1}+v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})v_{n-1}} - \frac{v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})^2v_n+1}{z-\xi_{nn}+v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})v_n}\right).
$$

Rearranging

$$
= \frac{v_1^{\mathbf{T}}R(z,\mathcal{A}_0)^2v_1+1}{\xi_{11}-z-v_0^{\mathbf{T}}R(z,\mathcal{A}_0)v_0} + \cdots + \frac{v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})^2v_{n-1}+1}{\xi_{n-1,n-1}-z-v_{n-1}^{\mathbf{T}}R(z,\mathcal{A}_{n-2})v_{n-1}} + \frac{v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})^2v_n+1}{\xi_{nn}-z-v_n^{\mathbf{T}}R(z,\mathcal{A}_{n-1})v_n}
$$

Collecting the n terms gives the desired summation

$$
\mathbf{Tr}R(z,\mathcal{A}_n) = \sum_{k=1}^{n}\frac{1+v_k^{\mathbf{T}}R(z,\mathcal{A}_{k-1})^2v_k}{\xi_{kk}-z-v_k^{\mathbf{T}}R(z,\mathcal{A}_{k-1})v_k}.
$$

With this result, we will now apply the **Stieljes Transform** and the formula for $\mathbf{Tr}R(z,\mathcal{A}_n)$ from **Lemma 2** to appropriately quantify the *Stieljes Transform.*

2.4 Representation of the Stieljes Transform

Theorem: Let $\hat{F}_n$ be the Stieljes transform of the empirial distribution function F_n of the eigenvalues of $\mathcal{A}_n/\sqrt{n}$. Then

$$\hat{F}_n(z) = \int_{t=0}^{1} g_n(t,z)dt, \text{ where } g_n(t,z) = \sum_{k=1}^{n} X_{kn}(z)\mathbf{1}_{\{(k-1)/n,k/n\}}(t), \text{ with } t \in [0,1)$$

where $\mathbf{1}_A$ is the indicator function of the set A

$$X_{1n}(z) = \left(\frac{\xi_{11}}{\sqrt{n}} - z\right)^{-1}$$

$$\vdots$$

$$X_{kn}(z) = \frac{1 + \sqrt{\frac{k}{k-1}}\int \left[x - z\sqrt{\frac{n}{k-1}}\right]^{-2} \mathrm{d}X_k(F_{k-1}(x))}{\frac{\xi_{kk}}{\sqrt{n}} - z - \int \left[x - z\sqrt{\frac{n}{k-1}}\right]^{-1} \mathrm{d}X_k(F_{k-1}(x))}$$

for $k = 2, \cdots, n$, and X_k is the Auxiliary function that we defined on **Page 3**.

Remark: From the expression $g_n(t,z)$, $\mathbf{1}_{\{(k-1)/n,k/n\}}$ by definition is the indicator function of the interval $[(k-1)/n, k/n]$. As we take higher n, the indicator functions will include more terms. From **Definition 1** of the auxiliary function $X_n(t)$ mentioned on the first page, we can use the auxiliary function to arrive to an expression for $X_{kn}(z)$. First, we will verify what the expression for $X_{11}(z)$ is, to the express $X_{kn}(z)$. If we take $k \equiv 1$, $X_{1n}(z)$ is the **Trace** of the resolvent that we proved in **Lemma 2**. In the most simple case for $k = 1$, the only eigenvalue is the first diagonal element ξ_{11} of $\mathcal{A}_1$:

$$\mathbf{Tr}R(z, \mathcal{A}_1) = X_{1n}(z) = \sum_{k=1}^{1} \xi_{ii} = \left(\frac{\xi_{11}}{\sqrt{n}} - z\right)^{-1}$$

From our expression for $X_{1n}(z)$, it is clear that increasing k by 1 with each step will lead us to include more eigenvalues λ_k^n in our summation. To this end, we end up **repeating key arguments** presented in the proof of *Lemma 2*. The same formula for $\mathbf{Tr}R(z, \mathcal{A}_n)$ holds, with the only difference being that we take the entries of the n by n submatrix to be normalized by $\sqrt{n}$. As a result, we will replace the entries ξ_{kk} of our matrix with the $\sqrt{n}$ normalization when appropriate, which only appears in the first term in the denominator. We now only have to deal with the terms in the numerator and denominator, respectively. [10]

Remark: *With a slight change of notation*, we will use these steps a few times throughout **Proof of Theorem**. Let $\mathcal{D}$ denote the matrix of eigenvalues, and $\mathcal{U}$ denote the matrix of eigenvectors, of $\mathcal{A}_n/\sqrt{n}$. We have that

$$\begin{aligned}\int f d\mu_n &= \frac{1}{n}\sum_j f(\mathcal{D}) \\ &= \frac{1}{n}\mathbf{Tr} f(\mathcal{D}_{jj}) \\ &= \frac{1}{n}\mathbf{Tr}(\,\mathcal{U}^{-1} f(\mathcal{D})\mathcal{U}) \\ &= \frac{1}{n}\mathbf{Tr}\left(f\left(\frac{\mathcal{A}_n}{\sqrt{n}}\right)\right).\end{aligned}$$

[10] I have described another way which I think could work in **IV** of the Supplementary Section.

After these steps, it is clear that we are reiterating steps of the **Proof for Lemma 2** that I have already mentioned in this **Remark**. These computations give the same result once we expand the products and insert the normalization $\sqrt{n}$ in the last diagonal entry of $\mathcal{A}_n/\sqrt{n}$. As we take the product over the n^{th} term, we will precisely obtain the factors of $\xi_{kk}/\sqrt{n}$ and $-z$ in the denominator. Finally, we can make use of the same argument that I have given to replace the **2 resolvent terms** with integrals, in addition to replacing the random vectors $v_k^{\mathbf{T}}$ and v_k with correct square roots.

Proof of Theorem: By definition, we can compute $\mathbf{Tr}R(z,\mathcal{A}_n)$. Because $R(z,\mathcal{A}_n) = (\mathcal{A}_n - zI_n)^{-1}$, we observe that $\mathbf{Tr}R(z,\mathcal{A}_n) = \sum_{k=1}^{n}(\lambda_k^{(n)} - z)^{-1}$. [11] From here, we will slightly modify the formula for $\mathbf{Tr}R(z,\mathcal{A}_n)$, in which we will be computing the resolvent of $\mathcal{A}_n/\sqrt{n}$, and then computing the **Trace** of this resolvent. We recount these steps:

$$\frac{1}{n}\mathbf{Tr}R\Big(z, \frac{\mathcal{A}_n}{\sqrt{n}}\Big) = \frac{1}{n}\sum_{k=1}^{n}\Big(\frac{\lambda_k^{(n)}}{\sqrt{n}} - z\Big)^{-1} = \int (x-z)^{-1} dF_n(x) = \hat{F}_n(z) \ .$$

With these manipulations, we can now make use of comments that we have mentioned in the **past Remark**. Namely, we will be interested in using the expression for $\mathbf{Tr}R(z,\mathcal{A}_n)$ from **Lemma 2** to the matrix $\mathcal{A}_n$ with each of its entries normalized by $\sqrt{n}$. If we would like to compute the **Trace** of the resolvent $R(z, \frac{\mathcal{A}_n}{\sqrt{n}})$, we will replace all entries of the summation with factors of $1/\sqrt{n}$ when necessary, in addition to including factors of $\sqrt{n}$ from the resolvent terms in the numerator and denominator. For convenience, we will denote this trace as $Y_{kn}(z)$; this notation is similar with that of $X_{kn}(z)$ that we analyzed in the **Remark** before stating **Theorem 1**. We have

$$Y_{kn}(z) = \frac{1 + \frac{1}{n} v_k^{\mathbf{T}} \Big[R(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}})\Big]^2 v_k}{\frac{\xi_{kk}}{\sqrt{n}} - z - \frac{1}{n} v_k^{\mathbf{T}} R(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}}) v_k} \ .$$

With this choice of $Y_{kn}(z)$, it is clear that taking the transform of the empirical distribution of eigenvalues $F_n(x)$ gives $\hat{F}_n(z)$. From our previous steps above, $\hat{F}_n(z) = \frac{1}{n}\sum_{k=1}^{n}\big(\frac{\lambda_k^{(n)}}{\sqrt{n}} - z\big)^{-1} \Rightarrow \hat{F}_n(z) = \frac{1}{n}\sum_{k=1}^{n} Y_{kn}(z)$. From properties of the **Trace**, particularly from **Lemma 2**, we know that $Y_{kn}(z)$ is precisely the Trace of the resolvent for $\mathcal{A}_n/\sqrt{n}$ because we have replaced all of the elements with a summation over the eigenvalues $\lambda_k^{(n)}/\sqrt{n}$ for $1 \le k \le n$. From our expression of the transform $\hat{F}_n(z)$, we will set $\hat{F}_n(z) = \int_{t=0}^{1} g_n(t,z)dt$ as stated in **Theorem 1**. So far, this is acceptable because we can take a step function $g_n(t,z) = Y_{\lfloor nt \rfloor + 1, n}(z)$ for $0 \le t < 1$. With these comments, it suffices to show that $Y_{kn} = X_{kn}$ for $1 < k \le n$, which will follow once we have shown that the numerators and denominators of Y_{kn} and X_{kn} are equal. We will show equality for the denominator, rewriting each term in both denominators, for convenience, below:

$$\overbrace{\frac{\xi_{kk}}{\sqrt{n}} - z - n^{-1} v_k^{\mathbf{T}} R\Big(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}}\Big) v_k}^{\text{Denominator of } Y_{kn}(z)} = \overbrace{\frac{\xi_{kk}}{\sqrt{n}} - z - \sqrt{\frac{k}{n}} \int \Big[x - z\sqrt{\frac{n}{k-1}}\Big]^{-1} \mathrm{d}X_k(F_{k-1}(x))}^{\text{Denominator of } X_{kn}(z)}$$

$$\Longleftrightarrow n^{-1} v_k^{\mathbf{T}} R\Big(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}}\Big) v_k = \sqrt{\frac{k}{n}} \int \Big[x - z\sqrt{\frac{n}{k-1}}\Big]^{-1} \mathrm{d}X_k(F_{k-1}(x))$$

2.4.1 Denominator Case

Observe that the two first terms in the denominators of Y_{kn} and X_{kn} are already the same, so we have to show that the remaining last term in the denominator is equal to finish a significant portion of the proof. We will show that these terms are equal [12], multiplying and dividing by $\sqrt{\frac{n}{k-1}}$:

[11] For reference, I have made use of *Arnold*'s proof of **Lemma 2** on page 189. To make the computations more clear, I have added more steps in. They include more detailed calculations of the factor of $(k(k-1))^{\frac{1}{2}}$ which occurs in **Definition 1** of the Auxiliary Function $X_n(t)$.

[12] I have concentrated on the steps in *Arnold*'s calculations, from page 191.

$$\xrightarrow[\text{from the denominator of } Y_{kn}(z)]{\text{Starting}} n^{-1} v_k^{\mathbf{T}} R\bigg(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}}\bigg) v_k$$
$$= \frac{\sqrt{\frac{n}{k-1}}}{\sqrt{\frac{n}{k-1}}}\bigg(n^{-1} v_k^{\mathbf{T}}\bigg[R\bigg(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}}\bigg)\bigg] v_k\bigg)$$
$$= \frac{\sqrt{\frac{n}{k-1}}}{\sqrt{\frac{n}{k-1}}}\bigg(n^{-1} v_k^{\mathbf{T}}\bigg[\frac{1}{\frac{\mathcal{A}_{k-1}}{\sqrt{n}} - zI_{k-1}}\bigg] v_k\bigg)$$
$$= n^{-1}\sqrt{\frac{n}{k-1}} v_k^{\mathbf{T}} \frac{1}{\sqrt{\frac{n}{k-1}}}\bigg[\frac{1}{\frac{\mathcal{A}_{k-1}}{\sqrt{n}} - zI_{k-1}}\bigg] v_k$$
$$= n^{-1}\sqrt{\frac{n}{k-1}} v_k^{\mathbf{T}}\bigg[\frac{1}{\sqrt{\frac{n}{k-1}}\frac{\mathcal{A}_{k-1}}{\sqrt{n}} - \sqrt{\frac{n}{k-1}} zI_{k-1}}\bigg] v_k$$
$$= n^{-1}\sqrt{\frac{n}{k-1}} v_k^{\mathbf{T}}\bigg[\frac{1}{\frac{\mathcal{A}_{k-1}}{\sqrt{k-1}} - \sqrt{\frac{n}{k-1}} zI_{k-1}}\bigg] v_k$$
$$= n^{-1}\sqrt{\frac{n}{k-1}} v_k^{\mathbf{T}}\bigg[R\bigg(z\sqrt{\frac{n}{k-1}}, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}}\bigg)\bigg] v_k\ .$$

Remark: I have fleshed out the computations that Arnold gives. He directly states that

$$n^{-1} v_k^{\mathbf{T}} R\bigg(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}}\bigg) v_k = n^{-1}\sqrt{\frac{n}{k-1}} v_k^{\mathbf{T}} R\bigg(z\sqrt{\frac{n}{k-1}}, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}}\bigg) v_k\ .$$

Although it may not be rigorously appropriate to "invert" the *resolvent* terms that we are given, I have illustrated the computation to show where all of the factors come from in the expression. The additional steps that I have given directly coincide with the result that he gives on **Page 190**.

We will now apply a similar result from a previous step that we stated near the beginning of the proof. We can express the *resolvent* $R\bigg(z\sqrt{\frac{n}{k-1}}, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}}\bigg)$ as an integral to rearrange terms further. Intuitively, we are adjusting terms slightly from the integral $\int (x-z)^{-1} dF_n(x)$ that we have given, in which we have that the entries of $\mathcal{A}_{k-1}$ are normalized by $\sqrt{k-1}$ from the resolvent that we have obtained above. Furthermore, we will include the factor of $\sqrt{\frac{n}{k-1}}$, in front of z, that we obtained after distributing $\sqrt{\frac{n}{k-1}}$ to get $R\bigg(z\sqrt{\frac{n}{k-1}}, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}}\bigg)$, again from previous steps. We multiply and divide by $\sqrt{\frac{k}{n}}$

$$\Rightarrow \frac{\sqrt{\frac{k}{n}}}{\sqrt{\frac{k}{n}}}\bigg(n^{-1}\sqrt{\frac{n}{k-1}} v_k^{\mathbf{T}}\bigg[R\bigg(z\sqrt{\frac{n}{k-1}}, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}}\bigg)\bigg] v_k\bigg)$$
$$= \sqrt{\frac{k}{n}}\bigg(\frac{1}{\sqrt{k(k-1)}} v_k^{\mathbf{T}} R\bigg(z\sqrt{\frac{n}{k-1}}, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}}\bigg) v_k\bigg)\ .$$

We can relate the terms in the parentheses to the *Stieljes Transform*. From properties of the *delta function*, [13] we can set these terms to the following transform:

[13]... from the property that integrating delta function gives a piecewise defined function.

$$\underbrace{\frac{1}{\sqrt{k(k-1)}} v_k^{\mathbf{T}} R\Big(z\sqrt{\frac{n}{k-1}}, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}}\Big) v_k}_{\int \left(\frac{x}{\sqrt{k-1}} - z\sqrt{\frac{n}{k-1}}\right)^{-1} d(k(k-1)^{-\frac{1}{2}} v_k^{\mathbf{T}} \mathcal{S}_{k-1}(x) v_k)}$$

In this equality, $\mathcal{S}_{k-1}$ is the **Spectral Matrix** of $\mathcal{A}_{k-1}$. By definition, the spectral matrix consists of the elements

$$s_{ij}^{(k-1)}(x) = \sum_{\lambda_r^{(k-1)} \leq x} u_{ir}^{(k-1)} \bar{u}_{jr}^{(k-1)} \in \mathcal{S}_{k-1} \ .$$

We will proceed with the calculation by rearranging terms further in $(k(k-1))^{-\frac{1}{2}} v_k^{\mathbf{T}} \mathcal{S}_{k-1}(x) v_k$. From the definition of the **Spectral Matrix**:

$$\begin{aligned} &\to (k(k-1))^{-\frac{1}{2}} v_k^{\mathbf{T}} \mathcal{S}_{k-1}(x) v_k \\ &= (k(k-1))^{-\frac{1}{2}} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} a_{ki} s_{ij}(x\sqrt{k-1}) a_{jk} \\ &= (k(k-1))^{-\frac{1}{2}} \sum_{\lambda_r^{(k-1)}/\sqrt{k-1} \leq x} \Big(\sum_{i=1}^{k-1} a_{ki} u_{ir}^{(k-1)} \bar{a}_{ki} \bar{u}_{jr}^{(k-1)} \Big) \\ &= (k(k-1))^{-\frac{1}{2}} \sum_{\lambda_r^{(k-1)}/\sqrt{k-1} \leq x} |v_k^{\mathbf{T}} u_r^{(k-1)}|^2 \\ &= X_k(F_{k-1}(x)) \ . \end{aligned}$$

These steps confirm that

$$n^{-1} v_k^{\mathbf{T}} R\Big(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}}\Big) = \sqrt{\frac{k}{n}} \int \Big(x - z\sqrt{\frac{n}{k-1}}\Big)^{-1} dX_k(F_{k-1}(x)) \ .$$

Remark: This equality relates to the form that $X_{kn}(z)$ takes, which we discussed. $dX_k(F_{k-1}(x))$ comes from the computations once we observe that the *Resolvent* can be related to the *Stieljes Transform*.

2.4.2 Numerator Case

The arguments for terms in the numerators of $Y_{kn}(z)$ and $X_{kn}(z)$ being equal can be treated analogously. In the case for the numerators, we begin by checking if there are already any terms that are equal. As in the previous case for the denominators of $Y_{kn}(z)$ and $X_{kn}(z)$, presumably the resolvent term $\frac{1}{n} v_k^{\mathbf{T}} [R(z, \mathcal{A}_{k-1}/\sqrt{n})]^2 v_k$ should be equal to $\sqrt{k/(k-1)} \int \left[x - z\sqrt{n/(k-1)}\right]^{-2} dX_k(F_{k-1}(x))$ if the numerators are to equal one another, because 1 is trivially equal to itself. To this end, we can rearrange terms [14] to show that this equality holds:

[14] **These** steps allow us to treat the case of the numerators. We will follow the computation that I have already given along similar lines.

$$
\begin{aligned}
&\xrightarrow[\text{from the numerator of } Y_{kn}(z)]{\text{Starting}} \frac{1}{n} v_k^{\mathbf{T}} \left[R\left(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}} \right) \right]^2 v_k \\
&= \frac{\sqrt{\frac{n}{k-1}}}{\sqrt{\frac{n}{k-1}}} \left(\frac{1}{n} v_k^{\mathbf{T}} \left[R\left(z, \frac{\mathcal{A}_{k-1}}{\sqrt{n}} \right) \right]^2 v_k \right) \\
&= \frac{\sqrt{\frac{n}{k-1}}}{\sqrt{\frac{n}{k-1}}} \left(\frac{1}{n} v_k^{\mathbf{T}} \left[\frac{1}{\frac{\mathcal{A}_{k-1}}{\sqrt{n}} - z I_{k-1}} \right]^2 v_k \right) \\
&= \frac{\sqrt{\frac{n}{k-1}}}{n} v_k^{\mathbf{T}} \frac{1}{\sqrt{\frac{n}{k-1}}} \left[\frac{1}{\frac{\mathcal{A}_{k-1}}{\sqrt{n}} - z I_{k-1}} \right] \left[\frac{1}{\frac{\mathcal{A}_{k-1}}{\sqrt{n}} - z I_{k-1}} \right] v_k \\
&= \frac{1}{n} v_k^{\mathbf{T}} \underbrace{\left[\frac{1}{\sqrt{\frac{n}{k-1}} \frac{\mathcal{A}_{k-1}}{\sqrt{n}} - \sqrt{\frac{n}{k-1}} z I_{k-1}} \right]}_{\underbrace{\frac{1}{\frac{\mathcal{A}_{k-1}}{\sqrt{k-1}} - \sqrt{\frac{n}{k-1}} z I_{k-1}}}_{R\left(\sqrt{\frac{k-1}{n}} z, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}} \right)}} \underbrace{\left[\frac{\sqrt{\frac{n}{k-1}}}{\frac{\mathcal{A}_{k-1}}{\sqrt{n}} - z I_{k-1}} \right]}_{\underbrace{\frac{1}{\sqrt{\frac{k-1}{n}} \frac{\mathcal{A}_{k-1}}{\sqrt{n}} - \sqrt{\frac{n}{k-1}} z I_{k-1}}}_{R\left(\sqrt{\frac{n}{k-1}} z, \frac{\mathcal{A}_{k-1}\sqrt{k-1}}{n} \right)}} v_k \ .
\end{aligned}
$$

We would now like to see if we can combine the two *Resolvent* terms to get a square. From the second term $R\left(\sqrt{\frac{n}{k-1}} z, \frac{\mathcal{A}_{k-1}\sqrt{k-1}}{n}\right)$, we can multiply the entries of $\mathcal{A}_{k-1}$ with an appropriate normalization so that both of the resolvent terms are identical, namely $R\left(\sqrt{\frac{n}{k-1}} z, \frac{\mathcal{A}_{k-1}}{\sqrt{k-1}} \right)$. We have:

$$
\begin{aligned}
\Rightarrow \frac{1}{n} v_k^{\mathbf{T}} R\left(z, \sqrt{\frac{n}{k-1}} \mathcal{A}_{k-1} \right) R\left(z, \sqrt{\frac{n}{k-1}} \mathcal{A}_{k-1} \right) \frac{n}{(k(k-1))^{\frac{1}{2}}} v_k \\
= \frac{1}{n} v_k^{\mathbf{T}} R\left(z, \sqrt{\frac{n}{k-1}} \mathcal{A}_{k-1} \right)^2 \frac{n}{(k(k-1))^{\frac{1}{2}}} v_k \\
= R\left(z, \sqrt{\frac{n}{k-1}} \mathcal{A}_{k-1} \right)^2 \frac{v_k^{\mathbf{T}} v_k}{(k(k-1))^{\frac{1}{2}}} \ .
\end{aligned}
$$

The numerator case told us that

$$
\frac{v_k^{\mathbf{T}} \mathcal{S}_{k-1} v_k}{(k(k-1))^{\frac{1}{2}}} = X_k(F_{k-1}(x)) \ .
$$

Replacing $\dfrac{v_k^{\mathbf{T}} \mathcal{S}_{k-1} v_k}{(k(k-1))^{\frac{1}{2}}}$ gives

$$
\int \left(x - \sqrt{\frac{n}{k-1}} z \right)^{-1} dX_k(F_{k-1}(x)) \ .
$$

From other steps that have been included thus far, we will **repeat the same rearrangements** that I have included in the denominator case. Namely, we can make use of properties of the *spectral matrix* $\mathcal{S}_{k-1}$ to rearrange terms to show that $k(k-1)^{-\frac{1}{2}} v_k^{\mathbf{T}} \mathcal{S}_{k-1}(x) v_k \equiv X_k(F_{k-1}(x))$. We have reached this goal in the last few steps that I have listed above. At this point, the computations in the numerator case are identical with those in the denominator case. With this result, we have that the numerator and denominator of $Y_{kn}(z)$ and $X_{kn}(z)$ are equal, affirming that $\hat{F}_n(z) = \int_{t=0}^{1} g_n(t,z)dz$, where $g_n(z)$ is the step function that we have given. This completes the **Proof**.

Consequences of Theorem 1: From our computations in the separate cases of the numerator and denominator, we have that the factor $(k(k-1))^{\frac{1}{2}}$ is the correct normalization for the **Auxiliary Function** that we introduced in **Definition 1**, on **Page 1**. Also, it establishes some of the reasons for introducing the *Stieljes Transform*. It plays an important role in allowing us to determine that terms in the numerator and denominator of $X_{kn}(z)$ and $Y_{kn}(z)$ can be expressed with integrals.

3 More Results before Main Theorem

3.1 Additional Lemmas

We will now state a few results before getting to the **Main Theorem**. Indeed, we will make use of the independence of the random entries $\{\xi_{ij}\}$. In particular, the entries being independent implies that $v_n^{\mathbf{T}}$ and the eigenvectors of $\mathcal{A}_{n-1}$ are independent.

Remark: We will study the *Stochastic Convergence* of $X_n(t)$. We will state **Definition 1** and **Lemmas 3, 3***. This collection of statements will be useful for the remaining part of the proof of the **Main Theorem**.

Throughout this section, we denote H as a *distribution function*. We will use H in the statement of the following **Lemma**.

3.1.1 Lemma 3

We have

$$\lim_n X_n(1) = \mathcal{C} < \infty \Longleftrightarrow \int x^2 dH < \infty \, .$$

If $\int x dH = 0$ and $\sigma^2 < \infty$, then $\mathcal{C} = \sigma^2$ and

$$\lim_n X_n(t) = \sigma^2 t \text{ in probability, } 0 \leq t \leq 1$$

Proof of 3: From **Definition 1**, we have an expression for the *Auxiliary Function*. After setting $t = 1$ in $X_n(t)$, we observe that the first assertion holds by an argument with the **Weak Law of Large Numbers**. We claim that

To prove the second assertion, we will define the entries $\bar{a}_{kn}$:

$$\bar{a}_{kn} = \begin{cases} a_{kn} \text{ for } |a_{kn}| \leq \sqrt{n-1} \\ 0 \text{ for } |a_{kn}| > \sqrt{n-1} \end{cases}$$

for $1 \leq k \leq n-1$. The $\bar{a}_{kn}$ are useful because we can obtain $\bar{X}_n(t)$ in an obvious way; we replace a_{kn} with $\bar{a}_{kn}$ for $1 \leq k \leq n-1$. This implies:

$$\mathbf{P}[|X_n(t) - \mathbf{E}\bar{X}_n(t)| > \epsilon] \leq \mathbf{P}[|\bar{X}_n(t) - \mathbf{E}\bar{X}_n(t)| > \epsilon] + \mathbf{P}[X_n(t) \neq \bar{X}_n(t)]$$

We can show that the right hand side of the inequality approaches 0:

$$\mathbf{P}[X_n(t) \neq \bar{X}_n(t)] \leq \sum_{k=1}^{n-1} \mathbf{P}[|a_{kn}| > \sqrt{n-1}]$$
$$= (n-1)\mathbf{P}[|a_{12}| > \sqrt{n-1}] \longrightarrow 0$$

The last inequality follows from the fact that $\sum \mathbf{P}[|a_{12}| > \sqrt{n}] < \infty$. This case happens iff $\sigma^2 < \infty$. By *Chebyshev's Inequality*:

$$\mathbf{P}[|\bar{X}_n(t) - \mathbf{E}\bar{X}_n(t)| > \epsilon] \leq \epsilon^{-2}(\mathbf{E}\bar{X}_n(t)^2 - (\mathbf{E}\bar{X}_n(t))^2)$$

From this inequality, **Proof of 3** will be complete if $\mathbf{E}\bar{X}_n(t) \longrightarrow \sigma^2 t$ and $(\mathbf{E}\bar{X}_n(t))^2 \longrightarrow \sigma^4 t^2$. **We will denote this as (1)** We know that:

$$\mathbf{E}\bar{X}_n(t) = (n(n-1))^{-\frac{1}{2}} \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \mathbf{E}(\bar{a}_{in}\bar{a}_{jn})\mathbf{E}(r_{ij})$$

In the summation:

$$r_{ij} = r_{ij}(t) = \sum_{k=1}^{\lfloor (n-1)t \rfloor} u_{ik}^{(n-1)} u_{jk}^{(n-1)}$$

We will substitute the relations

$$\sum_{i=1}^{n-1} r_{ii} = [(n-1)t]$$
$$\mathbf{E}\bar{a}_{in}\bar{a}_{jn} = \mathbf{E}\bar{a}_{in}^2 = \bar{m}^2 \text{ for } i = j$$
$$\mathbf{E}\bar{a}_{in}\bar{a}_{jn} = \bar{m}_1^2 \text{ for } i \neq j$$

into the summation from $\mathbf{E}\bar{X}_n(t)$. After substituting, we get:

$$\mathbf{E}\bar{X}_n(t) = (n(n-1))^{-\frac{1}{2}}([(n-1)t](\bar{m}_2 - \bar{m}_1^2) + \bar{m}_1^2 \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \mathbf{E}r_{ij}$$

By assumption, $\bar{m}_2 \longrightarrow \sigma^2 < \infty$ and $\bar{m}_1 = \mathrm{o}(n^{-\frac{1}{2}})$. From *On the asymptotic distribution of the eigenvalues of random matrices*, page 265, by Arnold, we know that .

Together, we can get $\mathbf{E}\bar{X}_n(t) \longrightarrow \sigma^2 t$. We can substitute the quantities from Arnold's statement into $\mathbf{E}(\bar{X}_n(t))$. With the trivial estimate, $|\sum\sum \mathbf{E}r_{ij}| \leq (n-1)^2$, we achieve part of (1). We will now estimate $(\mathbf{E}\bar{X}_n(t))^2$:

$$(\mathbf{E}\bar{X}_n(t))^2 = (n(n-1))^{-1} \sum_{i_1=1}^{n-1} \cdots \sum_{i_4=1}^{n-1} \mathbf{E}(\bar{a}_{i_1 n}\bar{a}_{i_2 n}\bar{a}_{i_3 n}\bar{a}_{i_4 n})\mathbf{E}(r_{i_1 i_2} r_{i_3 i_4})$$

From the indices of the summation, we have to consider 7 different cases of index degeneracy. From what we stated earlier, $\bar{m}_1 = \mathrm{o}(n^{-\frac{1}{2}})$, $\bar{m}_2 \longrightarrow \sigma^2$. Also, we even have that $\bar{m}_3 = \mathrm{o}(n^{\frac{1}{2}}$ and $\bar{m}_4 = \mathrm{o}(n)$. This follows from:

$$\int |x|^k dH < \infty \Rightarrow n^{\frac{k-r}{2}} \int_{|x| \leq \sqrt{n}} |x|^r dH = \mathrm{o}(1) \text{ for } r \geq k+1$$

From *Arnold's* computation on *page 264* of the same paper, we have that $(\mathbf{E}\bar{X}_n(t))^2 = \sigma^4 t^2 + \mathrm{o}(1)$. This completes **Proof of 3**.

Besides explicitly stating what $\mathcal{C}$ is, we will now return towards our goal of proving the **Main Theorem**. [15] We will state **Lemma 4**:

3.1.2 Lemma 4

Suppose that the conditions of the **Main Theorem** hold [16]. Then

(i) The sequence $\{g_n\}$ is sequentially compact in the space $\mathcal{B}([0,1])$ of bounded Borel measurable functions on $[0,1)$ endowed with the supremum topology.

(ii) All limit points of $\{g_n\}$ are continuous.

Proof of Lemma 4:
(i) follows from the fact that a bounded set $\mathcal{K} \subset \mathcal{B}([0,1])$ is *conditionally compact* (*sequentially compact* in metric spaces) if for every $\epsilon > 0$ there is a finite collection $\{S_1, \cdots, S_k\}$ of disjoint measurable sets with union $[0,1)$, and points $t_i \in S_i$, such that

$$\max_{i=1,\cdots,N} \sup_{t_i \in S_i} |g(t) - g(t_i)| < \epsilon \text{ for all } g \in K$$

We know that $K = \{g_n\}$ is bounded because $||g_n|| < \frac{1}{\mathrm{Im}(z)}$. The inequality follows from **Lemma 3***:

3.1.3 Lemma 5

Let $g(z) = \hat{\mu}(\alpha z)$, $\alpha > 0$, μ being a finite positive measure. Then for $h(z) = \frac{1+g'(z)}{\beta+z+g(z)}$, β real, we have:

$$|h(z)| \leq \frac{1}{\mathrm{Im}(z)} \;, |h'(z)| \leq \frac{3}{[\mathrm{Im}(z)]^2} \;.$$

Proof of 5: We have

$$\mathrm{Im}(g(z)) = \alpha \mathrm{Im}(z) \int |x - az|^{-2} d\mu \;,$$

and

[15] *Arnold* on **Page 193** of *On Wigner's Semicircle Law for the Eigenvalues of Random Matrices* also describes Almost Sure Convergence of $X_n(t)$. This convergence claims that $\lim_n X_n(1) = \mathcal{C} < \infty$ a.s. $\Leftrightarrow \int x^4 dH < \infty$. A similar conclusion to what is given in **Lemma 3** follows.

[16] This corresponds to *Arnold*'s Lemma 3 on page 191 of *Deterministic Version of Wigner's Semicircle Law.*

$$|\beta + z + g(z)| \geq \mathrm{Im}(z) + \mathrm{Im}(g(z)) \ ,$$
$$|1 + g'(z)| \leq 1 + \alpha \int |x - \alpha z|^{-2} d\mu = 1 + \frac{\mathrm{Im}(g(z))}{\mathrm{Im}(z)}$$

We observe that:

$$|g^{''}(z)| \leq 2\alpha^2 \int |x - az|^{-3} d\mu \leq \frac{2\mathrm{Im}(g(z))}{[\mathrm{Im}(z)]^2}$$

This gives:

$$|h(z)| \leq \frac{1}{\mathrm{Im}(z)}$$

Finally, differentiating the inequality gives:

$$|h'(z)| \leq \frac{3}{[\mathrm{Im}(z)]^2}$$

We will return to **Proof of Lemma 3**. With the result $||g_n|| < \frac{1}{\mathrm{Im}(z)}$, we will now subdivide $[0, 1)$ with the property that **(i)** gives. By the definition of compactness, the statements that we have listed will also hold uniformly in compact z sets.

From a previous result in **Theorem** on **Page 8**, we know what $X_{kn}(z)$ looks like. With $g_n(t, z) = X_{\lfloor nt \rfloor + 1, n}(z)$, we can calculate $g_n(0)$. We get

$$g_n(0) = X_{1n}(z) = \left(\frac{\xi_{11}}{\sqrt{n}} - z \right)^{-1} \longrightarrow -\frac{1}{z} \ .$$

We consider a neighborhood of $t = 0$. For every given $\epsilon > 0$, we can choose $n_1 = n_1(\epsilon)$ and $\tau = \tau(\epsilon)$ such that

$$\mathrm{sup}_{0 \leq t \leq \tau} |g_n(0) - g_n(t)| < \epsilon \text{ for all } n \geq n_1.$$

From properties of the *Stieljes Transform*, we achieve the following inequality:

$$\begin{aligned} \left| \int \left(x - z\sqrt{\frac{n}{\lfloor nt \rfloor}} \right)^{-1} dX_{\lfloor nt \rfloor + 1}(F_{\lfloor nt \rfloor}(x)) \right| &\leq \int \left| \left(x - z\sqrt{\frac{n}{\lfloor nt \rfloor}} \right)^{-1} \right| dX_{\lfloor nt \rfloor + 1}(F_{\lfloor nt \rfloor}(x)) \\ &= \int \left| \frac{1}{\left(x - z\sqrt{\frac{n}{\lfloor nt \rfloor}} \right)} \right| dX_{\lfloor nt \rfloor + 1}(F_{\lfloor nt \rfloor}(x)) \end{aligned}$$

From the properties of the *Stieljes Transform* that we have listed, it is possible to bound each of the terms separately to conclude that

$$\left| \int \left(x - z\sqrt{\frac{n}{\lfloor nt \rfloor}} \right)^{-1} dX_{\lfloor nt \rfloor+1}(F_{\lfloor nt \rfloor}(x)) \right| \leq \sqrt{\frac{\lfloor nt \rfloor}{n}} \frac{X_{\lfloor nt \rfloor+1}(1)}{\text{Im}(z)} .$$

Furthermore, from our first assumption of the **Main Theorem**, we also know that there exists a constant $\mathcal{C}_1$ so that $X_n(1) \leq \mathcal{C}_1$. For all n we have

$$\sqrt{\frac{\lfloor nt \rfloor + 1}{n}} \left| \int \left(x - z\sqrt{\frac{n}{\lfloor nt \rfloor}} \right)^{-1} dX_{\lfloor nt \rfloor+1}(F_{\lfloor nt \rfloor}(x) \right| \leq \mathcal{C}_1 \frac{t + \frac{2}{n}}{\text{Im}(z)} .$$

From the second assumption, we can choose another constant $\mathcal{C}_2$ so that $|\xi_{kk}/\sqrt{n}| \leq \mathcal{C}_2$ for all n. This implies that

$$\left| \frac{\xi_{\lfloor nt \rfloor+1, \lfloor nt \rfloor+1}}{\sqrt{n}} \right| \leq \mathcal{C}_2 \left(t + \frac{2}{n} \right)^{\frac{1}{2}} .$$

From the past 2 inequalities, we can make the Right hand side of each arbitrarily small by choosing n large and t small. With a similar argument, we can treat the case of the numerator. Together, these arguments give that $g_n(t)$ and $g_n(0)$ are sufficiently close to $-\frac{1}{z}$ for small t and large n.

We will now return to the remaining interval $(\tau, 1)$. We know that $X_n(t) \longrightarrow \mathcal{C}t$ uniformly in $[0, 1]$, because the limit function is continuous. This gives

$$X_n(F_{n-1}(x)) - \mathcal{C}F_{n-1}(x) \longrightarrow 0 .$$

Observe that the convergence is uniform in x. With these computations, we will also apply the **Fundamental Theorem of Grommer and Hamburger**. It allows us to conclude that $\hat{\mu}_n(z), \hat{\mu}_{n'}(z) \longrightarrow 0$, where $\{n'\}$ is a subsequence of $\{n\}$. The convergence is uniform in compact z sets. To do this, l=et $\mu_n \equiv X_n(F_{n-1}(x)) - \mathcal{C}F_{n-1}(x)$ be a finite measure [17]. We know that the total variation $||\mu_n||$ of μ_n is bounded because $||\mu + n|| \leq X_n(1) + \mathcal{C}$. This gives that $\hat{\mu}_n(z) \longrightarrow 0$, and that $\hat{\mu}'_z(z) \longrightarrow 0$. Then, we can pick $n_2 = n_2(\epsilon)$ so that

$$\text{sup}_{\tau < t < 1} |g_n(t, z) - \bar{g}_n(t, z)| < \frac{\epsilon}{2} \text{ for all } n \geq n_2 ,$$

where

$$\begin{aligned} \bar{g}_n(t, z) &= -\frac{1 + \mathcal{C} \displaystyle\int (x - z)^{-2} dF_{\lfloor nt \rfloor}(x)}{z + \mathcal{C}\sqrt{t} \displaystyle\int (x - \frac{z}{\sqrt{t}})^{-1} dF_{\lfloor nt \rfloor}(x)} \\ &= -\frac{1 + \mathcal{C}\hat{F}'_{\lfloor nt \rfloor}(z/\sqrt{t})}{z + \mathcal{C}\sqrt{t}\hat{F}_{\lfloor nt \rfloor}(x/\sqrt{t})} . \end{aligned}$$

Finally, we will prove that $\{\bar{g}_n\}$ has the property, namely that it is sequentially compact in $\mathcal{B}([0, 1])$. With previous results in **Lemmas 3,4**, we know what $\textbf{Tr}R(z, \mathcal{A}_k)$ is. Also, we have bounds for $|h(z)|$ and $|h'(z)|$. We can apply **Lemma**

[17] We will want to apply **Properties 1,2,3** of the *Stieljes Transform*.

4 because, after all, $h(z)$ that we defined resembles the expression for the **Trace** of the Resolvent that we proved earlier. Altogether, we will then

$$\hat{F}(z) - \hat{F}_1(z) = n^{-1} \frac{af'(z)}{1 + af(z)} .$$

is similar to the expression of the
This implies that

$$\hat{F}_n(z) = \hat{F}_{n-1}(z) + \frac{1}{n}\mathrm{O}(1) + \mathrm{O}\left(\frac{1}{n^2}\right) .$$

uniformly in compact z sets. Again, **properties** of the *Stieljes Transform* give that, for $\tau \leq t \leq s \leq 1$

$$\begin{aligned} \hat{F}_{\lfloor ns \rfloor}(z) - \hat{F}_{\lfloor nt \rfloor}(z) &= \mathrm{O}(1) \sum_{\lfloor nt \rfloor + 1}^{\lfloor ns \rfloor} \frac{1}{k} + \mathrm{o}(1) \\ &= \mathrm{O}(1) \log\left(\frac{s}{t}\right) + \mathrm{o}(1) . \end{aligned}$$

for $\tau \leq t \leq s \leq 1$. Now, we will choose $n_3 = n_3(\epsilon)$ such that $\mathrm{o}(1)$ is sufficiently small. From the expression

$$\mathrm{O}(1) \log\left(\frac{s}{t}\right) + \mathrm{o}(1)$$

we can pick a neighborhood so that

$$\hat{F}_{\lfloor ns \rfloor}\left(\frac{z}{\sqrt{s}}\right) - \hat{F}_{\lfloor nt \rfloor}\left(\frac{z}{\sqrt{t}}\right)$$

can be made arbitrarily small for all $n \geq n_3$. The same observations and argument apply to $\hat{F}'_{\lfloor nt \rfloor}$. Finitely many of these neighborhoods cover $(\tau, 1)$, so that there are finitely many disjoint intervals $\mathcal{I}_1, \cdots, \mathcal{I}_k$ with union $(\tau, 1)$ and points $t \in \mathcal{I}_k$ satisfying

$$\mathrm{max}_{i=1,\cdots,n} \mathrm{sup}_{t \in \mathcal{I}_i} |\bar{g}_n(t) - \bar{g}_n(t_i)| < \frac{\epsilon}{2} \text{ for all } n \geq n_3.$$

We can continue getting a new decomposition of $[0, 1)$ by adding points k/p to a subinterval $1 \leq k < p < n_0 = \mathrm{Max}(n_1, n_2, n_3)$. The desired inequality continues to hold for all n.

(ii) follows from the statement of equicontinuity. Let $g_n \longrightarrow g$. We have that

$$|g(t) - g(s)| \leq |g(t) - g_{n'}(t)| + |g(s) - g_{n'}(s) + |g_{n'}(t) - g_{n'}(s)| .$$

From the difference, the first and second terms can be made arbitrarily small for large n'. Similarly, the third term can be made arbitrarily small for small $|t - s|$, from a previous result $\hat{F}_{\lfloor ns \rfloor}(z) - \hat{F}_{\lfloor nt \rfloor}(z)| = \mathrm{O}(1)\log(\frac{s}{t}) + \mathrm{o}(1)$. For **(ii)**, this proves that g is continuous.

- Note for myself

We will apply the definition of *equicontinuity*, along with sections of the proof of the *Arzela-Ascoli Theorem*; a version that I am referring to is given on **Page 312** of the Way of Analysis. A main idea of the proof that is given here is to pick elements along the diagonal to get a *subsequence* that converges at all points x_k of a countable dense subset of the domain. Therefore, continuing to take diagonal subsequences of $\{f_{11}, f_{22}, f_{33}, \cdots\}$ gives a subsequence that converges at all of these points at the same time.

From the diagonal subsequence $\{f_{11}, f_{22}, f_{33}, \cdots\}$, any subsequence that is chosen after the k^{th} row $f_{k1}, f_{k2}, f_{k3}, \cdots$ is a subsequence of the k^{th} row. **For the second step**, we will break up the difference $|f_j(x) - f_k(x)|$ with a standard argument using the definitions of equicontinuity, etc to control $|f_j(x) - f_j(x_p)|$, $|f_j(x_p) - f_k(x_p)|$, and $|f_k(x_p) - f_k(x)|$. Change variables from this statement to match what we want to show in the **Main Theorem**.

3.2 Main Theorem

Main Theorem, Wigner Semicircle Law: Let $\mathcal{A}$ be an ∞ by ∞ matrix as described on **Page 1**. That is, we choose each ξ_{ij} at random such that the entries have mean 0 and variances σ and 0 when $i = j$ and $i \neq j$, respectively. From properties of the *Stochastic Convergence* of $X_n(t)$ in **Lemma 3**, it is natural to study the limit of $X_n(t)$ as $n \longrightarrow \infty$. Suppose that

$$\lim_{n\longrightarrow\infty} X_n(t) = \mathcal{C}t,$$

and

$$\lim_{n\longrightarrow\infty} \frac{\xi_{kk}}{\sqrt{n}} = 0.$$

From **Lemma 3**, we know that $\mathcal{C} = \sigma^2$. [18] From **conditions i, ii**, we conclude that

$$\lim_{n\longrightarrow\infty} F_n(x) = W(x, \mathcal{C})$$

where W is absolutely continuous with density

$$w(x, \mathcal{C}) = \begin{cases} (2\mathcal{C})^{-1}(4\mathcal{C} - x^2)^{\frac{1}{2}} & \text{for } |x| \leq 2\sqrt{\mathcal{C}} \\ 0 & \text{for } |x| > 2\sqrt{\mathcal{C}}\ . \end{cases}$$

Remark: We can interpret **Condition (i)** geometrically. It means that

$$X_n(1) = (n(n-1))^{\frac{1}{2}} \sum_{k=1}^{n-1} |\xi_{kn}|^2 \longrightarrow \mathcal{C}\ .$$

The length of all projections of the new row $v_n^{\mathbf{T}}$ onto the eigenvectors $u_k^{(n-1)}$ of $\mathcal{A}_{n-1}$ must be asymptotically independent of k and equal to $\mathcal{C}$.

In the last **Proof**, we will have to solve an integro-differential equation after a few manipulations. We can do this by introducing a change of variables $h(t, z)$. With **Definition 1** and **Lemmas 3,3***, we will form a convergence subsequence. Then, we will define $h(t, z)$ to solve a differential equation for the *Wigner Distribution*. [19]

[18] Recall that $\mathcal{C} < \infty$ and $0 \leq t \leq 1$.

[19] The definition of $h(t, z)$ is given on **Page 195** of *Arnold*.

Proof of Main Theorem: We will use the definition of compactness, applied to the sequence of distribution functions $\{F_n\}$. From our first remark following the proof, we know that the compact subsequence $\{n'\} \subset \{n\}$ exists and that the applying the *Stieljes Transform* preserves convergence, ie $\hat{F}_{n'} \longrightarrow \hat{F}$ as $n \longrightarrow \infty$. But we know from **Theorem 1** that the Stieljes Transform of F_n is $\hat{F}_n(z) = \int_{t=0}^{1} g_n(t,z)\mathrm{d}t$. As in our previous statement, $g_n(t,z) = \sum_{k=1}^{n} X_{kn}(z)\mathbf{1}_{[(k-1)/n,k/n]}(t)$.

We will apply the definition of compactness. From $\{n'\}$, we can choose another uniformly convergent subsequence $\{n''\}$. With $\{n''\}$, we would want $\lim_{n\longrightarrow\infty} g_{n''}(t,z) = g(t,z)$ uniformly for all $t \in (0,1)$. The uniform convergence implies that there exists a corresponding $F_{n''}(t)$, with $F_{n''}(t) \longrightarrow F$. We set

$$g(t,z) = \begin{cases} -\dfrac{1}{z} & \text{for } t = 0 \\ \dfrac{-1+\mathcal{C}\hat{F}_{t'}}{z+\mathcal{C}\sqrt{t}\hat{F}_t(z/\sqrt{t})} & \text{for } t \in (0,1) \ . \end{cases}$$

Setting $t = 1$ and $F_t = F$, we can apply the Dominated Convergence Theorem:

$$\hat{F}(z) = \int_{t=0}^{1} g(t,z)dt$$

From our definition of $g(t,z)$, we have

$$\hat{F}_t(z) = -\int_{s=0}^{1} \frac{-1+\mathcal{C}\hat{F}_{t'}(z/\sqrt{t})}{z+\mathcal{C}\sqrt{t}\hat{F}_t(z/\sqrt{t})} ds \text{ for } 0 < t \leq 1 \ .$$

Multiplying and dividing $\hat{F}_t(z)$ by $\sqrt{t}$ and setting $z \equiv z/\sqrt{t}$ gives

$$\begin{aligned} \frac{\sqrt{t}}{\sqrt{t}}\hat{F}_t\left(\frac{z}{\sqrt{t}}\right) &= -\int_{s=0}^{1} \frac{-1+\mathcal{C}\hat{F}_{t'}(z/\sqrt{t})}{z+\mathcal{C}\sqrt{t}\hat{F}_t(z/\sqrt{t})} ds \\ \Rightarrow \sqrt{t}\hat{F}_t\left(\frac{z}{\sqrt{t}}\right) &= -\sqrt{t}\int_{s=0}^{1} \frac{-1+\mathcal{C}\hat{F}_{t'}(z/t)}{z+\mathcal{C}\sqrt{t}\hat{F}_t(z/t)} ds \end{aligned}$$

From our previous steps, we introduce a change of variables, $h(t,z) = \sqrt{t}\hat{F}_t(z/\sqrt{t})$. It gives

$$h(t,z) = -\int_{s=0}^{t} \frac{1+\mathcal{C}(\partial/\partial z)h(s,z)}{z+\mathcal{C}h(s,z)} ds \text{ for } 0 < t \leq 1 \ .$$

From **Lemma 3**, g is continuous. Thus, h is differentiable at t. Differentiating the previous expression for $0 \leq t \leq 1$ gives

$$[z+\mathcal{C}h(t,z)]\frac{\partial}{\partial t}h(t,z) + \mathcal{C}\frac{\partial}{\partial z}h(t,z) = -1 \ .$$

- Note for myself

This PDE is a **Cauchy Problem**. From available methods, we can solve the equation by making a system of equations from the coefficients.

3.3 Solving the PDE

The partial differential equation is first-order quasi-linear hyperbolic. We want to determine a solution with the initial condition $h(1,z) = \hat{F}(z)$. Another constraint on the solution is that it must satisfy is $h(0,z) = 0$ because $|h(t,z)| \leq \frac{t}{\mathrm{Im}(z)}$. We will use the coefficients of the PDE to form a system of first order differential equations [20]

$$\begin{cases} \dot{t} = z + \mathcal{C}h(t,z) \\ \dot{z} = \mathcal{C} \ . \end{cases}$$

Solving each equation of the system gives

$$\begin{aligned} \frac{dt}{d\alpha} &= z + \mathcal{C}h(t,z) \\ t(\alpha) &= \int [z + \mathcal{C}h(t,z)] d\alpha \ . \end{aligned}$$

The second one is even more trivial. For initial conditions α_0, α_1, we have that

$$t(\alpha) = z\alpha + \alpha_0 + \mathcal{C} \int h(t,z) d\alpha \ , \text{ and } z(\alpha) = \mathcal{C}\alpha + \alpha_1 \ .$$

From this set up, we can substitute initial conditions, observing that **characteristic curves** to obtain that the solution is

$$h(t,z) = (2\mathcal{C})^{-1}[-z + (z^2 - 4\mathcal{C}t)]^{\frac{1}{2}} \ .$$

By definition, this implies that

$$\hat{F}_t(z) = (2\mathcal{C})^{-1}[-z + (z^2 - 4\mathcal{C})^{\frac{1}{2}}] \ .$$

The result $\hat{F}_t = \hat{F}$ independently of t implies that $\hat{F}$ is the unique limit point of $\{\hat{F}_n\}$. We have that

$$\lim_n g_n(t,z) = g(t,z) = -(z^2 - 4\mathcal{C}t)^{-\frac{1}{2}} \ ,$$

and

$$\begin{aligned} \lim_n \hat{F}_n(z) = \hat{F}(z) &= \int_{t=0}^{1} g(t,z) dt \\ &= (2\mathcal{C})^{-1}[-z + (z^2 - 4\mathcal{C})^{\frac{1}{2}}] \ . \end{aligned}$$

From properties of the *Stieljes Transform*, $F_n \longrightarrow F$. Similarly, property of the Transform gives that $F = W(\cdot, \mathcal{C})$. This completes **Proof of Main Theorem**.

Now that we have proven the Wigner Semicircle Law, we will provide 2 examples to give intuition on the types of matrices that would satisfy the properties that we have discussed.

[20] Yes, yes $\dot{z}$ usually denotes derivative with respect to time. But I will denote the differentiation with respect to another dummy variable α to avoid some abuse with the t that's already there.

4 Brief Examples

4.1 Example 0

From the Introduction to Random Matrices, we can readily form an ensemble of random matrices converging to the Wigner Semicircle Distribution. We can satisfy **Condition (ii)** directly by setting $\xi_{nn} = 0$ for all n; from linear algebra, we would then be able to calculate the $n-1$ remaining components of v_n by solving the $n-1$ equations of $v_n^{\mathbf{T}} u_k^{(n-1)}$. This condition would then be trivially satisfied. We can verify that **Condition (i)** holds by referring to the computations that I have given on **Page** . The Stochastic Convergence of $X_n(t)$ allows us to conclude that the limit in **Condition (i)** exists and is finite. The Semicircle Law holds for $\mathcal{C} \equiv 1$.

4.2 Example 1

We take a block matrix $\mathcal{C}$:

$$\mathcal{C} = \begin{bmatrix} A & B \\ B^{\mathbf{T}} & D \end{bmatrix}$$

Within $\mathcal{C}$, A is m by m symmetric, B is m by $(n-m)$, and D is $(n-m)$ by $(n-m)$ diagonal. Clearly, $B^{\mathbf{T}}$ is $(n-m)$ by m. With these observations in mind, we can show that the empirical spectral functions associated with $\mathcal{C}$, under the conditions that I will provide, does not converge to the Wigner Semicircle Law. To begin, suppose that the entries of A,B and D has the standard normal distribution. Then, we set $m \equiv n/2$. From this choice of m and n, we see that the condition that I stated for the convergence to the Wigner Semicircle Law does not hold; explicitly, the condition relating to the convergence that state which is not fulfilled is

$$\frac{1}{n}\sum_{i=1}^{n} |\mathcal{B}_i^2 - 1| \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

$\mathcal{B}_i$ is a row sum of the form

$$\mathcal{B}_i^2 = n^{-1}\sum_{j=1}^{n} \sigma_{ij}^2$$

From the block matrix $\mathcal{C}$, we can pick row sums $\mathcal{B}_i^2$ to see that the convergence for this matrix does not hold. Another way to realize that the eigenvalues associated to $\mathcal{C}$ do not converge to the Semicircle Distribution follows. Visually, we can *histogram* the eigenvalues to see their distribution on $[-2, 2]$

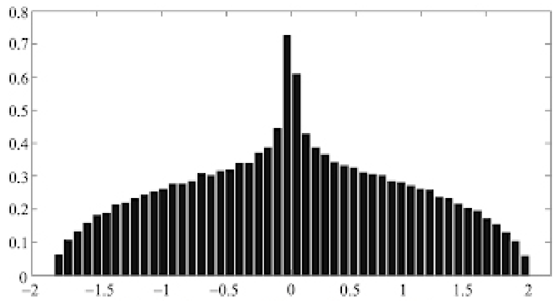


Fig. 1. *Spectrum of the random matrix* $\mathbf{X}_n$.

We see that the rough shape of this distribution does not resemble that of a semicircle in the UHP. Nevertheless, here is a plot of the eigenvalues of a matrix that would converge to the universal Semicircle Distribution

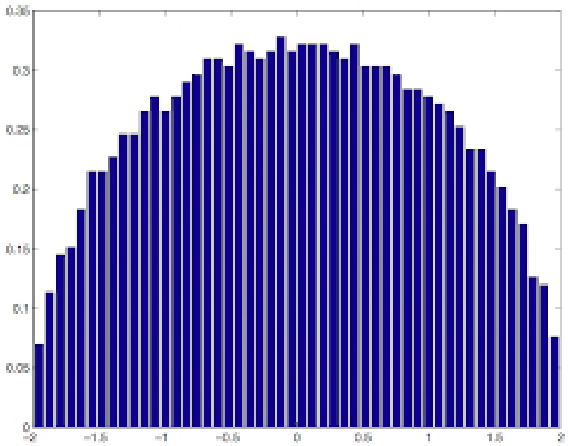


Figure 1: Distribution of Eigenvalues

These images were altered with Photoshop Elements 13 Editor.

4.3 Example 2

We will end by giving a nice (but brief!) computational example. One generalization of the *Wigner Semicircle Law*, the *parabolic probability distribution*, is given by

$$f_{X,Y,Z}(x, y, z) = \frac{3}{4R^3}(R^2 - x^2)$$

when $-R \leq x \leq R$, and $f(x) = 0$ if $|x| > R$. For $f = \frac{3}{4\pi}$, the joint distribution is

$$f_X(x) = \int_0^{\pi} \int_0^{2\pi} \int_0^{R} f_{X,Y,Z}(x, y, z) dr \sin(\theta) d\theta d\phi = 1 \ .$$

Directly, one could compute the marginal distributions which are given by

$$\begin{aligned} f_X(x) &= \int_{-\sqrt{1-y^2-x^2}}^{\sqrt{1-y^2-x^2}} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f_{X,Y,Z}(x, y, z) dx dy dz \\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 2\sqrt{1 - y^2 - x^2} dy \\ &= \frac{3}{4}(1 - x^2) \ . \end{aligned}$$

5 Conclusion

In this exposition, we began with an overview of the distribution and auxiliary functions $X_n(t)$ and $F_n(x)$. We described several characteristics of these functions, all of which was important in our final proof of the Semicircle Law. Also, characterizing properties of the *resolvent* was crucial in several proofs that appeared earlier on. With **Lemma 2** and **Theorem 1**, we were able to use observations along these lines to finish our Proof of the Main Theorem. We finished the exposition with a **Corollary**, and a few lighter examples.

6 Supplementary Section

6.1 Directions for Research

From the list of References that I have given in **Section V**, there are several mathematicians who are currently trying to understand how particular universality classes of matrices can be used to understand different ensembles of random matrices. Also, it is of great interest to understand, at a more fundamental level, how the *eigenvalue gap* and even single eigenvalues contribute to the universal distribution that several different types of physical phenomena converge in distribution to.

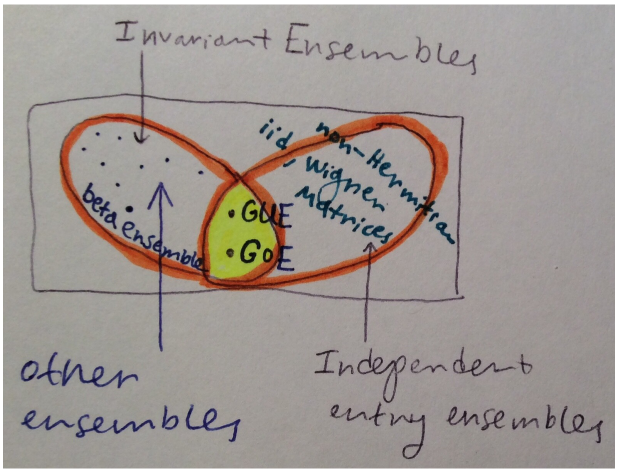
From the resources that I used, people who are working along these lines include Lazlo Erdos and Horng-Tzer Yau.

6.2 Further Remarks

6.2.1 Footnote 1 on Page 1

We can explain how there are universal classes of matrices which all have the same the distribution because we can replace one random matrix with another one from the *universal class* to obtain the same distribution. In particular, if we look at the class of invariant ensembles, this class of Hermitian matrices is invariant under the action of a Lie group. There exists unitary matrices such that conjugating the original matrix W_n gives a new random matrix with the same distribution. There are several other classes, including the *Gaussian Orthogonal Esnemble* and *Beta Ensemble*, which are invariant in similar ways.

On the other hand, there is another class of ensembles. These ensembles are called independent entry ensembles. Although these matrices may not have any conjugation invariance, the entries are *jointly independent* of each other. One of these ensembles is the Wigner Matrices, which includes GUE and GOE as special cases. There are even similar results for non-Hermitian iid matrices, polynomials with random coefficients, random regular graphs, and even random Schrodinger operators. For reference, I have included the diagram below to illustrate the points that I have given:



6.2.2 Footnote 4 on Page 5

From our previous remarks, the symmetric Bernoulli matrices are a strong opposite towards the Gaussian matrices in the diagram that I have included. The symmetric Bernoulli matrices are invariant under permutations, and are discrete and therefore have no distribution. But they converge to the same distribution that the Gaussian matrices do, which is the Wigner distribution.

We can analyze the spectrum on which the n eigenvalues live by using the Law of Large Numbers. Indeed, dividing by the $\sqrt{n}$ normalization is appropriate because if we have $1/n\big(\sum_{i=1}^{n} \lambda_i(W_n)^2\big)$, this is equal to $1/n\big(\mathbf{Tr}W_n^2\big)$ by definition; then, we have that this quantity converges to 1 because $1/n^2\big(\sum_{i=1}^{n}\sum_{j=1}^{n} |\xi_{ij}|^2\big) \longrightarrow 1$. This implies that the eigenvalues $\lambda_k^{(n)}$ have order $\mathcal{O}(1)$. Heuristically, because the eigenvalues together have $\mathcal{O}(1)$, and there are n of them from $\mathcal{A}_n$, each has $\mathcal{O}(1/n)$. We can separate these different magnitudes of the orders $\mathcal{O}$ by labeling them either as the *macroscopic*, *mesoscopic* or *microscopic*, to denote the respective layers of orders between the interval, between the eigenvalues $\lambda_i^{(n)}$ and $\lambda_{i+1}^{(n)}$ (*the eigenvalue gap*), or at the level of the eigenvalues themselves. The middle layer which has the *eigenvalue gap* has eigenvalues on the interval $n^{-1+\epsilon} < \lambda_i^{(k)} < n^{\epsilon}$. In this exposition, our statement and proof of the Wigner Semicircle Law are fundamental macroscopic results; by and large, we are accounting for is not on the precise level of eigenvalues, but rather on the interval itself. In terms of the spectrum, we can histogram the eigenvalues. Explicitly, we know that the eigenvalues are necessarily random because the entries of the matrix are.

The Wigner Semicircle Law accurately reflects how collecting the number of eigenvalues in subintervals of $[-2, 2]$ immediately shows that we obtain the Semicircle Distribution. We could start to subdivide $[-2, 2]$ into 4 subintervals of *unit* length. This could be helpful but as we subdivide the interval into finer partitions, we would be able to more clearly see how each eigenvalue would contribute to the universal distribution that we are obtaining. In $[-2, 2]$, we can take any subinterval $\mathcal{I}$ and we will denote $N_{\mathcal{I}}$ as the number of eigenvalues of a random matrix within $\mathcal{I}$. With this $\mathcal{I}$, we can normalize by a factor of n, which is equivalent to giving the **area** of the *Semicircle Distribution* along this interval. Directly integrating the semicircle distribution, allowing for an **error term** with order o(1), gives $\frac{1}{n}N_{\mathcal{I}}$.

It is good to mention what it is

$$\frac{1}{n}N_{\mathcal{I}} = \int_{\mathcal{I}} W(x, \mathcal{C})dx + \text{o}(1) \text{ , again where } \mathcal{I} \subset [-2, 2] \text{ .}$$

From this expression, it ends up that the error $\text{o}(1) \longrightarrow 0$ as $n \longrightarrow \infty$. This serves as a more general comment to reflect some of the intuition.

6.2.3 Footnote 5 on Page 5

We can go a little further to detail another relationship if we define the **classical location** of the eigenvalues of the matrix. On $[-2, 2]$, the classical location of an eigenvalue eigenvalue λ_i, λ_i^{cl}, from the equation

$$\int_i^{\lambda_i^{\text{cl}}} W(x, \mathcal{C}) = \frac{1}{n}$$

Visually, this corresponds to integrating the Semicircle on $[-2, 2]$ until we get $1/n$ amount of area.

With these more loose observations, we can then use the location of the eigenvalue that we would have histogrammed to study properties of the universal distribution. This relates to any footnote.

6.3 Geometric Interpretation of the Dot Product of Random Vectors

Although we can choose two vectors at random to compute the square of their dot product, which accounts for the term $|v_k^{\mathbf{T}} v_k^{(n-1)}|^2$ in the summation, we can also make use of random principles to interpret this quantity geometrically.

For convenience, we will take these random vectors to have unit length. To do this, we will consider the $(n-1)$-sphere $\mathcal{S}^{n-1} \subseteq \mathbb{R}^n$, where $\mathcal{S}^{n-1} = \{x \in \mathbb{R}^n : ||x|| = r\}$. If we take $n = 2$ and $r \equiv 1$, the vectors that we will be choosing at random in the Cartesian plane amount to picking a random point anywhere on the 2-sphere $\mathcal{S}^1$. But we know that the angles on the circle are evenly spaced, so each angle within the circle has the same likelihood of being chosen, with the angle between the random vectors $v_k^{\mathbf{T}}$ and $v_k^{(n-1)}$ varying between $0°$ and $180°$. The case $n = 2$ is undoubtedly more complicated, as we are considering geometric arrangements within the sphere to analyze whether random vectors are orthogonal. In a similar way for the $n = 1$ case, we are again picking a random point on the sphere. More specifically, for the $n = 2$ case, $\mathbf{P}(\gamma$ points to any circle $\mathcal{C}$ with latitude $\mathcal{L}) \propto \mathcal{C}_\alpha$, where γ is a random vector, $\mathcal{C}$ is a circle represented by a vertical "slice" of the sphere, $\mathcal{C}_\alpha$ is the circumference of a circle within the sphere corresponding to the angle α, and $0° \leq \alpha \leq 90°$. Intuitively, choosing a random point on the sphere, which immediately gives a random vector γ, leads us to conclude that γ is much more likely to point towards the equator of the sphere because the circumference $\mathcal{C}_{90°} \equiv \max_{0° \leq i \leq 90°} \mathcal{C}_i$. **Please refer to the illustrations on slides # of my presentation**.

Indeed, making use of these observations is useful in generalizing qualities of the dot product of the random vectors as n increases. On average, increasing the dimension n gives more possibility for any pair of random vectors to be orthogonal.

Increasing n also decreases the standard deviation of angles between random vectors, in which the distribution of angles between random vectors concentrates around $90°$. For general n, we can directly compute the variance of this distribution to show that as $n \longrightarrow \infty$, the angles between random vectors in higher dimensional space will most frequently be $90°$. Explicitly, the *cosine* between angles of random vectors, has mean 0 and standard deviation $1/\sqrt{n}$. [21] Clearly, the variance $1/\sqrt{n} \longrightarrow 0$ as $n \longrightarrow \infty$. Formally this gives a **Proposition**. [22]

6.4 Proposition

The distribution of the cosine of angles between random vectors in n- dimensional space has mean 0 and standard deviation $\frac{1}{\sqrt{n}}$.

Proof of Proposition: We can show that this claim holds with the following. We will consider the *surface area* of sphere between height h between t and $t+\delta t$, where δt is an infinitesimally small increment. Between the heights t and $t+\delta t$, we can take a $(D-2)$-sphere, $\mathcal{S}^{D-2}$, with height δt and radius $r = \sqrt{1-t^2}$ to calculate the probability distribution. This geometric representation of the sphere with the height and radius that I have given slightly resembles a *frustrum* with slope $1/\sqrt{1-r^2}$. It naturally follows that $f_D(u)$, the distribution of the dot product between random unit vectors, is proportional to

$$f_D(u)du \propto \frac{(\sqrt{1-t^2})^{D-2}}{\sqrt{1-t^2}}dt = (1-t^2)^{(D-3)/2}dt\ .$$

We know that this expression holds because it is a ratio of the radius and height of the $(D-2)$ sphere $\mathcal{S}^{D-2}$. More rigorously, we obtain this ratio

To make the expression more clear, we can introduce a change of variables for t, setting $v = \frac{t+1}{2}$ for $v \in [0,1]$:

$$\begin{aligned}
(1-t^2)^{(D-3)/2}dt \xrightarrow[v=\frac{t+1}{2} \text{ or } t=2v-1]{\text{Setting}} & (1-(2v-1)^2)^{(D-3)/2}d(2v-1) \\
&= (1-(4v^2-4v+1))^{(D-3)/2}d(2v-1) \\
&= (-4v^2+4v)^{(D-3)/2}d(2v-1) \\
&= (2^2(v-v^2))^{(D-3)/2}d(2v-1) \\
&= (2^2)^{(D-3)/2}(v-v^2)^{(D-3)/2}d(2v-1) \\
\xrightarrow[\text{the differential } d(2v-1)\equiv d(v)]{\text{Setting}} & 2^{D-3}(v-v^2)^{(D-3)/2}dv \propto f_D(u)du
\end{aligned}$$

These computations give a normalizing constant 2^{D-3}, and that $(2^2)^{(D-3)/2}(v-v^2)^{(D-3)/2}d(2v-1) \propto f_D(u)$ From these rearrangements, we also claim that the change of variables $v = (t+1)/2$ has a Beta distribution of $((D-1)/2, (D-1)/2)$. From properties of the *Beta Distribution*, we know that it is proportional to the same constant that I have previously expressed. Therefore, the distributions are the same because they are proportional to the same constant. **By definition**, we can compute the mean, median and mode of the *Beta Distribution* depending on the specific value of D.

Altogether, we can now quantify the limiting behavior of this distribution because we know several characteristics of the *Beta Distribution*. Notably, the limiting behavior of this distribution follows once we integrate the expression that I have given in the last step above. We integrate this expression to obtain a **constant of proportionality** depending on the *Gamma Function*:

$$\xRightarrow[\text{previous work}]{\text{From}} \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi}\Gamma(\frac{D-1}{2})}$$

[21] The *Johnson-Lindenstrauss Lemma* gives a nice result which is closely related to this discussion.

[22] Alternatively, we could have shown that the expectation $\mathbf{E}(\frac{|v_1 v_2|^2}{|v_1|^2|v_2|^2}) = 1/\sqrt{n}$, where v_1 and v_2 are random vectors. With this calculation, we are considering the expectation of $\cos^2(\theta)$, where θ is the angle between v_1 and v_2. From the definition of the *dot product*, this is equivalent to the result that we will show later, in which random vectors are more likely to be orthogonal as n becomes arbitrarily large. In the more general case, one would calculate higher moments of $\cos^{2k}(\theta)$, which are similar to the moment calculations that will be mentioned on the next page.

Remark: So far, one could explicitly rearrange the expression to integrate it by parts.

This gives one constant of proportionality from the *Gamma Function.* I obtained the constant directly by evaluating the integral with Wolfram Alpha. With this computation, we get the mean of the distribution. For the variance, we can obtain another similar constant by calculating higher moments of $t^k f_D(t)$. These calculations would give a similar result, leading us to conclude (from properties of the beta distribution), that the variance of the distribution approaches 0.

With elementary techniques it is clear to integrate this expression, from which we may conclude that the variance of the *Beta Distribution* that I listed earlier is $1/D$, which precisely matches with other comments that were made earlier; the angle between random vectors, has variance $1/\sqrt{n}$ which directly coincides with the variance that we have obtained from calculation of higher moments. As D or $n \longrightarrow \infty$, the variance will vanish. With these results, we are able to conclude, by *Chebyshev's Theorem*, that the proportion of the *Beta Distribution* becomes concentrated around $t \approx 0$. Finally, we can determine the *limiting* distribution by analyzing the density of the standard distribution, for t arbitrarily small:

$$\log\left(f_D\left(\frac{t}{\sqrt{D}}\right)\right) = \mathrm{C}(D) + \frac{D-3}{2}\log\left(1 - \frac{t^2}{D}\right) = \mathrm{C}(D) - \left(\frac{1}{2} + \frac{3}{2D}\right)t^2 + \mathcal{O}\left(\frac{t^4}{D}\right) \xRightarrow[D\to\infty]{\text{Taking}} \mathrm{C} - \frac{t^2}{2}$$

In the steps above, the constant C represents the logarithmic constant of integration. From the expression that we obtained by taking $D \to \infty$, we have that the logarithmic density approaches normality, in which it equals $-t^2/2$, at a rate $\mathcal{O}(1/D)$. **Please refer to slide # for Mathematica Plots of the density of this distribution for 3 values of D: 4,6,10.** Consequently, we have shown that as the dimension of our space becomes larger, we will more frequently obtain pairs of vectors in this higher dimensional space that are *orthogonal* to one another. This pattern could also be observed from the variance that was mentioned earlier, in which adding a random vector gets us $1/n$ of the square distance closer to the vectors being orthogonal. Continuing to add more random vectors eventually gives that the correlation $r = n/n \equiv 1$. From our choice of the iid entries $\{\xi\}_{ij}$, the random vectors that we are picking are certainly independent of each other.

This geometric interpretation relates to the submatrix $\mathcal{A}_n$ because as $n \to \infty$, the probability that row vectors in this submatrix are orthogonal will increase. More broadly, this relates to our interpretation of the empirical distribution of eigenvalues of the ∞ by ∞ matrix $\mathcal{A}$ that we will introduce in **Definition 2**, in which taking n unit vectors and arranging them within an n by n matrix ensures that the matrix is nearly orthogonal; at some level, this will qualify our choice for the $\sqrt{n}$ normalization. Being able to geometrically interpret how the dot product of *row vectors* behaves for increasingly larger submatrices is helpful for our discussion that will follow.

6.5 Statement of the Perron-Frobenius Theorem for Regular Matrices

Theorem: Suppose that $A \in \mathbb{R}^{nxn}$ is nonnegative and regular, *i.e.*, $A^k > 0$ for some k. Then

1: There is an eigenvalue λ_{pf} of A that is real and positive, with positive left and right eigenvectors.

2: For any other eigenvalue λ, we have $|\lambda| < \lambda_{pf}$.

3: The eigenvalue λ_{pf} is simple, *i.e.*, has multiplicity one, and corresponds to a 1x1 Jordan block.

The eigenvalue λ_{pf} is called the *Perron-Frobenius* (PF) eigenvalue of A.

6.6 Alternate Approach for the $X_{kn}(z)$ functions in Theorem 1

We could also compute $X_{kn}(z)$ with brute force. We will sum over eigenvalues $\lambda_k^{(n)}$, for $k = 1, \cdots, n$, from $\mathcal{A}_n$:

$$\begin{aligned}\mathbf{Tr}R(z, \mathcal{A}_n) = X_{kn}(z) = \sum_{k=1}^{n} \xi_{ii} &= \left(\frac{\xi_{11}}{\sqrt{n}} - z\right)^{-1} + \left(\frac{\xi_{22}}{\sqrt{n}} - z\right)^{-1} + \cdots + \left(\frac{\xi_{nn}}{\sqrt{n}} - z\right)^{-1} \\ &= \frac{1}{\left(\frac{\lambda_1^{(n)}}{\sqrt{n}} - z\right)^{-1}} + \frac{1}{\left(\frac{\lambda_2^{(n)}}{\sqrt{n}} - z\right)^{-1}} + \cdots + \frac{1}{\left(\frac{\lambda_k^{(n)}}{\sqrt{n}} - z\right)^{-1}}\end{aligned}$$

Obtaining common denominators:

$$= \frac{\overbrace{\left(\frac{\lambda_2^{(n)}}{\sqrt{n}}-z\right)^{-1}\left(\frac{\lambda_3^{(n)}}{\sqrt{n}}-z\right)^{-1}\cdots\left(\frac{\lambda_k^{(n)}}{\sqrt{n}}-z\right)^{-1}}^{\text{First Term}}+\overbrace{\left(\frac{\lambda_1^{(n)}}{\sqrt{n}}-z\right)^{-1}\left(\frac{\lambda_3^{(n)}}{\sqrt{n}}-z\right)^{-1}\cdots\left(\frac{\lambda_k^{(n)}}{\sqrt{n}}-z\right)^{-1}}^{\text{Second Term}}+\overbrace{\cdots}^{k-2\text{ remaining terms}}}{\prod_{k=1}^{n}\left(\frac{\lambda_k^{(n)}}{\sqrt{n}}-z\right)^{-1}}$$

where the numerator of the k^{th} term is $\prod_{i\neq k}(\lambda_i^{(n)}/\sqrt{n}-z)^{-1}$. That is, we include all terms $(\lambda_i^{(n)}/\sqrt{n}-z)^{-1}$ for $1\leq i\leq k-1$ because $(\lambda_k^{(n)}/\sqrt{n}-z)^{-1}$ is already in the denominator of the k^{th} term; the same observation holds for all k terms. With the product index i, we express the numerator in simpler terms with the common denominator $\prod_{k=1}^{n}\left(\frac{\lambda_k^{(n)}}{\sqrt{n}}-z\right)^{-1}$:

$$\Longrightarrow \frac{\sum_{k=1}^{n}\left(\prod_{i\neq k}^{n}\left(\frac{\lambda_i^{(n)}}{\sqrt{n}}-z\right)^{-1}\right)}{\prod_{k=1}^{n}\left(\frac{\lambda_k^{(n)}}{\sqrt{n}}-z\right)^{-1}}$$

I think that this statement could be similar to what we ended up using in **Proof of Theorem 1**.

6.7 Statement of Corollary 2

Corollary [23]: Let entries X_{jk}, $1\leq j\leq k\leq n$, of the random matrix $\mathbf{X}_n$ satisfy the conditions.[24]

$$\begin{aligned}
\mathbf{E}(X_{ij}|\mathcal{F}^{(i,j)})=0,\ \frac{1}{n^2}\sum_{i,j=1}^{n}\mathbf{E}|\mathbf{E}(X_{ij}^2|\mathcal{F}^{(i,j)})-\sigma_{ij}^2| &\longrightarrow 0 \text{ as } n\longrightarrow\infty\ ,\\
\frac{1}{n^2}\sum_{i,j=1}^{n}\mathbf{E}X_{ij}^2\mathbf{1}_{(|X_{ij}|\geq\tau\sqrt{n})} &\longrightarrow 0 \text{ as } n\longrightarrow\infty\ ,\\
\frac{1}{n}\sum_{i=1}^{n}|\mathcal{B}_i^2-1| &\longrightarrow 0 \text{ as } n\longrightarrow\infty\ ,\\
\max_{1\leq i\leq n}\mathcal{B}_i &\leq C\ ,\\
\max_{1\leq i\leq n}|\mathcal{B}_i^2-1| &\longrightarrow 0 \text{ as } n\longrightarrow\infty\ .
\end{aligned}$$

Observe that C is some absolute constant. Observe that this constant is not the same as $\mathcal{C}$. Then $\sup_x|F^{\mathbf{X}_n}(x)-G(x)|\longrightarrow 0$ as $n\longrightarrow\infty$.

6.8 Mathematica and Java Implementation

I took the plots of the Semicircle Distribution for the values of D, $4,6,10$, directly from the **Wigner Distribution** function in Mathematica. On **Page 4**, I referred to these plots to clarify the steps that I was taking in the computations.

I was interested in looking at more plots. I either changed some Mathematica code slightly by adding a Manipulate tab or changed some readily available Java code for plotting the distribution.

[23]From **Page 25** of *Gotze, Naumov and Tikhomirov.*
[24]Condition **4** in the third line was already mentioned in the first example.

6.9 Acknowledgments

Thank you to Professor Hubbard for having met with me 6 times as of April 17, 2018.

6.10 References

(1) Gotze, F., Naumov, A.A. and Tikhomirov, A.N. Limit Theorems for Two Classes of Random Matrices with Dependent Entries. Theory Probab. Appl., 2015, 59.1, 23-29.

(2) Credner, K.H. and Stolz, M. Wigner Theorems for Random Matrices with Dependent Entries: Ensembles Associated to Symmetric Spaces and Sample Covariance Matrices. Comm. in Probab., 2008, 13, 401-414.

(3) Erdos, L., Schlein, B. and Yau, H.T. Local Semicircle Law and Complete Delocalization for Wigner Random Matrices. Commun. Math. Phys., 2009, 641-655.

(4) Arnold, Ludwig. Deterministic Version of Wigner's Semicircle Law for the Distribution of Matrix Eigenvalues. Linear Algebra and its Applications, 1976, 13, 185-199.

(5) Arnold, Ludwig. On Wigner's Semicircle Law for the Eigenvalues of Random Matrices. Springer-Verlag, 1971, 191-198.

(6) Arnold, Ludwig. On the Asymptotic Distribution of the Eigenvalues of Random Matrices.Journal of Mathematical Analysis and Applications, 1967, 20, 262-268.

(7) Juhasz, F. On the Spectrum of a Random Graph. Algebraic Methods in Graph Theory, 1978, 25, 313-316.

(8) Kornyik, M. and Michaletzky, G. Wigner matrices, the moments of roots of Hermite polynomials and the semicircle law. Journal of Approximation Theory, 2016, 211, 29-41.

(9) Eichelsbacher, P. and Stolz, M. Large deviations for random matrix ensembles in mesoscopic physics. 2006, 1-21.

(10) Knowles, A. and Jun, Y. Eigenvector distribution of Wigner matrices. Probab. Theory Relat. Fields, 2013, 155, 543-582.

(11) Bai, Z.D. and Yin, Y.Q. Convergence to the Semicircle Law. The Annals of Probability, 1988, 16, 863-875.