CONVERGENCE RATES OF SPECTRAL DISTRIBUTION OF RANDOM INNER PRODUCT KERNEL MATRICES

A Dissertation
Submitted to
the Temple University Graduate Board

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

by
Nayeong Kong
May, 2018

Examining Committee Members:

Brian Rider, Advisory Chair, Department of Mathematics

Wei Shih Yang, TU Department of Mathematics

Shiferaw S. Berhanu, TU Department of Mathematics

Subhadeep Mukhopadhyay, TU Department of Statistical Science

ABSTRACT

CONVERGENCE RATES OF SPECTRAL DISTRIBUTION OF RANDOM INNER PRODUCT KERNEL MATRICES

Nayeong Kong
DOCTOR OF PHILOSOPHY

Temple University, May, 2018

Professor Brian Rider, Chair

This dissertation has two parts.

In the first part, we focus on random inner product kernel matrices. Under various assumptions, many authors have proved that the limiting empirical spectral distribution (ESD) of such matrices A converges to the Marchenko-Pastur distribution. Here, we establish the corresponding rate of convergence. The strategy is as follows. First, we show that for $z = u + iv \in \mathbb{C},\ v > 0$, the distance between the Stieltjes transform $m_A(z)$ of ESD of matrix A and Machenko-Pastur distribution $m(z)$ is of order $\mathcal{O}\left(\frac{\log n}{nv}\right)$. Next, we prove the Kolmogorov distance between ESD of matrix A and Marchenko-Pastur distribution is of order $\mathcal{O}\left(\sqrt[3]{\frac{\log n}{n}}\right)$. It is the less sharp rate for much more general class of matrices. This uses a Berry-Esseen type bound that has been employed for similar purposes for other families of random matrics.

In the second part, random geometric graphs on the unit sphere are considered. Observing that adjacency matrices of these graphs can be thought of as random inner product matrices, we are able to use an idea of Cheng-Singer to establish the limiting for the ESD of these adjacency matrices.

ACKNOWLEDGEMENTS

I am mostly grateful to my advisor Professor Brian Rider for his generous support and invaluable remarks on my work. Also, I'm indebted to him for everything he has taught me. I would like to thank the committee members for their consideration and care taken in reading. I'm very grateful to all of my professors at Temple University for what they taught me. A special thanks goes to my husband Seonguk for his support, encouragement, and understanding throughout my graduate education.

To myself,

Nayeong Kong

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Historical Review on Spectrum of Random Matrix Theory

The rapid development of modern science and technology brought statisticians up against the task of analyzing data in large dimension. In 1928, Wishart introduced random matrix theory [41] in the field of multi-variate statistics. To understand the asymptotic behavior as the number of samples increase, he needed a special matrix from samples so that statisticians focused on a random matrix defined as follows: If $X_1, X_2, \cdots, X_n$ are independent identically distributed (iid) Gaussian vectors with mean zero and covariance $\Sigma X_1 X_1^T$, then a $n \times n$ sample covariance matrix (or Wishart matrix) is defined as

$$S = \frac{1}{n} \sum_{i=1}^{n} X_i^T X_i. \tag{1.1}$$

The probability distribution of eigenvalues of S (Wishart distribution) and its density function have played a crucial role in advancing statistical analysis, information theory and many branches of physics [41].

In 1950, spectral analysis of large dimensional random matrices was developed by Wigner [36], [37], [39], and [40]. To understand the behavior of large dimensional random matrices, Wigner focused on a special type of matrix: $n \times n$ real symmetric random matrix $W_n = (w_{ij})_{1 \le i,j \le n}$ (Wigner matrix) of the form:

(1) For $1 \le i < j < \infty$, w_{ij} are independent identically distributed (i.i.d),

(2) For $1 \le i \le n$ w_{ii} are i.i.d real random variables,

(3) $\mathbb{E}[w_{12}] = 0, \mathbb{E}[w_{12}^2] = \sigma_1^2, \mathbb{E}[w_{11}] = 0, \mathbb{E}[w_{11}^2] < \sigma_2^2$.

To figure out asymptotic behavior of random matrices, we focus on a special ensemble of it: Let us consider a $n \times n$ matrix A with eigenvalues λ_i, $i = 1, 2, \cdots, n$. If all these eigenvalues are real, we can define an one-dimensional distribution function

$$F_A(x) = \frac{1}{n}\#\{i \le n : \lambda_i \le x\}, \quad \# : \text{number of i's},$$

called the empirical spectral distribution (ESD) of the matrix A.

In 1955, Wigner proved Wigner's semi-circle law: If W_n is a $n \times n$ Wigner matrix, then $\frac{1}{\sqrt{n}}\mathbb{E}[F_{W_n}(x)]$ converges to a semicircle law $F_{\mathrm{SC},\sigma^2}(x)$ whose density function is

$$\frac{d}{dx}F_{\mathrm{SC},\sigma^2}(x) = \begin{cases} \frac{1}{2\pi\sigma}\sqrt{4\sigma^2 - x^2}, & |x| \le 2\sigma, \\ 0, & \text{otherwise}, \end{cases}$$

where $F_{\mathrm{SC},\sigma^2}(x)$ is commonly referred as the limiting spectral distribution (LSD) of W_n.

The study of the spectrum of large random matrices, since Wigner's semicircle law, has been an active research area motivated by applications such as quantum physics, signal processing, numerical linear algebra, statistical inference, among others. One of important results is the Marcenko-Pastur (M.P.) law [27] for the spectrum of random matrices of the form (1.1). Marcenko-Pastur distribution with index γ is defined by

$$\rho_{MP,\gamma}(x) := 1_{\gamma>1}\delta_0(x) + \frac{1}{2\pi x\gamma}\sqrt{(b-x)(x-a)}1_{[a,b]}(x), \tag{1.2}$$

where $a := (1-\sqrt{\gamma})^2$ and $b := (1+\sqrt{\gamma})^2$. In [27], Marcenko and Pastur proved that if $n, p \to \infty$ with $n/p \to \gamma \in (0, 1]$, then the ESD $F_S(x)$ of a

Wishart matrix S defined in (1.1) weekly converged to a distribution with density function $\rho_{MP,\gamma}(x)$ (See (1.2)) almost surely.

In the recent years, the general task of pattern analysis for the machine learning is to find and study general types of relations (for example clusters, rankings, principal components, correlations, classifications) in datasets. The kernel methods are a class of algorithms in pattern analysis and there has been a significant advance in pattern analysis via the methods. To use kernel methods in a random matrix area, probabilists have developed new type of random matrices called a random kernel matrix. Especially, the matrix whose entry depends merely on the inner product is considered. It is called an inner product kernel matrix defined as follows: If $X_1, X_2, \cdots .X_n$ are iid n-dimensional Gaussian random vectors with mean 0 and variance $\frac{1}{n}$, then the matrix A is defined by

$$A = (A_{ij})_{1\leq i,j\leq n} = \begin{cases} f(X_i^T X_j, n), & i \neq j, \\ 0, & i = j, \end{cases} \tag{1.3}$$

where $f(x,n)$ is a real valued function which may depend on n and differentiable at $x = 0$. Note that it is a general type of (1.1).

Until now, there have been many endeavors to find the limiting of the ESD of random matrices of the form (1.3). It was started from El Karoui's results in [15] which investigated the spectrum of kernel random matrices by using the operator norm $||A||_2$ of a $n \times n$ random matrix A defined by

$$||A||_2 = \max_{1\leq i\leq n} \sqrt{\lambda_i(A^T A)}, \quad \lambda_i : \text{ eigenvalues of } A^T A.$$

He proved that if the function f defined in (1.3) is differentiable at 0 and three-times differentiable at 1, then

$$||A - B||_2 \to 0 \quad \text{in probability as } n \to \infty,$$

where (i,j)-th entry of matrix B is

$$B_{ij} = [f(1) - f(0) - f'(0)]I_n + f'(0)(X_i^T X_j). \tag{1.4}$$

We call the matrix B as a sample covariance matrix and there have been also many results which proved the limiting spectral distribution of the matrix. The first success was due to Marcenko and Pastur in [26] named as Marcenko -Pastur law and next there has been a significant literature rediscovering and extending this theorem, with contributions by Bai and Yin [4], Grenander and Silverstein [21], Jonsson [25] and Wachter [34].

Especially, Yin [42] proved the limiting spectral distribution of B using combinatorial techniques when all entries of B are real and iid. For the complex iid case, Silverstein [32] showed that the limiting of Stieltjes transform of ESD of B satisfied a differential equation and extended the Marcenko-Pastur law. Also, Cheng and Singer [11] proved the limiting of ESD of random inner matrix A with the following conditions: Consider the function f in (1.3) and define the normalized function of it as

$$k(x,n) = \sqrt{n} f(\frac{x}{\sqrt{n}}, n). \tag{1.5}$$

If we denote μ as the probability density function of $\sqrt{n}X_1^T X_2$ provided that $f \in L^2(d\mu)$, then we can express $k(x,n)$ as the series:

$$k(x,n) = \sum_{k\geq 0} a_{k,n} p_{k,n}(x), \tag{1.6}$$

where $p_{k,n}(x)$ is k^{th} orthogonal polynomial with respect to μ and the coefficient $a_{k,n}$ can be written as

$$a_{k,n} = \int_{\mathbb{R}} k(x,n) p_{k,n}(x) d\mu(x).$$

Now we assume that the function f satisfies the following three conditions:

(Normalization) There exists $c \in [0,\infty)$ such that

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} |a_{i,n}|^2 = c.$$

(Uniform convergence) For any $\epsilon > 0$, there exists $L = L(\epsilon) > 0$ such that the following holds for enoughly large n:

$$\sum_{i>L} |a_{i,n}|^2 \leq \epsilon.$$

(Scaling) There exists $a \in [0, \infty)$ such that

$$\lim_{n\to\infty} a_{1,n} = a.$$

Then Cheng and Singer proved the following results.

Theorem 1.1.1 (Cheng and Singer [11])**.** *Let A be a random matrix defined as* (1.3) *and assume that the function $k(x, n)$ defined by* (1.5) *satisfies the three conditions,* ***(Normalization)****,* ***(Uniform convergence)*** *and* ***(Scaling)****. Then the Stieltjes transform of ESD (Empirical Spectral Distribution) of the matrix A, denoted by $m_A(z)$, converges weakly to $m(z)$ which satisfies the following equation.*

$$-\frac{1}{m(z)} = z + a(1 - \frac{1}{1 + am(z)}) + (c - a^2)m(z). \tag{1.7}$$

From now we keep up with this historical facts and obtain a more advanced result in Theorem 4.2.4.

1.2 Description of Main Idea and Strategy

In the previous section, we recognized that many probabilists computed the limiting of ESD for the random matrix A defined in (1.3). One of the important problems in applications of asymptotic theorems of spectral analysis of large dimensional random matrices is advancing the convergent rate of the ESD. So we will improve the rate of convergence for ESD of the random matrix A defined in (1.3). To do so, we first consider the Stieltjes transform defined as follows: For a probability measure $d\mu$ on $\mathbb{R}$, its Stieltjes transform is defined by

$$m(z) = \int_{\mathbb{R}} \frac{1}{x - z} d\mu(x), \qquad \mathcal{I}(z) > 0,$$

where $\mathcal{I}(z)$ is the imaginary part of complex number z. If A is defined as (1.3) and $\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A)$ are real eigenvalues of matrix A, then Stieltjes transform of ESD of A can be written as

$$m_A(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i(A) - z} = \frac{1}{n} \operatorname{Tr}(A - zI)^{-1}, \qquad \mathcal{I}(z) > 0,$$

where $\operatorname{Tr} M$ is a trace of the matrix M.

Now the main goal is to show that for fixed $z = u + iv \in \mathbb{C}$ with $v > 0$, the expectation $\mathbb{E}m_A(z)$ converges to $m(z)$ as $n \to \infty$, where $m_A(z)$ and $m(z)$ are the Stietjes transforms of ESD for the random matrix A and Marcenko-Pastur distribution with index 1 defined on (1.2), repectively. In particular, for fixed $z = u + iv \in \mathbb{C}$ with $v > 0$, we prove the following rate of convergence:

$$|\mathbb{E}m_A(z) - m(z)| = \mathcal{O}\Big(\frac{\log n}{nv}\Big). \tag{1.8}$$

Also, we will choose the sequence $v = v_n$ later which allows to approach zero as n tends to ∞.

The main strategy to show (1.8) is as follows: We divide into two steps. Let B be a $n \times n$ random matrix which is defined as

$$B_{ij} = f(0) + f'(0)X_i^T X_j, \tag{1.9}$$

where $X_1, X_2, \cdots, X_n$ are iid standard Gaussian vectors and the function $f(x) = f(x, n)$ is the same function as (1.3).

The first step is to prove the estimate

$$\mathbb{E}\Big|m_A(z) - m_B(z)\Big| = \mathcal{O}\Big(\frac{\log n}{nv}\Big), \quad z = u + iv \in \mathbb{C},\ v > 0, \tag{1.10}$$

which will be Theorem 3.2.2. The second step is to show the estimate

$$\Big|\mathbb{E}m_B(z) - m(z)\Big| = \mathcal{O}\Big(\frac{1}{nv}\Big), \quad z = u + iv \in \mathbb{C},\ v > 0, \tag{1.11}$$

which will be Theorem 3.3.3. These two estimates prove (1.8).

For the first step, since $X_1, X_2, \cdots, X_n$ are iid standard Gaussian vectors, using Taylor series and the generating function, we obtain the estimate,

$$\Pr\left\{|X_i^T X_j| > \sqrt{\frac{4\log n}{n}}\right\} \le \frac{2}{n^2}, \tag{1.12}$$

which will play an important role in showing the estimate (1.10). Here the main point is assuming that $X_1, X_2, \cdots, X_n$ are iid standard Gaussian vectors to handle the tail part of expectation of differences of two Stieltjes transforms.

For the second step, there are many papers which prove a rate of convergence of ESD for the random matrix B defined on (1.9). However, the random matrix in the papers is much more general than the matrix B defined on (1.9). In other word, the matrix B has the same form but it is assumed that X_i is a random vector with entries $X_{ij},\ j = 1, \cdots, n$, as iid standard normal variables such that

$$\mathbb{E}[X_{ij}] = 0, \quad \mathbb{E}[X_{ij}^2] = \frac{1}{n}. \tag{1.13}$$

(This is much more general random vector than Gaussian random vector). In [2], Bai and Miao assumed the following: Let $M = (x_{ij})$ be a real symmetric matrix with

$$\begin{aligned}
&(1)\ \mathbb{E}[x_{ij}] = 0, \mathbb{E}[x_{ij}^2] = 1, \quad 1 \leq i \leq p, 1 \leq j \leq n,\\
&(2)\ \sup_{i,j} \mathbb{E}[|x_{ij}^8|] < \infty,\\
&(3)\ \sum_{i,j} \mathbb{E}[x_{ij}^8] I_{|x_{ij}| > \epsilon\sqrt{n}} = \mathcal{O}(n^2),\ \text{for any } \epsilon > 0,
\end{aligned}$$

and $F_M(x)$ be the ESD of M. They proved that the rates of convergence to $\rho_{MP,\gamma}(x)$ defined on (1.2) were $\mathcal{O}(n^{-1/4})$ and $\mathcal{O}(n^{-5/48})$ according to $\gamma \neq 1$ or $\gamma = 1$. These rates were also established for the convergence in probability of $F_B(x)$ to $\rho_{MP,1}(x)$. In 2010, Bai [6] developed the rate of convergence of the ESD of the random matrix B to $\rho_{MP,\gamma}(x)$: If a Hermitian matrix $M = (x_{ij}),\ 1 \leq i, j \leq n$ satisfies the following conditions:

$$\begin{aligned}
&(1)\ \mathbb{E}[x_{ij}] = 0, \mathbb{E}[x_{ij}^2] = 1, \quad 1 \leq i < j \leq n,\\
&(2)\ \mathbb{E}[|x_{ii}^2|] = \sigma^2, 1 \leq i \leq n,\\
&(3)\ \sup_{i,j} \mathbb{E}[x_{ij}^8]], \mathbb{E}[x_{ii}^3] \leq C < \infty,
\end{aligned}$$

with $\sigma^2 \leq C$ for some constant $C > 0$, then the ESD of M converges to $\rho_{MP,1}(x)$ with a rate $\mathcal{O}(n^{-1/2})$. This rate worked for the convergence in probability of $F_B(x)$ to $\rho_{MP,1}(x)$. Moreover, Gotze and Tikhomirov [18] proved that the ESD of Wishart matrix $W = \frac{1}{n}X^TX$ where $X = (w_{ij})$ with

$$(1)\ \mathbb{E}[w_{ij}] = 0, \mathbb{E}[w_{ij}^2] = 1, \quad 1 \le i, j \le n,$$
$$(2)\ \Pr\{|w_{ij}| > t\} \le \frac{1}{M}\exp\{-t^M\},$$

converged to $\rho_{MP,1}(x)$ with a rate $\mathcal{O}\Big(\frac{(\log n)^{4+\frac{4}{M}}}{n}\Big)$.

In this dissertation, we investigate the convergence rate of $m_B(z)$ to $m(z)$ for any $z = u + iv \in \mathbb{C}$ with $v > 0$ assuming that X_i is a random vector with entries $X_{ij},\ j = 1, \cdots, n$, as iid standard normal variables with (1.13). The main idea is to use the diagonal entries of resolvent for random matrix B_n which comes from O'Rourke's paper [29]: Let us consider the resolvent of B_n

$$R_n(z) = (B_n - zI)^{-1}.$$

Then the diagonal entries of $R_n(z)$ can be expressed by

$$\Big(R_n(z)\Big)_{kk} = \frac{1}{f'(0)\sum_{s=1}^{n} x_{ks}^2 - z - B_k^T R_n^k(z) B_k}, \quad k = 1, \cdots, n, \tag{1.14}$$

where B_k is the k-th row of B_n except k-th entry. The diagonal entries (1.14) lead to the rate (See Theorem 3.3.3):

$$\Bigg|\mathbb{E}m_B(z) - \frac{1}{-z - zf'(0)\mathbb{E}m_B(z)}\Bigg| = \mathcal{O}\Big(\frac{1}{nv}\Big). \tag{1.15}$$

Using the uniqueness and formula of Stieltjes transform for Marchenko-Pastur distribution, we get

$$|\mathbb{E}m_B(z) - m(z)| = \mathcal{O}\Big(\frac{1}{nv}\Big).$$

By using above two steps, we are able to obtain the following theorem and our desired rate of convergence.

Theorem 1.2.1. *Let A_n be $n \times n$ random matrices defined as* (1.3) *and $m_{A_n}(z)$ be the Stieltjes transform of the ESD $F_{A_n}(x)$ of A_n. Then for any $z = u + iv,\ v > 0$,*

$$\Bigg|\mathbb{E}m_{A_n}(z) - m(z)\Bigg| = \mathcal{O}\Big(\frac{\log n}{nv}\Big),$$

where $m(z)$ is the Stieltjes transform of Marchenko-Pastur distribution defined on (1.2) *with* $\gamma = 1$.

Next, we compute a rate of convergence of expected ESD of A to the distribution

$$G(x) = \int_{-\infty}^{x} \rho_{MP,\gamma}(t)dt, \tag{1.16}$$

using the Bai's idea in [1] and Theorem 1.2.1. Indeed, it was proved in [1] that the bound for Kolmogorov distance,

$$||F - G|| = \sup_x |F(x) - G(x)|,$$

was obtained by a Berry-Esseen type inequality whose right side has the integral term having a finite interval and the integrand as the absolute difference between their Stieltjes transforms. Then we have the following main result.

Theorem 1.2.2 (Main Result). *If A is a random inner product kernel matrix defined in* (1.3) *and $F_A(x)$ is the ESD of A then*

$$||\mathbb{E}F_A(x) - G(x)|| = \sup_x |\mathbb{E}F_A(x) - G(x)| = \mathcal{O}\Big(\sqrt[3]{\frac{\log n}{n}}\Big),$$

where

$$G(x) = \int_{-\infty}^{x} \rho_{MP,1}(x)dx,$$

and $\rho_{MP,1}(x)$ defined in (1.16).

1.3 Beyond the Gaussian

For convenience, we assumed that all vectors $X_1, X_2, \cdots, X_n$ are iid Gaussian with mean 0 and variance $\frac{1}{n}$. In Chapter 3, we prove that if $X_1, X_2, \cdots, X_n$ are Gaussian then we obtain the right tail part of $X_i^T X_j$,

$$\Pr\left\{|X_i^T X_j| > \sqrt{\frac{4\log n}{n}}\right\} \leq \frac{2}{n}. \tag{1.17}$$

Using (1.17), we have improved a rate of convergence of $m_{A_n}(z)$ to $m_{B_n}(z)$ in the Theorem 3.2.2. However, it is known that the condition that the vectors are Gaussian is too strong and used only to obtain the upper bound (1.17). Thus, assuming that $X_1, X_2, \cdots, X_n$ are iid random vectors with mean 0 and variance $\frac{1}{n}$ and have the upper bound (1.17) with the condition,

$$\sup_{i,j} \mathbb{E}[x_{i,j}^4] = \mathcal{O}\Big(\frac{1}{n^2}\Big),$$

where $x_{i,j}$ is a j-th entry of the vector X_i, we are able to extend the main results in Theorem 3.2.2 for more general random inner product matrix A_n defined in (1.3) with above random vectors.

1.4 Random Graph on Unit Ball

The theory of random graphs was founded in the late 1950s by Erdös and Rényi [16]. The work of Watts and Strogatz [35] and Barabási and Albert [7] at the end of the last century initiated new interest in this field. The subject is at the intersection between graph theory and probability theory.

Based on simulation results, there are interesting differences between the spectrum of the adjacency matrices of the different classes of graph. It is proposed to use the spectrum of the adjacency matrix as a tool to indicate which model may be most appropriate for a particular real-world graph such as transportation, the Internet, social networks and neural networks [17].

In the second part of this thesis, we investigate the spectral properties of the adjacency matrix of a random geometric graph on the unit ball. The random geometric graph $G(n, r_n)$ defined as follows: Consider uniformly distributed n vertices $X_1, X_2, \cdots, X_n$ on the unit sphere,

$$S^{n-1} = \{X = (x_1, x_2, \cdots, x_n) : {x_1}^2 + {x_2}^2 + \cdots + {x_n}^2 = 1\} \subset \mathbb{R}^n.$$

We define that two points are connected if their Euclidean distance is less than the radius r_n. Recently, this class of graphs has attracted attention from a model for wireless networks [17].

Now we focus on studying the ESD of the adjacency matrix of a random geometric graph which has n-uniformly distributed vertex on the unit $(n-1)$ sphere in $\mathbb{R}^n$. Then the adjacency matrix is $A = (A_{ij})_{1\leq i,j\leq n}$ defined as

$$\begin{aligned} A_{ij} &= \begin{cases} 1 & \text{if } ||X_i - X_j|| \leq r_n, \\ 0 & \text{if } ||X_i - X_j|| \geq r_n, \end{cases} \\ &= f(||X_i - X_j||, n), \end{aligned} \tag{1.18}$$

where

$$f(x,n) = \chi_{[0,r_n)}(x,n). \tag{1.19}$$

We call this type of matrix random as Euclidean matrix. Probabilists recognize that it is not easy to investigate this random euclidean matrix and so there have had little results computing the limiting of ESD of the matrix. However, since $X_1, X_2, \cdots, X_n$ are on the unit sphere for this matrix, then we think of the Euclidean norm $||X_i - X_j||$ as

$$||X_i - X_j|| = \sqrt{2 - 2{X_i}^T X_j}. \tag{1.20}$$

Using (1.20), the random matrix A in (1.18) can be changed by

$$\begin{aligned} A = (A_{ij})_{1\leq i,j\leq n} &= \begin{cases} 1 & \text{if } {X_i}^T X_j \geq 1 - \frac{{r_n}^2}{2}, \\ 0 & \text{if } {X_i}^T X_j < 1 - \frac{{r_n}^2}{2} \end{cases}, \\ &= \chi_{[1-\frac{{r_n}^2}{2},\infty)}({X_i}^T X_j, n). \end{aligned} \tag{1.21}$$

Thus, instead of the Euclidean matrix, we think of the adjacency matrix A as a kind of random inner product matrix with the function f as

$$f(x,n) = \chi_{[1-\frac{{r_n}^2}{2},\infty)}(x,n). \tag{1.22}$$

There have been many results which compute the limiting of ESD of a random inner product matrix defined by (1.21) and (1.22) assuming that $X_1, X_2, \cdots, X_n$ are i.i.d. random vectors uniformly distributed on S^{n-1}. In [23], Jiang proved that ESD of A converged to $\frac{1}{2f'(2)}\rho_{MP,1}\left(\frac{x-f(0)+f(2)-2f'(2)}{2f'(2)}\right)$ with probability 1. In the paper, Jiang assumed that $f(x)$ was defined on $[0,4]$, $f''(2)$ was existed and $f'(2) \neq 0$. Whereas, the function f defined as (1.22) does not satisfy this condition so that we need to use more general result proved by Cheng and Singer [11]. In this thesis, we show that the function f defined (1.22) satisfies conditions in [11], and so the ESD for the adjacency matrix converges to the semi-circle distribution with scaling window a constant $c > 0$.

1.5 Outline of the Thesis

The main tools proving the theorems and some basic consequences are introduced in Section 2. In Section 3, we compute the rate of convergence of differences of two stieltjes transforms. To do so, we divided into two steps and show each steps using generating function and resolvent of the matrix. Based on the result in Section 3, we compute the difference between the probability functions for ESD of the matrix A and $\rho_{MP,\gamma}(x)$ by using Berry-Essen type inequality in Section 4. Finally, in Section 5, we introduce the random geometric graph $G(n, r_n)$ on the unit sphere and show the limiting of ESD of the adjacency matrix of $G(n, r_n)$.

CHAPTER 2

PRELIMINARIES

Based on the definition of the Stieltjes transform of ESD of a random matrix A, we review basic facts for the trace of an inverse matrix which are essential for us to prove our results. Most facts in this chapter came from the classical linear algebra theory, especially the matrix computation.

2.1 Random Inner-Product Kernel Matrix

Let $X_1, X_2, \ldots, X_n$ be iid standard Gaussian vectors in $\mathbb{R}^n$ i.e., for each $1 \le i \le n$,

$$X_i = (X_{i1}, X_{i2}, \ldots, X_{in})^T, \quad X_{ij} \sim \mathcal{N}(0, \frac{1}{n}).$$

Since for any $1 \le i, j \le n$, X_{ij} is a Gaussian random variable with mean 0 and variance $\frac{1}{n}$, one can know that

$$\Pr\{|X_{ij}| > t\} = 2\sqrt{\frac{n}{2\pi}} \int_t^\infty \exp\Big\{ -\frac{nx^2}{2} \Big\} dx \le \exp\Big\{ -\frac{n}{2} t^2 \Big\}, \quad t \ge 0. \quad (2.1)$$

We consider a special random kernel matrix defined as follows:

Definition 2.1.1. A_n is called a $n \times n$ symmetric random kernel matrix if its entries are

$$(A_n)_{ij} = \begin{cases} f(X_i^T X_j, n), & i \neq j, \\ 0, & i = j, \end{cases} \quad (2.2)$$

where $f(x,n)$ is a real valued function which depends on n.

Definition 2.1.2. Suppose that the eigenvalues of the matrix A_n are λ_1, λ_2, $\cdots$, λ_n, then Empirical Spectral Distribution (ESD) of A_n, $F_{A_n}(x)$ is defined by

$$F_{A_n}(x) = \frac{1}{n}\#\{j \leq n : \lambda_j \leq x\}, \quad x \in \mathbb{R} \quad \# : \text{number of i's.} \tag{2.3}$$

There are a few methods to compute the limiting of ESD of random matrices and one of them is to use the Stieltjes transform of ESD.

2.2 Stieltjes Transform

Next, we define the Stieltjes Transform, which is an useful tool to compute the limiting of the ESD of random matrices.

Definition 2.2.1. For a probability measure $d\mu$ on $\mathbb{R}$, its Stieltjes Transform $m_\mu(z)$ is defined by

$$m_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\mu(x), \quad z \in \mathbb{C},$$

where the imaginary part of z is positive, i.e., $\mathcal{I}(z) > 0$.

Now, we have basic remarks about the Stieltjes Transform.

Lemma 2.2.2. *Suppose $\mathcal{I}(z) > 0$, then we have $\mathcal{I}(m_\mu(z)) > 0$.*

Proof. It is clear that for $z = a + bi,\ b > 0$,

$$\begin{aligned}\mathcal{I}(m_\mu(z)) =& \mathcal{I}\Big(\int_{\mathbb{R}} \frac{1}{x-z} d\mu(x)\Big) = \mathcal{I}\Big(\int_{\mathbb{R}} \frac{(x-a)+bi}{(x-a)^2+b^2} d\mu(x)\Big) \\ =& b\int_{\mathbb{R}} \frac{1}{(x-a)^2+b^2} d\mu(x) > 0.\end{aligned}$$

□

Lemma 2.2.3. *For a random matrix A_n with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, the Stieltjes Transform of the ESD of A_n, say, $m_{A_n}(z)$ is*

$$m_{A_n}(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i - z} = \frac{1}{n}\operatorname{Tr}(A_n - zI_n)^{-1}, \qquad \mathcal{I}(z) > 0. \tag{2.4}$$

Corollary 2.2.4. *The Stieltjes Transform* (2.4) *is uniformly bounded for fixed* $z \in \mathbb{C}$. *Specially, if* $z = u + vi$ *is fixed with* $v > 0$, *then*

$$|m_{A_n}(z)| \leq \frac{1}{v}. \tag{2.5}$$

Proof. By the definition of the Stieltjes transform, it is clear that

$$|m_{A_n}(z)| \leq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|\lambda_i - z|} \leq \frac{1}{n}\sum_{i=1}^{n}\frac{1}{v} = \frac{1}{v}. \tag{2.6}$$

□

It is well-known that the ESD of a random inner product kernel matrix defined on (2.3) converges weakly to a Marcenko-Pastur distribution, (See [11]). Here, we denote a Marcenko-Pastur distribution with index $\gamma = 1$ by

$$\rho_{MP,1}(x) := \frac{1}{2\pi x}\sqrt{x(2-x)}1_{[0,2]}(x). \tag{2.7}$$

Theorem 2.2.5. *Assume that* $X_1, X_2, \ldots, X_n$ *are* n*-independent identically distributed vectors in* $\mathbb{R}^n$ *with* $\mathbb{E}[X_i] = 0$ *and* $\mathbb{E}[X_i^2] = 1$ *for each* $1 \leq i \leq n$. *Also, suppose that* f *is a real valued function and differentiable at* 0. *If* A_n *is a random matrix defined as Definition 2.1, then the ESD of* A_n *weakly converges to*

$$\frac{1}{|f'(0)|}\rho_{MP,1}\Big(\frac{x-\alpha}{f'(0)}\Big), \qquad \alpha = -f(0) - f'(0),\ f'(0) \neq 0, \tag{2.8}$$

almost surely. If $f'(0) = 0$, *then the ESD of* A *weakly converges to the Dirac measure* $\delta_\alpha(x)$ *at* α.

2.3 Inverse Matrix and Resolvent

In this section, based on basic linear algebra, we introduce some useful results below.

Let $A = (a_{ij}), 1 \leq i, j \leq n$ be an $n \times n$ matrix. After deleting the k-th row and column from A, we have a $(n-1) \times (n-1)$ submatrix A_k of A. Then we obtain the following useful theorem.

Theorem 2.3.1. *[See [22]] Let A be nonsingular and U, B and V may be rectangular. Then*

$$(A + UBV)^{-1} = A^{-1} - A^{-1}U(I + BVA^{-1}U)^{-1}BVA^{-1}. \tag{2.9}$$

Theorem 2.3.2. *[6] If both A and A_k, $k = 1, 2, \cdots, n$ are Hermitian nonsingular, and if we write $A^{-1} = (A^{ij}), 1 \le i, j \le n$, then*

$$a^{kk} = \frac{1}{a_{kk} - a_k A_k^{-1} a_k^T}, \tag{2.10}$$

and hence

$$\mathrm{Tr}(A^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - a_k A_k^{-1} a_k^T}, \tag{2.11}$$

where a_{kk} is the k-th diagonal entry of A, A_k is a $(n-1) \times (n-1)$ submatrix of A defined as above, a_k is the vector obtained from k-th row of A by deleting k-th entry.

In this thesis, we frequently consider the resolvent of a Hermitian matrix A, i.e, $(A - zI_n)^{-1}$, where z is a complex number with positive imaginary part. In this case, by (2.11), we have

$$\mathrm{Tr}(A - zI_n)^{-1} = \sum_{k=1}^{n} \frac{1}{a_{kk} - z - a_k(A_k - zI_{n-1})^{-1}a_k^T}. \tag{2.12}$$

Now, we need the inverse of partitioned matrix formula to compute the difference of traces of a matrix A and its submatrix A_k for each $k = 1, 2, \cdots, n$. Suppose that the matrix P is nonsingular positive definite and has the partition as given by $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$. Then the inverse of P has the form

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}. \tag{2.13}$$

We use the indentity (2.13) to prove the following theorem.

Theorem 2.3.3. *If the matrix A and A_k defined as above are both nonsingular and Hermitian then*

$$\operatorname{Tr}(A^{-1}) - \operatorname{Tr}(A_k^{-1}) = \frac{1 + a_k A_k^{-2} a_k^T}{a_{kk} - a_k A_k^{-1} a_k^T}, \tag{2.14}$$

where a_k is obtained from k-th row of A by deleting k-th entry.

Proof. Note that by (2.13),

$$\begin{aligned}
\operatorname{Tr} A^{-1} &= \operatorname{Tr} \begin{pmatrix} A_k & a_k^T \\ a_k & a_{kk} \end{pmatrix}^{-1} \\
&= \operatorname{Tr} \begin{pmatrix} (A_k - a_k^T a_{kk}^{-1} a_k)^{-1} & -A_k^{-1} a_k^T (a_{kk} - a_k A_k^{-1} a_k^T)^{-1} \\ -a_{kk}^{-1} a_k (A_k - a_k^T a_{kk}^{-1} a_k)^{-1} & (a_{kk} - a_k A_k^{-1} a_k^T)^{-1} \end{pmatrix} \\
&= \operatorname{Tr}(A_k - a_k^T a_{kk}^{-1} a_k)^{-1} + \operatorname{Tr}(a_{kk} - a_k A_k^{-1} a_k^T)^{-1}
\end{aligned}$$

By Theorem 2.3.1, we have

$$\begin{aligned}
\operatorname{Tr}(A_k - a_k^T a_{kk}^{-1} a_k)^{-1} &= \operatorname{Tr} A_k^{-1} - \operatorname{Tr} A_k^{-1} a_k^T (I - a_{kk}^{-1} a_k A_k^{-1} a_k^T)^{-1} (-a_{kk}^{-1}) a_k A_k^T \\
&= \operatorname{Tr} A_k^{-1} + (a_{kk} - a_k A_k^{-1} a_k^T)^{-1} \operatorname{Tr}(a_k A_k^{-2} a_k^T).
\end{aligned}$$

This completes the proof. □

From Theorem 2.3.3, we obtain the upper bound of the difference between traces of resolvents of a matrix A and its submatrix A_k for each $k = 1, 2, \cdots, n$.

Theorem 2.3.4. *Let $z = u + iv, v > 0$ and let A be a $n \times n$ Hermitian matrix. Then*

$$|\operatorname{Tr}(A - zI_n)^{-1} - \operatorname{Tr}(A_k - zI_{n-1})^{-1}| \le \frac{1}{v}. \tag{2.15}$$

Proof. From (2.14), we have

$$\begin{aligned}
&|\operatorname{Tr}(A - zI_n)^{-1} - \operatorname{Tr}(A_k - zI_{n-1})^{-1}| = \\
&\qquad \left| \frac{1 + (a_k - ze_k)(A_k - zI_{n-1})^{-2}(a_k - ze_k)^T}{(a_{kk} - z) - (a_k - ze_k)(A_k - zI_{n-1})^{-1}(a_k - ze_k)^T} \right|,
\end{aligned}$$

where $e_k = (0, \cdots, \underbrace{1}_{k-th}, \cdots, 0)^T$ is the elementary vector. Put $\lambda_1, \lambda_2, \cdots, \lambda_{n-1}$ as the eigenvalues of matrix A_k. Then we can denote

$$A_k = B\,\mathrm{diag}[\lambda_1, \lambda_2, \cdots, \lambda_{n-1}]B^T,$$

where B is an $(n-1) \times (n-1)$ unitary matrix and $\mathrm{diag}[\lambda_1, \lambda_2, \cdots, \lambda_{n-1}]$ is a diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \cdots, \lambda_{n-1}$. Since $(a_k - ze_k)B = (y_1, y_2, \cdots, y_{n-1})$ and $z = u + iv$, we have

$$\begin{aligned}
&|1{+}(a_k - ze_k)(A_k - zI_{n-1})^{-2}(a_k - ze_k)^T|\\
&\quad= \left|1 + \sum_{i=1}^{n-1} |y_i|^2(\lambda_i - z)^{-2}\right|\\
&\quad\le 1 + \sum_{i=1}^{n-1} |y_i|^2((\lambda_i - u)^2 + v^2)^{-1}\\
&\quad= 1 + (a_k - ze_k)((A_k - uI_{n-1})^2 + v^2 I_{n-1})^{-1}(a_k - ze_k)^T.
\end{aligned}$$

Note that if $\mathbf{C}^2 + \mathbf{D}^2$ is nonsingular for any two Hermitian matrices $\mathbf{C}$ and $\mathbf{D}$, then

$$(\mathbf{C} + i\mathbf{D})^{-1} = (\mathbf{C} - i\mathbf{D})(\mathbf{C}^2 + \mathbf{D}^2)^{-1}. \tag{2.16}$$

It follows from (2.16) that

$$\begin{aligned}
&\mathcal{I}(a_{kk} - z - (a_k - ze_k)(A_k - zI_{n-1})^{-1}(a_k - ze_k)^T)\\
=&v(1 + (a_k - ze_k)((A_k - uI_{n-1})^2 + v^2 I_{n-1})^{-1}(a_k - ze_k)^T),
\end{aligned}$$

which proves (2.15). □

CHAPTER 3

RATE OF LIMITING TO STIELTJES TRANSFORM OF MP DISTRIBUTION

3.1 Description of Main Result and Ideas

Let A_n be a $n \times n$ random inner product kernel matrix defined by

$$(A_n)_{ij} = \begin{cases} f(X_i^T X_j, n), & i \neq j, \\ 0, & i = j, \end{cases} \tag{3.1}$$

where $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a twice differentiable function at $x = 0$, i.e., $f''(0)$ exists.

In this Chapter, the main purpose is to show that for fixed $z = u + iv \in \mathbb{C}$ with $v > 0$, $\mathbb{E}m_{A_n}(z)$ converges to $m(z)$ as $n \to \infty$, where $m_{A_n}(z)$ and $m(z)$ are the Stieltjes transforms of ESD, $F_{A_n}(x)$ of A_n and Marcenko-Pastur distribution defined on (2.3) and (2.7), respectively. In particular, we prove the following rate of convergence: For fixed $z = u + vi \in \mathbb{C},\ v > 0$,

$$|\mathbb{E}m_{A_n}(z) - m(z)| = \mathcal{O}\Big(\frac{\log n}{nv}\Big). \tag{3.2}$$

Also, if $z \in \{u+iv \in \mathbb{C} : |u| \leq \alpha,\ 0 < v_n < v\}$, then we obtain

$$|\mathbb{E}m_{A_n}(z) - m(z)| = \mathcal{O}\Big(\frac{\log n}{nv_n}\Big).$$

This means that the rate of convergence of $\mathbb{E}m_{A_n}(z)$ to $m(z)$ depends on the rate of convergence of $v = v_n$ to 0 as $n \to \infty$. We will choose the sequence v_n later which allows to approach zero.

The outline to show the estimate (3.2) is the following. In the first step, we define a matrix B_n by

$$(B_n)_{ij} = f(0) + f'(0)X_i^T X_j, \tag{3.3}$$

where $X_1, X_2, \ldots, X_n$ are iid standard Gaussian vectors and the function f is the same function as the definition of A_n. In Theorem 3.2.2, we prove the estimate: For all $z = u + iv,\ v > 0$,

$$\mathbb{E}\Big|m_{A_n}(z) - m_{B_n}(z)\Big| = \mathcal{O}\Big(\frac{\log n}{nv}\Big), \quad n \to \infty. \tag{3.4}$$

Second step is to show that $m_{B_n}(z)$ converges to the Stietljes transform $m(z)$ of Marcenko-Pastur distribution in (2.7) as $n \to \infty$, i.e., for any $z = u+iv,\ v > 0$,

$$\Big|\mathbb{E}m_{B_n}(z) - m(z)\Big| = \mathcal{O}\Big(\frac{1}{nv}\Big). \tag{3.5}$$

This is Corollary 3.3.4. Below, together we have the following theorem.

Theorem 3.1.1. *Let A_n be $n \times n$ random matrices defined as* (3.1) *and $m_{A_n}(z)$ be the Stieltjes transform of the function $F_{A_n}(x)$ defined on* (2.3)*. Then we have for any $z = u + iv,\ v > 0$,*

$$\Big|\mathbb{E}m_{A_n}(z) - m(z)\Big| = \mathcal{O}\Big(\frac{\log n}{nv}\Big),$$

where $m(z)$ is the Stieltjes transform of Marchenko-Pastur distribution defined on (2.7)*.*

3.2 Rate of Convergence of $m_{A_n}(z)$ to $m_{B_n}(z)$

We start with the following Lemma.

Lemma 3.2.1. *Let $X_1, X_2, \dots, X_n$ be iid Gaussian vectors in $\mathbb{R}^n$ with $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1^2] = 1$. Then we have*

$$\Pr\left\{|X_i^T X_j| > \sqrt{\frac{4\log n}{n}}\right\} \leq \frac{2}{n^2}. \tag{3.6}$$

Proof. Note that for any $\epsilon > 0$,

$$\Pr\left\{|X_i^T X_j| > \epsilon\right\} = \Pr\left\{|nX_i^T X_j| > n\epsilon\right\} = \Pr\left\{\Big|\sum_{k=1}^{n} Z_k\Big| > n\epsilon\right\}, \tag{3.7}$$

where Z_k is the product of two independent Gaussian variables with mean 0 and variance $\frac{1}{n}$. By a simple expectation, we have that

$$M_{Z_k}(t) = \mathbb{E}[e^{tZ_k}] = \frac{1}{\sqrt{1-t^2}}, \quad t \in (-1, 1). \tag{3.8}$$

Hence, by the independence of Z_k's, we obtain

$$M_{\sum_{k=1}^{n} Z_k}(t) = \Big(\frac{1}{\sqrt{1-t^2}}\Big)^n. \tag{3.9}$$

Optimizing over t and using the assumption that Z_k are independent, it follows from (3.7), (3.9) and Generic Chernoff Bound that

$$\begin{aligned}\Pr\left\{\sum_{k=1}^{n} Z_k > n\epsilon\right\} &= \Pr\left\{e^{t\sum_{k=1}^{n} Z_k} > e^{tn\epsilon}\right\} \\ &\leq \min_{0<t<1} \frac{1}{e^{tn\epsilon}} \mathbb{E}[e^{t\sum_{k=1}^{n} Z_k}] \\ &= \min_{0<t<1} \frac{1}{e^{tn\epsilon}} \Big(\frac{1}{\sqrt{1-t^2}}\Big)^n.\end{aligned} \tag{3.10}$$

Similarly,

$$\begin{aligned}\Pr\left\{\sum_{k=1}^{n} Z_k < -n\epsilon\right\} &= \Pr\left\{e^{-tn\epsilon\sum_{k=1}^{n} Z_k} > e^{tn\epsilon}\right\} \\ &\leq \min_{0<t<1} \frac{1}{e^{tn\epsilon}} \mathbb{E}[e^{-t\sum_{k=1}^{n} Z_k}] \\ &= \min_{0<t<1} \frac{1}{e^{tn\epsilon}} \Big(\frac{1}{\sqrt{1-(-t)^2}}\Big)^n.\end{aligned} \tag{3.11}$$

Since for $0 < t < 1$,

$$\log(1-t) \le -t,$$

then for large n in $\mathbb{N}$,

$$\begin{aligned}\frac{1}{e^{tn\epsilon}}\Big(\frac{1}{\sqrt{1-t^2}}\Big)^n =& e^{-tn\epsilon}(1-t^2)^{-\frac{n}{2}} \\ =& e^{-tn\epsilon}e^{\log(1-t^2)^{-\frac{n}{2}}} \\ =& e^{-nt\epsilon-\frac{n}{2}\log(1-t^2)} \\ \le& e^{-tn\epsilon+\frac{n}{2}t^2} \\ =& e^{\frac{n}{2}\Big((t-\epsilon)^2-\epsilon^2\Big)}.\end{aligned}$$

Choosing $t = \epsilon$, we get

$$\min_{0<t<1}\frac{1}{e^{tn\epsilon}}\Big(\frac{1}{\sqrt{1-t^2}}\Big)^n \le \min_{0<t<1} e^{\frac{n}{2}\Big((t-\epsilon)^2-\epsilon^2\Big)} \le e^{-\frac{n}{2}\epsilon^2}. \tag{3.12}$$

Thus (3.10), (3.11) and (3.12) give us that

$$\Pr\Big\{|X_i^T X_j| > \epsilon\Big\} = 2e^{-\frac{n}{2}\epsilon^2}. \tag{3.13}$$

Choosing

$$\epsilon = \sqrt{\frac{4\log n}{n}},$$

we have by (3.13)

$$\Pr\left\{|X_i^T X_j| > \sqrt{\frac{4\log n}{n}}\right\} \le \frac{2}{n^2}.$$

Thus, we get our desired estimate (3.6). □

For convenience, we denote A_n as A and B_n as B. Using Lemma 3.2.1, we prove the next Theorem which completes the first step.

Theorem 3.2.2. *Let A and B be $n \times n$ random matrices defined as* (3.1) *and* (3.3). *Then we have for any $z = u + iv \in \mathbb{C}$, $v > 0$,*

$$\mathbb{E}\Big|m_A(z) - m_B(z)\Big| = \mathcal{O}\Big(\frac{\log n}{nv}\Big). \tag{3.14}$$

Proof. Let $\{\lambda_i(A)\}_{i=1}^n$ be eigenvalues of matrix A and $\{\lambda_i(B)\}_{i=1}^n$ be eigenvalues of B. Using Cauchy-Schwarz Inequality, we have for $z = u+iv \in \mathbb{C},\ v > 0$,

$$\begin{aligned}
|m_A(z) - m_B(z)|^2 &\le \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{(\lambda_i(A) - z)} - \frac{1}{(\lambda_i(B) - z)} \right|^2 \\
&\le \left| \frac{1}{n} \sum_{i=1}^n \frac{(\lambda_i(A) - \lambda_i(B))^{\frac{1}{2}}}{(\lambda_i(A) - z)} \frac{(\lambda_i(A) - \lambda_i(B))^{\frac{1}{2}}}{(\lambda_i(B) - z)} \right|^2 \\
&\le \frac{1}{n^2} \sum_{i=1}^n \left| \frac{\lambda_i(A) - \lambda_i(B)}{(\lambda_i(A) - z)^2} \right| \sum_{i=1}^n \left| \frac{\lambda_i(A) - \lambda_i(B)}{(\lambda_i(B) - z)^2} \right| \\
&\le \frac{1}{n^2 v^4} \left(\sum_{i=1}^n |\lambda_i(A) - \lambda_i(B)| \right)^2. \qquad (3.15)
\end{aligned}$$

Using Hoffman-Wielandt inequality,

$$\sum_{i=1}^n |\lambda_i(A) - \lambda_i(B)|^2 \le \mathrm{Tr}(A - B)^2,$$

it follows from (3.15) that

$$|m_A(z) - m_B(z)|^2 \le \frac{1}{n^2 v^2} \left(\sum_{i=1}^n \sum_{j=1}^n |A_{ij} - B_{ij}|^2 \right)^2. \qquad (3.16)$$

Next, let $\mathcal{F}$ be an event that there exist $i \neq j$ such that

$$|X_i^T X_j| > \sqrt{\frac{4 \log n}{n}}.$$

It is clear from Lemma 3.2.1 that

$$\Pr(\mathcal{F}) \le \mathcal{O}\Big(\frac{1}{n^2}\Big). \qquad (3.17)$$

Notice that

$$\begin{aligned}
\mathbb{E}\Big| m_A(z) - m_B(z) \Big|^2 \le& \mathbb{E}\Big[1_{\mathcal{F}} |m_A(z) - m_B(z)|^2 \Big] + \mathbb{E}\Big[1_{\mathcal{F}^c} |m_A(z) - m_B(z)|^2 \Big] \\
\le& M_1 + M_2, \qquad (3.18)
\end{aligned}$$

where

$$M_1 = \mathbb{E}\Big[1_{\mathcal{F}} |m_A(z) - m_B(z)|^2 \Big], \quad M_2 = \mathbb{E}\Big[1_{\mathcal{F}^c} |m_A(z) - m_B(z)|^2 \Big].$$

Since, by Remark 2.2.4, we have

$$|m_A(z)| \le \frac{1}{v}, \quad |m_B(z)| \le \frac{1}{v},$$

then we obtain

$$\begin{aligned} M_1 &\le \mathbb{E}\Big[1_{\mathcal{F}}\big(|m_A(z)|^2 + 2|m_A(z)||m_B(z)| + |m_B(z)|^2\big)\Big] \\ &\le \mathcal{O}\left(\Pr(\mathcal{F})\frac{1}{v^2}\right) = \mathcal{O}\left(\frac{(\log n)^2}{n^2v^2}\right). \end{aligned} \tag{3.19}$$

Also, since f is twice differentiable at $x = 0$, then it comes from the Taylor polynomial that there exists a constant $C = C(|x|) > 0$ such that

$$|f(x) - f(0) - f'(0)x| \le C|x|^2. \tag{3.20}$$

This implies that

$$\begin{aligned} M_2 &\le \mathbb{E}\left[1_{\mathcal{F}^c}\frac{1}{n^2v^2}\sum_{i\neq j}|A_{ij} - B_{ij}|^2\right] \\ &= \mathbb{E}\left[1_{\mathcal{F}^c}\frac{1}{n^2v^2}\sum_{i\neq j}|f(X_i^TX_j) - f(0) - f'(0)X_i^TX_j|^2\right] \\ &\le \mathbb{E}\left[1_{\mathcal{F}^c}\frac{C}{n^2}\sum_{i\neq j}|X_i^TX_j|^4\right] \\ &\le \mathbb{E}\left[\frac{C}{n^2v^2}\sum_{i\neq j}\frac{(\log n)^2}{n^2}\right] \\ &= \mathcal{O}\left(\frac{1}{n^2v^2}\frac{(\log n)^2}{n^2}n(n-1)\right) \\ &= \mathcal{O}\left(\frac{(\log n)^2}{n^2v^2}\right). \end{aligned} \tag{3.21}$$

Therefore, using (3.19), (3.21) and Cauchy-Schwarz Inequality, we obtain

$$\begin{aligned} \mathbb{E}\Big|m_A(z) - m_B(z)\Big| &\le \Big(\mathbb{E}\Big|m_A(z) - m_B(z)\Big|^2\Big)^{\frac{1}{2}} \\ &\le \Big(M_1 + M_2\Big)^{\frac{1}{2}} \\ &= \mathcal{O}\Big(\frac{\log n}{nv}\Big), \end{aligned}$$

which proves (3.14). $\square$

3.3 Rate of Convergence of $m_{B_n}(z)$ to $m(z)$

Now, we start with the second step. First remark is the following fact:

Remark 3.3.1. If $f'(0) = 0$, then the matrix B_n is a nonrandom matrix which has ESD as a function

$$F_{B_n}(x) = \delta_{f(0)}(x), \tag{3.22}$$

where $\delta_a(x)$ is the Dirac measure.

Suppose that $f'(0) \neq 0$, then the matrix B_n is a linear function of a sample covariance matrix which has been more extensively studied. Since the eigenvalues of B_n are translations of the eigenvalues of $f'(0)X_i^T X_j$, for convenience, we consider

$$(B_n)_{ij} = f'(0)X_i^T X_j, \tag{3.23}$$

where $i = 1, \cdots, n$, X_i is a random vector with entries X_{ij}, $j = 1, \cdots, n$, defined by (1.13).

So far there have been several papers which showed the convergence of ESD of B_n defined in (3.23) to Marchenko Pastur Distribution using Stieltjes Transforms of distributions as the introduction.

Now, we prove the estimate of difference between the two Stieltjes transforms. Notice that this way is different from computing the rate of convergence in the sense of probability convergence. A key point comes from the following results.

Proposition 3.3.2. *Assume that $m(z)$ is the Stieltjes transform of*

$$\frac{1}{|f'(0)|}\rho_{MP,1}\Big(\frac{x+f'(0)}{f'(0)}\Big), \quad \rho_{MP,1}(x) \text{ is defined on (2.7).}$$

Then $m(z)$ satisfies the following quadratic equation,

$$-\frac{1}{m(z)} = z - \frac{f'(0)}{1+f'(0)m(z)}.$$

From this we have the equation,

$$m(z) = \frac{1}{-z - zf'(0)m(z)}. \tag{3.24}$$

Our proof is based on using an idea in O'Rourke's paper [29]. The idea is to use the diagonal entries of resolvent for the random matrix B_n defined in (3.23). Let us consider the resolvent

$$R_n(z) = (B_n - zI)^{-1}, \qquad z = u + iv \in \mathbb{C} \quad v > 0.$$

Then we can write

$$\Big(R_n(z)\Big)_{kk} = \frac{1}{f'(0)\sum_{s=1}^n x_{ks}^2 - z - B_k^T R_n^k(z) B_k}, \quad k = 1, \cdots, n$$

where B_k is the k-th row of B_n except k-th entry. For each $k = 1, 2, \cdots, n$, we define

$$a_k := \frac{1}{\Big(R_n(z)\Big)_{kk}}. \tag{3.25}$$

Then we have the following theorem.

Theorem 3.3.3. *For $z = u + iv \in \mathbb{C},\ v > 0,$ we have*

$$\left|\mathbb{E}m_{B_n}(z) - \frac{1}{-z - zf'(0)\mathbb{E}m_{B_n}(z)}\right| = \mathcal{O}\Big(\frac{1}{nv}\Big). \tag{3.26}$$

Proof. Let us consider

$$B_n = \Big(f'(0) X_i^T X_j\Big)_{1 \le i,j \le n} = f'(0) \sum_{s=1}^n x_{is} x_{js},$$

where X_i is n-dimensional vector by

$$X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix}, \quad i = 1, 2, \cdots, n.$$

Define the resolvent of B_n by,

$$R_n(z) = \Big(B_n - zI_n\Big)^{-1},$$

and for $k = 1, 2, \cdots, n$, its diagonal entries,

$$\big(R_n(z)\big)_{kk} = \frac{1}{f'(0)\sum\limits_{s=1}^n x_{ks}^2 - z - B_k^T R_n^k(z) B_k}, \tag{3.27}$$

where $R_n^k(z) = (B_n^k - zI_{n-1})^{-1}$ and $(n-1)$-dimensional vector

$$\begin{aligned} B_k^T =& f'(0)\Big(X_k^T X_1, \cdots, X_k^T X_{k-1}, X_k^T X_{k+1}, \cdots, X_k^T X_n\Big) \\ =& f'(0) X_k^T \Big(X_1, \cdots, X_{k-1}, X_{k+1}, \cdots, X_n\Big)_{n\times(n-1)} \\ =& f'(0) X_k^T X_{n-1}^k. \end{aligned} \tag{3.28}$$

Here, we denote the $n \times (n-1)$ matrix X_{n-1}^k by

$$X_{n-1}^k = \Big(X_1, \cdots, X_{k-1}, X_{k+1}, \cdots, X_n\Big)_{n\times(n-1)}. \tag{3.29}$$

Based on the definition of a_k in (3.25), we have

$$a_k = f'(0)\sum_{s=1}^{n} x_{ks}^2 - z - B_k^T R_n^k(z) B_k. \tag{3.30}$$

It is easy to see from the notation (3.28) that

$$\mathbb{E}\Big[B_k^T R_n^k(z) B_k\Big] = (f'(0))^2 \mathbb{E}\Big[\sum_{\substack{s,r=1\\ s,r\neq k}}^{n} X_k^T X_s \big(R_n^k(z)\big)_{sr} X_r^T X_k\Big]. \tag{3.31}$$

Notice that

$$\begin{aligned} & \mathbb{E}\Big[X_k^T X_s \big(R_n^k(z)\big)_{sr} X_r^T X_k\Big] \\ =& \sum_{q_1,q_2=1}^{n} \mathbb{E}\Big[x_{kq_1} x_{sq_1} x_{k1_2} x_{rq_2} \big(R_n^k(z)\big)_{sr}\Big] \\ =& \sum_{q_1,q_2=1}^{n} \mathbb{E}\Big[x_{kq_1} x_{kq_2}\Big] E\Big[x_{sq_1} x_{rq_2} \big(R_n^k(z)\big)_{sr}\Big] \\ =& \sum_{q=1}^{n} \mathbb{E}\Big[(x_{kq})^2\Big] \mathbb{E}\Big[x_{sq} x_{rq} \big(R_n^k(z)\big)_{sr}\Big] \\ =& \frac{1}{n}\sum_{q=1}^{n} \mathbb{E}\Big[x_{sq} x_{rq} \big(R_n^k(z)\big)_{sr}\Big]. \end{aligned} \tag{3.32}$$

It follows from (3.32) that

$$\mathbb{E}\Big[B_k^T R_n^k(z) B_k\Big] = \frac{(f'(0))^2}{n} \sum_{\substack{s,r=1\\ s,r\neq k}}^{n} \sum_{q=1}^{n} \mathbb{E}\Big[x_{sq} x_{rq} \big(R_n^k(z)\big)_{sr}\Big]. \tag{3.33}$$

Now, let us compute the estimate for $\mathbb{E}|a_k - \mathbb{E}a_k|$. Indeed, we have for $k = 1, \cdots, n$,

$$\begin{aligned}&\mathbb{E}|a_k - \mathbb{E}a_k| \\ \leq&\mathbb{E}|X_k^T X_k - 1| + \mathbb{E}\Big|B_k^T R_n^k(z) B_k - \mathbb{E}\Big[B_k^T R_n^k(z) B_k\Big]\Big| \end{aligned} \tag{3.34}$$

Then for the first term of (3.34), by Cauchy-Schwarz inequality, we have

$$\begin{aligned}&\mathbb{E}|X_k^T X_k - 1| \\ \leq&\Big(\mathbb{E}\Big[\sum_{s,r=1}^{n}\big(x_{ks}^2 - \frac{1}{n}\big)\big(x_{kr}^2 - \frac{1}{n}\big)\Big]\Big)^{1/2} \\ =&\begin{cases}\sum_{s=1}^{n}\mathbb{E}[x_{ks}^4] - \frac{2}{n}\sum_{s=1}^{n}\mathbb{E}[x_{ks}^2] + \frac{1}{n}, & s = r, \\ \sum_{s\neq r=1}^{n}\mathbb{E}[x_{ks}^2 x_{kr}^2] - \frac{1}{n}\sum_{s=1}^{n}\mathbb{E}[x_{ks}^2] - \frac{1}{n}\sum_{r=1}^{n}\mathbb{E}[x_{kr}^2] + \frac{1}{n}, & s \neq r.\end{cases} \\ =&\begin{cases}\frac{\sqrt{2}}{n}, & s = r, \\ 0, & s \neq r.\end{cases}\end{aligned} \tag{3.35}$$

Next let us consider the second term of (3.34). Indeed, by the notation (3.29), we think of $B_k^T R_n^k(z) B_k$ as

$$B_k^T R_n^k(z) B_k = (f'(0))^2 X_k^T X_{n-1}^k R_n^k(z) (X_{n-1}^k)^T X_k, \tag{3.36}$$

where $X_{n-1}^k R_n^k(z)(X_{n-1}^k)^T$ is a $n \times n$ matrix. It follows from (3.33) and (3.36) that

$$\mathbb{E}\Big[B_k^T R_n^k(z) B_k\Big] = \frac{(f'(0))^2}{n}\mathbb{E}\Big[\operatorname{Tr}\Big(X_{n-1}^k R_n^k(z)(X_{n-1}^k)^T\Big)\Big].$$

Also, we have

$$\begin{aligned}&\mathbb{E}\Big[\operatorname{Tr}\Big(X_{n-1}^k R_n^k(z)(X_{n-1}^k)^T\Big)\Big] \\ =&\mathbb{E}\Big[\operatorname{Tr}\Big((X_{n-1}^k)^T X_{n-1}^k R_n^k(z)\Big)\Big] \\ =&\mathbb{E}\Big[\frac{1}{f'(0)}\operatorname{Tr}\Big(I_{n-1} + zR_n^k(z)\Big)\Big] \\ =&\frac{(n-1)}{f'(0)} + \frac{z}{f'(0)}\mathbb{E}\Big[\operatorname{Tr} R_n^k(z)\Big].\end{aligned} \tag{3.37}$$

Since

$$\sum_{\substack{s,r=1\\s,r\neq k}}^{n} \mathbb{E}\Big[\Big(\big(R_n^k(z)\big)_{sr}\Big)^2\Big] \leq \mathbb{E}\Big[\operatorname{Tr}\big(R_n^k(z)\big)^2\Big], \tag{3.38}$$

then

$$\begin{aligned}
&\mathbb{E}\Big|B_k^T R_n^k(z) B_k - \mathbb{E}\Big[B_k^T R_n^k(z) B_k\Big]\Big|^2 \\
=&\frac{f'(0)^2}{n^2}\mathbb{E}\Big(B_k^T R_n^k(z) B_k\Big)^2 - \Big(\mathbb{E}\Big[B_k^T R_n^k(z) B_k\Big]\Big)^2 \\
\leq& f'(0)|z|\mathbb{E}\Big[\frac{1}{n^2}\operatorname{Tr}\big(R_n^k(z)\big)^2\Big] - \Big(\mathbb{E}\Big[\frac{1}{n}\operatorname{Tr} R_n^k(z)\Big]\Big)^2 \\
=& f'(0)|z|\mathbb{E}\Big|\frac{1}{n}\operatorname{Tr} R_n^k(z) - \mathbb{E}\Big[\frac{1}{n}\operatorname{Tr} R_n^k(z)\Big]\Big|^2 \\
=&\mathcal{O}\Big(\frac{1}{n^3}\Big).
\end{aligned} \tag{3.39}$$

It comes from (3.34), (3.35) and (3.39) that

$$\sup_k \mathbb{E}|a_k - \mathbb{E}a_k| = \mathcal{O}\Big(\frac{1}{n}\Big). \tag{3.40}$$

Next, (3.37) enables us to show that for large n,

$$\begin{aligned}
\mathbb{E}a_k =&\mathbb{E}\Big[f'(0)\sum_{s=1}^{n} x_{ks}^2 - z - B_k^T R_n^k(z) B_k\Big] \\
=& f'(0) - z - \mathbb{E}\Big[B_k^T R_n^k(z) B_k\Big] \\
=& f'(0) - z - \frac{(f'(0))^2}{n}\mathbb{E}\Big[\operatorname{Tr}\Big(X_{n-1}^k R_n^k(z)(X_{n-1}^k)^T\Big)\Big] \\
=& f'(0) - z - \frac{(n-1)}{n}f'(0) - \frac{zf'(0)}{n}\mathbb{E}\Big[\operatorname{Tr} R_n^k(z)\Big] \\
=& -z - \frac{zf'(0)}{n}\mathbb{E}\Big[\operatorname{Tr} R_n^k(z)\Big].
\end{aligned} \tag{3.41}$$

Using (3.37), (3.41) and Cauchy's Interlacing Theorem,

$$\Big|\frac{1}{n}\operatorname{Tr} R_n^k(z) - \frac{1}{n}\operatorname{Tr} R_n(z)\Big| = \mathcal{O}\Big(\frac{1}{nv}\Big),$$

we get for all $k = 1, 2, \cdots, n$,

$$\begin{aligned}
&\Big|\mathbb{E}a_k - \big(-z - zf'(0)\mathbb{E}m_{B_n}(z)\big)\Big| \\
\leq&\big|zf'(0)\big|\Big|\frac{1}{n}\operatorname{Tr} R_n^k(z) - \frac{1}{n}\operatorname{Tr} R_n(z)\Big| \\
=&\mathcal{O}\Big(\frac{1}{nv}\Big).
\end{aligned} \tag{3.42}$$

Therefore, (3.40) and (3.42) give that for some constant $C > 0$,

$$\begin{aligned}
&\left|\mathbb{E}m_{B_n}(z) - \frac{1}{-z - zf'(0)\mathbb{E}m_{B_n}(z)}\right| \\
\leq &\left|\mathbb{E}m_{B_n}(z) - \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\mathbb{E}a_k}\right| + \left|\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\mathbb{E}a_k} - \frac{1}{-z - zf'(0)\mathbb{E}m_{B_n}(z)}\right| \\
\leq &\sup_k \mathbb{E}|a_k - \mathbb{E}a_k| + C\sup_k \left|\mathbb{E}a_k - \big(-z - zf'(0)\mathbb{E}m_{B_n}(z)\big)\right| \\
= &\mathcal{O}\Big(\frac{1}{nv}\Big),
\end{aligned}$$

which completes the proof. □

Corollary 3.3.4. *For $z = u + vi,\ v > 0$, we have*

$$|\mathbb{E}m_{B_n}(z) - m(z)| = \mathcal{O}\Big(\frac{1}{nv}\Big). \tag{3.43}$$

Proof. By Theorem 3.3.3, we have for all $z = u + vi,\ v > 0$,

$$\mathbb{E}m_{B_n}(z) = \frac{1}{-z - zf'(0)\mathbb{E}m_{B_n}(z)} + \mathcal{O}\Big(\frac{1}{nv}\Big). \tag{3.44}$$

Fix $z_0 = u_0 + iv_0$. Since $|\mathbb{E}m_{B_n}(z)| \leq \frac{1}{v_0}$, by using a compactness argument, one can see that there is a convergent subsequence such that

$$|\mathbb{E}m_{B_{n_k}}(z_0) - s(z_0)| = \mathcal{O}\Big(\frac{1}{n_k}\Big). \tag{3.45}$$

Next, since the matrix $f'(0)X_i^TX_j$ are positive semi-definite, then $\Im(z\mathbb{E}m_{B_n}) \geq 0$ for all $\Im z > 0$ and so $\Im\big(z_0 s(z_0)\big) \geq 0$. Thus, it follows from (3.24) that

$$s(z_0) = m(z_0).$$

Since every convergent subsequences of $\mathbb{E}m_{B_n}$ converges to the same limit, we obtain by (3.44) and (3.45),

$$|\mathbb{E}m_{B_n}(z_0) - m(z_0)| = \mathcal{O}\Big(\frac{1}{nv}\Big).$$

Finally, since z_0 is arbitrary, we get (3.43). □

3.4 Beyond the Gaussian

In the previous chapters, we assumed that all vectors $X_1, X_2, \cdots, X_n$ were iid Gaussian vectors with mean 0 and variance $\frac{1}{n}$ for convenience. In this section, we prove the rate of limiting to Stieltjes transform of MP distribution without the assumption that they are Gaussian distribution. Instead, we assume the following sub-Gaussian condition on $X_i^T X_j$: For any $\epsilon > 0$,

$$\Pr\left(|X_i^T X_j| > \epsilon\right) \leq 2e^{-\frac{n}{2}\epsilon^2}. \tag{3.46}$$

Under the new assumption, let us consider the random matrix A defined as (3.1) and the random matrix B defined as (3.3). Then we have the same results as Theorem 3.2.2 and Theorem 3.3.3.

Theorem 3.4.1. *Let A and B be $n \times n$ matrices defined as* (3.1) *and* (3.3)*. Then for any $z = u + iv \in \mathbb{C}, v > 0$,*

$$\mathbb{E}\Big[m_A(z) - m_B(z)\Big] = \mathcal{O}\Big(\frac{\log n}{nv}\Big).$$

Proof. Let us consider $\mathcal{F}$ as an event that there exist $i \neq j$ such that

$$|X_i^T X_j| > \sqrt{\frac{4\log n}{n}}.$$

Using (3.46), we have

$$\Pr(\mathcal{F}) \leq \mathcal{O}\Big(\frac{1}{n^2}\Big). \tag{3.47}$$

Considering as the proof of Theorem 3.2.2, (3.47) completes the proof. □

Based on the sub-Gaussian condition on $X_i^T X_j$, we have

$$\sup_{i,j} \mathbb{E}[x_{i,j}^4] = \mathcal{O}\Big(\frac{1}{n^2}\Big). \tag{3.48}$$

From this, we have the following theorem.

Theorem 3.4.2. *For $z = u + iv \in \mathbb{C},\ v > 0,$ we have*

$$\left|\mathbb{E}m_B(z) - \frac{1}{-z - zf'(0)\mathbb{E}m_B(z)}\right| = \mathcal{O}\Big(\frac{1}{nv}\Big) \tag{3.49}$$

Proof. Using (3.48), we have

$$
\begin{aligned}
&\mathbb{E}|X_k^T X_k - 1| \\
\leq &\Big(\mathbb{E}\Big[\sum_{s,r=1}^{n}\big(x_{ks}^2 - \frac{1}{n}\big)\big(x_{kr}^2 - \frac{1}{n}\big)\Big]\Big)^{1/2} \\
= &\begin{cases} \sum_{s=1}^{n}\mathbb{E}[x_{ks}^4] - \frac{2}{n}\sum_{s=1}^{n}\mathbb{E}[x_{ks}^2] + \frac{1}{n}, & s = r, \\ \sum_{s\neq r=1}^{n}\mathbb{E}[x_{ks}^2 x_{kr}^2] - \frac{1}{n}\sum_{s=1}^{n}\mathbb{E}[x_{ks}^2] - \frac{1}{n}\sum_{r=1}^{n}\mathbb{E}[x_{kr}^2] + \frac{1}{n}, & s \neq r. \end{cases} \\
= &\begin{cases} \mathcal{O}(\frac{1}{n}), & s = r, \\ 0, & s \neq r. \end{cases}
\end{aligned} \tag{3.50}
$$

Considering as the proof of Theorem 3.3.3, the estimate (3.50) completes the proof. □

CHAPTER 4

KOLMOGOROV DISTANCE BETWEEN ESD FUNCTIONS

Let $A = (A_{ij})_{1\le i,j\le n}$ be a $n\times n$ random inner product kernel matrix defined by

$$A_{ij} = \begin{cases} f(X_i^T X_j, n), & i \neq j, \\ 0, & i = j, \end{cases} \tag{4.1}$$

where $f : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is a twice differentiable function at $x = 0$, i.e., $f''(0)$ exists. Also, we define a matrix B by

$$B_{ij} = f(0) + f'(0) X_i^T X_j, \tag{4.2}$$

where $X_1, X_2, \ldots, X_n$ are iid standard Gaussian vectors and the function f is the same function as the definition of A.

In Chapter 3, we showed that Stieltjes transform $m_A(z)$ converged to $m(z)$ for all $z = u + vi,\ v > 0$ with rate $\mathcal{O}\Big(\frac{\log n}{n}\Big)$ in Theorem 3.3.4. In this Chapter, using the result, we prove the rate of Kolmogorov distance between the expected ESD $\mathbb{E}F_A(x)$ of a random matrix A defined in (4.1) and Marchenko-Pastur distribution

$$G(x) := \int_{-\infty}^{x} \rho_{MP,1}(x)dx. \tag{4.3}$$

where $\rho_{MP,1}(x)$ defined in (1.2) is of order $\mathcal{O}\Big(\sqrt[3]{\frac{\log n}{n}}\Big)$. For convenience, we assume that $f(0) = 0$.

4.1 Kolmogorov Distance

To prove the result, we need the following Berry-Esseen type inequality proved in [5].

Proposition 4.1.1. *(**Bai inequality**) [See [5]] Let F be a distribution function and let G be a function of bounded variation satisfying*

$$\int_{-\infty}^{\infty} |F(x) - G(x)|dx < \infty. \tag{4.4}$$

Denote the Stieltjes transforms by $m_F(z)$ and $m_G(z)$ respectively, where $z = u+iv \in \mathbb{C}$ with $v > 0$. Then we have the following estimate for the Kolmogorov distance:

$$\begin{aligned}\|F - G\| =& \sup_x |F(x) - G(x)| \\ \leq& \frac{1}{\pi(1-\zeta)(2\beta-1)}\Bigg(\int_{-a}^{a} |m_F(z) - m_G(z)|du \\ &+ 2\pi\frac{1}{v}\int_{|x|>b} |F(x) - G(x)|dx \\ &+ \frac{1}{v}\sup_x \int_{|u|<2v\epsilon} |G(x+u) - G(x)|du\Bigg),\end{aligned}$$

where the constants $a > b > 0$, $\beta > 0$ and $\epsilon > 0$ are restricted by

$$\beta = \frac{1}{\pi}\int_{|u|\leq\epsilon} \frac{1}{u^2+1}du > \frac{1}{2},$$

and

$$\zeta = \frac{4b}{\pi(a-b)(2\beta-1)} \in (0,1).$$

Let us consider $G(x)$ defined as (4.3). By Proposition 4.1.1 with $a = 25, b = 5$, we have the upper bound for Kolmogorov distance between $\mathbb{E}F_A(x)$ and

$G(x)$,

$$\|\mathbb{E}F_A - G\| \leq \frac{1}{\pi(1-\zeta)(2\beta-1)}\Bigg(\int_{-25}^{25} |\mathbb{E}m_A(u+iv) - m(u+iv)|du + 2\pi\frac{1}{v}\int_{|x|>5} |\mathbb{E}F_A(x) - G(x)|dx + \frac{1}{v}\sup_x \int_{|t|<2v\epsilon} |G(x+t) - G(x)|dt\Bigg), \quad (4.5)$$

where $\beta, \epsilon > 0$ are restricted by

$$\beta = \frac{1}{\pi}\int_{|u|\leq\epsilon} \frac{1}{u^2+1}du > \frac{1}{2}, \quad \zeta = \frac{1}{\pi(2\beta-1)} \in (0,1).$$

Now, we prove the Lemmas to estimate the upper bounds for above three terms on the right-hand side in the inequality (4.5).

Lemma 4.1.2. *For any $z = u + vi$, $v > 0$, there exists a constant $C_1 > 0$ such that*

$$\int_{-25}^{25} |\mathbb{E}m_A(u+iv) - m(u+iv)|du \leq C_1\frac{\log n}{nv}.$$

Proof. By Theorem 3.1.1, there exists a constant $C > 0$ such that

$$\int_{-25}^{25} |\mathbb{E}m_A(u+iv) - m(u+iv)|du \leq \int_{-25}^{25} C\frac{\log n}{nv}du \leq 50C\frac{\log n}{nv}.$$

□

Lemma 4.1.3. *For any $z = u + vi$, $v > 0$, there exists a constant $C_2 > 0$ such that*

$$\frac{1}{v}\sup_x \int_{|t|<2v\epsilon} |G(x+t) - G(x)|dt \leq C_2\sqrt{v}.$$

Proof. By the definition of the function $G(x)$, we have

$$\int_{-2v\epsilon}^{2v\epsilon} |G(x+t) - G(x)|dt = 2\int_0^{2v\epsilon} G(x+t) - G(x)dt = 2\int_0^{2v\epsilon}\int_x^{x+t} \rho_{MP,1}(s)\ ds\ dt. \quad (4.6)$$

Put

$$\phi(x) = \int_x^{x+t} \rho_{MP,1}(s)\ ds.$$

Since

$$\frac{\sqrt{x(4-x)}}{x} \leq \frac{2}{\sqrt{x}}, \quad 0 < x < 4,$$

then we have

$$\begin{aligned} \phi(x) &\leq \int_x^{x+t} \frac{1}{\pi\sqrt{s}} ds \\ &= \frac{2}{\pi}(\sqrt{x+t} - \sqrt{x}) \\ &= \frac{2t}{\pi(\sqrt{x+t} + \sqrt{x})} \\ &\leq \frac{2t}{\pi\sqrt{t}} = \frac{2}{\pi}\sqrt{t}. \end{aligned} \tag{4.7}$$

Notice that

$$2\int_0^{2v\epsilon} \frac{2}{\pi}\sqrt{t} dt = \frac{8}{3\pi}(2v\epsilon)^{3/2} = \frac{8}{3\pi}(2\epsilon)^{3/2} v^{3/2}. \tag{4.8}$$

Thus, it follows from (4.2), (4.7) and (4.8) that

$$\frac{1}{v} \sup_x \int_{|t|<2v\epsilon} |G(x+t) - G(x)| dt \leq \frac{8}{3\pi}(2\epsilon)^{3/2}\sqrt{v}.$$

□

Next, we obtain an upper bound of the second term in (4.5). To do so, we investigate the distribution of the largest eigenvalue of the matrix A. The following Proposition gives us to get our required estimate for the eigenvalue.

Proposition 4.1.4. *Consider a random matrix B defined in* (4.2) *and the largest eigenvalue of B, denoted by $\lambda_{\max}(B)$. Then for any $x > 0$, we have*

$$\Pr\{\lambda_{\max}(B) > 4 + x\} \leq \frac{1}{8\sqrt{\pi}\sqrt{nx}} \exp\Big(-\frac{8nx}{3}\Big).$$

Proof. In [24], Johnstone proved that there was a constant $C > 0$ such that

$$\Pr\{\lambda_{\max}(B) > 4 + x\} = 1 - F_1(2\sqrt[3]{2}\sqrt[3]{n^2}x) + C\frac{1}{n^{2/3}} \exp(-2\sqrt[3]{2}\sqrt[3]{n^2}x), \tag{4.9}$$

where $F_1(x)$ is a Tracy-Widom distribution defined as

$$F_1(x) = \exp\Big(-\frac{1}{2}\int_x^\infty q(t) + (x-t)q^2(t)dt\Big).$$

Here, $q(t)$ is a solution of the Second Painlevé differential equation:

$$\frac{d^2q(t)}{dt^2} = tq(t) + 2q^3(t), \qquad q(t) \sim \frac{t^{-1/4}}{2\sqrt{\pi}}\exp(-\frac{2t^{3/2}}{3}) \text{ as } t \to \infty.$$

Then we have

$$F_1(x) = \exp\Big(-\frac{1}{2}\int_x^\infty q(t)dt\Big)\exp\Big(-\frac{1}{2}\int_x^\infty q^2(t)dt\Big).$$

Since

$$\int_x^\infty q(t)dt \le \frac{e^{-\frac{4x^{3/2}}{3}}}{16\pi x^{3/2}}, \tag{4.10}$$

and

$$\int_x^\infty q^2(t)dt \le \frac{e^{-\frac{2x^{3/2}}{3}}}{2\sqrt{\pi}x^{3/4}}, \tag{4.11}$$

then we obtain the estimate

$$1 - F_1(\sqrt[3]{2}\sqrt[3]{n^2}x) \le \frac{1}{8\sqrt{\pi}\sqrt{nx}}\exp\Big(-\frac{8nx}{3}\Big).$$

This completes the proof. □

Lemma 4.1.5. *For any $z = u + vi,\ v > 0$,*

$$2\pi\frac{1}{v}\int_{|x|>5}|\mathbb{E}F_A(x) - G(x)|dx = \mathcal{O}\Big(\frac{n^2}{e^n}\Big).$$

Proof. It is obvious from the definition of $G(x)$ that

$$G(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x > 4. \end{cases} \tag{4.12}$$

Since A is positive-definite, it is clear that

$$\int_{-\infty}^{-5}|\mathbb{E}F_A(x)|dx = 0. \tag{4.13}$$

Next, let us consider the largest eigenvalue $\lambda_{\max}(A)$ of A and the largest eigenvalue $\lambda_{\max}(B)$ of B defined in (4.2). Note that for all $x > 0$,

$$\begin{aligned} &\Pr(\lambda_{\max}(A) > 5 + x) \\ \leq &\Pr\Big(|\lambda_{\max}(A) - \lambda_{\max}(B)| > 1 + \frac{x}{2}\Big) + \Pr\Big(\lambda_{\max}(B) > 4 + \frac{x}{2}\Big). \end{aligned} \tag{4.14}$$

Using Hoffman-Wielandt inequality, we have

$$\begin{aligned} |\lambda_{\max}(A) - \lambda_{\max}(B)|^2 &\leq \sum_{i=1}^{n} |\lambda_i(A) - \lambda_i(B)|^2 \\ &\leq \mathrm{Tr}(A - B)^2 \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{n} |A_{ij} - B_{ij}|^2. \end{aligned} \tag{4.15}$$

Based on the choice of f in the definition of matrix A, it comes from the Taylor polynomial that there exists a constant $D_1 > 0$ such that

$$|A_{ij} - B_{ij}| \leq D_1 |X_i^T X_j|^2. \tag{4.16}$$

It follows from (4.15) and (4.16) that

$$\begin{aligned} &\Pr\Big(\lambda_{\max}(A) - \lambda_{\max}(B) > 1 + \frac{x}{2}\Big) \\ \leq &\Pr\Big(|\lambda_{\max}(A) - \lambda_{\max}(B)|^2 > 1 + \frac{x}{2}\Big) \\ \leq &\Pr\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} D_1^2 |X_i^T X_j|^4 > 1 + \frac{x}{2}\Big) \\ \leq &D_2 n^2 \Pr\Big(|X_i^T X_j|^4 > 1 + \frac{x}{2}\Big) \\ \leq &D_2 n^2 \Pr\Big(|X_i^T X_j| > 1 + \frac{x}{2}\Big) \\ \leq &D_2 n^2 \exp\Big(-\frac{n}{2}(1 + \frac{x}{2})^2\Big), \end{aligned} \tag{4.17}$$

for some constant $D_2 > 0$. Thus, it follows from (4.14), (4.17) and Proposition 4.1.4 that

$$\begin{aligned} &\Pr(\lambda_{\max}(A) > 5 + x) \\ \leq &D_2 n^2 \exp\Big(-\frac{n}{2}\big(1 + \frac{x}{2}\big)^2\Big) + \frac{1}{8\sqrt{\pi}\sqrt{nx}} \exp\Big(-\frac{8nx}{3}\Big). \end{aligned} \tag{4.18}$$

Since

$$1 - F_A(x) \le \chi_{(-\infty,\ \lambda_{\max}(A))}(x) \quad \text{for} \quad x \ge 0,$$

then there exists a constant $D > 0$ such that

$$\begin{aligned}\int_5^{\infty} |\mathbb{E}F_A(x) - 1| dx &\le \int_5^{\infty} \Pr(\lambda_{\max}(A) > x) dx \\ &\le Dn^2 \exp(-n). \end{aligned} \tag{4.19}$$

Therefore, (4.12), (4.13) and (4.19) enable us to see that

$$\begin{aligned}&\int_{|x|>5} |\mathbb{E}F_A(x) - G(x)| dx \\ &= \int_{-\infty}^{-5} |\mathbb{E}F_A(x) - G(x)| dx + \int_5^{\infty} |\mathbb{E}F_A(x) - G(x)| dx. \\ &= \int_{-\infty}^{-5} |\mathbb{E}F_A(x)| dx + \int_5^{\infty} |\mathbb{E}F_A(x) - 1| dx = \mathcal{O}\Big(\frac{n^2}{e^n}\Big),\end{aligned}$$

which completes the proof. □

Actually, it follows from Lemma 4.1.5 that

$$\begin{aligned}\int_{-\infty}^{\infty} |\mathbb{E}F_A(x) - G(x)| dx &= \int_{|x|\le 5} |\mathbb{E}F_A(x) - G(x)| dx + \int_{|x|>5} |\mathbb{E}F_A(x) - G(x)| dx \\ &= \int_{|x|\le 5} |\mathbb{E}F_A(x) - G(x)| dx + \mathcal{O}\Big(\frac{n^2}{e^n}\Big) < \infty,\end{aligned}$$

which satisfies the condition of Proposition 4.1.1. Then we can prove the following Theorem to get our desired rate.

Theorem 4.1.6. *If A is a random inner product kernel matrix defined in* (4.1) *and $F_A(x)$ is the ESD of A then*

$$\|\mathbb{E}F_A(x) - G(x)\| = \sup_x |\mathbb{E}F_A(x) - G(x)| = \mathcal{O}\left(\sqrt[3]{\frac{\log n}{n}}\right), \tag{4.20}$$

where $G(x) = \int_{-\infty}^{x} \rho_{MP,1}(t)dt$.

Proof. By Lemma 4.1.2, Lemma 4.1.3 and Lemma 4.1.5, we have

$$||\mathbb{E}F_A(x) - G(x)|| = \mathcal{O}\Big(\frac{\log n}{nv}\Big) + \mathcal{O}(\sqrt{v}).$$

Now taking $v = v_n = \sqrt[3]{\left(\frac{\log n}{n}\right)^2} \to 0$ as $n \to \infty$, then we obatin the rate (4.23). □

4.2 Beyond the Gaussian

In this section, we assume that all vectors $X_1, X_2, \cdots, X_n$ are iid with mean 0 and variance $\frac{1}{n}$. To compute the rate of limiting to MP distribution without the assumption that they are Gaussian distributions, we need the following sub-Gaussian condition on $X_i^T X_j$: For any $\epsilon > 0$,

$$\Pr\Big(|X_i^T X_j| > \epsilon\Big) \le 2e^{-\frac{n}{2}\epsilon^2}. \tag{4.21}$$

Under this new assumption, let us consider the random matrix A defined as (4.1) and the random matrix B defined as (4.2). Then we have the same results as Theorem 4.2.4.

First of all, to prove Proposition 4.2.2, we use Soshnikov's result in [33]. In this paper, he focused on Wishart matrices with the following four conditions:

(1) For $1 \le i < j \le n$, $\mathbb{E}[x_{ij}] = 0, \mathbb{E}[x_{ij}^2] = 1$,

(2) The random variables x_{ij} have symmetric laws of distribution,

(3) All moments of these random variables are finite,

(4) The distributions of x_{ij} decay at infinity at least as fast as Gaussian distribution,

$$\mathbb{E}[x_{ij}^{2k}] \le (Ck)^k,\ C > 0.$$

Here, our matrix satisfies automatically the first three conditions so that it suffices to prove the fourth condition using (4.21).

Lemma 4.2.1. *Under the condition* (4.21)*, for any positive integer $k \ge 1$,*

$$\mathbb{E}[x_{ij}^{2k}] \le 2\Big(2\frac{k}{n}\Big)^k,$$

where x_{ij} is $j-th$ entry of the vector X_i.

Proof. It is clear that

$$\begin{aligned}\mathbb{E}[x_{ij}^{2k}] \le \mathbb{E}\big[|X_i^T X_j|^{2k}\big] &= \int_0^\infty \Pr\{|X_i^T X_j|^{2k} > t\}dt \\ &= \int_0^\infty \Pr\{|X_i^T X_j| > t^{\frac{1}{2k}}\}dt \\ &\le 2\int_0^\infty \exp\Big(-\frac{n}{2}t^{\frac{1}{k}}\Big)dt \\ &= 2\Big(2\frac{1}{n}\Big)^k k\Gamma(k) \le 2\Big(2\frac{k}{n}\Big)^k.\end{aligned}$$

□

Lemma 4.2.1 leads to the following Proposition.

Proposition 4.2.2. *Consider a random matrix B defined in* (4.2) *and the largest eigenvalue of B, denoted by $\lambda_{\max}(B)$. Then for any $x > 0$, we have*

$$\Pr\{\lambda_{\max}(B) > 4 + x\} \le \frac{1}{8\sqrt{\pi}\sqrt{nx}} \exp\Big(-\frac{8nx}{3}\Big).$$

Proof. Under the above assumption, Soshnikov proved the following fact in [33]: There exists a constant $C > 0$ such that

$$\Pr\{\lambda_{\max}(B) > 4 + x\} = 1 - F_1(2\sqrt[3]{2}\sqrt[3]{n^2}x) + C\exp(-2\sqrt[3]{2}\sqrt[3]{n^2}x). \tag{4.22}$$

where $F_1(x)$ is a Tracy-Widom distribution defined as

$$F_1(x) = \exp\Big(-\frac{1}{2}\int_x^\infty q(t) + (x-t)q^2(t)dt\Big).$$

Here, $q(t)$ is a solution of the Second Painlevé differential equation:

$$\frac{d^2q(t)}{dt^2} = tq(t) + 2q^3(t), \qquad q(t) \sim \frac{t^{-1/4}}{2\sqrt{\pi}} \exp\Big(-\frac{2t^{3/2}}{3}\Big) \text{ as } t \to \infty.$$

Then we have again

$$F_1(x) = \exp\Big(-\frac{1}{2}\int_x^\infty q(t)dt\Big)\exp\Big(-\frac{1}{2}\int_x^\infty q^2(t)dt\Big).$$

Based on the definition of $F_1(x)$, we obtain the estimate

$$1 - F_1(\sqrt[3]{2}\sqrt[3]{n^2}x) \le \frac{1}{8\sqrt{\pi}\sqrt{nx}} \exp\Big(-\frac{8nx}{3}\Big).$$

This completes the proof. □

Using Proposition 4.2.2 and similar argument of the proof for Lemma 4.1.5, we have the following Lemma.

Lemma 4.2.3. *For any* $z = u + vi,\ v > 0,$

$$2\pi\frac{1}{v}\int_{|x|>5}|\mathbb{E}F_A(x) - G(x)|dx = \mathcal{O}\Big(\frac{n^2}{e^n}\Big).$$

Notice that it follows from Lemma 4.2.3 that

$$\begin{aligned}\int_{-\infty}^{\infty}|\mathbb{E}F_A(x) - G(x)|dx &= \int_{|x|\le 5}|\mathbb{E}F_A(x) - G(x)|dx + \int_{|x|>5}|\mathbb{E}F_A(x) - G(x)|dx \\ &= \int_{|x|\le 5}|\mathbb{E}F_A(x) - G(x)|dx + \mathcal{O}\Big(\frac{n^2}{e^n}\Big),\end{aligned}$$

which satisfies the condition of Proposition 4.1.1. Then we can prove the following Theorem.

Theorem 4.2.4. *If A is a random inner product kernel matrix defined in* (4.1) *and $F_A(x)$ is the ESD of A then*

$$\|\mathbb{E}F_A(x) - G(x)\| = \sup_x|\mathbb{E}F_A(x) - G(x)| = \mathcal{O}\left(\sqrt[3]{\frac{\log n}{n}}\right), \tag{4.23}$$

where $G(x) = \int_{-\infty}^{x}\rho_{MP,1}(t)dt$.

Proof. By Lemma 4.1.2, Lemma 4.1.3 and Lemma 4.2.3, we have

$$||\mathbb{E}F_A(x) - G(x)|| = \mathcal{O}\Big(\frac{\log n}{nv}\Big) + \mathcal{O}(\sqrt{v}).$$

Now taking $v = v_n = \sqrt[3]{\left(\frac{\log n}{n}\right)^2} \to 0$ as $n \to \infty$, then we obatin the rate (4.23). □

CHAPTER 5

SPECTRAL DISTRIBUTIONS OF ADJACENCY MARIX OF RANDOM GRAPH ON UNIT SPHERE

5.1 Random Graph on Unit Sphere and its Adjacency Matrix

Let us consider n vertices $X_1, X_2, \cdots, X_n$ on the unit sphere

$$S^{n-1} = \left\{ X = (x_1, x_2, \cdots, x_n) : {x_1}^2 + {x_2}^2 + \cdots + {x_n}^2 = 1 \right\} \subset \mathbb{R}^n,$$

and assume that these vertices are uniformly distributed on S^{n-1}.

Now, we define a Random Geometric Graph on the unit sphere in n-dimensional space and its corresponding adjacent matrix.

Definition 5.1.1. Assume that vertices X_i and X_j are connected if the Euclidean distance $||X_i - X_j|| \leq r_n$ where r_n is a constant which may depend on n. We call this graph as a Random Geometric Graph on the unit sphere in n-dimensional space.

Definition 5.1.2. The matrix $A = (A_{ij})_{1\le i,j\le n}$ is called an adjacency matrix if it is defined by

$$A_{ij} = \begin{cases} 1 & \text{if } ||X_i - X_j|| \le r_n, \\ 0 & \text{if } ||X_i - X_j|| > r_n, \end{cases}$$

where X_i are vertices on the sphere defined above.

Remark 5.1.3. If (X_i)'s are random vectors located on the unit sphere in $\mathbb{R}^n$, then we have

$$||X_i|| = 1, \quad \mathbb{E}[X_i] = 0.$$

Definition 5.1.4. Consider a $n \times n$ random matrix $K = (K_{i,j})_{1\le i,j\le n}$ defined by

$$K_{i,j} = \begin{cases} f(g(X_i, X_j), n) & \text{if } i \ne j, \\ 0 & \text{if } i = j, \end{cases}$$

where $X_1, X_2, \cdots, X_n$ are random vectors in $\mathbb{R}^n$ and $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a real valued function. Then the matrix K is called as a random kernel matrix. Especially, if $g(X_i, X_j) = X_i X_j^T$ then we call the matrix K as a random inner product kernel matrix and if $g(X_i, X_j) = ||X_i - X_j||$ where $||\cdot||$ is a Euclidean measure then we call the matrix K as a random Euclidean kernel matrix.

In [15], El Karoui introduced a random kernel matrix and in [11], Cheng and Singer focused on random inner product kernel matrix. In [13], Yen Do and Van Vu investigated the limiting of spectrum of random kernel matrices for the following cases:

(Case 1) $g(X_i, X_j) = X_i X_j^T$ and f doesn't depend on n.

(Case 2) $g(X_i, X_j) = ||X_i - X_j||$ and f doesn't depend on n.

(Case 3) $g(X_i, X_j) = X_i X_j^T$ and f depends on n.

Let's go back to and focus on adjacency matrices of Random Geometric Graphs in Definition 5.1.1. For the random matrix A in Definintion 5.1.2, the functions g and f can be considered as

$$g(X, Y) = ||X - Y||, \quad f(x, n) = \chi_{[0,r_n]}(x), \tag{5.1}$$

respectively. In this case, f depends on the dimension n and g is an Euclidean kernel function. At first glance, it looks that this random matrix A can not be applied to (Case 1), (Case 2) nor (Case 3). However, since $X_1, X_2, \cdots, X_n$ are on the unit sphere, then we can write

$$||X_i - X_j|| = \sqrt{2 - 2{X_i}^T X_j}. \tag{5.2}$$

In the same way, Definition 5.1.2 can be written as

$$\begin{aligned} A = (A_{ij})_{1\le i,j\le n} &= \begin{cases} 1 & \text{if } {X_i}^T X_j \ge 1 - \frac{{r_n}^2}{2} \\ 0 & \text{if } {X_i}^T X_j < 1 - \frac{{r_n}^2}{2} \end{cases} \\ &= \chi_{[1-\frac{{r_n}^2}{2},\infty)}({X_i}^T X_j, n). \end{aligned} \tag{5.3}$$

So, one can see that the matrix A defined as (5.3) have the functions

$$g(X,Y) = \sqrt{2 - 2{X_i}^T X_j}, \quad f(x,n) = \chi_{[1-\frac{{r_n}^2}{2},\infty)}(x,n), \tag{5.4}$$

and we can say the matrix A is in **(Case 3)**. Also, it is easy to see that the function $f(x,n)$ is differentiable at $x = 0$ and

$$f(2,n) = f'(2,n) = 0. \tag{5.5}$$

In [23], Jiang proved the limiting of ESD of a random inner product kernel matrix defined as:

$$B_{ij} = g(X_i^T X_j, n),$$

where the function g satisfies

$$g''(2,n) \text{ exists and } \quad g'(2,n) \neq 0.$$

Unfortunately, the function f having the condition (5.5) doesn't satisfy the conditions in [23] so that we need more general results. Nevertheless, using Cheng and Singer's idea in [11], we are able to obtain our desired result (See Theorem 5.6.2). In the next section, we introduce relative results proved by them.

5.2 Result of Cheng and Singer

Let us consider random matrices $A = (A_{ij})_{1\le i,j\le n}$ generated from iid random vectors $X_1, X_2, \cdots, X_n$ in $\mathbb{R}^n$ and a real valued function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ as

$$A_{ij} = \begin{cases} f\big(g(X_i, X_j), n\big) & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases} \tag{5.6}$$

where the function $g(X_i, X_j)$ is a kernel function. As mentioned before, we define the ESD of the matrix A by,

$$\rho_A(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i}(x), \tag{5.7}$$

where $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ are eigenvalues of A and the Stieltjes transform of $d\rho_A(x)$ is

$$m_A(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i(A) - z} = \frac{1}{n}\mathrm{Tr}(A - zI)^{-1}, \quad \mathcal{I}(z) > 0.$$

Next, assume that $X_1, X_2, \cdots, X_n$ are iid random vectors in $\mathbb{R}^n$ whose coordinates are independent copies of a random variable Z with the following assumptions:

1. $\mathbb{E}[Z] = 0, \quad \mathbb{E}[Z^2] = \frac{1}{n}$.

2. For all $k > 0$, then there exists a constant $c_k > 0$ such that

$$\mathbb{E}\big[|Z|^k\big] \le c_k n^{-\frac{k}{2}}.$$

By using above assumptions, we define a new random variable.

Definition 5.2.1. We define for each n

$$\epsilon_n = \sqrt{n}X^T Y, \quad n \in \mathbb{N},$$

where X and Y are independent copies of random vector X_1.

Remark 5.2.2. It is easy to see that

$$\mathbb{E}\big[\epsilon_n\big] = \mathbb{E}\big[\sqrt{n}X^TY\big] = \sum_{i=1}^{n}\sqrt{n}\mathbb{E}\big[X_i\big]\mathbb{E}\big[Y_i\big] = 0,$$

and

$$\begin{aligned}\mathbb{E}\big[|\epsilon_n|^2\big] =& n\mathbb{E}\big[|X^TYX^TY|\big]\\ =& n\sum_{i,j=1}^{n}\mathbb{E}\big[|X_iX_j||Y_iY_j|\big]\\ =& n\sum_{i=1}^{n}\mathbb{E}\big[|X_i|^2\big]\mathbb{E}\big[|Y_i|^2\big] = 1.\end{aligned}$$

Now, let us consider the k^{th}-orthogonal polynomial $p_k(x)$ in $L^2(d\mu)$ with respect to a probability measure $\mu(x)$,

$$\int_{\mathbb{R}} p_s(x)p_t(x)d\mu(x) = \begin{cases} 0, & \text{if } \ s \neq t, \\ 1 & \text{if } \ s = t. \end{cases}$$

Since $L^2(d\mu)$ is a Hilbert Space, we have the formal expansion of f in $L^2(d\mu)$:

$$f(x) = \sum_{k=0}^{\infty} a_k p_k(x) \quad \text{where} \quad a_k = \int_{\mathbb{R}} f(x)p_k(x)d\mu(x). \tag{5.8}$$

Let us assume that $f \in L^2(d\mu)$ is a function defined as (5.6) and we define a normalized function

$$k(x,n) = \sqrt{n}f(\frac{x}{\sqrt{n}}, n). \tag{5.9}$$

We can write this $k(x,n)$ as a sum of orthogonal polynomials $p_{k,n}(x)$ with coefficient $a_{k,n}$ as (5.8).

Also, assume that f satisfies the following three conditions: **(Normalization)**, **(Uniform Convergence)**, **(Scaling)** (See Sections 5.4, .5.5, 5.6).

Yen Do and Van Vu [13] showed that one could extend the Chen-Singer result beyond the Gaussian case considered in their papers. For the function f which satisfies above three conditions, Cheng and Singer proved the following results.

Theorem 5.2.3 (Cheng and Singer [11]). *Let A be a random matrix defined as* (5.6) *and assume that the function $k(x,n)$ defined by* (5.9) *satisfies the three conditions,* ***(Normalization)****,* ***(Uniform Convergence)*** *and* ***(Scaling)****. Then the Stieltjes transform of ESD (Empirical Spectral Distribution) of the matrix A, denoted by $m_A(z)$, converges weakly to $m(z)$ which satisfies the following equation,*

$$-\frac{1}{m(z)} = z + a(1 - \frac{1}{1+am(z)}) + (c-a^2)m(z). \tag{5.10}$$

Remark 5.2.4. This limiting spectral distribution defined in (5.10) is no longer MP when $c \neq a^2$. Actually, for our problem, we have $c \neq a^2$ and it follows the semi-circle law.

Remark 5.2.5. **(Uniform Convergence)** holds if the measure is compactly supported. It follows that the convergence of the orthogonal expansion in $L^2(dm)$ holds automatically if X_i's are Gaussian or bounded. In the general case when completeness of the orthogonal polynomials is not guaranteed, **(Uniform Convergence)** has to be checked carefully for both the convergence of the expansion for each n and the uniformity of the convergence over n large.

5.3 Main Result

Let's go back to the ESD of adjacency matrix of a random geometric graph on the unit sphere. In Section 5.1, one can see the adjacent matrix $A = (A_{ij})_{1\leq i,j\leq n}$ of a random geometric graph on the unit shpere as a random matrix with the entries,

$$\begin{aligned}(A_{ij}) &= \begin{cases} 1 & \text{if } {X_i}^T X_j \geq 1 - \frac{{r_n}^2}{2} \\ 0 & \text{if } {X_i}^T X_j < 1 - \frac{{r_n}^2}{2} \end{cases} \\ &= \chi_{[1-\frac{{r_n}^2}{2},\infty)}({X_i}^T X_j, n), \qquad n \in \mathbb{N}. \end{aligned} \tag{5.11}$$

where $X_i's$ are random vectors with $\|X_i\| = 1,\ i = 1, 2, \cdots, n$. According to Section 5.2, we define the normalized function $k(x, n)$ by

$$k(x,n) = \sqrt{n}\chi_{[\sqrt{n}(1-\frac{{r_n}^2}{2}),\infty)}(x,n) = \begin{cases} \sqrt{n} & \text{if } x \geq (1 - \frac{{r_n}^2}{2})\sqrt{n}, \\ 0 & \text{if } x < (1 - \frac{{r_n}^2}{2})\sqrt{n}, \end{cases}$$

where r_n is a sequence with $r_n \to \sqrt{2}$.

To apply for **(Case 3)** and get the result as Theorem 5.2.3 with the random matrix A in Section 5.1, it is suffices to prove that the function $k(x, n)$ satisfies the three conditions, **(Normalization)**, **(Uniform Convergence)** and **(Scaling)** in Section 5.2.

To do so, we need to do several steps. First, from the volume of the sphere, we get the explicit formula of the probability density function of $\epsilon_n = \sqrt{n}{X_i}^T X_j$ defined in Definition 5.2.2.

Lemma 5.3.1. *Let us consider a random variable*

$$\epsilon_n = \sqrt{n}{X_i}^T X_j, \quad n \in \mathbb{N}. \tag{5.12}$$

Then for any $n \in \mathbb{N}$, the probability density function of ϵ_n, $q_n(x)$, is

$$q_n(x) = \begin{cases} \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \left(1 - \frac{x^2}{n}\right)^{\frac{n-3}{2}}, & \text{if } -\sqrt{n} \leq x \leq \sqrt{n} \\ 0 & , \text{otherwise.} \end{cases} \tag{5.13}$$

Proof. Since X_i, X_j are uniformly distributed on the unit ball S^{n-1}, the inner product $X_i^T X_j$ is related to the volume of the ball. Let us consider the subset of sphere S^{n-1},

$$S_x^{n-1} = \left\{X = (x_1, x_2, \cdots, x_n) \in S^{n-1} : x_n \leq x\right\}, \quad n \in \mathbb{N}.$$

Using volume of sphere $S^{n-1}(r) \subset \mathbb{R}^n$ with radius r

$$Vol\Big(S^{n-1}(r)\Big) = \frac{2\sqrt{\pi^n}}{\Gamma(\frac{n}{2})} r^n,$$

we obtain the volume of the subset S_x^{n-1} for $-1 \le x \le 1$:

$$\begin{aligned} Vol(S_x^{n-1}) &= \int_{-1}^{x} Vol\Big(S^{n-2}(\sqrt{1-u^2})\Big)\frac{1}{\sqrt{1-u^2}}du \\ &= \int_{-1}^{x} \frac{2\pi^{\frac{n-1}{2}}}{\Gamma\Big(\frac{n-1}{2}\Big)}\Big(1-u^2\Big)^{\frac{n-3}{2}}du. \end{aligned}$$

If $x \ge 1$ or $x \le -1$ then $Vol(S_x^{n-1}) = 0$ so that $q_n(x) = 0$.

Next, let $X_i, X_j \in S^{n-1}$. After rotation, we may assume that

$$X_j = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \quad X_i = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

This implies that

$$X_i^T X_j = u_n.$$

Since $\epsilon_n = \sqrt{n} X_i^T X_j$, then it follows from the volume of the sphere that for any $n \in \mathbb{N}$,

$$\begin{aligned} q_n(x) =& \frac{d}{dx}\Pr\Big(\sqrt{n}X_i^T X_j \le x\Big) \\ =& \frac{d}{dx}\Pr\Big(\sqrt{n}u_n \le x\Big) \\ =& \frac{d}{dx}\frac{Vol\Big(S^{n-1}_{\frac{x}{\sqrt{n}}}\Big)}{Vol(S^{n-1})} \\ =& \frac{1}{\sqrt{n\pi}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\Big(1-\frac{x^2}{n}\Big)^{\frac{n-3}{2}}. \end{aligned}$$

which completes proof. □

From Lemma 5.3.1, we have the following Proposition.

Proposition 5.3.2. *Let $q_n(x)$ be defined as* (5.13)*. Then for any $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} q_n(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Proof. Using the formula of gamma function:

$$\lim_{n\to\infty}\frac{\Gamma(n+\alpha)}{\Gamma(n)n^{\alpha}}=1,\quad \alpha\in\mathbb{C},$$

we get

$$\lim_{n\to\infty}\frac{1}{\sqrt{n\pi}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}=\frac{1}{\sqrt{2\pi}}. \tag{5.14}$$

Also, it is easy to see that

$$\lim_{n\to\infty}(1-\frac{x^2}{n})^{\frac{n-3}{2}}=\lim_{n\to\infty}\Big((1-\frac{x^2}{n})^{-\frac{n}{x^2}}\Big)^{-\frac{x^2}{2}}\lim_{n\to\infty}(1-\frac{x^2}{n})^{-\frac{3}{2}}=e^{-\frac{x^2}{2}}. \tag{5.15}$$

Thus, it follows from (5.14) and (5.15) that for all $x\in\mathbb{R}^n$,

$$\lim_{n\to\infty}q_n(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

□

Let us define a set of orthonormal polynomials of $q_n(x)$ by $\{\ P_{l,n}(x):\ l=1,2,\cdots\ \}$:

$$\int_{\mathbb{R}}P_{i,n}(x)P_{j,n}(x)q_n(x)dx=\delta_{ij},\quad n\in\mathbb{N}, \tag{5.16}$$

where δ_{ij} is 1 when $i=j$ and 0 otherwise. Let us define coefficients of the normalized function $k(x,n)$ by

$$a_{l,n}=\int_{\mathbb{R}}k(x,n)P_{l,n}(x)q_n(x)dx.$$

Proposition 5.3.3. *Put the sequence r_n as*

$$r_n=\sqrt{2\Big(1-\sqrt{\frac{2\log n-\log(\log n)}{n}}\Big)},$$

then we have

$$\int_{\mathbb{R}}k(x,n)^2e^{-\frac{1}{2}x^2}dx\to\frac{1}{2},\quad n\to\infty.$$

Proof. Choose the sequence r_n as

$$r_n = \sqrt{2\Big(1 - \sqrt{\frac{2\log n - \log(\log n)}{n}}\Big)},$$

which comes from the idea of W-Lambert function (See [12] in details). Notice that

$$\Big(1 - \frac{r_n^2}{2}\Big)\sqrt{n} = \sqrt{2\log n - \log(\log n)} \to \infty, \quad n \to \infty.$$

Also, using the sequence r_n, we have

$$-\frac{1}{2}\Big(1 - \frac{r_n^2}{2}\Big)^2 n = -\log n + \frac{1}{2}\log(\log n).$$

It implies that

$$\begin{aligned}\int_{\mathbb{R}} k(x,n)^2 e^{-\frac{1}{2}x^2} dx &= \int_{\left(1-\frac{r_n^2}{2}\right)\sqrt{n}}^{\infty} n e^{-\frac{1}{2}x^2} dx \\ &= n\frac{\sqrt{\pi}}{2}\Big[\frac{e^{-\frac{1}{2}\left(1-\frac{r_n^2}{2}\right)^2 n}}{\left(1-\frac{r_n^2}{2}\right)\sqrt{n}} + R_1(\big(1-\frac{r_n^2}{2}\big)\sqrt{n})\Big]\end{aligned}$$

where $R_1(\big(1-\frac{r_n^2}{2}\big)\sqrt{n}) = \frac{-1}{\sqrt{\pi}}\int_{\left(1-\frac{r_n^2}{2}\right)\sqrt{n}}^{\infty} t^{-2}e^{-t^2}\, dt$. Since

$$R_1(\big(1-\frac{r_n^2}{2}\big)\sqrt{n}) = \mathcal{O}\Big(\big(1-\frac{r_n^2}{2}\big)\sqrt{n}^{\,-1} e^{-n\left(1-\frac{r_n^2}{2}\right)^2}\Big),$$

we have $R_1(\big(1-\frac{r_n^2}{2}\big)\sqrt{n}) \to 0$ as $n \to \infty$. This implies that

$$\int_{\mathbb{R}} k(x,n)^2 e^{-\frac{1}{2}x^2} dx \to \frac{1}{2}, \quad n \to \infty.$$

□

As a corollary of the above Proposition 5.3.3, we can compute limiting of the following integral.

Corollary 5.3.4. *Put the sequence* r_n *as*

$$r_n = \sqrt{2\Big(1 - \sqrt{\frac{2\log n - \log(\log n)}{n}}\Big)},$$

then we have

$$\int_{\mathbb{R}} k(x,n)^2\ q_n(x) dx \to \frac{1}{2\sqrt{2\pi}}, \quad n \to \infty,$$

where $q_n(x)$ *is defined as* (5.13).

Proof. Since the fact that

$$\lim_{n\to\infty} q_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

there exists $N > 0$ such that for any $n > N$, $\left| q_n(x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right| < \frac{1}{n^2}$. From this,

$$\begin{aligned}
&\int_{\mathbb{R}} k(x,n)^2 \; q_n(x)dx - \int_{\mathbb{R}} k(x,n)^2 \; \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\
&\leq \int_{\left(1-\frac{r_n^2}{2}\right)\sqrt{n}}^{\sqrt{n}} nq_n(x)dx - \int_{\left(1-\frac{r_n^2}{2}\right)\sqrt{n}}^{\sqrt{n}} ne^{-\frac{1}{2}x^2} dx - \int_{\sqrt{n}}^{\infty} ne^{-\frac{1}{2}x^2} dx \\
&\leq \int_{\left(1-\frac{r_n^2}{2}\right)\sqrt{n}}^{\sqrt{n}} n \left| q_n(x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right| dx \\
&\leq \int_{\left(1-\frac{r_n^2}{2}\right)\sqrt{n}}^{\sqrt{n}} n \; \frac{1}{n^2} dx \\
&= \frac{\sqrt{n}}{n} - \left(1 - \frac{r_n^2}{2}\right) \frac{\sqrt{n}}{n}
\end{aligned}$$

for any $n > N$. Since $\lim_{n\to\infty} \frac{\sqrt{n}}{n} - \left(1 - \frac{r_n^2}{2}\right) \frac{\sqrt{n}}{n} = 0$, this completes the proof. □

From this fact, the constant c in the Theorem 5.2.3 is $c = \frac{1}{2\sqrt{2\pi}}$.

Next, we consider $\{h_l(x) : l = 1, 2, \cdots\}$ called normalized Hermite polynomials as an orthonormal polynomials of standard normal distribution,

$$\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Define the coefficients $b_{l,n}$ of the normalized function $k(x,n)$ by

$$b_{l,n} = \int_{\mathbb{R}} k(x,n) h_l(x) \mu(x) dx. \tag{5.17}$$

Lemma 5.3.5. *For each $n \in \mathbb{N}$, the function $k(x,n)$ belongs to $L^2(\mu(x), \mathbb{R}^n)$ (See [11]). In particular, for each $n \in \mathbb{N}$,*

$$\sum_{l=1}^{\infty} b_{l,n}^2 = \int_{\mathbb{R}} k(x,n)^2 \mu(x) dx < \infty.$$

Proof. It is obvious from Proposition 5.3.3 that

$$\begin{aligned}\int_{\mathbb{R}} k(x,n)^2\mu(x)dx =& \frac{1}{\sqrt{2\pi}}\int_{x>(1-r_n)\sqrt{n}} n\mu(x)dx \\ \leq& \frac{1}{\sqrt{2\pi}}\frac{\sqrt{\log n}}{\sqrt{2\log n - \log(\log n)}} < \infty.\end{aligned}$$

□

Lemma 5.3.6. *[See Lemma C.1 in [11]]*
Suppose that

$$\int_{\mathbb{R}} k(x,n)^2|q_n(x) - \mu(x)|dx \to 0, \qquad n \to \infty, \tag{5.18}$$

then, for each $l = 1, 2, \cdots,$ *we have*

$$|b_{l,n} - a_{l,n}| \to 0, \quad n \to \infty. \tag{5.19}$$

Lemma 5.3.6 shows that the for each l, two coefficients $a_{l,n}$ and $b_{l,n}$ are close to each other when $n \to \infty$ and enables us to check that the function $k(x,n)$ satisfies the conditions with the coefficients $b_{l,n}$ instead of $a_{l,n}$. Actually, the coefficients $b_{l,n}$ come from Hermite polynomials and standard normal distribution which are very well-known functions and we are able to deal with much better features of the functions than the polynomials $P_{l,n}(x)$ and distribution $q_n(x)$. To use Lemma 5.3.6, we need to show (5.18) with our normalized function $k(x,n)$. Indeed, it is obvious from Proposition 5.3.2 that

$$\int_{\mathbb{R}} k(x,n)^2|q_n(x) - \mu(x)|dx \to 0, \qquad n \to \infty.$$

Now, we show the function $k(k,n)$ satisfies the three conditions in the following three sections.

5.4 Normalization

(Normalization) There exists $c \in [0,\infty)$ such that

$$\lim_{n\to\infty}\sum_{i=1}^{\infty}|a_{i,n}|^2 = c.$$

From Lemma 5.3.5 and 5.3.6, we have the following proposition.

Proposition 5.4.1. *Let* $b_{l,n}$ *be a coefficient defined as* (5.17). *Then there exists* $c \in [0, \infty)$ *such that*

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} |b_{i,n}|^2 = c.$$

Proof. Suppose that for sufficiently large n,

$$\sum_{l=1}^{\infty} b_{l,n}^2 < \infty.$$

Then for any finite $L > 0$,

$$\sum_{l=L+1}^{\infty} b_{l,n}^2 \to \sum_{l=L+1}^{\infty} a_{l,n}^2, \quad n \to \infty.$$

Since

$$\sum_{l=1}^{\infty} a_{l,n}^2 = \int_{\mathbb{R}} k(x,n)^2 q_n(x)dx,$$

and

$$\sum_{l=1}^{\infty} b_{l,n}^2 = \int_{\mathbb{R}} k(x,n)^2 \mu(x)dx,$$

then it follows from Lemma 5.3.5 and Lemma 5.3.6 that for any finite $L > 0$,

$$\begin{aligned} \sum_{l=L+1}^{\infty} b_{l,n}^2 &= \sum_{n=1}^{\infty} b_{l,n}^2 - \sum_{l=1}^{L} b_{l,n}^2 \\ &\to \sum_{n=1}^{\infty} a_{l,n}^2 - \sum_{l=1}^{L} a_{l,n}^2 = \sum_{l=L+1}^{\infty} a_{l,n}^2 \quad n \to \infty. \end{aligned}$$

Now it suffices to show that

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} b_{i,n}^2 < \infty. \tag{5.20}$$

But (5.20) can be proved immediately from Lemma 5.3.5 that

$$c = \lim_{n\to\infty} \sum_{i=0}^{\infty} b_{i,n}^2 \le \frac{1}{4\pi}.$$

□

5.5 Uniform Convergence

(Uniform Convergence) For any $\epsilon > 0$, there exists $L = L(\epsilon) > 0$ such that the following holds for enoughly large n:

$$\sum_{i>L} |a_{i,n}|^2 \leq \epsilon.$$

Proposition 5.5.1. *For any $\epsilon > 0$, there exist $N = N(\epsilon) > 0$ and $L = L(\epsilon) > 0$ such that for all $n \geq N$,*

$$\sum_{l=L}^{\infty} b_{l,n}^2 < \epsilon.$$

Proof. Let $H_l(x)$ be a hermitian polynomial with respect to $\mathcal{N}(0,1)$ and define the function $h(x)$ by

$$h_l(x) := \frac{1}{\sqrt{l!}} H_l(x), \quad l = 1, 2, 3 \cdots .$$

By using the identity

$$H_l(x) = xH_{l-1}(x) - (l-1)H_{l-2}(x),$$

we get

$$\begin{aligned} & x^2 H_{l-1}^2(x) \\ = & \Big(H_l(x) + (l-1)H_{l-2}(x) \Big)^2 \\ = & H_l^2(x) + 2(l-1)H_l(x)H_{l-2}(x) + (l-1)^2 H_{l-2}^2(x). \end{aligned}$$

In [8], Bonan and Clark proved that

$$\max_{x \in \mathbb{R}} h_l^2(x) e^{-\frac{x^2}{2}} \leq \frac{d}{2^l} \frac{1}{l^{\frac{1}{6}}}, \tag{5.21}$$

for some constant $d > 0$ which doesn't depend on l. It follows from (5.21) that

$$\begin{aligned} & \max_{x \in \mathbb{R}} x^2 h_{l-1}^2(x) e^{-\frac{x^2}{2}} \\ \leq & \frac{l!}{(l-1)!} \max_{x \in \mathbb{R}} h_l^2(x) e^{-\frac{x^2}{2}} + \frac{2(l-1)\sqrt{l!}\sqrt{(l-2)!}}{(l-1)!} \max_{x \in \mathbb{R}} h_l(x) h_{l-2}(x) e^{-\frac{x^2}{2}} \\ & + \frac{(l-1)^2 (l-2)!}{(l-1)!} \max_{x \in \mathbb{R}} h_{l-2}^2(x) e^{-\frac{x^2}{2}} \\ \leq & C_1 \frac{l!}{(l-1)!} \frac{d}{2^l} \frac{1}{l^{\frac{1}{6}}}, \end{aligned} \tag{5.22}$$

where $C_1 > 0$ does not depend on l. Also, one can see that for a sufficiently large n and r_n defined in Proposition 5.3.3,

$$\begin{aligned}\frac{1}{(1-\frac{r_n^2}{2})^2}e^{-\frac{(1-\frac{r_n^2}{2})^2 n}{2}} =& \frac{n}{2\log n - \log\log n}e^{-\log n + \frac{1}{2}\log(\log n)} \\ =& \frac{\sqrt{\log n}}{\log n - \log(\log n)} \leq 1. \end{aligned} \tag{5.23}$$

It follows from (5.22) and (5.23) that for $\eta_n = (1 - \frac{r_n^2}{2})\sqrt{n}$,

$$\begin{aligned} b_{l,n}^2 =& \Big(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} k(t,n) h_l(x) e^{-\frac{x^2}{2}} dt \Big)^2 \\ =& \Big(\int_{\eta_n}^{\infty} \frac{1}{\sqrt{2\pi}} \sqrt{n} h_l(x) e^{-\frac{x^2}{2}} dt \Big)^2 \\ =& \Big(\frac{1}{\sqrt{2\pi}} \frac{\sqrt{(l-1)!}}{\sqrt{l!}} \sqrt{n} h_{l-1}(\eta_n) e^{-\frac{\eta_n^2}{2}} \Big)^2 \\ =& \frac{1}{2\pi} \frac{1}{(1-\frac{r_n^2}{2})^2} e^{-\frac{\eta_n^2}{2}} \frac{(l-1)!}{l!} \eta_n^2 h_{l-1}^2(\eta_n) e^{-\frac{\eta_n^2}{2}} \leq C \frac{1}{2^l} \frac{1}{l^{\frac{1}{6}}}, \end{aligned}$$

where $C > 0$ does not depend on l. Thus, for all $\epsilon > 0$, there exist $N = N(\epsilon)$, and $L = L(\epsilon)$ which do not depend on n such that for all $n \geq \mathbb{N}$,

$$\sum_{l=L}^{\infty} b_{l,n}^2 = C \sum_{l=L}^{\infty} \frac{1}{2^l} \frac{1}{l^{\frac{1}{6}}} < \epsilon.$$

□

5.6 Scaling

(Scaling) There exists $a \in [0, \infty)$ such that

$$\lim_{n\to\infty} a_{1,n} = a.$$

Proposition 5.6.1.

$$\lim_{n\to\infty} b_{1,n} = 0.$$

Proof. Note that

$$\begin{aligned} b_{1,n} =& \int_{\mathbb{R}} k(x,n) x\mu(x) dx \\ =& \frac{1}{2\pi}\sqrt{n}\int_{(1-\frac{r_n^2}{2})\sqrt{n}} xe^{-\frac{1}{2}x^2}dx \\ =& \frac{1}{2\pi}\sqrt{n}\int_{\frac{1}{2}(1-\frac{r_n^2}{2})^2 n} e^{-t}dt \\ =& \frac{1}{2\pi}\sqrt{n}(-e^{-t})\Big]_{\frac{1}{2}(1-\frac{r_n^2}{2})^2 n}^{\infty} \\ =& \frac{1}{2\pi}\sqrt{n}e^{\frac{1}{2}(1-\frac{r_n^2}{2})^2 n} = \frac{\sqrt{\log n}}{\sqrt{n}} \to 0, \quad n \to \infty. \end{aligned}$$

By Lemma 5.18, $a_{1,n} \to 0$ as $n \to \infty$. □

From the above Lemma, we have the constant $a = 0$ in Theorem 5.2.3. Considering as in Theorem 5.2.3 ($c = \frac{1}{2\sqrt{2\pi}}$, $a = 0$), we have the following Theorem immediately.

Theorem 5.6.2. *Consider a random matrix A defined as (5.11) with $\rho_A(x)$, the ESD of A, defined as (5.7). Then for any $z \in \mathbb{C}$, the stieltjes transform $m_A(z)$ of $\rho_A(x)$, converges weakly to*

$$m(z) = \frac{-z + \sqrt{z^2 - 4\frac{1}{2\sqrt{2\pi}}}}{2},$$

as $n \to \infty$. Moreover,

$$\rho_A(x) \to \frac{1}{2\pi}\sqrt{\frac{2}{\sqrt{2\pi}} - x^2} \quad \textit{almost surely as } n \to \infty.$$

REFERENCES

[1] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.*, 21(2):625–648, 1993.

[2] Z. D. Bai, Baiqi Miao, and Jhishen Tsay. A note on the convergence rate of the spectral distributions of large random matrices. *Statist. Probab. Lett.*, 34(1):95–101, 1997.

[3] Z. D. Bai and Jack W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004.

[4] Z. D. Bai and Y. Q. Yin. Convergence to the semicircle law. *Ann. Probab.*, 16(2):863–875, 1988.

[5] Zhidong Bai, Jiang Hu, Guangming Pan, and Wang Zhou. A note on rate of convergence in probability to semicircular law. *Electron. J. Probab.*, 16:no. 88, 2439–2451, 2011.

[6] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer Series in Statistics. Springer, New York, second edition, 2010.

[7] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[8] Stanford S. Bonan and Dean S. Clark. Estimates of the orthogonal polynomials with weight $\exp(-x^m)$, m an even positive integer. *J. Approx. Theory*, 46(4):408–410, 1986.

[9] D. L. Burkholder. Distribution function inequalities for martingales. *Ann. Probability*, 1:19–42, 1973.

[10] W. Cauer. The Poisson integral for functions with positive real part. *Bull. Amer. Math. Soc.*, 38(10):713–717, 1932.

[11] Xiuyuan Cheng and Amit Singer. The spectrum of random inner-product kernel matrices. *Random Matrices Theory Appl.*, 2(4):1350010, 47, 2013.

[12] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Adv. Comput. Math.*, 5(4):329–359, 1996.

[13] Yen Do and Van Vu. The spectrum of random kernel matrices: universality results for rough and varying kernels. *Random Matrices Theory Appl.*, 2(3):1350005, 29, 2013.

[14] Laure Dumaz and Bálint Virág. The right tail exponent of the Tracy-Widom β distribution. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(4):915–933, 2013.

[15] Noureddine El Karoui. The spectrum of kernel random matrices. *Ann. Statist.*, 38(1):1–50, 2010.

[16] P. Erd″ os and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[17] I. Farkas, I. Derényi, H. Jeong, Z. Néda, Z. N. Oltvai, E. Ravasz, A. Schubert, A.-L. Barabási, and T. Vicsek. Networks in life: scaling properties and eigenvalue spectra. *Phys. A*, 314(1-4):25–34, 2002. Horizons in complex systems (Messina, 2001).

[18] F. Götze and A. Tikhomirov. On the Rate of Convergence to the Marchenko–Pastur Distribution. *ArXiv e-prints*, October 2011.

[19] F. Götze and A. N. Tikhomirov. The rate of convergence of spectra of sample covariance matrices. *Teor. Veroyatn. Primen.*, 54(1):196–206, 2009.

[20] Friedrich Götze and Alexandre Tikhomirov. On the rate of convergence to the semi-circular law. In *High dimensional probability VI*, volume 66 of *Progr. Probab.*, pages 139–165. Birkhäuser/Springer, Basel, 2013.

[21] Ulf Grenander and Jack W. Silverstein. Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.*, 32(2):499–519, 1977.

[22] H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Rev.*, 23(1):53–60, 1981.

[23] Tiefeng Jiang. Distributions of eigenvalues of large Euclidean matrices generated from l_p balls and spheres. *Linear Algebra Appl.*, 473:14–36, 2015.

[24] Iain M. Johnstone. Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.*, 36(6):2638–2716, 2008.

[25] Dag Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12(1):1–38, 1982.

[26] V. A. Marčenko and L. A. Pastur. The spectrum of random matrices. *Teor. Funkciĭ Funkcional. Anal. i Priložen. Vyp.*, 4:122–145, 1967.

[27] V A Marenko and L A Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

[28] Amy Nyberg, Thilo Gross, and Kevin E. Bassler. Mesoscopic structures and the Laplacian spectra of random geometric graphs. *J. Complex Netw.*, 3(4):543–551, 2015.

[29] Sean O'Rourke. A note on the Marchenko-Pastur law for a class of random matrices with dependent entries. *Electron. Commun. Probab.*, 17:no. 28, 13, 2012.

[30] Bernhard Scholkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[31] John H. Schwarz. The generalized Stieltjes transform and its inverse. *J. Math. Phys.*, 46(1):013501, 8, 2005.

[32] Jack W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995.

[33] Alexander Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6):1033–1056, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[34] Kenneth W. Wachter. The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probability*, 6(1):1–18, 1978.

[35] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of /'small-world/' networks. *Nature*, 393(6684):440–442, June 1998.

[36] Eugene P. Wigner. On a class of analytic functions from the quantum theory of collisions. *Annals of Mathematics*, 53(1):36–67, 1951.

[37] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955.

[38] Eugene P. Wigner. Lower limit for the energy derivative of the scattering phase shift. *Phys. Rev. (2)*, 98:145–147, 1955.

[39] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

[40] Eugene P. Wigner. Random matrices in physics. *SIAM Review*, 9(1):1–23, 1967.

[41] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):32–52, 1928.

[42] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20(1):50–68, 1986.