UNIVERSITY OF CALIFORNIA

SANTA CRUZ

EIGENVALUES OF LARGE DIMENSIONAL RANDOM MATRICES

A thesis submitted in partial satisfaction of the

requirements for the degree of

MASTER OF ARTS

in

MATHEMATICS

by

Brendan Shea Sullivan

September 2017

The Thesis of Brendan Shea Sullivan
is approved:

Torsten Ehrhardt, Chair

Viktor Ginzburg

Francois Monard

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

Abstract

Eigenvalues of Large Dimensional Random Matrices

by

Brendan Shea Sullivan

This paper demonstrates an introduction to the statistical distribution of eigenvalues in Random Matrix theory. Using mathematical analysis and probabilistic measure theory instead of statistical methods, we are able to draw conclusions on large dimensional cases and as our dimensions of the random matrices tend to infinity. Applications of large-dimensional random matrices occur in the study of heavy-nuclei atoms, where Eigenvalues express some physical measurement or observation at a distinct state of a quantum-mechanical system. This specifically motivates our study of Wigner Matrices. Classical limit theorems from statistics can fail in the large-dimensional case of a covariance matrix. By using methods from combinatorics and complex analysis, we are able to draw multiple conclusions on its spectral distributions. The Spectral distributions that arise allow for boundedness to occur on extreme eigenvalues.

Acknowledgments

I want to thank my younger brothers for all the support they have given me through the years.

Chapter 1

Introduction

This paper is concerned with random matrices. Specifically, we are concerned with the statistical distributions of the eigenvalues of random matrices. When looking at smaller, finite dimensional random matrices, it makes sense to look at them through classical limit theorems and basic statistics gained from samples. However, when dimension tends to infinity (or large) we are unsure if classical limit theorems remain true.

This paper reframes arguments made on two kinds of large dimensional random matrices: the Wigner matrices, and the sample Covariance matrices. Wigner matrices have a direct application to the studies of quantum mechanics and nuclear physics, hence being named after the physicist Eugene Wigner. Sample Covariance matrices are fundamental in statistical applications to economics and research science, where measurements for principal component analysis, discrimination analysis, and hypothesis testing are done. We the paper focuses on the statistical behavior of the eigenvalues of these types of matrices (the "spectral distribution"), in the large dimensional case. In doing so, we are applying some of the consequences of random matrix theory from the Gaussian Unitary Ensemble.

In the following chapter, we will look at the statistical distribution of the eigenvalues of Wigner matrices. We use the moment convergence theorem to find that the empirical spectral distribution of Wigner matrices tends to Wigner's Semi-Circle Law, a probability density function that graphs to the shape of a semi-circle. In Chapter 3, we analogously look at the ESD of sample covariance matrices. We find that the ESD tends to what is called the Marcenko-Pastur Law, via the Stieltje's transform and some theorems from complex analysis.

In chapter 4 we move on to the limiting spectral distributions of our two types of random matrices. We prove that the eigenvalues tend to be bounded with respect to their standard deviations found in their probability density functions. In chapter 5, we conclude our paper with how the study of these forms relate to the Gaussian Unitary Ensembles (GUE).

As a preliminary, when dealing with Probability as a measure space, with properties such as Expectation and the probability of an event happening, it will be assumed that the space in question is a general bounded set of a product space of real or complex random variables (which will be specified).

Chapter 2

Wigner Matrices and the Semicircular Law

2.1 Wigner Matrices

A Wigner Matrix, named after Eugene Wigner, is defined as a Real Symmetric Matrix, or a Hermitian Matrix in the complex case. Wigner Matrices are a type of random matrix, since the entries of the symmetric (or hermitian) matrix are random variables. Furthermore, the upper-triangular entries of the matrix are iid complex random variables with mean 0 and variance 1. The diagonal entries are iid real variables, independent of the upper triangular entries, with bounded mean and variance. Wigner's original motivation for studying random matrices of this type lies in quantum mechanical systems. In Nuclear Physics, we can describe such a system as an eigenvalue problem

$$H\Psi_n = E_n\Psi_n \tag{2.1}$$

Where H is a Hermitian operator (a sample Wigner Matrix) Ψ_n is an eigenfunction and E_n is the corresponding eigenvalue. On the Left-hand side of the equation, the Hermitian operator represents a physical measurement of a system at a given point. The different eigenfunctions can characterize the different states of the system, and the eigenvalue that corresponds is a value to be measured for the system in that specific state.If we were to consider the system as an atomic nucleus, the Hermitian operator is the "Hamiltonian" and the eigenvalues represent the different energy level states.

For a heavy nuclei atom, we would naturally have a large number of states, energy levels, and this is where large dimensional random matrices come into play. Instead of working with one, very large Hermitian operator, a sequence of random matrices, suffices to model the distribution of the eigenvalues. Each element of the sequence contributes a finite amount of eigenvalues to the total distribution, called the empirical spectral distribution. The middle part of this spectrum (ignoring outliers) might model an infinite spectrum, as needed to understand an atomic nucleus.

In this Chapter, we will show via the moment method for iid variables that the Empirical Spectral distribution of Wigner Matrices converges to the semicircular law as the size of our matrix approaches infinity. This proof is lengthy, and has also been done by Bai and Silverstein [1], Liu [9], and Edelman [6].

The Empirical Spectral Definition (ESD) is defined as follows,

Definition 2.1.1 *The Empirical Spectral Distribution of any square matrix A is* *a probability distribution function P that puts equal mass on each eigenvalue of A.*

The Wigner Semicircular law is a probability density function centered at the origin defined as $f(x) = \frac{2}{\pi R^2}\sqrt{R^2 - x^2}$ For $x \in R$ and $f(x) = 0$ for X not in R.

2.2 The Moment Method

Before we show that the Empirical Spectral distribution for a sequence Wigner Matrices converges to the semicircular law, we must first define and show the moment method.

In Random Matrix Theory, we use the Moment Convergence Theorem. Suppose F_n denotes a sequence of probability distribution functions. The Moment Convergence Theorem checks what conditions determine the convergence of moments for all fixed orders, which is a weak converge of the sequence F_n. Specifically, we will utilize the Carleman Condition to show a unique convergence to the semicircular law.

Let the k-th moment of distribution function F_n be denoted by

$\beta_{n,k} = \beta_k(F_n) := \int x^k dF_n(x)$

The Carleman condition is a condition made for the Carleman's Theorem.

Theorem 2.2.1 *Carleman's theorem: Let $\beta_k = \beta_k(F)$ be the sequence of moments of the distribution function F. If the Carleman condition*

$$\sum \beta_{2k}^{\frac{-1}{2k}} = \infty$$

is satisfied, then F is uniquely determined by the moment sequence β_k

Proof. Let$F$ and G be two distribution functions with the common moment sequence β_k satisfying the Carleman condition. Let $f(t)$ and $g(t)$ be the characteristic functions of F and G, respectively. By the uniqueness of characteristic functions, we need only prove that f(t) = g(t) for all $t > 0$.

We have the relation:

$$\beta_{2k}^{\frac{1}{2k}} \leq \beta_{2k+2}^{\frac{1}{2k+2}}$$

From here, we can see that the Carleman condition is equivalent to

$$\sum_{k=1}^{\infty} 2^k \beta_{2^k}^{-2^{-k}} = \infty$$

And for any integer $n \geq 6$ and $k \geq 1$, define

$$h_{n,k} = n^{-1} 2^k (\beta_{2^k}^4 \beta_{2^{k+1}}^{\frac{5}{2}})^{2^{-k}}$$

Now we want to show that for any n,

$$\sum_{k=1}^{\infty} h_{n,k} = \infty$$

Let $c < 1/2$ be a positive constant. Define

$$\mathcal{K}_1 = \{1\} \cup \left\{ k : \beta_{2^k}^{2^{-k}} \geq c\beta_{2^{k=1}}^{-k-1} \right\} \text{ and}$$

$$\mathcal{K}_2 = \{k \notin \mathcal{K}_1\} = \left\{ k : \beta_{2^k}^{2^{-k}} < c\beta_{2^{k+1}}^{2^{-k-1}} \right\}$$

First, we want to show that,

$$\sum_{k \in \mathcal{K}_1} 2^k \beta_{2^{k+1}}^{-2^{-k-1}} = \infty$$

Suppose $k \in \mathcal{K}_1$ and $k+1, ...k+s \in \mathcal{K}_2$. Then,

$$\beta_{2^{k+s+1}}^{-2^{-k-s-1}} < c\beta_{2^{k+s}}^{-2^{-k-s}} < ... < c^s \beta_{2^{k+1}}^{-2^{-k-1}}$$

From here, we see $\mathcal{K}_1$is nonempty, we can construct

$$\sum_{k\in\mathcal{K}_2} 2^k \beta_{2^{k+1}}^{-2^{-k-1}} \leq \frac{1}{1-2c} \sum_{k\in\mathcal{K}_1} 2^k \beta_{2^{k+1}}^{-2^{-k-1}}$$

This inequality, along with

$$\sum_{k=1}^{\infty} 2^k \beta_{2^k}^{-2^{-k}} = \infty, \text{ show that} \tag{2.2}$$

$$\sum_{k\in\mathcal{K}_1} 2^k \beta_{2^{k+1}}^{-2^{-k-1}} = \infty \tag{2.3}$$

Now, for each $k \in \mathcal{K}_1$ we get

$$h_{n,k} \geq c^4 n^- 1 2^k \beta_{2^{k+1}}^{-2^{-k-1}}$$

and then by (3);

$$\sum_{k\in\mathcal{K}_1} 2^k \beta_{2^{k+1}}^{-2^{-k-1}} = \infty$$

So for each fixed n, we have

$$\sum_{k=1}^{\infty} h_{n,k} \geq c^4 n^- 1 \sum_{k\in\mathcal{K}_1} 2^k \beta_{2^{k+1}}^{-2^{-k-1}} = \infty$$

Now we have for any $t > 0$, there exists an integer m, such that

$$t_{n,m-1} \leq t \leq t_{n,m} \qquad \text{where} \tag{2.4}$$

$$t_{n,j} = h_{n,1} + ... + h_{n,j} \qquad \text{for } j = 1, 2, ...m-1. \tag{2.5}$$

Notation-wise, $h_{n,m} = t - t_{n,m-1}, t_{n,0} = 0$ and $t_{n,m} = t$ Let

$$H = F - G, q_{n,1}(x) = exp(ih_{n,1}x - 1 - ih_{n,1}x) \text{ and}$$

$$q_{n,k}(x) = (\prod_{j=1}^{k-1}(1+ih_{n,j}x+...+\frac{(ih_{n,j}x)^{2^j-1}}{(2^j-1})!)\times(exp(ih_{n,k}x)-1-ih_{n,k}x-...-\frac{(ih_{n,k}x)^{2^k-1}}{(2^k-1})!)$$

. For $k \leq m$, by an extension of Riesz's lemma, we have

$$|q_{n,k}(x)| \leq Q_{n,k}(x) := (\prod_{j=1}^{k-1}(1 + h_{n,j}|x| + ... + \frac{(h_{n,j}|x|)^{2^j-1}}{(2^j-1)!}))\frac{(h_{n,k}|x|)^{2^k}}{(2^k)!}$$

Since, $\int x^j H(dx) = 0$, we get

$$|f(t) - g(t)| = |\int_{\infty}^{\infty} e^{itx} H(dx)|$$

$$= |\sum_{k \le m} \int_{\infty}^{\infty} exp[i(t - t_{n,k})x] q_{n,k}(x) H(dx)|$$

$$\le \sum_{k \le m} \int_{\infty}^{\infty} Q_{n,k}(x)(F(dx) + G(dx)) = 2 \sum_{k \le m} \int_{\infty}^{\infty} Q_{n,k}(x) F(dx)$$

Now, we can expand $Q_{n,k}(x)$. This gives us the form

$$\frac{h_{n,1}^{v_1}}{v_1!} \dots \frac{h_{n,k-1}^{v_{k-1}}}{v_{k-1}!} \frac{h_{n,k}^{2^k} |x|^v}{(2^k)!}$$

Where $v = v_1 + \dots v_{k-1} + 2^k$ and $0 \le v_j \le 2_j - 1$ By definition of $h_{n,k}$, the integral of this term is bounded above by

$$\frac{n^{-v} 2^{\mu} \beta_2^{2v_1} \beta_4^{4^{-1}(4v_2 - 5v_1)} \dots \beta_{2^{k-1}}^{2^{-k+1}(4v_{k-1} - 5v_{k-2})}}{v_1! v_2! \dots v_{k-1}!} \times \frac{\beta_{2k}^{2^{-k}(2^{k+2} - 5v_{k-1})} \beta_{2^{k+1}}^{2^{-k-1} v - 5/2}}{(2^k)!}$$

Here, $\mu = v_1 + 2v_2 + \dots (k-1)v_{k-1} + k2^k$ and note that,

$$4v_1 + (4v_2 - 5v_1) + \dots + (4v_{k-1} - 5v_{k-2}) + (2^{k+2} - 5v_{k-1}) = 2^{k+2} + 2^k - v > 0$$

Then, if we use an extension of Holder's inequality $\beta_{2^k} \le \beta_{2^{k+1}}^{2^{-k-1+s}}$ we get,

$$\beta_2^{2v_1} \beta_4^{4^{-1}(4v_2 - 5v_1)} \dots \beta_{2^{k-1}}^{2^{-k+1}(4v_{k-1} - 5v_{k-2})} \beta_{2^k}^{2^{-k}(2^{k+2} - 5v_{k-1})}$$

$$\le \beta_{2^{k+1}}^{2^{-k-1}(4v_1 + (5v_2 - 5v_1) + \dots + (4v_{k-1} - 5v_{k-2} + 2^{k+2} - 5v_{k-1}))}$$

$= \beta_{2^{k+1}}^{5/2 - 2^{-k-1} v}$ which implies that,

$$\frac{h_{n,1}^{v_1}}{v_1!} \dots \frac{h_{n,k-1}^{v_{k-1}}}{v_{k-1}!} \frac{h_{n,k}^{2^k} \beta_v}{(2^k)!} \le \frac{n^{-v} 2^{\mu}}{v_1! v_2! \dots v_{k-1}! (2^k)!}$$

and since $v \ge 2^k$, we can use following bounds for the integral,

$$\int_{-\infty}^{\infty} Q_{n,k}(x) F(dx) \le \sum_{v_1 + \dots + v_{k-1} + 2^k = v} \frac{(n^{-1}2)^{v_1} \dots (n^{-1} 2^{k-1})^{v_{k-1}} (n^{-1} 2^k)^{2^k}}{v_1! v_2! \dots v_{k-1}! (2^k)!}$$

$$\leq \sum_{v_1+...+v_{k-1}+2^k=v} \frac{(n^{-1}2)^{v_1}...(n^{-1}2^k)^{v_k}}{v_1!v_2!...v_{k-1}!(2^k)!}$$

$$= \sum_{v=2^k}^{\infty} (n^{-1}(2+...+2^k))^v/v! \leq \sum_{v=2^k}^{\infty} (2e/n)^v$$

$$= (2e/n)^{2^k} \frac{n}{n-2e}$$ After substituting this into

$$2\sum_{k\leq m}\int_{\infty}^{\infty} Q_{n,k}(x)F(dx)$$ we get

$$|f(t)-g(t)| \leq \frac{n}{n-2e}\sum_{k=1}^{\infty}(2e/n)^{2^k}$$ which converges to zero as n goes to infinity.

This is equivalent to Carleman's theorem, and the theorem is now proven. We can now move on to using converging moments for the semicircular law. □

2.3 Moments of the Semicircular law

Let β_k denote the k-th moment of the semicircular law. We have the following theorem.

Theorem 2.3.1 *For $k = 0, 1, 2, ...$, we have*

$$\beta_{2k} = \frac{1}{k+1}\binom{2k}{k}, \quad \beta_{2k+1} = 0 \tag{2.6}$$

Proof. Since the semicircular distribution is symmetric about 0, so β_{2k+1} being 0 is trivial. For β_{2k}, we know that

$$\beta_{2k} = \frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}dx$$

$$= \frac{1}{\pi}\int_{0}^{2} x^{2k}\sqrt{4-x^2}dx$$

$$= \frac{2^{2k+1}}{\pi} \int_0^1 y^{k-1/2} \sqrt{1-y} dy$$

(by letting $x = 2\sqrt{y}$)

$$= \frac{2^{2k+1}}{\pi} \frac{\Gamma(k+1/2)\Gamma(3/2)}{\Gamma(k+2)} = \frac{1}{k+1} \binom{2k}{k}$$

Which is the desired result. □

2.4 Semicircular Law for Wigner Matrices

Up until now, we have established the convergence of moments, and have used that to show the convergence of the moments of the semicircular law. For a family of Wigner Matrices, (a sequence of Wigner Matrices that changes in size, (for n=1, 2, ...) we can construct an empirical distribution function F^{W_n} that shows the occurrence of eigenvalues for the Wigner Matrices. Using the moment method, as the size tends to infinity, we will demonstrate that our function F^{W_n}, known as the empirical spectral distribution converges to the semicircular law.

Theorem 2.4.1 *Suppose that X_n is an $n \times n$ Hermitian matrix whose diagonal entries are independently-identically-distributed (iid) real random variables and those above the diagonal are iid complex random variables with variance $\sigma^2 = 1$. Then, the ESD of $W_n = \frac{1}{\sqrt{n}} X_n$ converges to the semicircular law almost surely.*

The convergence of Wigner Matrices is extremely powerful, and the proof is lengthy.The proof has also been done at length by Bai and Silversteing [1]. For clarity, this theorem does not depend on the mean of extradiagonal terms, however does require each moment to be finite. In order to prove this theorem using moment convergence, we must first perturb our Wigner Matrices in a certain manner. The following steps will describe how we alter them, and we will also show how these perturbations do not change the Limiting Spectral Distribution (LSD) of our family of Wigner Matrices. This will allow for a more general convergence of ESD's for normalized Wigner Matrices to the semicircular law, rather than just specific cases where we have iid random variables with mean zero and variance 1.

Our first alter will be removing the diagonal elements on the Wigner Matrix. Let $\widetilde{W_n}$ be the matrix obtained from W_n by replacing the diagonal elements with zeroes. If we can show that these two matrices are "asymptotically equivalent" (their LSD's are the same) then we can use $\widetilde{W_n}$ equivalently in showing the ESD of W_n.

Let N_n be the cardinality of the set $|x_{ii}| \geq \sqrt[4]{n}$. Let us change the diagonal elements of W_n by $\frac{1}{\sqrt{n}} x_{ii} I(|x_{ii}| < \sqrt[4]{n})$ Where I is the characteristic function. Let us call this new matrix $\widehat{W_n}$. Bai and Silverstein [1] proved the following Corollaries we will now apply:

Corollary 2.4.2 *Let A and B be two $n \times n$ normal matrices with ESD's F^A and F^B. Then,*

$$L^3(F^A, F^B) \leq \frac{1}{n} tr[(A-B)(A-B)^*].$$

Where $L(F, G)$ is the Levi distance between two functions, and is defined as:

$$L(F, G) = inf\{\epsilon > 0 : F(x - \epsilon) - \epsilon \leq G(x) \leq F(x + \epsilon) + \epsilon, \forall x\}$$

We now get the following inequality:

$$L^3(F^{\widehat{W_n}}, F^{\widetilde{W_n}})$$
$$\leq \frac{1}{n} tr[(\widehat{W_n} - \widetilde{W_n})^2]$$
$$\leq \frac{1}{n^2} \sum_{i=1}^{n} |x_i i|^2 I(|x_i i| < \sqrt[4]{n}) \leq \frac{1}{\sqrt{n}}$$

From linear algebra, we use the following corollary:

Corollary 2.4.3 *Let A and B be two $n \times n$ hermitian matrices. Then,*

$$||F^A - F^B|| \leq \frac{1}{n} rank(A - B)$$

Then we have,

$$||F^{W_n} - F^{\widetilde{W_n}}|| \leq \frac{N_n}{n}$$

Now, all we must do is show that $\frac{N_n}{n}$ converges to zero almost surely. Define $p_n = P(x_{11} \geq \sqrt[4]{n})$ which converges to zero. By Bernstein's inequality, we know that for any $\epsilon > 0$,

$$P(N_n \geq \epsilon n) = P(\sum_{i=1}^{n} (I(|x_{ii}| \geq \sqrt[4]{n}) - p_n) \geq (\epsilon - p_n)n)$$
$$\leq 2exp(-(\epsilon - p_n)^2 n^2 / 2[np_n + (\epsilon - p_n)n]) \leq 2e^{-bn}$$

For some positive constant, $b > 0$. This means that it does in fact converge to zero almost surely, and that our perturbation does not affect the LSD of our Wigner Matrix.

The next step on our sample Wigner Matrix is we must now truncate the variables for any fixed positive constant C, at C. We can express it as $x_{ij(C)} = x_{ij}I(|x_{ij}| \leq C)$. We can replace these elements in W_n (Our Wigner that has already had its diagonal elements altered) and call this new Wigner $W_{n(C)}$. For clarification, the our new Wigner has zeroes for the diagonal elements and the off-diagonal elements are $\frac{1}{\sqrt{n}}x_{ij(C)}$. We must now show that the LSD remains unchanged for our new Wigner Matrix.

Equivalently, this means that we want to show for $W_{n(C)} = \frac{1}{\sqrt{n}}X_{n(C)}$ that for any constant C,

$$\limsup_n L^3(F^{W_n}, F^{W_{n(C)}}) \leq E(|x_{11}|^2 I(|x_{11}| > C))$$

almost surely.

By Corollary 2.4, as well as the law of large numbers, we know that

$$L^3(F^{W_n}, F^{W_n(C)}) \leq \frac{2}{n^2}(\sum_{1 \leq i \leq j \leq n} |x_{ij}|^2 I(|x_{11}| > C)) \quad \text{Which converges to,} \tag{2.7}$$

$$E(|x_{11}|^2 I(|x_{11} > C)) \tag{2.8}$$

Now, C can be made extremely large, so our expectation here can be arbitrarily small. When proving the ESD semicircular law, we consider the entries of our matrix to be uniformly bounded to avoid confusion.

After altering the diagonals and truncating the rest of our matrix, the next step is centralization. Using Corollary 2.5, we can apply it to our Perturbed Wigner Matrix $W_{n(C)}$:

$$||F^{W_{n(C)}} - F^{W_{n(C)}-a11'}|| \leq \frac{1}{n}$$

Where $a = \frac{1}{\sqrt{n}}\mathcal{R}(E(x_{12(C)}))$ and $a11'$ is a rank 1 submatrix of our Wigner matrix. Also, if we consider corollary 2.4, we get:

$$L(F^{W_(n(C))-\mathcal{R}E(W_{n(C)})}, F^{W_{n(C)}-a11'}) \leq \frac{|\mathcal{R}(E(x_{12(C)}))|^2}{n}$$

Which converges to zero almost surely for at least the real parts. This means that truncation works at least when working with a Real-valued case. We can now consider the complex case, by looking solely at the imaginary part. Note from linear algebra, we have the following lemma on skew-symmetric matrices:

Lemma 2.4.4 *Let A_n be an $n \times n$ skew-symmetric matrix whose elements above the diagonal are 1 and those below are -1. Then, the eigenvalues of A_n are $\lambda_k = icot(\pi(2k-1)/2n)$ for $k = 1, 2, ..., n$. The eigenvectors associated are $u_k = \frac{1}{\sqrt{n}}(1, \rho_k, ..., \rho_k^{n-1})'$, where $\rho_k = (\lambda_k - 1)/(\lambda_k + 1) = exp(-i\pi(2k-1)/n)$.*

Now, we can apply this lemma to the imaginary part. Let $b = E\mathcal{I}(x_{12(C)})$. Then $E\mathcal{I}(W_{n(C)}) = ibA_n$. By lemma 2.6, we know the eigenvalues of $i\mathcal{I}(E(W_{n(C)}))$ are $ib\lambda_k = -n^{-1/2}bcot(\pi(2k-1)/2n)$ Now, consider the spectral decomposition of $A_n = U_n D_n U_n^*$ if we rewrite $\mathcal{I}(E(W_{n(C)})) = B_1 + B_2$ where $B_j = -\frac{1}{\sqrt{n}}bU_n D_n U_n^*, j = 1, 2,$ where U_n is a unitary matrix, D_n is a diagonal of the eigenvalues, and

$$D_{n1} = D_n - D_{n2} = diag[0, ..., 0, \lambda_{[n^{3/4}]}, \lambda_{[n^{3/4}]+1}, ...\lambda_{n-[n^{3/4}]}, 0, ..., 0]$$

Now, for any Hermitian Matrix, using corollary 2.4, we have

$$L^3(F^C, F^{C-B_1}) \leq \frac{1}{n^2} \sum_{n^{3/4} \leq k \leq n-n^{3/4}} cot^2(\pi(2k-1)/2n) < \frac{2}{nsin^2(n^{-1/4}\pi)}$$

which converges to zero.

And by Corollary 2.5:

$$||F^C - F^{C-B_2}|| \leq \frac{2n^{3/4}}{n}$$

which also converges to zero.

In summation, centralization of a Wigner Matrix results in the following:

$$L(F^{W_{n(C)}}, F^{W_{n(C)}-E(W_{n(C)})}) = o(1)$$

We can now move on to the final step of "preparing" our Wigner Matrices for the moment method. Let the variance of C be defined as $\sigma^2(C) = Var(x_{11(C)}$ and let $\widetilde{W_n} = \sigma^{-1}(C)(W_{n(C)} - E(W_{n(C)}))$. From here, we see that the off-diagonal entries of $\sqrt{n}\widetilde{W_n}$ are $\hat{x}_{kj} = \sigma^{-1}(C)(x_{kj(C)} - E(x_{kj(C)}))$.

Apply Corollary 2.4, and we get,

$$L^3(F^{\widetilde{W}_n}, F^{W_{n(C)}-E(W_{n(C)})}) \leq \frac{2(\sigma(C)-1)^2}{n^2\sigma^2(C)} \sum_{1 \leq i \leq j \leq n} |x_{kj(C)} - E(x_{kj(C)})|^2$$

which converges to $(\sigma(C) - 1)^2$ almost surely.

If C is large, then our solution above can be made arbitrarily small.

After altering our matrices in such fashion, we have seen that the Limiting Spectral distribution is unchanged. We now can use Moment convergence to show the semicircular distribution of our eigenvalues.

For clarification, let W_n be our Wigner Matrix after we have enacted the previous alterations to it. Let x_{ij} be the variables changed respectively. We know that the semicircular distribution satisfies the Riesz condition, so if we show that the moments of the spectral distribution converge to the moments of the semicircular distribution almost surely, this will suffice.

Let $\beta_k(W_n)$ be the k-th moment of the ESD of W_n:

$$\beta_k(F^{W_n}) = \int x^k dF^{W_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^k = \frac{1}{n} tr(W_n^k) = \frac{1}{n^{1+k/2}} tr(X_n^k)$$

Where λ_i's are the eigenvalues of the matrix $W_n, X(i) = x_{i_1 i_2}, x_{i_2 i_3}, ... x_{i_k i_1}$ where $i = (i_1, ... i_k)$ and the summation $\sum_i$ runs over all possibilities that $i \in [1, ... n]^k$. By applying the moment convergence theorem, we can complete this proof for iid variables by showing two parts.

The first part being $E[\beta_k(W_n)]$ converges to the k-th moment of the semicircular distribution, which is (2.6).

The second part is that for each fixed k, $\sum_n Var[\beta_k(W_n)] < \infty$

The proof to part 1 ($E[\beta_k(W_n)] \to \beta_k$):

We have

$$E[\beta_k(W_n)] = \frac{1}{n^{1+k/2}} \sum EX(i)$$

And for each vector i, we can construct a Γ Graph G(i). Let X(i) = X(G(i)).

This summation is taken over all of our sequences: $i = (i_1, i_2, ...i_k) \in \{1, 2, ..., n\}^k$. We know that isomorphic graphs correspond to equal terms, so we first need to group terms according to each isomorphism classes, then split $E[\beta_k(W_n)]$ into three sums, based on their categories. So,

$$E[\beta_k(W_n)] = S_1 + S_2 + S_3$$

Where

$$S_j = n^{-1-k/2} \sum_{\Gamma(k,t)\in C_j} \sum_{G(i)\in\Gamma(k,t)} E[XG(i)]$$

. Here the first (outer) series is taken on all canonical $\Gamma(k,t)$ -graphs in category j and the second (inner) summation is taken on all isomorphic graphs for a given canonical graph. By definition of graph category, along with the entries of our perturbed matrices, we have $S_2 = 0$. Furthermore, since our random variables are bounded by C, the number of isomorphic graphs are less than n^t and $t \leq (k+1)/2$. From here, we can conclude that $|S_3| \leq n^{-1-k/2}O(n^t) = o(1)$

If k = 2m-1, then $S_1 = 0$. Now, let k= 2m. Since each edge coincides with an edge of opposite direction, each term in S_1 is $(E|x_{12}|^2)^m = 1$. Therefore we get:

$$S_1 = n^{-1-m} \sum_{\Gamma(2m,t)\in C_1} n(n-1)...(n-m) = \beta_{2m}(1-\frac{1}{m})...(1-\frac{m}{n}) \to \beta_{2m}$$

Now tht we have show part 1, we can move on to part 2, (that $Var(\beta_k(W_n))$ is finite for all k-natural numbers). We have:

$$Var(\beta_k(W_n)) = E[|\beta_k(W_n)|^2] - |E[\beta_k(W_n)]|^2 = \frac{1}{n^{2+k}} \sum * \{E[X(i)X(j)] - E[X(i)]E[X(j)]\}$$

,

Where $i = (i_1, ..., i_k), j = (j_1, ..., j_k)$ and $\sum *$ is taken over all possibilities for $i, j \in \{1, ...n\}^k$. Here, we see that $\beta_k(W_n)$ is real, meaning the RHS is meaningful, despite X(i) and X(j) being complex. Using i and j, let G(i) and G(j) be two different graphs as in part 1. If no edges coincide between the two graphs, then X(i) is independent of X(j). Therefore, the corresponding term in the sum is 0. If we combine both graphs and call it G, and we get one edge, then E[X(i)X(j)] = E[X(i)]E[X(j)] = 0, then the corresponding term in our summation is also 0.Let G have no single edges and the graph of non-coincidental edges has a cycle. Then the noncoincidental vertices of $G \leq k$. Now, consider the same case but with no cycle, then there exists at least one edge with coincidence multiplicity ≥ 4, and the number of noncoincidental vertices is still not above k. From here, we can also see that each term is $\leq 2C^{2k}n^{-2-k}$. As a result, we have

$$Var(\beta_k(W_n)) \leq K_k C^{2k} n^{-2}$$

Where K_k is a constant with respect to k. This shows that our variance is always finite. Part 2 is now verified, thus showing via moments that our ESD for Wigner matrices with iid variables converge to the Semicircular law.

Chapter 3

Spectral distribution of Covariance Matrices

3.1 Covariance Matrices

Besides the Physical applications Random Matrices have when looking at Wigner Matrices, there are many statistical and economic uses for Random Matrices. A sample Covariance Matrix is fundamental in statistics, and can also be used in Random Matrix theory. In economics, covariance matrices play a key role in Principal Component Analysis. In Statistics, it is used in hypothesis testing, factor analysis, and discrimination analysis. Many of these tests use eigenvalues.

Before continuing with the large dimensional spectral analysis of Covariance Matrices, we must first define what a sample covariance matrix is. Given $\{x_{jk}, j, k = 1, 2, ...\}$ is a double array of iid complex random variables, with mean 0 and variance σ^2. Let $x_j = (x_{1j}, ..., x_{pj})'$ and $X = (x_1, ..., x_n)$. The sample covariance matrix is defined by

$$S = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})(x_k - \bar{x})^*$$

where $\bar{x} = \frac{1}{n}\sum x_j$ When dealing with large dimensional cases, (such as ours) we usually just define S as the following:

$$S = \frac{1}{n}\sum_{k=1}^{n} x_k x_k^* = \frac{1}{n}XX^*$$

We are able to make this substitttion because $\bar{x}\bar{x}^*$ is a rank 1 matrix, so its deletion will not affect its Limiting Spectral distribution because of this theorem:

Theorem 3.1.1 *Let A and B be two $p \times n$ complex matrices, then*

$$||F^{AA^*} - F^{BB^*}|| \leq \frac{1}{p} rank(A - B)$$

Proof. Let $C = B - A$. Let $rank(C) = k$ Then, we know that for any nonnegative integer $i \leq p - k$, we have

$$\sigma_{i+k+1}(A) \leq \sigma_{i+1}(B) \qquad \text{and}$$
$$\sigma_{i+k+1}(B) \leq \sigma_{i+1}(A)$$

Thus, for any $x \in (\sigma_{i+1}(B), \sigma_i(B))$ we get,

$$F^{BB^*}(x) = 1 - \frac{i}{p} = 1 - \frac{i+k}{p} + \frac{k}{p} \leq F^{AA^*}(x) + \frac{k}{p}$$

Therefore, for all x,

$$F^{BB^*}(x) - F^{AA^*}(x) \leq \frac{k}{p} \qquad \text{and similarly,}$$
$$F^{AA^*}(x) - F^{BB^*}(x) \leq \frac{k}{p}$$

□

When discussing spectral distributions of covariance matrices, we assume that our dimension p tends to infinity proportionally to the degrees of freedom n. In other words, p/n converges to some positive constant. For sample Covariance Matrices of large dimension, we will show that the spectral distribution of them tend to a certain structure, named after its founders, the Marcenko-Pastur Law.

3.2 The M-P Law

The M-P law is a density function $F_y(x)$,

$$p_y(x) = \begin{cases} \frac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)} & \text{when } x \in (a,b) \\ 0 & \text{else} \end{cases}$$

This function has a point mass at $1 - 1/y$ at the origin if $y > 1$, where $a = \sigma^2(1-\sqrt{y})^2$ and $b = \sigma^2(1+\sqrt{y})^2$. y is the constant we use to show the dimension of our smaple size ratio index (p/n). σ^2 is the scale parameter, and if $\sigma^2 = 1$, the M-P law function is known as the standard M-P law (which is what we will be using in alignment with our earlier stipulations on our matrix perturbations). To help get an idea of determining how the ESD of covariance matrices tends to the M-P law, we must first show what the moments would be as before.

Theorem 3.2.1 *The k-th moment of the M-P law is as follows:*

$$\beta_k = \sum_{r=0}^{k-1} \frac{1}{r+1}\binom{k}{r}\binom{k-1}{r} y^r$$

Proof. We can show this directly:

$$\beta_k = \frac{1}{2\pi y}\int_a^b x^{k-1}\sqrt{(b-x)(x-a)}dx$$

let $x = 1 + y + z$. then

$$\beta_k = \frac{1}{2\pi y}\int_{-2\sqrt{y}}^{2\sqrt{y}} (1+y+z)^{k-1}\sqrt{4y-z^2}dz$$

$$= \frac{1}{2\pi y}\sum_{l=0}^{k-1}\binom{k-1}{l}(1+y)^{k-1-l}\int_{-2\sqrt{y}}^{2\sqrt{y}}(1+y+z)^{k-1}z^l\sqrt{4y-z^2}dz$$

$$= \frac{1}{2\pi y}\sum_{l=0}^{k-1}\binom{k-1}{l}(1+y)^{k-1-l}\int_{-2\sqrt{y}}^{2\sqrt{y}} z^l\sqrt{4y-z^2}dz$$

Let $z = 2\sqrt{y}u$,

$$= \frac{1}{2\pi y}\sum_{l=0}^{(k-1)/2}\binom{k-1}{2l}(1+y)^{k-1-2l}(4y)^{l+1}\int_{-1}^{1} u^{2l}\sqrt{1-u^2}du$$

let $u = \sqrt{w}$

$$= \frac{1}{2\pi y}\sum_{l=0}^{(k-1)/2}\binom{k-1}{2l}(1+y)^{k-1-2l}(4y)^{l+1}\int_0^1 w^{l-1/2}\sqrt{1-w}dw$$

$$= \sum_{l=0}^{(k-1)/2}\frac{(k-1)!}{l!(l+1)!(k-1-2l)!}y^l(1+y)^{k-1-2l}$$

$$= \sum_{l=0}^{(k-1)/2}\sum_{s=0}^{k-1-l}\frac{(k-1)!}{l!(l+1)!s!(k-1-2l-s)!}y^{l+s}$$

$$= \sum_{r=0}^{(k-1)}\sum_{r=l}^{k-1-l}\frac{(k-1)!}{l!(l+1)!(r-l)!(k-1-l-r)!}y^r$$

$$= \frac{1}{k}\sum_{r=0}^{(k-1)}\binom{k}{r}y^r\sum_{l=0}^{min(r,k-1-r)}\binom{s}{l}\binom{k-r}{k-r-l-1}$$

$$= \frac{1}{k}\sum_{r=0}^{(k-1)}\binom{k}{r}\binom{k}{r+1}y^r = \sum_{r=0}^{k-1}\frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}y^r$$

By definition, we know $\beta_{2k} \le b^{2k} = (1+\sqrt{y})^{4k}$. Therefore, the Carleman condition is satisfied, and the theorem is proven. □

3.3 M-P Law for iid Covariance Matrices

Now, we can look at the Empirical Spectral distribution of the sample covariance matrix with iid variable entries. Work on these kinds of statistical distributions has been done by Bai [1], Pereira [2], and Bose [8].

Theorem 3.3.1 *Suppose that $\{x_{ij}\}$ are iid complex random variables with mean 0 and variance σ^2. Assume p/n tends to some positive finite constant. Then with probability one, F^S tends to the M-P law.*

Before we begin the proof, it must be noted that as before in chapter 2, we prepare our sample covariance matrix. As implied by theorems and corollaries in chapter 2, our ESD's and LSD's remain unchanged. There are currently two ways to show this distribution; via MCT and the Stieltje's transform. Our first proof to theorem 3.3 will be with the moment convergence theorem. From calculus, we know that

$$\beta_k(S_n) = \int x^k F^{S_n}(dx) = p^{-1}n^{-k} \sum_{i_1,\ldots i_k} \sum_{j_1,\ldots j_k} x_{i_1j_1}\bar{x}_{i_2j_1}x_{i_2j_2}\ldots x_{i_kj_k}\bar{x}_{i_1j_k}$$
$$= p^{-1}n^{-k}\sum_{ij} X_{G(i,j)}$$

Where $G(i,j)$ are Δ-graphs. To show the convergence of our ESD of S_n we need to show two parts:

$$E(\beta_k(S_n)) = p^{-1}n^{-k}\sum_{ij} E(x_{G(i,j)}) = \sum_{r=0}^{k-1} \frac{y_n^r}{r+1}\binom{k}{r}\binom{k-1}{r} + O(n^{-1}) \quad (3.1)$$

$$Var(\beta_k(S_n)) = p^{-2}n^{-2k} \sum_{i_1,j_1,i_2,j_2} [E(x_{G_1(i_1,j_1)}x_{G_2(i_2j_2)} - E(x_{G_1(i_1,j_1)})E(x_{G_2(i_2j_2)}))] = O(n^{-2}) \quad (3.2)$$

with $y_n = p/n$ and graphs G_1 and G_2 are defined by (i_1, j_1) and (i_2, j_2) respectively.

In order to prove the equality in 3.1, we know that two terms will be equal if their graphs are isomorphic. This means we can rewrite our equation as follows:

$$E(\beta_k(S_n)) = p^{-1}n^{-k} \sum_{\Delta(k,r,s)} p(p-1)...(p-r)n(n-1)...(n-s+1)E(X_{\Delta(k,r,s)})$$

Where this summation is done over canonical $\Delta(k, r, s)$-graphs. We split the sum in three parts, according to $\Delta_1(k, r)$, $\Delta_2(k, r, s)$, and $\Delta_3(k, r, s)$ types of graphs. We know that $\Delta_2(k, r, s)$ contains at least one single edge, so the expection of that graph is zero. This means that,

$$S_2 = p^{-1}n^{-k} \sum_{\Delta_2(k,r,s)} p(p-1)...(p-r)n(n-1)...(n-s+1)E(X_{\Delta_2(k,r,s)}) = 0$$

Now, we can just look at the other two graphs. Consider $\Delta_3(k, r, s)$. We know $r + s < k$. Since the variable $x_{\Delta(k,r,s)}$ is bounded by $(2C/\bar{\sigma})^{2k}$, we conclude that,

$$S_3 = p^{-1}n^{-k} \sum_{\Delta_3(k,r,s)} p(p-1)...(p-r)n(n-1)...(n-s+1)E(X_{\Delta(k,r,s)} = O(n^{-1}).$$

We now only need to evaluate S_1. Consider the graph in $\Delta_1(k, r)$ (with $s = k-r$), each pair of coincident edges has a down edge and an up edge. Let us specify one with the edge (j_a, i_a) which will coincide with (i_a, j_a). The pair of edges together correspond to the expectation $E(|X_{i_a,j_a}|^2) = 1$ (from our earlier stipulations). Therefore, $E(X_{\Delta_1(k,r)}) = 1$. From here, we can apply lemmas from graph theory and combinatorics, as discussed in the appendix:

$$\begin{aligned} S_1 &= p^{-1}n^{-k} \sum_{\Delta_1(k,r)} p(p-1)...(p-r)n(n-1)...(n-s+1)E(X_{\Delta_1(k,r)}) \\ &= \sum_{r=0}^{k-1} \frac{y_n^r}{r+1} \binom{k}{r} \binom{k-1}{r} + O(n^{-1}) \\ &= \beta_k + o(1) \end{aligned}$$

Here, $y_n = p/n$ and y is some positive constant that y_n converges to. The equality in the first part is now shown.

As for equation 3.2, remember

$$Var(\beta_k(S_n)) = p^{-2}n^{-2k}\sum_{ij}[E(X_{G_1(i_1j_1)}X_{G_2(i_2j_2)}) - E(X_{G_1(i_1j_1)})E(X_{G_2(i_2j_2)})]$$

This is a parallel to our proof of the semicircle structure of the ESD's in chapter 2. We can see that the number of noncoincident vertices in G is less than or equal to $2k$. Since the terms are bounded, this verifies and concludes the proof. We have now verified, by moment convergence, that the ESD of a sample covariance matrix with iid entries converges to the M-P law.

3.4 Generalization to the non-iid Case

We can generalize to the non-iid case of covariance matrices by applying Stieltjes transforms to sample covariance matrices. It is practical to consider such a case, where entries of X_n are independent, but not necessarily identically distributed. Here, we have the following theorem to prove,

Theorem 3.4.1 *Suppose that, for each n, the entries of X are independent complex variables with a common mean μ and variance σ^2. Assume that p/n converges to some finite positive constant y, and that, for any $\eta > 0$,*

$$\frac{1}{\eta^2 np}\sum_{jk} E(|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \geq \eta\sqrt{n})) \longrightarrow 0$$

Before we begin the proof for theorem 3.4, we need to show what the Stieltje's transform is for the M-P law. As in Chapter 2, and earlier in Chapter 3, we assume the same method of truncation, centralization and rescaling on our sample covariance matrix. Let $z = u + iv$ with $v > 0$ and $s(z)$ be the Stieltje's transform of the M-P law.

Lemma 3.4.2 $s(z) = \frac{\sigma^2(1-y)-z+\sqrt{(z-\sigma^2-y\sigma^2)^2-4y\sigma^4}}{2yz\sigma^2}$

Proof. If $y < 1$ we have

$$s(z) = \int_a^b \frac{1}{x-z}\frac{1}{2\pi xy\sigma^2}\sqrt{(b-x)(x-a)}dx$$

where $a = \sigma^2(1-\sqrt{y})^2$ and $b = \sigma^2(1+\sqrt{y})^2$. If we let $x = \sigma^2(1+y+2\sqrt{y}cos(w))$ and $\zeta = e^{iw}$ this gives us,

$$s(z) = \int_0^\pi \frac{2}{\pi}\frac{1}{(1+y+2\sqrt{y}cos(w))(\sigma^2(1+y+2\sqrt{y}cos(w))-z)}sin^2(w)dw$$

$$= \frac{1}{4i\pi}\oint_{|\zeta|=1}\frac{(\zeta^2-1)^2}{\zeta((1+y)\zeta+\sqrt{y}(\zeta^2+1))(\sigma^2(1+y)\zeta+\sqrt{y}\sigma^2(\zeta^2+1)-z\zeta)}d\zeta$$

From here we use the Residue Theorem, this integrand function has five simple poles at,

$$\zeta_0 = 0$$

$$\zeta_1 = \frac{-(1+y)+(1-y)}{2\sqrt{y}}$$

,

$$\zeta_2 = \frac{-(1+y)-(1-y)}{2\sqrt{y}}$$

,

$$\zeta_3 = \frac{-\sigma^2(1+y)+z+\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}}{2\sigma^2\sqrt{y}}$$

$$\zeta_4 = \frac{-\sigma^2(1+y)+z-\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}}{2\sigma^2\sqrt{y}}$$

The residues from these are,

$$\frac{1}{y\sigma^2}, -\frac{1-y}{yz}, \frac{1-y}{yz}, \frac{1}{\sigma^2yz}\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}, -\frac{1}{\sigma^2yz}\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}$$

Respectively. We see that $\zeta_3\zeta_4 = 1$. We also see that by the definition of radical complex numbers, the real and imaginary part of $\sqrt{\sigma^4(1-y)^2-2\sigma^2(1+y)z+z^2}$

and $\sigma^2(1+y)+z$ have the same signs, meaning $|\zeta_3| > 1$, $|\zeta_4| < 1$. Furthermore, $|\zeta_1| = |-\sqrt{y}| < 1$ and $|\zeta_2| = |-1/\sqrt{y}| > 1$. By Cauchy integration, we get

$$s(z) = \frac{1}{2}(\frac{1}{y\sigma^2} - \frac{1}{\sigma^2 yz}\sqrt{\sigma^4(1-y)^2 - 2\sigma^2(1+y)z + z^2} - \frac{1-y}{yz})$$
$$= \frac{\sigma^2(1-y) - z + \sqrt{(z-\sigma^2-y\sigma^2) - 4y\sigma^4}}{2yz\sigma^2}$$

Since $y < 1$, this verifies the lemma. More generally, for other cases of y, the M-P law has point mass $1 - 1/y$ at zero, and $s(z)$ equals the integrand above, with $(1-y)/yz$ which gives us residues $|\zeta_3| = |-\sqrt{y}| > 1$ and $|\zeta_4| = |-1/\sqrt{y}| < 1$. This means, that the residue for ζ_4 would be counted, and the equation of the lemma would still hold. If $y = 1$, we yield the same equation by continuity. □

We are now ready to show via Stieltjes transform that the ESD of S_n converges to the M-P law for the more general, non-iid case. Let the Stieltjes transform of the ESD be denoted by $s_n(z)$. Define,

$$s_n(z) = \frac{1}{p}tr(S_n - zI_p)^{-1}$$

In order to complete the proof, we need to accomplish three parts:

(i)For any fixed $z \in \mathbb{C}^+$, $s_n(z) - Es_n(z) \to 0$ a.s.

(ii)For any fixed $z \in \mathbb{C}^+$, $Es_n(z) \to s(z)$, the Stieltjes transform of the M-P law.

(iii)Except for a null set, $s_n(z) \to s(z)$ for every $z \in \mathbb{C}^+$

From here, we see the third part is implied by the first two.

For the first part, we need to show almost sure convergence of the random part. Let $E_k(\cdot)$ denote the conditional expectation given by $\{x_{k+1}, ...x_n\}$. Since

$$(A + \alpha\beta^*)^{-1} = A^{-1} - \frac{A^{-1}\alpha\beta^* A^{-1}}{1 + \beta^* A^{-1}\alpha}$$

We can use this and get,

$$\begin{aligned} s_n(z) - Es_n(z) &= \frac{1}{p}\sum_{k=1}^{n}[E_k tr(S_n - zI_p)^{-1} - E_{k-1} tr(S_n - zI_p)^{-1}] \\ &= \frac{1}{p}\sum_{k=1}^{n}\gamma_k, \end{aligned}$$

The following lemmas are found and proven in Bai [1]: We can now apply the following Lemma,

Lemma 3.4.3 *If the matrix A and A_k the k-th submatrix of A of order $(n-1)$, are both nonsingular and symmetric, then*

$$tr(A^{-1}) - tr(A_k^{-1}) = \frac{1 + \alpha_k^t A_k^{-2}\alpha_k}{\alpha_{kk} - \alpha_k^t A_k^{-1}\alpha_k}$$

With a similar situation with the Hermitian in the complex case.

By applying this, we get

$$\begin{aligned} \gamma_k &= (E_k - E_{k-1})[tr(S_n - zI_p)^{-1} - tr(S_{nk} - zI_p)^{-1}] \\ &= -[E_k - E_{k-2}]\frac{x_k^*(S_{nk} - zI_p)^{-2}x_k}{1 + x_k^*(S_{nk} - zI_p)^{-1}x_k} \end{aligned}$$

With $S_{nk} = S_n x_k x_k^*$, we see,

$$\begin{aligned} &|\frac{x_k^*(S_{nk} - zI_p)^{-2}x_k}{1 + x_k^*(S_{nk} - zI_p)^{-1}x_k}| \\ &\le \frac{x_k^*((S_{nk} - uI_p)^2 + v^2 i_p)^{-1}x_k}{Im(1 + x_k^*(S_{nk} - zI_p)^{-1}x_k)} = \frac{1}{v} \end{aligned}$$

.

Here we see that $\{\gamma_k\}$ forms a sequence of bounded martingale differences, and we can now apply this lemma,

Lemma 3.4.4 *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing $\sigma - field\{\mathcal{F}_k\}$. Then for $p > 1$,*

$$E|\sum X_k|^p \leq K_p E(\sum |X_k|^2)^{p/2}$$

The proof to the above lemma can be found in the appendix. Letting $p = 4$, we see,

$$E|s_n(z) - E(s_n(z)|^4 \leq \frac{K_4}{p^4} E(\sum_{k=1}^{n} |\gamma_k|^2)^2 \leq \frac{4K_4 n^2}{v^4 p^4} = O(n^{-2})$$

Citing Borel-Cantelli lemma, the first part of the proof is now complete. We now must show the convergence of the mean,

$$Es_n(z) \to s(z)$$

where $s(z)$ is defined as before, with $\sigma^2 = 1$. Here we use the following theorem,

Theorem 3.4.5 *If both A and A_k, $k = 1, 2, ..., n$ are nonsingular, and we write $A^{-1} = [a^{kl}]$ then*

$$a^{kk} = \frac{1}{a_{kk} - \alpha_k^t A_k^{-1} \beta_k}$$
$$tr(A^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \alpha_k^t A_k^{-1} \beta_k}$$

Where a_{kk} is the k-th diagonal entry of A, A_k is defined above, α_k^t is the vector from the k-th row of A by deleting the k-th entry, and β_k is the vector from the k-ht column by deleting the k-th entry.

This theorem gives us,

$$s_n(z) \frac{1}{p} \sum_{k=1}^{p} \frac{1}{\frac{1}{n}\alpha_k^t \bar{\alpha}_k - z - \frac{1}{n^2}\alpha_k^t X_k^* (\frac{1}{n} X_k X_k^* - zI_{p-1})^{-1} X_k \bar{\alpha}_k}$$

Where X_k is the matrix obtained from X with the k-th row removed and α_k^t is the k-th row of X. Then, let

$$\epsilon_k = \frac{1}{n}\alpha_k^t\bar{\alpha}_k - 1 - \frac{1}{n^2}\alpha_k^t X_k^*(\frac{1}{n}X_kX_k^* - zI_p^{-1})^{-1}X_k\bar{\alpha}_k + y_n + y_nzEs_n(z)$$

And as before, $y_n = p/n$. Now,

$$Es_n(z) = \frac{1}{1 - z - y_n - y_nzEs_n(z)} + \delta_n$$

With $\delta_n = \frac{-1}{p}\sum_{k=1}^{p} E(\frac{\epsilon_k}{(1-z-y_n-y_nzEs_n(z))(1-z-y_n-y_nzEs_n(z)+\epsilon_k)})$

If we solve $Es_N(z)$ we get two solutions,

$$s_1(z) = \frac{1}{2y_nz}(1 - z - y_n + y_nz\delta_n + \sqrt{(1 - z - y_n - y_nz\delta_n)^2 - 4y_nz})$$
$$s_2(z) = \frac{1}{2y_nz}(1 - z - y_n + y_nz\delta_n - \sqrt{(1 - z - y_n - y_nz\delta_n)^2 - 4y_nz})$$

Which means, we need to show that $Es_n(z) = s_1(z)$, and $\delta_n \to 0$. As $v \to \infty$, we know that $Es_n(z) \to 0$ and $\delta_n \to 0$. This implies that $Es_n(z) = s_1(z)$ for all z with a large imaginary part.

Assume that this is not the case for all $z \in \mathbb{C}^+$, by continuity of s_1 and s_2 there exists a $z_0 \in \mathbb{C}^+$ such that $s_1(z_0) = s_2(z_0)$ which shows

$$1 - z_0 - y_n - y_nz\delta_n)^2 - 4y_nz_0(1 + \delta_n(1 - z_0 - y_n)) = 0$$

Implying,

$$Es_n(z_0) = s_1(z_0) = \frac{1 - z_0 - y_n + y_nz_0\delta_n}{2y_nz_0}$$
$$= \frac{1 - z_0 - y_n}{y_nz_0} + \frac{1}{y_n + z_0 - 1 + y_nz_0Es_n(z_0)}$$

Looking back at properties of the Stieltjes transform, we know that for any transform $s(z)$ of probability F on $\mathbb{R}^+$ and positive y, we have

$$Im(y + z - 1 + yzs(z)) = Im(z - 1 + \int_0^\infty \frac{yxdF(x)}{x - z})$$

$$= v(1 + \int_0^\infty \frac{yxdF(x)}{(x-u)^2 + v^2}) > 0$$

. This means that the imaginary part of the second term must be negative. If $y_n \leq 1$, it is apparent that $Im(1-z_0-y_n)/(y_n z_0) < 0$ which implies the imaginary part of the expectation is negative at z_0. This is a contradiction, since a Stieltjes transform should have a positive imaginary part. This verifies our conclusion for $y_n \leq 1$. In general, by the previous equations, we have

$$y_n + z_0 - 1 + y_n z_0 E s_n(z_0) = \sqrt{y_n z_0}$$

Let $\tilde{s}_n(z)$ be the Stieltjes transform for our scaled covariance matrix $\frac{1}{n}X^*X$. We know that $\frac{1}{n}X^*X$ and S_n have the same non-zero eigenvalues, we have the relation between s_n and $\tilde{s}_n$ given by

$$s_n(z) = y_n^{-1}\tilde{s}_n(z) - \frac{1-1/y_n}{z}$$

Which is true for all values of y_n. This gives us

$$y_n - 1 + y_n z_0 E s_n(z_0) = z_0 E\tilde{s}_n(z_0)$$

By substitution, we get

$$1 + E\tilde{s}_n(z_0) = \sqrt{y}/\sqrt{z_0}$$

This is a contradiction, since the LHS has positive imaginary part, and the RHS has a negative imaginary part. This verifies the convergence of the expectation. We now must look at the convergence of $\delta_n \to 0$.

Let,

$$\delta_n = -\frac{1}{p}\sum_{k=1}^{p}(\frac{E\epsilon_k}{(1-z-y_n-y_n zEs_n(z))^2})$$

$$+\frac{1}{p}\sum_{k=1}^{p}E(\frac{\epsilon_k^2}{(1-z-y_n-y_n zEs_n(z))^2(1-z-y_n-y_n zEs_n(z)+\epsilon_k)})$$

$$= J_1 + J_2$$

We see from here that,

$$|E\epsilon_k| = |-\frac{1}{n^2}EtrX_k^*(\frac{1}{n}X_kX_k^* - zI_{p-1})^{-1}X_k + y_n + y_nzEs_n(z)|$$
$$= |\frac{-1}{n}Etr(\frac{1}{n}X_kX_k^* - zI_{p-1})^{-1}\frac{1}{n}X_kX_k^* + y_n + y_nzEs_n(z)|$$
$$\leq \frac{1}{n} + \frac{|z|y_n}{n}E|tr(\frac{1}{n}X_kX_k^* - zI_{p-1})^{-1} - s_n(z)|$$
$$\leq \frac{1}{n} + \frac{|z|y_n}{nv} \to 0.$$

Showing J_1 converges to zero. As for J_2,

$$Im(1 - z - y_n - y_nzEs_n(z) + \epsilon_k)$$
$$= Im(\frac{1}{n}\alpha_k^t\bar{\alpha_k} - z - \frac{1}{n^2}\alpha_k^tX_k^*(\frac{1}{n}X_kX_k^* - zI_{p-1})^{-1}X_k\bar{\alpha_k})$$
$$= -v(1 + \frac{1}{n^2}\alpha_k^tX_k^*[(\frac{1}{n}X_kX_k^* - uI_{p-1})^2 + v^2I_{p-1}]^{-1}X_k\bar{\alpha_k}) < -v,$$

Combining this with the Stieltjes transform,

$$|J_2| \leq \frac{1}{pv^3}\sum_{k=1}^{p}E|\epsilon_k|^2$$

$$= \frac{1}{pv^3}\sum_{k=1}^{p}[E|\epsilon_k - \tilde{E}(\epsilon_k)|^2 + E|\tilde{E}(\epsilon_k) - E(\epsilon_k)|^2 + (E(\epsilon_k))^2]$$

Where $\tilde{E}(\dot{})$ is the denotes the conditional expectation. We have shown J_1 converges to 0. Let $A = (a_{ij}) = I_n - \frac{1}{n}X_k^*(\frac{1}{n}X_kX_k^* - zI_{p-1}^{-1}X_k$. Then,

$$\epsilon_k - \tilde{E}(\epsilon_k) = \frac{1}{n}(\sum_{i=1}^{n}a_{ii}(|x_{ki}|^2 - 1) + \sum_{i\neq j}a_{ij}x_{ki}\bar{x}_{kj})$$

Which gives us,

$$\frac{1}{n^2}\tilde{E}|\epsilon_k^t - \tilde{E}\epsilon_k|^2 \leq \frac{\eta_n^2}{v^2} + \frac{2}{nv^2}$$

Which is true since $|a_{ii}| \leq v^{-1}$. Using Martingale decomposition, we see that,

$$E|\tilde{E}\epsilon_k - E\epsilon_k|^2 = \frac{|z|^2y^2}{n^2}E|tr(\frac{1}{n}X_kX_k^* - zI_{p-1})^{-1} - Etr(\frac{1}{n}X_kX_k^* - zI_{p-1})^{-1}|^2$$

$\leq \frac{|z|^2 y^2}{nv^2} \to 0.$

This completes the proof, using the three estimations and this demonstrates the mean convergence of the ESD of S_n.

chapterExtreme Eigenvalues Chapter 4 takes our study of spectral distributions of random matrices in a different direction. Rather than study the Empirical Spectral Distribution (ESD) of certain kinds of random matrices, we will be looking at limits of extreme eigenvalues. In other words, when looking at the Limiting Spectral Distributions (LSD). It makes sense to focus on the lower and upper bounds of the eigenvalues as the dimension tends to infinity.

Definition 3.4.6 *The Limiting Spectral Distribution is the weak limit of the sequence of Empirical Spectral Distributions of a sequence of square random matrices.*

In the following sections, we will look at the bounds of our LSD's for the matrices previous discussed in this paper: Wigner and Covariance Matrices. This topic has largely been covered by Liu [9], and Yin and Bai [7].

3.5 Extreme Eigenvalues of Wigner Matrices

This section will cover the complex iid variable case of a scaled Wigner matrix ($\sqrt{n}W_n$), on the limits of Extreme Eigenvalues. Here we have the following theorems to prove, which have also been proved in Bai [1]:

Theorem 3.5.1 *Suppose the diagonal entries of the Wigner matrix ($\sqrt{n}W_n$) are* *iid real random variables, and the entries above them are iid complex random variables, all independent. Then, the largest eigenvalue of W tends to some positive number c_1 with probability 1 iff the following five conditions are true:*

- $E((X_{11}^{+})^2) < \infty$
- $E(x_{12}) \in \mathbb{R}, E(x_{12}) \leq 0$
- $E(|x_{12} - E(x_{12})|^2) = \sigma^2$
- $E(|x_{12}^4|) < \infty$
- $c_1 = 2\sigma$ *Here,* $x^+ = max(x, 0)$.

We can use the conditions and implication of theorem 4.2 to find the limit of the smallest eigenvalues of Wigner Matrix, which gives way to the next theorem.

Theorem 3.5.2 *Suppose that the diagonal entries of our Wigner matrix W_n are iid real random variables, with entries above the diagonal being iid complex random variables, with all variables independent of each other. Then, the largest eigenvalue of W tends to c_1 and the smallest eigenvalue tends to c_2 with probability 1 iff the following five conditions are true:*

- $E(x_{11}^2) < \infty$
- $E(x_{12}) = 0$
- $E(|x_{12}|^2) = \sigma^2$
- $E(|x_{12}^4|) < \infty$
- $c_1 = 2\sigma, c_2 = -2\sigma$

Also from Theorem 4.2, using the same proof, we can see the weak convergence of the extreme eigenvalue of a large Wigner Matrix, which gives us the following theorem.

Theorem 3.5.3 *Suppose that the diagonal entries of the scaled Wigner matrix* $\sqrt{n}W_n = (x_{ij})$ *are iid real random variables, and the entries above the diagonal are iid complex random variables, and all variables are independent of each other. Then, the largest eigenvalue of* W *tends to* $c > 0$ *in probability iff the following conditions hold true:*

- $P(x_{11}^{+} > \sqrt{n}) = o(n^{-1})$
- $E(x_{12}) \in \mathbb{R}, \leq 0$
- $E(|x_{12} - E(x_{12})|^2) = \sigma^2$
- $P(|x_{12}| > \sqrt{n}) = o(n^{-2})$
- $c = 2\sigma$

In order to prove theorem 4.2, we use the assumptions of our Wigner matrix from chapter 2, which allows us to be given the following property,

$$\liminf_{n\to\infty} \lambda_n(W) \geq 2,$$

almost surely.

This means that to prove this theorem, we only need to show,

$$\limsup_{n\to\infty} \lambda_n(W) \leq 2$$

almost surely.

Here we break down each of the conditions and their implications, along with some combinatorics, will allow us to conclude the sufficiency of conditions for theorem 4.2.

Our first condition implies that $\limsup \frac{1}{\sqrt{n}} \max_k \leq x_{kk}^+ = 0$ almost surely. Note the relation,

$$\lambda_{\max}(W) = \frac{1}{\sqrt{n}} \max_{||z||=1} (\sum_{j,k} z_j \bar{z}_k x_{jk})$$

$$= \max_{||z||=1} [\frac{1}{\sqrt{n}} \sum_{j \neq k} z_j \bar{z}_k (x_{jk} - E(x_{jk})) + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |z_k|^2 x_{kk}$$

$$+ Re(E(x_{12})) \frac{1}{\sqrt{n}} \sum_{j \neq k} z_j \bar{z}_k]$$

$$\leq \max_{||z||=1} (\frac{1}{\sqrt{n}} \sum_{j \neq k} z_j \bar{z}_k (x_{jk} - E(x_{jk})) + \frac{1}{\sqrt{n}} \max_k (x_{kk}^+ - Re(E(x_{12}))))$$

$$\leq \lambda_m ax(\tilde{(}W)) + o_{a.s.}(1),$$

Here $\tilde{W}_n$ is the matrix whose diagonal entries are zero and off-diagonal entries are $\frac{1}{\sqrt{n}}(x_{ij} - E(x_{ij}))$. We now only need to show the upper bound for our max eigenvalue for $\tilde{W}$, since we have shown before that the LSD remains unchanged with $\tilde{W}$.

As in previous chapters, we truncate the off-diagonal elements. By the fourth condition of theorem 4.2, for any $\delta > 0$, we know,

$$\sum_{k=1}^{\infty} \delta^{-2} 2^k E|x_{12}|^2 I(|x_{12}| \geq \delta 2^{k/2}) < \infty.$$

Now, choose a decreasing sequence of δ_n's that converge to zero, such that,

$$\sum_{k=1}^{\infty} \delta_{2^k}^{-2} 2^k E|x_{12}|^2 I(|x_{12}| \geq \delta_{2^k} 2^{k/2}) < \infty.$$

Let

$$\tilde{W} = \frac{1}{\sqrt{n}} (x_{jk} I(|x_{jk}| \leq \delta_n \sqrt{n})) \tag{3.3}$$

. Then by 3.3, we get,

$$P(W \neq \tilde{W}, i.o.) = \lim_{k \to \infty} P(\cup_{n=2^k} \cup_{1 \leq i < j \leq n} (|x_{jk} \geq \delta_n \sqrt{n})) \leq 0$$

By our choice of δ_n, we get that the maximum eigenvalue of the expectation of $\tilde{W}$ converges to zero. This means we only need to consider the upper limit of the largest eigenvalue of the difference between $\tilde{W}$ and $E(\tilde{W})$. As before, we assume W is a truncated and recentralized matrix, so $\sqrt{n}W_n = (x_{ij})$ and we have the following conditions about our Wigner Matrix:

- $x_{ii} = 0$,
- $E(x_{ij}) = 0, \sigma_n^2 = E(|x_{ij}|^2) \leq 1$, for $i \neq j$
- $|x_{ij}| \leq \delta_n \sqrt{n}$, for $i \neq j$
- $E|x_{ij}^l| \leq b(\delta_n \sqrt{n})^{l-3}$, for some constant $b > 0$ and $i \neq j$, $l \geq 3$

We will prove,

$$\limsup_{n \to \infty} \lambda_n(W) \leq 2$$

almost surely. Under the four above assumptions, with iid entries. We know, for any integer k and real number $\eta > 2$, we have:

$$P(\lambda_{max}(W_n) \geq \eta) \leq P(tr[(W_n)^k] \geq \eta^k) \leq \eta^{-k} E(tr(W_n)^k).$$

To complete the proof, let k be a sequence of even integers $k = k_n = 2s$ with the properties:

$$k / \log n \to \infty$$

$$k\delta_n^{1/3} / \log n \to 0$$

We now must show that the RHS of our above inequality is summable. This means we estimate,

$$E(tr(W^k)) = n^{-k/2} \sum_{i_1,\ldots,i_k} E(x_{i_1,i_2} x_{i_2 i_3} \ldots x_{i_k i_1})$$
$$= n^{-k/2} \sum_G \sum_i E(x_G(i)),$$

Where the of G are $\Gamma(k)$ -graphs as defined in earlier sections. We can now look at the different types of sections based on edges. A T_1 edge is if $f(a+1) = \max(f(1), ..., f(a)) + 1$ the edge, $e_a = (f(a), f(a+1))$. This kind of edge leads to a new vertex in the path of $e_1, ..., e_a$. A T_3 edge if it coincides with a T_1 that is single until the T_3 appears. It is irregular if there is only one T_1 single up to a–if it does not coincide with any other edges in the chain up to a. All other T_3 edges are called regular. All other edges are T_4. A T_2 edge is merely the first appearance of a T_4.

To show our estimations, we need the following lemmas from Bai and Silverstein [1].

Lemma 3.5.4 *Let $(f(a), ..., f(c))$ be a chain such that the edge $(f(a), f(a+1))$ is a T_1 single up to c. Then there is a T_2 edge contained in the chain $(f(a), ..., f(c))$.*

Lemma 3.5.5 *Let t denote the number of T_2 edges and s denote the number of innovations in the chain $(f(1), ...f(a))$ that are single up to a and have a vertex coincident with $f(a)$. Then $s \leq t + 1$.*

Lemma 3.5.6 *The number of regular T_3 edges is not greater than twice the number of T_2 edges.*

Using these lemmas, we can now prove our original theorem regarding the extreme eigenvalues of Wigner Matrices. We begin with our estimation from earlier. If G has a single edge, then our corresponding terms are zero. So, we only need to estimate terms corresponding to Γ_1 and Γ_3 graphs. Suppose that there are r innovations ($r \leq s$) and t T_2 edges in the graph G. Then, there are r T_3 edges, $k-2r$, T_4 edges and $r+1$ noncoincident vertices.

From here, we see that the number of graphs of each isomorphic class is less than n^{r+1} and the expectation corresponding to each canonical graph is not larger than $b^t(\delta_n\sqrt{n})^{k-2r-t}$. Then, we need to estimate the number of canonical graphs. We first note, there are at most $\binom{k}{r}$ ways to choose r edges out of the total to be r T_1 edges. Then, there are at most $\binom{k-r}{r}$ ways to choose r edges out of the rest of the $k-r$ edges for the r T_3 edges. Then the rest of the $k-2r$ edges are for the T_4 edges. For a T_1, we have the relation $f(l) = \max f(1), ..., f(l+1)+1$, the only way to plot the T_1 is unique, after the subgraph before this edge is plotted. Irregular T_3 edges have only one single T_1 edge to be matched. Therefore, there is only one way to plot it when the subgraph prior to this T_3 edge is plotted. By lemma 4.6, there are at most $t+1$ T_1 edges to be matched by a regular T_3 edge. This means that each regular T_3 edge has at most $t+1$ ways to plot it. By lemma 4.7, we have at most $2t$ regular T_3 edges.

This means there are at most $(t+1)^{2t} \leq (t+1)^{2(k-2r)}$ ways to plot the regular T_3 edges. Now, we have to look at the T_4 edges. We know for each T_4 edge, we have at most $(r+1)^2 < k^2$ ways to determine its two vertices. Then, we have at most $\binom{k^2}{t}$ ways to plot t T_2 edges. After all the t positions of T_4 are determined, there are at most $t^{k-2r} < (t+1)^{k-2r}$ ways to distribute the $k-2r$

T_4 edges. We now have the following inequality:

$$
\begin{aligned}
&E(tr(W)^k) \\
&\le n^{-k/2}\sum_{r=1}^{k/2}\sum_{t=0}^{k-2r} n^{r+1}\binom{k}{r}\binom{k-r}{r}\binom{k^2}{t}(t+1)^{3(k-2r)}b^t(\sqrt{n}\delta_n)^{k-2r-t} \\
&\quad \le \sum_{r=1}^{k/2}\sum_{t=0}^{k-2r} n^{r+1}n\binom{k}{2r}\binom{2r}{r}(t+1)^{3(k-2r)}[bk^2/(\sqrt{n}\delta_n)]^t\delta_n^{k-2r} \\
&\qquad\quad \le n^2b^{-1}\sum_{r=1}^{k/2}\binom{k}{2r}2^{2r}\delta_n^{k-2r}(\frac{3(k-2r)}{\log n\delta_n/bk^2}^{3(k-2r)} \\
&\qquad\quad \le n^2[2+(10\delta_n^{1/3}k/\log n)^3] = n^2[2+o(1)]^k
\end{aligned}
$$

With $a = t+1$, and since $\delta_n^{1/3}k/\log n \to 0$, we can use this to obtain,

$$P(\lambda_{max}(W_n) \ge \eta) \le n^2(2+o(1)/\eta)^{-k}$$

which is summable since $k/\log n \to \infty$. The proof is now complete. We now have an understanding on the extreme eigenvalues for Wigner Matrices, which imply the other theorems on it in this section.

3.6 Extreme Eigenvalues of Covariance Matrices

Now that we have shown the limits of extreme eigenvalues of Wigner Matrices, we can move on to looking at the same characteristics for Sample Covariance Matrices. Analogous to our last section, we introduce the following theorems:

Theorem 3.6.1 *Suppose the entries of $X_n = (x_{jkn}, j \le p, k \le n)$ are independent, but not necessarily iid. If the entries satisfy the following:*

$$E(x_{jkn}) = 0$$

$$|x_{jkn}| \leq \sqrt{n}\delta_n$$

$$\max_{j,k} |E|X_{jkn}|^2 - \sigma^2| \to 0$$

as $n \to \infty$

$$E|x_{jkn}|^l \leq b(\sqrt{n}\delta_n)^{l-3}$$

for all $l \geq 3$ *Where* $\delta_n \to 0$ *and* $b > 0$. *Let* $S_n = \frac{1}{n}X_nX_n^*$.. *Then, for any* $x > \epsilon > 0$ *and integers* $j, k \geq 2$, *we have*

$$P(\lambda_{max}(S_n) \geq \sigma^2(1+\sqrt{y})^2 + x) \leq Cn^{-k}(\sigma^2(1+\sqrt{y})^2 + x - \epsilon)^{-k}$$

for some constant $C > 0$.

From here, we construct the main theorem of this section,

Theorem 3.6.2 *Assume the entries of* $\{x_{ij}\}$ *are a double array of iid complex random variables with mean zero, variance* σ^2 *and finite fourth moment. Let* $X_n = (x_{ij}; i \leq p, j \leq n)$ *be the p x n matric of the upper left corner of the double array. If* $p/n \to y \in (0,1)$, *then with probability 1, we have*

$$-2\sqrt{y}\sigma^2 \leq \liminf_{n\to\infty} \lambda_{min}(S_n - \sigma^2(1+y)I_n)$$

$$\leq \limsup_{n\to\infty} \lambda_{max}(S_n - \sigma^2(1+y)I_n) \leq 2\sqrt{y}\sigma^2$$

In order to prove theorem 4.9, we use several lemmas from Bai and Silverstein [1]. The illustration of the proof is that that we estimate the spectral norm of the power matrix, $(S_n - \sigma^2(1+y)I)^l$. In the first step, we split this matrix into several different matrices, with a key matrix being noted as $T_n(l)$. This matrix is defined in the following lemma, proved by Bai and Silverstein [1], which gives us the estimate of the norm. This leads us to estimate the norm of the rest of power matrix.

Lemma 3.6.3 *Define,*

$$T_n(l) = n^{-l}(\sum{}' x_{av_1}\bar{x}_{u_1v_1}x_{u_1v_2}\bar{x}_{u_2v_2}...x_{u_{l-1}v_l}\bar{x}_{bv_l})$$

Here the summation runs for $v_1, ..., v_l = 1, 2, ..., n$ *and* $u_1, ..., u_{l-1} = 1, 2, ..., p$ *with the constraints* $a \neq u_1$*,* $u_1 \neq u_2...$ $u_{l-1} \neq b$ ... *and similary for* v_i*. By theorem 4.9's conditions, we get,*

$$\limsup_{n\to\infty} ||T_n(l)|| \leq (2l+1)(l+1)y^{(l-1)/2}\sigma^{2l}$$

almost surely.

Once we consider this lemma, we can then apply another lemma using the conditions from our theorem:

Lemma 3.6.4 *Under the conditions of theorem 4.9, we get,*

- $\limsup ||Y_n^{(1)}|| \leq \sqrt{y}\sigma$
- $\limsup ||Y_n^{(2)}|| \leq \sqrt{E|x_{11}|^4}$ *almost surely,*
- $\limsup ||Y_n^{(f)}|| = 0$ *almost surely, for all* $f > 2$

For clarification, we will prove this using lemma 4.10.

Proof. We have

$$||Y_n^{(1)}||^2 \leq ||T_n(1)|| + \frac{1}{n}\max_{i\leq p}\sum_{j=1}^{n}|x_{ij}|^2$$

Since, $||Y_n^{(2)}||^2 \leq tr(Y_n^{(2)}Y_n^{(2)*})$, we get that

$$||Y_n^{(2)}||^2 \leq^{-2} \sum_{ij}|x_{ij}|^4 \to yE(|x_{11}|^4)$$

For the case of $f > 2$, we get

$||Y_n^{(f)}||^2 \leq n^{-f} \sum_{ij} |x_{ij}|^{2f} \to 0$ almost surely. □

To continue the estimation of the the spectral norm, we use the next lemma.

Lemma 3.6.5 *As before, using conditions in 4.9, we have*

$$T_n T_n(k) = T_n(k+1) + y\sigma^2 T_n(k) + y\sigma^4 T_n(k-1) + o(1)$$

almost surely.

Proof. Assume $\sigma = 1$ without loss of generality, as in chapters 2 and 3. By our previous lemma,

$$T_n(k) = Y_n(Y_n^* \odot Y_n^* \odot ... \odot Y_n^*) \quad \text{(k times)}$$

$$-[\text{diag}\,(Y_nY_n^*)]T_n(k-1) + Y_n^{(3)} \odot (Y_n^* \odot Y_n^* \odot ... \odot Y_n^*)$$

$$= Y_n(Y_n^* \odot Y_n^* \odot ... \odot Y_n^*) - T_n(k-1) + o(1) \quad \text{almost surely, and analogously,}$$

$$T_n(k+1) = Y_n(Y_n^* \odot Y_n^* \odot ... \odot Y_n^*) \quad \text{(k times)}$$

$$-[\text{diag}\,(Y_nY_n^*)]T_n(k) + o(1) \quad \text{almost surely.}$$

$$= Y_nY_n^*T_n(k) - Y_n\text{diag}\,(Y_n^*Y_n^*)(Y_n^* \odot Y_n^* \odot ... \odot Y_n^*)$$

$$-\text{diag}\,(Y_nY_n^*)T_n(k) + o(1)$$

$$= T_nT_n(k) - yY_n(Y_n^* \odot Y_n^* \odot ... \odot Y_n^*) + o(1)$$

$= T_nT_n(k) - y(T_n(k) + T_n(k-1)) + o(1)$ almost surely. □

The Lemma is as follows,

Lemma 3.6.6 *Under the same conditions, we have*

$$(T_n - y\sigma^2 I_p)^k = \sum_{r=0}^{k} (-1)^{r+1} \sigma^{2(k-r)} T(r) \sum_{t-0}^{[(k-r)/2]} C_i(k,r) y^{k-r-i} + o(1)$$

With $|C_i(k,r)| \leq 2^k$.

Proof. Let $k = 1$, then $T(0) = I$ which is trivial with $C_0(1,1) = 1$ and $C_0(1,0) = 1$, by induction, this is true in general for all natural k. □

We can now apply these lemmas to prove our theorem. Assume $\sigma^2 = 1$. We know that,

$$||S_n - I_p - T_n|| \leq \max_{i \leq p} |\sum_{j=1}^{n} (|x_{ij}|^2 - 1)| \to 0$$

almost surely. Therefore, what is now left to show is that

$$\limsup ||T_n - yI_p|| \leq 2\sqrt{y}$$

By the previous lemmas, for any fixed k,

$$\limsup ||T_n - yI_p||^k \leq Ck^4 2^k y^{(k-1)/2}$$

Implying, $\limsup ||T_n - yI_p|| \leq C^{1/k} k^{4/k} 2 y^{(k-1)/2k}$

When $k \to \infty$, we see the estimation and bounds hold true, concluding the proof to this theorem.

Chapter 4

Gaussian Unitary Ensembles

So far in this text, we have covered the distribution of eigenvalues and the bounds of maximum eigenvalues for Wigner matrices and the sample covariance matrices. In the previous chapter, we looked at what were the extreme eigenvalues of certain kinds of random matrices: Wigner matrices and Sample Covariance Matrices. The upper bound for Eigenvalues of any matrix have always been statistically and computationally relevant; its applications have ranged from image buffering, high-energy states of Hamiltonians, to finding conditional numbers. In other words, extreme eigenvalues are worth noting, and in following suit, the limiting distribution of normalized extreme eigenvalues a are important.

We introduce the Gaussian Unitary Ensemble, which relates directly to Wigner Matrices in the sense that it is the set of $n \times n$ Hermitian matrices. The Gaussian Unitary Ensemble is a family of statistical distributions of $n \times n$ Hermitian Matrices. We use the term unitary to describe them because after spectral decomposition, we know that these matrices are invariant under unitary conjugation. For further reading on this topic, the reader can be pointed to Edelman [3], Fyaodorov [4], and Rezakhanlou [5].

Another property of the GUE is that the linearly independent elements of our Hermitian matrix are also statistically indepndent. This means that it can be written as a product of functions depending only on the independent matrix elements. The Gaussian Measure is described by the following equation,

$$P(H) = \frac{1}{Z_u} e^{\frac{1}{2}TrH^2}$$

On the space of $n \times n$ Hermitian matrices, where Z_u is a normalization constant.

The joint distribution functions of the eigenvalues of our $n \times n$ hermitian matrices $P_N^{(2)}$ is given by,

$$P_N^{(2)}(dx_1, ...dx_N) = \bar{C}_N^{(2)} 1_{x_1 \leq ... \leq x_N} |\Delta(x)|^2 \prod_{i=1}^{N} e^{2x_i^2/4} dx_1...dx_N$$

where

$$\bar{C}_N^{(2)} = (\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} |\Delta(x)|^2 \prod_{i=1}^{N} e^{-2x_i^2/4} dx_i)^{-1}$$

The significance of this topic is that it leads on to further reading on topics such as orthogonal polynomials, which were not introduced in this paper. These are used not only in understanding the Wigner Semi-Circle law, but also asymptotic spectral distributions as $n \to \infty$.

Chapter 5

Appendix

In this appendix we will define and introduce the various structures and lemmas used in the earlier text.

5.1 General Probability Definitions

Definition 5.1.1 *A Sequence of random variables is said to converge in distribution (weakly converge) if*

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for every number $x \in \mathbb{R}$ at which F is continuous, where F and F_n are cumulative distribution functions of random variables X_n and X

Definition 5.1.2 *A sequence $\{X_n\}$ of random variables is said to converge in mean (converge strongly) towards X if for $\epsilon > 0$*

$$\lim_{n \to \infty} Pr(|X_n - X| \geq \epsilon) = 0$$

.

5.2 Some lemmas from Combinatorics and graph theory

We now introduce some definitions and lemmas from Combinatorics/graph theory:

Definition 5.2.1 *A Γ graph is a triple (E, V, F) where E is the set of edges, V is the set of vertices, and F is a function where $F : E \to V x V$.*

Here are some properties on Γ graphs: If $F(e) = (v_1, v_2)$, the vertices v_1, v_2 are called the ends of the edge e, v_1 is the initial of e, and v_2 is the terminal of e. If $v_1 = v_2$, then the edge e is a loop. If two edges have the same set of ends, then they are said to be coincident.

The Graphs used in these texts have the following properties: Its vertex set is $V = \{1, ..., t\}$ Its edge set is $E = \{e_1, ..., e_k\}$ There exists a function g from $\{1, 2, ..., k\}$ onto $\{1, ..., t\}$ satisfying $g(1) = 1$ and $g(i) \leq \max\{g(1), ...g(i-1)\} + 1$ for $1 < i \leq k$. $F(e_i) = (g(i), g(i+1))$, for $i = 1, ..., k$ with convention $g(k+1) = g(1) = 1$. From here, we look at the isomorphism class of a Γ graph.

Lemma 5.2.2 *Each isomorphism class contains $n(n-1)...(n-t+1)$ $\Gamma(k, t)$ graphs.*

We can now classify canonical gamma graphs into three categories:

Category 1 is when each edge is coincident with other edge of opposite direction and the graph of noncoincident edges forms a tree.

Category 2 is all canonical graphs that have at least one single edge, such as an edge not coincident with any other edges.

Category 3 is all others.

Lemma 5.2.3 *A* $\Gamma_3(k,t)$ *graph,* $t \leq \frac{k+1}{2}$

Lemma 5.2.4 *The number of* $\Gamma_1(2m)$ *graphs is* $\frac{1}{m+1}\binom{2m}{m}$

5.3 More Theorems and Lemmas

Theorem 5.3.1 *The Borel Cantelli Theorem states that if the sum of the probabilities of* E_n *is finite, then the probability that infinitely many of them occur is 0.*

Theorem 5.3.2 *Let* F *and* F_n *be cumulative distribution functions. If* F_n *converges weakly to* F*, then*

$$\int_{\mathbb{R}} g(x)dF_n(x)$$

converges as n approaches infinity to

$$\int_{\mathbb{R}} g(x)dF(x)$$

for each bounded, continuous function g*.*

Lemma 5.3.3 *Let* A_n *be an* $n \times n$ *skew-symmetric matrix whose elements above the diagonal are 1, and those below the diagonal are -1. Then, the eigenvalues of* A_n *are* $\lambda_k = icot(\pi(2k-1)/2n)$ *for* $k = 1, 2, ..., n$*. The eigenvector associate with* λ_k *is* $u_k = \frac{1}{\sqrt{n}}(1, \rho_k, ..., \rho_k^{n-1})$ *where* $\rho_k = (\lambda_k - 1)/(\lambda_k + 1)$*.*

The proof to this lemma can be found in [1] and is done by Bai and Silverstein.

Definition 5.3.4 *A stochastic series X on a probability space* $(\Omega, \mathcal{F}, P)$ *is a Martingale Difference Sequence if it satisfies the following:*

$$E|X_t| < \infty$$

$$E[X_{t+1}|\mathcal{F}_t]$$

$$= 0$$

Lemma 5.3.5 *Let* $\{X_k\}$ *be a complex martingale difference sequence with respect to the increasing* $\sigma - field\{\mathcal{F}_k\}$. *Then for* $p > 1$,

$$E|\sum X_k|^p \leq K_p E(\sum |X_k|^2)^{p/2}$$

Proof. Both $\{Re(X_k)\}$ and $\{Im(X_k)\}$ are martingale difference sequences. Therefore we have the following inequality:

$$E|\sum X_k|^p \leq C_p[E|\sum Re(X_k)|^p + E|Im(X_k)|^p]$$

$$\leq C_p[K_p E(\sum |Re(X_k)|^2)^{p/2} + K_p E(\sum |Im(X_k)|^2)^{p/2}]$$

$$\leq 2C_p K_p E(\sum |X_k|^2)^{p/2}$$

Where $C_p = 2^{p-1}$. □

Bibliography

[1] Zhidong Bai, Jack W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer Series in Statistics. Second Edition.

[2] Tiago Pereira *Lectures on Random Matrices*, Department of Mathematics, Imperial College London.

[3] Alan Edelman, N. Raj Rao, *Random matrix theory*, Cambridge University Press

[4] Yan V. Fyodorov *Introduction to the Random Matrix Theory: Gaussian Unitary Ensemble and Beyond*, Department of Mathematical Sciences, Brunel University.

[5] Fraydoun Rezakhanlou *Lectures on Random Matrices*, Department of Mathematics, UC Berkeley.

[6] Alan Edelman *The Eigenvalues of Random Matrices, Experiments with the Classical Ensembles*, Lecture, MIT.

[7] Yanqing Yin, Zhidong Bai, Jiang Hu *On the Limit of Extreme Eigenvalues of Large Dimensional Random Quaternion Matrices* , Researchgate Publication.

[8] Arup Bose, Sourav Chatterjee, Sreela Gangopadhyay *Limiting Spectral Distributions of Large Dimensional Random Matrices*, Indian Statistical Institute, Kolkata.

[9] Yi-Kai Liu *Statistical Behavior of the Eigenvalues of Random Matrices* , Princeton University.