SPECTRAL ANALYSIS OF LARGE DIMENSIONAL RANDOM MATRICES

ZHANG LIXIN

(B.S., JILIN UNIV.)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2006

Acknowledgements

It is a great pleasure to say thanks to those people who have made this thesis possible.

Above all, I would like to thank my supervisor Prof. Bai ZhiDong for his supervision of my doctoral research work. I wish to express my sincere appreciation to him for his enthusiasm and devotion in teaching me scientific knowledge and for his passion and patience in giving me help at all time. I would like to thank the Department of Statistics and Applied Probability of the National University of Singapore for providing scholarships to pursue research studies and opportunities to learn with outstanding academicians. During my graduate studies here, I have successively learned valuable courses on statistics from, besides my supervisor, Prof. Chan Hock Peng, Prof. Chen SongXi, Prof. Chen ZeHua, Prof. Gan Fah Fatt, Prof. Wang YouGan. I am grateful to them.

I am grateful to the many teachers who have given me dedicated instructions during my long lasting student life. I am especially grateful to my former supervisor Prof. Shi NingZhong for his encouragement and care which are especially specious in the earlier time after I changed my major from mathematics to statistics. I am grateful to my undergraduate teachers who have taught me mathematics in JiLin University of China: Prof. Hu ZongCai, Prof. Jiang ChunLan, Prof. Ji YouQing, Prof. Yan ZiQian, Prof. Yuan YongJiu, Prof. Zhou QinDe. I would like to thank my high school teachers, my junior high school teachers and my primary school teachers. Although I am not able to present all their names in list here, their instructions are very much appreciated.

Lastly and most importantly, I wish to express my forever love, respect and gratitude to my parents, Zhang Jin and He WenYan. They brought up me with unselfish devotion, gave me all-out support to receive education and granted me specious assistance through difficult times. To them, I dedicate this thesis.

Contents

Summary

The thesis is concerned with finding the limiting spectral distributions of three classes of large dimensional random matrices.

The first class of matrices we considered are large dimensional Wigner type random matrices taking the form $A_n = (1/\sqrt{n})W_n T_n$, where W_n is a classical Wigner matrix and T_n is a nonnegative definite matrix. By using the Stieltjes transform method, we prove the convergence of the empirical spectral distributions of the Wigner type matrices, derive some analytical properties possessed by the limiting spectral distribution, and present calculation of the density function when the matrix T_n has some given forms. We also present a moment method to prove the existence of the limiting spectral distribution with explicit form of the limiting moments.

The second class of matrices we considered is a general form of large dimensional sample covariance matrices having the form $B_n = (1/N)T_{2n}^{1/2} X_n T_{1n} X_n^* T_{2n}^{1/2}$, where T_{2n} is nonnegative definite and T_{1n} is Hermitian. Existing work on this class of matrices is confined to the special cases where T_{1n} is an identity matrix or T_{1n} and T_{2n} are both diagonal matrices. The class of matrices have important applications in many fields and so systematic investigations of their spectral properties are valuable. In view of the important role played by the Stieltjes transform method in the spectral analysis of random matrices, we investigated a way to manipulate the Stieltjes transform method on the class of general sample covariance matrices so

that systematic investigations of the spectral properties of this class of matrices can be carried out with the aid of this powerful method. In the thesis, we accomplished in proving the empirical spectral distributions of the general sample covariance matrices converge weakly to a non-random limiting spectral distribution whose Stieltjes transform is uniquely determined by a system of equations.

The third class of matrices we considered are large dimensional sparse random matrices taking the form of the Hadamard products of a normalized sample covariance matrix and a sparsing matrix. We prove the empirical spectral distributions of this class of matrices converge weakly to the semicircle law. This result is consistent with other findings in the field. Our main achievement is, by imposing suitable conditions on the moments of the entries in the sparsing matrix instead of letting them be just independent and identically distributed Bernoulli trials, we present a new sparseness scheme of the matrices so that the sparsing factors may not be of zero-one form nor homogeneous. We establish our proof by means of the moment method. Based on our finding, we conjecture the result can be generalized to consider the Hadamard products of a normalized sample covariance matrices with some statistical correlation assumed and a sparsing matrix.

In summary, this thesis presents a collection of theoretical results which provide fundamental solutions to finding the limiting spectral distribution for three important classes of random matrices and furnish elementary material for future development of the spectral analysis of these three classes of matrices.

List of Figures

Chapter 1

Introduction

The present thesis is devoted to limit theorems on eigenvalues of large dimensional random matrices. The subject is widely known as random matrix theory, which is concerned with statistical analysis of asymptotic properties of eigenvalues and eigenvectors for high-dimensional random matrices. In the recent decades, random matrix theory has attracted considerable interest in a variety of areas, due to the high emergence rate of high-dimensional data in modern technological developments and the rich mathematical essence contained in the theory.

Literature Review on Random Matrix Theory

The very beginning of random matrix theory dates back to the momentous work of Wishart in 1928 (Wishart (1928)), which motivated the formation of multivariate statistical analysis. Wishart derived, for independent and identically distributed n-dimensional normal (or, Gaussian) random vectors x_1, x_2, $\cdots$, x_N, the precise expression of the joint probability density function of the random matrix $S = (1/N)\sum_{i=1}^{N} x_i x_i^*$. For the Wishart matrix, the joint probability density function of its ordered and unordered strictly positive eigenvalues as well as the density function of its kth largest eigenvalue for any integer k were later found (Fisher (1939), Hsu (1939), Girshick (1939), Roy (1939), Khatri (1964,1969) and Gao and Smith (2000)). These results play a significant role in not only multivariate statistical analysis but also in applied areas like information theory, communications engineering and many branches of physics.

Spectral analysis of large dimensional random matrices was initiated in the area of nucleus physics by Wigner in the 1950's. At that time, theoretical analysis of low-lying excited states of complex nuclei achieved great success, but the same analytical methods were not applicable for analyzing the highly excited states. The reason was because the base of the methods, level assignments, cannot be carried forward to the case when the order of magnitude of the levels becomes very high. Indeed, in view of the considerable complexity of the systems, a reasonable alternative way is to use statistical mechanics. In the searching for a suitable statistical mechanics, Wigner initially considered statistical distributions for energy levels of complex nuclei (Wigner (1951)) and later produced the idea of using large random matrices to model statistical properties of the energy levels (Wigner (1955)). In fact, system Hamiltonians can be reasonably represented by Hermitian matrices and so naturally it was expected energy levels of complex nuclei, viewed as a complex quantum system, can be described by eigenvalues of the matrices. There is the underlying philosophy, explained clearly by Dyson in his famous work (Dyson

(1962)) and well accepted in the field, that when physical systems are sufficiently complex, their detailed structure can be renounced in which case statistical theory describing their generic behavior can be used. In case of the complex nuclei, a renouncement of their detailed structure means admitting, provided that there is a large number of particles in a complex nucleus interacting with each other according to unknown laws, all possible laws of interaction are equally probable. Therefore the prescribed Hermitian matrices should be, in statistical alphabet, the sample space of a Hermitian random matrix.

The random matrix Wigner investigated is $n \times n$ real symmetric matrix $W_n = [w_{ij}]$ whose entries on and above the diagonal are independent Gaussian random variables with mean 0 and variance σ^2 for non-diagonal entries and $2\sigma^2$ for the diagonal ones. In physics, it is referred to as the Gaussian ensemble, in which case its sample space is nominated. Note that since complex nuclei are very complicated, the dimension n of the matrix W_n is very large. So in using W_n, or any other random matrix suitably defined, to model complex nuclei, limiting statistical behavior as $n \to \infty$ of the eigenvalues of W_n are considered appropriate for describing generic properties of the energy levels of the nuclei. For the matrix W_n, Wigner proved as $n \to \infty$ the expected empirical spectral distribution (ESD) of $W_n/\sqrt{n}$ converges weakly to the semicircle law whose density function is given by

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

where for any $n \times n$ matrix A_n having real eigenvalues only, the empirical spectral distribution (ESD), denoted by $F^{A_n}(x)$, of A_n refers to the empirical distribution of its eigenvalues, *i.e.*

$$F^{A_n}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(\lambda_i(A_n)\le x)},$$

with $\lambda_i(A_n)$, $i = 1, \cdots, n$ denoting the eigenvalues of A_n. The semicircle law $F_{sc,\sigma^2}(x)$ is commonly referred to as the limiting spectral distribution (LSD) of W_n. This convention of course applies to any prescribed matrix A_n.

Note that the LSDs describe distributions of the eigenvalues of random matrices over their whole spectrum domains and so are said in the literature of physics to be global spectral distributions of eigenvalues of random matrices. When a random matrix is a full characterization of a real system, such as the sample covariance matrices for channels in wireless communications, global spectral distribution contains a great deal of information for understanding statistical properties of the system. However, in physics, due to limitations imposed by the complexity of the problems, random matrices can only be viewed as gross mutilations of real systems (Dyson (1962) p.141). As a consequence, the so-called local spectral statistics provide more reliable results for physical problems. Classical problems of random matrix theory in physics concern partition function of the eigenvalues, distribution function of spacing between nearest-neighbor eigenvalues and correlation function of k eigenvalues for any positive integer k. For Wigner's Gaussian ensemble, these problems as well as the joint density function of the eigenvalues were settled in Thomas and Porter (1956) and Gurevich and Pevsner (1957), Rosenzweig and Porter (1960), Mehta (1960), Mehta and Gaudin (1960) and Guadin (1961).

The achievement of Wigner and his colleagues convinced theoretical physicists that although a statistical mechanics of random matrices is mathematically demanding, it is indeed a solvable model for theoretical analysis of nucleus physics. However, except the finding that the semicircle law does not show any similarity to observed spectra of a real nucleus (Wigner (1967)), it was also noticed that the definition of Wigner's Gaussian ensemble has some arbitrary segment which is not expected to be present in a real physical system (Rosenzweig and Porter (1960), Dyson (1962), Bronk (1964)). Motivated by this weakness of Gaussian ensemble and also the success achieved on random matrix based statistical mechanics, Dyson contributed his very influential work in 1962 (Dyson(1962)).

Besides clarifying the underlying philosophy of random matrix theory in physics, Dyson introduced three new ensembles of matrices which turned out to be the most important component of the theory today. They are well known as the Gaussian orthogonal ensemble (GOE), the Gaussian unitary ensemble (GUE) and the Gaussian symplectic ensemble (GSE). Although Dyson started from the same point as Wigner, that is, an ensemble of matrices represent an ensemble of systems, the connection to the systems are not the same. Since what is needed is that the eigenvalues of the matrices are distributed equally as the energy levels, Dyson straightforwardly assumed that, for the GOE case for example, there is an $N \times N$ unitary matrix S with eigenvalues $[\exp(i\theta_j)]$, $j = 1, \cdots, N$, distributed around the unit circle, with which his basic statistical hypothesis is just "the behavior of n consecutive levels of an actual system, where n small compared with the total number of levels, is statistically equivalent to the behavior in the ensemble E_1 of n consecutive angles θ_j on the unit circle, where n is small compared with N"

(Dyson (1962) p.141). Here E_1 just stands for the GOE. The connection of the matrix S to the system, usually represented by its Hamiltonian, was left vague, but an important point is the matrix type represents the system symmetries. For example, the GOE represents a system invariance property under space rotations or under time reversal and even spin. The GUE and GSE then respectively represent systems having odd spin, invariance property under time reversal, but no rotational symmetry, and systems without invariance property under time reversal. Also, for each of the three ensembles, Dyson calculated the joint density function of the eigenvalues, the partition function, the level spacing distribution and the level correlation function.

Nowadays, there are totally eleven different ensembles of matrices in the random matrix theory of physics. Their definitions all obey the principle Dyson has adopted, that is, matrix type should be consistent with the underlying physical symmetries. For example, the chiral Gaussian orthogonal ensemble, the chiral Gaussian unitary ensemble and the chiral Gaussian symplectic ensemble were defined in accordance with the so-called chiral symmetry and its spontaneous breaking. These symmetries characterize the spectrum of the quantum chromodynamics Dirac operator, while the chiral ensembles are representing this operator (Verbaarschot and Wetting (2000)). The other five ensembles are the four Oppermann-Altland-Zirnbaner ensembles for description of disordered superconductors (Oppermann (1990) and Altland and Zirnbauer (1996)) and the Ginibre ensemble for the distribution of poles of S-matrices (Ginibre (1965)). Except for Ginibre's ensemble, all the other ten ensembles are of Hermitian nature. In fact, it was found in Zirnbauer(1996) that there is a one to one correspondence between the ten Hermitian ensembles and the large families of Cartan's symmetric spaces.

Random matrix theory is very fruitful in physics. As a solvable and reliable model for theoretical analysis, it was deliberately used to solve a diversity of physical problems. Besides the description of energy levels of complex nuclei, there are also its applications in the description of the Euclidean Dirac operator in QCD, the description of universal conductance fluctuations, or more generally, in theoretical nucleus physics, in low-lying energy theory of QCD, in the theory of disordered conductance, in solid state physics theory, in mesoscopic physics theory, in quantum chaos and in quantum gravity. Many powerful mathematical tools were exploited and invented to deal with various kinds of matrix integrations. Among others, it is worthy to mention the orthogonal polynomial method, the Riemann-Hilbert method and the supersymmetric method. By means of them the classical problems such as those mentioned for the Gaussian ensemble of Wigner were all systematically discovered and rediscovered for the various ensembles of matrices. A recent significant result is on the distribution of the largest eigenvalue of the GOE, GUE and GSE matrices (Tracy and Widom (1993,1994,1996)). A good reference list can be found in the recent review work Forrester, Snaith and Verbaarschot (2003).

These results have surprisingly far-reaching implications and applications in areas other than physics. Encouraging findings have appeared, in an increasing number, in financial correlations (Laloux et al [51], Plerou et al (2001)), portfolio optimization (Pafka and Kondor (2004), Potters, Bouchaud and Laloux (2005)), data analysis (Achlioptas [1]) and RNA folding (Vernizzi and Orland (2005), Barash

(2004)). And the most fascinating news should be the finding in number theory. In this area, one of the most important unsolved problem is the Riemann hypothesis, which says that all the non-trivial zeros of the Riemann zeta function lie on a critical line in the complex plain $z = 1/2 + iv$. Now it has been shown, partially but with quite a deal of evidence, these zeros demonstrate the same spectral properties of the eigenvalues of the GUE matrices. Up to now, mathematically rigorous proof has been established for connecting the two-point correlation function of the zeros of zeta functions of varieties over finite fields and the eigenvalues of the GUE matrices (Sanak and Katz (1998)). Further advances are still looked forward to. Nonetheless, great attention has been drawn from mathematicians in number theory on employing random matrix theory to predict important quantities closely related to the Riemann hypothesis (See references [52]-[61] in Forrester et al (2003) and more recent works summarized in [101] of the present thesis).

The result on the largest eigenvalue of the GOE, GUE and GSE matrices also have profound consequences in many other areas. These so-called Tracy-Widom laws are found to describe simultaneously, in combinatorics the limit laws for the length of the longest increasing subsequence in a random permutation, in many growth processes the fluctuations about their limiting shape, in random tilings the fluctuations about the limiting circle of the boundary between the temperature zone and the polar zone in an Aztec diamond tiling, in queuing theory, the limiting distribution of the departure time, appropriately normalized, of a customer k from the last queue n in a series of n single-server queues with unlimited waiting space and a first-in first-out service rule, and in statistics the asymptotic distribution of the largest eigenvalue of the Wigner matrix and the largest singular value of the sample covariance matrix under the assumptions in the literature of probability for these matrices (Tracy and Widom (2002), Soshnikov (1999, 2002)).

From the above review, it can be seen that random matrix theory developed in physics has achieved marvellous success. Two important factors contribute to this great success. first, the Gaussian assumption put on those ensembles constructed in physics play a significant role. The assumption provides for all those ensembles explicit expression of the joint density function of their eigenvalues and so makes possible the discussions of very deep and fine statistical properties, such as the correlation function of eigenvalues, be developed through calculating various matrix integrals. However, there is one virtue mostly valued by every area about the random matrix theory, that is, the so-called universality. Generally speaking, this virtue means results in the random matrix theory obtained in the limit sense do not depend on the specific distributions of the matrix elements. Thus to claim those results derived under the Gaussian assumption on their random matrix models behave with universality, physicists examine further the validity of the same results with a change of the so-called potentials governing the trace in the power part of the exponential in the joint density function of the various Gaussian ensembles. Arguments on this aspect are usually said to be universality theorems. However, one can see these type arguments are not enough for asserting real universality.

Furthermore, one of the most serious problems demonstrated by the universality theorems is sometimes with different choices of the potential different limit laws emerge up, as were shown in many cases. For example, when the limit law for the largest eigenvalue of the GUE matrices is examined on its universality, it was proved by finely tuning the potential new universality laws, other than the Tracy-Widom law for the GUE, were obtained (Deift et al. (1999)). Thus the notion of universality needs some more refinement works in the random matrix theory of physics, since the density functions of the matrices do show their effect and unfortunately there is no complete understanding, for a particular physical model, on how many different consequences can possibly be found by choosing different potentials on the density function. This breaking phenomenon of universality is also a reminder that, in applied areas, if the Gaussian assumption does not hold, then more attention should be put on the universality arguments. However, due to the lack of statistical meaning of the potential, in case that the universality problem is inquired, it is also hard to test in a statistical way whether the data at hand are generated from the random matrix model specified by the potential.

Secondly, in the success of random matrix theory of physics, the various ensembles constructed in the field play an essential role. The very appreciable quality of these ensembles is they are intimately rooted in the very foundation of mathematics of symmetric spaces. This accounts for at least partly today's remarkable connection of random matrix theory in main branches of pure mathematics. In fact, physical ensembles were originally constructed to represent the symmetry properties of certain physical operators. For some, if not all, of them, the global distribution of their eigenvalues is already known. For instance, the eigenvalues of GOE are distributed on the unit circle while the eigenvalues of the Ginibre ensemble are distributed on the unit disc. Only local spectral statistics are of interest in the literature of physics. This, in many situations, is in contrast with the necessities of other applied areas. In applied areas, very often the needed random matrix models are straightforwardly posted by actual world problems. They are roughly known by their general properties such as their matrix forms and the the existence of certain moments of their elements, but not on the distribution of their eigenvalues. Rather, the distributions of their eigenvalues, or more generally, the statistical properties of global spectral statistics, are of central interest. In these cases, the random matrix models in physics are lacking in this regard.

In conclusion, developments of random matrix theory have been impressively successful in physics and the success has brought new insights into many mathematical problems arising from various branches of mathematics. The main impetus of this achievement seems due to various matrix integrations that have played the role of bridges connecting together originally independent problems. The success of random matrix theory of physics shows that random matrices can be very powerful and versatile tools to deal with the nowadays more and more complicated scientific problems. However, the two most important factors contributing to this success also induced some limitations on applying the theory to other applied areas. The first limitation is more essential since it lies in the theoretical foundation of the theory. That is, the various ensembles in the theory directly constructed for solving physical problems by investigating local spectral behavior are not consistent with most practically needed random matrices which take on certain forms implicitly or explicitly determined by actual world problems. The other limitation is that, in most results in the theory, the Gaussian assumption is crucial and the universality arguments are not enough. This results that once the Gaussian assumption failed, the breaking phenomenon of the universality theorems and the difficulty of testing the universality family specified by a potential will also lead an application to false results. The limitations indicate that in applications of random matrix theory, more attention should be put on using correct random matrix models.

To have a representative random matrix model is crucial to any application of random matrix theory in applied areas. In some cases, this needs constructing random matrix models suitable for the problems at hand. Then the stimulating principle in physics of reflecting certain invariant properties of real systems can be helpful. In some other cases, however, as the random matrix model has been determined by the actual problem, to make effective use of random matrix theory will mean to resort to the random matrix theory developed in probability theory. This is another important area where spectral analysis of large dimensional random matrices has gotten significant achievements. A distinctive property of the random matrix theory in probability is the random matrices are all studied under very general assumptions which usually express themselves as conditions on existence of certain moments of the matrix elements. This quality clearly represents the universality virtue which is expected from the random matrix theory. Moreover, the source of the various random matrices studied in the literature of probability are either from classical statistical methods or straightforwardly from practical problems. Thus they have a clear understanding in either statistics or other applied areas. This helps the wide applications of the theory in a diversity of applied areas. Indeed, in every area where statistical methods are of use, the random matrix theory in probability can find applications. A simple example below shows an effective application of random matrix theory in probability in wireless communications.

In wireless communications, an effective theory on performance of wireless channels is most extensively built on the following channel model:

$$y = Hx + n,$$

where x is the K-dimensional input vector, y is the N-dimensional output vector, n is the N-dimensional noise vector. The $N \times K$ matrix H is the so-called channel matrix and is random. This channel model has applications in many different areas of wireless communications. In different applications H has different interpretations and takes on different assumptions. In the simplest case, H consists of independent and identically distributed (i.i.d.) entries. This case happens with, for example, a single-user narrow-band channel with K and N antennas at transmitter and receiver respectively or direct-sequence code-division multiple-access (CDMA) channel not subject to fading. In many other cases, the entries of H are not i.i.d. any more.

The wide use of random matrix theory in wireless communications is due to the significant role played in the field by the singular values of the random channel matrix H, or equivalently, the eigenvalues of the random matrix H^*H. In fact, fundamental performance measures like channel capacity and minimum mean-square-error (MMSE) can be expressed as functionals of the eigenvalues of H^*H. For example, assuming constraints $Enn^* = \sigma_0^2 I$ and $Ex^*x \leq KP$, the channel capacity is expressed by

$$\begin{aligned} C &= \frac{1}{N} E \log\det\left(I + \frac{P}{\sigma_0^2} HH^*\right) \\ &= E \int \log\left(1 + \frac{P}{\sigma_0^2}\lambda\right) dF^{HH^*}(\lambda), \end{aligned}$$

and the MMSE is expressed as

$$\begin{aligned} MMSE &= \frac{1}{K} Etr\left(I + \frac{P}{\sigma_0^2} H^*H\right)^{-1} \\ &= E \int \frac{1}{1 + \frac{P}{\sigma_0^2}\lambda} dF^{H^*H}(\lambda), \end{aligned}$$

where $F^{HH^*}(x)$ and $F^{H^*H}(x)$ denote respectively the ESD of HH^* and H^*H.

When K and N are both large with their ratio approaching a positive constant, say c, the limit of the channel capacity and the MMSE in the almost sure sense can be derived from the known results on the so-called sample covariance matrix in the random matrix theory of probability. In fact, assuming the simplest case that H consists of i.i.d. entries with mean 0 and variance $1/N$, by the result proven first in Marcĕnko and Pastur (1967), with probability one as $K \to \infty$ with $K/N \to c$, the ESD of H^*H converges weakly to the Marcĕnko and Pastur law with density function

$$\frac{d}{dx} F^c_{M-P}(x) = \begin{cases} \frac{1}{2\pi x c}\sqrt{(x-a)(b-x)}, & a < x < b, \\ 0, & otherwise, \end{cases}$$

and, if $c > 1$, an additional point mass of $(1 - 1/c)$, where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$. Using properties of weak convergence in classical probability theory, this then indicates that the limit of the channel capacity and the MMSE should be given by

$$c \int \log \left(1 + \frac{P}{\sigma_0^2} \lambda \right) dF^c_{M-P}(\lambda),$$

and

$$\int \frac{1}{1 + \frac{P}{\sigma_0^2} \lambda} dF^c_{M-P}(\lambda).$$

And indeed, the second one is true since the integrand is bounded and continuous, but the first one needs further consideration since the log function is not bounded. To make rigorous for the first one, a bound argument is needed on the largest eigenvalue of the matrix H^*H. But in the random matrix theory, there is also a known result proven in Yin, Bai and Krishnaiah (1988) that if and only if the fourth moment of the entries of H is finite the largest eigenvalue of H^*H converges almost surely to $b = (1 + \sqrt{c})^2$. Therefore, large number laws on the channel capacity and the MMSE are thus established with the aids of random matrix theory in probability. By calculating the two integrals above, one then knows asymptotically the two performance measures important for the channel model. For the precise results of these integrals and for more details on applications of random matrix theory to wireless communications, see the monograph Tulino and Verdú (2005).

To the satisfaction of engineers, we see that asymptotic results are obtained with the aid of random matrix theory. For example, they have universality property of not being sensitive to the distribution of the random matrix entries. In case of a single-user multi-antenna link, this means the asymptotic results hold for any type of fading statistics, and in case of the CDMA channel, this means restricting the

CDMA waveforms to be binary valued incurs no loss in capacity asymptotically (Tulino and Verdú (2005)). Also, since the asymptotic results are shown in the almost sure sense, in experimental observations, one realization is sufficient to obtain the convergence to the deterministic limit. Also, of practical interest is the knowledge of the convergence rate of the channel capacity and the MMSE to their limits and the distribution of fluctuations of the channel capacity and the MMSE around their limits. These two problems can be fully solved by the central limit theorems on analytical functionals of eigenvalues of sample covariance matrices proven in Bai and Silverstein (2004). As was shown, the convergence rate is $O(n^{-1})$ and the fluctuations follow normal distribution with mean and variance explicitly expressed. Therefore, we see how random matrix theory can help in applied areas.

In the following, when we say random matrix theory in probability we are referring to spectral analysis of large dimensional random matrices, but the classical theory on the Wishart matrix of course is a big component of the whole theory developed on random matrices in the literature of probability. The importance of studying spectral properties of large random matrices for the development of statistics was well stated in Bai and Silverstein (2004). That is, highly developed computational techniques make possible systematic collection, conservation and computation of data of very high dimension, but classical statistical analysis methods have limitations and weakness to deal with them. Sometimes the existing statistical analysis methods simply do not apply to high dimensional data. For instance, if significance test is considered for the difference of the means of two k-dimensional populations based on two samples of sizes n_1 and n_2 taken respectively from the two populations, then classical statistical inference methods using the T^2 statistic of Hotelling or the best linear discriminator of Fisher are undefined whenever $k > n_1+n_2-2$ (Dempster (1958, 1960)). In this case, it is an unavoidable task to develop statistical analysis methods relevant to high dimensional data.

On the other hand, although sometimes classical statistical methods can still be carried out on high dimensional data, the outcomes may deviate from what should be expected to be true. This phenomenon is relating to the underlying philosophy of classical statistical methods. Generally speaking, classical limit theorems fundamental to multivariate statistical inference are developed under a hypothesis that the vector dimension is fixed and the sample size increases to infinity. This hypothesis is also commonly said to be the hypothesis of large sample, since in practical experimental work requiring a multivariate inference technique, these limit theorems are expected to behave well in the case that compared with the vector dimension of the data, the sample size should be overwhelmingly large. However, when the vector dimension is large, an overwhelmingly large sample size becomes a tremendous magnitude and is unattainable in most situations. As a result, classical statistical methods in multivariate analysis are used when the hypothesis of large sample is not satisfied. But, from the following example presented in Bai and Silverstein (2004), it can be seen such use may induce very serious errors.

Consider a statistic constructed from the sample covariance matrix S_n. Here $S_n = (1/N)X_n^*X_n$, where X_n is $N \times n$ consisting of i.i.d. standard normal random variables. Define statistic $L_N = \ln\det S_N$. This is an important statistic used in classical multivariate statistical inference. Under the large sample hypothesis, namely n is fixed and N tends to infinity, $\sqrt{N/n}L_N$ converges weakly to normal distribution with mean 0 and variance 2. However, when the hypothesis is violated with $n/N \to c \in (0,1)$, using the results in Marcĕnko and Pastur (1967) and Yin, Bai and Krishnaiah (1988), it can be seen $(1/n)L_N \to d(c) \equiv (1-1/c)\ln(1-c)-1 < 0$ which implying $\sqrt{N/n}L_N \to -\infty$. Thus, of course, in this case neglecting the nature of high dimension of the data to use classical statistical inference method based on asymptotic normality of $\sqrt{N/n}L_N$ will cause serious error. Indeed, such performance loss of classical statistical methods on high dimensional data has been examined early in Bai and Saranadasa (1996) and referred to as an effect of high dimension. As stated in Bai and Silverstein (2004), the above example got an solution on its asymptotic distribution as a by-product of the main result in the paper, which will be reviewed in the sequel. Specifically, the normalized statistic $L_N - nd(n/N)$ converges in distribution to a normal random variable with mean $\frac{1}{2}\ln(1-c)$ and variance $-2\ln(1-c)$. The example thus exhibits both the need and the value of spectral analysis of large dimensional random matrices.

Generalizations of Wigner's matrix were the first tide in the developments of random matrix theory in probability. As already noted in Wigner (1958), the semicircle law is the LSD of a much more general symmetric random matrix model which satisfies only that the entries on and above the diagonal are independent and the entries have symmetric distribution function with variance σ^2 for the non-diagonal entries and $2\sigma^2$ for the diagonal ones and all higher moments uniformly bounded. This claim motivated the interest of relaxing the conditions on the matrix to the most possible extent (Grenander (1963), Arnold (1967,1971)). In the important review work on random matrix theory in probability Bai (1999), it was shown two general assumptions can be used to define the matrix which mostly extends the matrix of Wigner. Let $W_n = [w_{ij}]$ be $n \times n$ whose entries on and above the diagonal are i.i.d. complex random variables with a common mean and variance σ^2, or let $W_n = [w_{ij}]$ be $n \times n$ Hermitian whose entries on and above the diagonal are independent complex random variables with a common mean satisfying the Lindeberg type condition for any $\delta > 0$ as $n \to \infty$

$$\frac{1}{\delta^2 n^2} \sum_{ij} E|w_{ij}|^2 I_{(|w_{ij}| > \delta\sqrt{n})} \to 0. \tag{1.1}$$

Note that the second assumption is more general than the first one. It was shown in Bai (1999) that, under either assumption, as $n \to \infty$ with probability one the ESD of $\frac{1}{\sqrt{n}} W_n$ converges weakly to the semicircle law.

The convergence rate of the expected ESD of the normalized Wigner matrix $\frac{1}{\sqrt{n}} W_n$, under the first assumption above with the additional condition that fourth moments uniformly bounded in n, was shown to be not slower than $O(n^{-1/4})$ in Bai (1993a). This problem is one of the toughest problems since the inception of random matrix theory. Bai's work developed a method of discussing convergence rates of ESDs through establishing a Berry-Esseen type inequality in terms of the Stieltjes transforms. The result was later improved in Bai, Miao and Tsay (1999) by assuming a slightly milder condition but confirming the convergence rate of the expected ESD and the convergence rate in the sense of in probability to be both not slower than $O(n^{-1/3})$. It can be expected further improvements in the future since the conjectured ideal convergence rate can achieve $O(n^{-1})$.

The limiting behavior of the largest eigenvalue of $\frac{1}{\sqrt{n}}W_n$ is another important aspect in the spectral analysis of large dimensional Wigner matrices. A sufficient and necessary condition for the largest eigenvalue of $\frac{1}{\sqrt{n}}W_n$ to converge almost surely to a finite constant was given in Bai and Yin (1988b). The asymptotic distribution of the largest eigenvalue of the Wigner matrix was recently solved in Soshnikov (1999). As was shown, the limit laws for the largest eigenvalue of the real and complex Wigner matrix are respectively the Tracy-Widom laws for GOE and GUE.

Central limit theorems concerning analytic functionals of eigenvalues of the Wigner matrix were shown recently in Bai and Yao (2005). The paper continued the same type of arguments developed in the significant work Bai and Silverstein (2004) on the sample covariance matrices. Mainly, let $F_n(x)$ and $F(x)$ denote respectively the ESD of $\frac{1}{\sqrt{n}}W_n$ and the semicircle law. Define the so-called spectral empirical process as

$$G_n(f) = n\int f(x)[F_n(x) - F(x)]dx, f \in \mathcal{A},$$

where $\mathcal{A}$ is the set of functions analytic on an open set enclosing the support of the semicircle law. Then it was shown the spectral empirical process is tight, and under appropriate conditions on the moments of the entries of the Wigner matrix, converges weakly to a Gaussian process. Central limit theorems concerning $[G_n(f_1), \cdots, G_n(f_k)]$ are therefore consequences of the convergence of finite dimensional distributions of the process.

The random matrices most extensively investigated in the literature of probability are the so-called sample covariance type matrices. The pioneering work in this aspect is Marcěnko and Pastur (1967). The random matrix they considered takes the form $A_n + X_n^* T_n X_n$. Here A_n, T_n, X_n are independent of each other, A_n is $n \times n$ Hermitian, T_n is $n \times n$ diagonal, and X_n is $N \times n$ consisting of i.i.d. random variables. They proved under certain conditions the ESD of the matrix converges weakly to a non-random limit. Their method was then original. Before their work, the main methodology in the field was the moment method which proves the convergence of ESDs by showing the convergence of their moments. However, they adopted the method of proving the convergence of the ESDs through proving the convergence of their Stieltjes transforms. Many later works studied the random matrix $A_n + X_n^* T_n X_n$. Examples are Grenander and Silverstein (1977), Jonsson (1982) and Wachter (1978).

Marcěnko and Pastur's problem was later reconsidered in Silverstein and Bai (1995) with milder conditions imposed on the underlying random variables. Under the assumptions that X_n is consisting of i.i.d. random variables with mean 0 and variance σ^2, T_n is diagonal with, almost surely, its ESD converging weakly to a probability distribution function (p.d.f.) and A_n is Hermitian with, almost surely, its ESD converging vaguely to a non-random limit, it was shown with probability one as $n \to \infty$ while $N = N(n)$ with $n/N \to c > 0$, the ESD of $A_n + X_n^* T_n X_n$ converges vaguely to a non-random limit whose Stieltjes transform satisfies a uniquely solvable equation. The authors continued with and modified Marcěnko and Pastur's method. Note that in Marcěnko and Pastur (1967), the convergence of the

Stieltjes transforms was shown by constructing a stochastic function involving a parameter t and consequently the limit of the Stieltjes transforms was the solution to a partial differential equation at $t = 1$. Silverstein and Bai's method still relies on showing the convergence of the Stieltjes transforms but, with the aids of fundamental matrix properties and classical probability theory, is more straightforward while providing a clear understanding of the convergence process of the Stieltjes transforms. Indeed, the method was later further developed by the authors to investigate more complicated problems and nowadays the method is widely known as the Stieltjes transform method.

The so-called sample covariance matrix in random matrix theory of probability is, as we have indicated in our previous review, of the form $S_n = (1/N)X_n^* X_n$, which is the special case of the random matrix studied by Marcĕnko and Pastur when $A_n = O_{n\times n}$ and $T_n = I_{n\times n}$. This explains why the LSD of the sample covariance matrix is called the Marcĕnko and Pastur law. A proof via the moment method of the convergence of the ESD of S_n can be found in Bai (1999). In Bai (1999), two assumptions on S_n were considered. One assumption is to require X_n to be composed of i.i.d. entries with mean 0 and variance σ^2. The other assumption is to require X_n to be composed of independent entries with mean 0 and variance σ^2 satisfying the following Lindeberg type condition for any $\delta > 0$ as $n \to \infty$

$$\frac{1}{\delta^2 nN} \sum_{ij} E\left(|x_{ij}|^2 I_{(|x_{ij}|>\delta\sqrt{n})}\right) \to 0. \tag{1.2}$$

Under either assumption, it was shown as $n \to \infty$ with $n/N \to c > 0$ the k-th moment of the ESD of S_n converges almost surely to the k-th moment of the

Marcĕnko and Pastur law. Since it is easy to check the Marcĕnko and Pastur law satisfies the Carleman condition which confirms it to be determined by its moments, it then follows the ESD of S_n must converge to the Marcĕnko and Pastur law. This is indeed the main scheme of using the moment method to show the convergence of ESDs and identify the LSD.

The convergence rate of the ESD of S_n to the Marcĕnko and Pastur law was established altogether with that for the Wigner matrix in Bai (1993b). As is shown, the convergence rate of the expected ESD of S_n is not slower than $O(n^{-1/4})$. Further improvements are still looked forward to as the conjectured ideal rate is still as fast as $O(n^{-1})$.

Concerning the limiting behavior of the largest eigenvalue of S_n, as for the Wigner matrix, results are known on both its almost sure convergence and its asymptotic distribution. In fact, almost sure convergence has been shown for both of the extreme eigenvalues, the largest and the smallest eigenvalues, of S_n. Under the first assumption of Bai (1999) on S_n with an additional condition of finite fourth moment of x_{11}, it was shown in Yin, Bai and Krishnaiah (1988) that the largest eigenvalue of S_n converges almost surely to $\sigma^2(1+\sqrt{c})^2$, the largest number of the support of the Marcĕnko and Pastur law. Later in Bai, Silverstein and Yin (1988) it was confirmed further the condition of finite fourth moment is also necessary for the convergence of the largest eigenvalue. The convergence of the smallest eigenvalue of S_n was solved in Bai and Yin (1993). The result in this work indeed established simultaneously the convergence of the largest eigenvalue and the convergence of the smallest eigenvalue. Specifically, under the same condition as in the case of the largest eigenvalue, it was shown

$$-2\sqrt{c}\sigma^2 \le \liminf_{n\to\infty} \lambda_{min}(S_n - \sigma^2(1+c)I) \le \limsup_{n\to\infty} \lambda_{max}(S_n - \sigma^2(1+c)I) \le 2\sqrt{c}\sigma^2,$$

where of course $\lambda_{min}(\cdot)$ and $\lambda_{max}(\cdot)$ respectively denote the largest eigenvalue and the smallest eigenvalue of the matrix $(\cdot)$. For the largest eigenvalue of S_n, its asymptotic distribution is also known. Mainly, it was shown in Johnstone (2001) that if X_n is consisting of i.i.d. standard normal random variables, then as $n \to \infty$ with $n/N \to c \ge 1$, the normalized largest eigenvalue $(\lambda_{max}(S_n)-\mu_n)/\sigma_n$ converges in distribution to the Tracy-Widom law for the GOE. Here the normalization constants are $\mu_n = (\sqrt{N-1}+\sqrt{n})^2$ and $\sigma_n = (\sqrt{N-1}+\sqrt{n})((N-1)^{-1/2}+n^{-1/2})^{1/3}$. When the entries of X_n are i.i.d. complex standard normal, then the asymptotic distribution becomes the Tracy-Widom law for the GUE. The normalized constants will also need a slight modification.

Most of the known results on the sample covariance type matrices concern random matrices taking the form $B_n = (1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$, where X_n is $N \times n$ consisting of i.i.d. random variables, T_n is $n \times n$ nonnegative definite, and X_n, T_n are independent. This type random matrices are representative for a large class of matrices which are of importance to multivariate statistical analysis. first, the sample covariance matrix S_n is the special case of B_n when $T_n = I_n$. More generally, when T_n is taken to be non-random while the entries in X_n are taken to be i.i.d. with mean 0 and variance 1, the matrix B_n is the sample covariance matrix of the N i.i.d. n-dimensional samples $(1/\sqrt{N})T_n^{1/2}\vec{x}_1, \cdots, (1/\sqrt{N})T_n^{1/2}\vec{x}_N$ with mean vector zero and variance matrix T_n, where the vector $\vec{x}_i$ denotes the

i-th column of the matrix X_n^*. Then, the Wishart matrix and the F-matrix, both crucially important to multivariate statistical methods, can be modelled by the matrix B_n. In fact, a Wishart matrix is the special case of S_n when the entries in X_n are i.i.d. normal random variables. The F-matrix is the special case of B_n when X_n is taken to be composed of i.i.d. normal random variables while T_n is taken to the inverse of another Wishart matrix independent of X_n. These account for the wide applications of spectral analysis results on B_n in areas as diverse as time series analysis, high-dimensional statistical inference methods, neural network theory and wireless communications. Motivated by this prominent conceptual and practical value of B_n, in random matrix theory in probability, spectral analysis results on B_n are the most significant and matured.

The convergence of the ESD of the matrix B_n has been well studied in the field. first, the result was established by using the moment method in Yin and Krishnaiah (1983) and Yin (1986). The latter work was done under a more general condition following the arguments developed in the former one. Specifically, it was shown if X_n is consisting of i.i.d. entries with finite second moment, T_n is such that its ESD with probability one converges weakly to a p.d.f. H, and certain additional conditions hold, then with probability one the ESD of B_n converges weakly to a non-random limiting distribution function. Due to the use of the moment method, finding the limits of the moments of the ESD of B_n is the core of the arguments. This involved a rather complicated combinatorial derivation, but the argument bears then a value to combinatorics also. As is indicated in some works in wireless communications, M$\ddot{u}$ller and verd$\acute{u}$ (2001) for example, the moments of the LSD are useful in the real-time implementation of the linear MMSE detector to compute the coefficients of the Yule-Walker equations. The additional conditions in Yin (1986) require that the moments of H satisfy the Carleman condition and that the moment of the ESD of T_n converges to that of H for every order. It was later shown in Bai (1999), they can be avoided by applying the truncation and centralization techniques to the ESD of T_n. Moreover, Bai (1999) also extended the result in the sense that the convergence of the ESD of the matrix $(1/N)X_n^* X_n T_n$ was proved, where the matrix T_n is Hermitian. Note that when T_n is nonnegative definite, the eigenvalues of the two matrices, B_n and $(1/N)X_n^* X_n T_n$, are exactly the same.

It turns out for better understanding of the spectral properties of the matrix B_n very important is to develop a proof by using the Stieltjes transform method to show the convergence of the ESD. This was obtained in Silverstein (1995), to which Silverstein and Bai (1995) is an important related work. Silverstein (1995) proved that if X_n is $N \times n$ consisting of i.i.d. entries with finite second moment, T_n is nonnegative definite with its ESD almost surely converging weakly to a p.d.f. H and X_n, T_n are independent, then with probability one as $n \to \infty$ while $N = N(n)$ with $n/N \to c > 0$, the ESD of B_n converges weakly to a non-random p.d.f.. This LSD is given by an equation to which its Stieltjes transform is the unique solution. An important point is, via the equation, analytical properties of the LSD can be derived. This is one advantage, but by no means all, of using the Stieltjes transform method. Analytic properties of the LSD of B_n were derived in Silverstein and Choi (1995). They mainly proved the LSD is continuously differentiable at any point on the real line except the origin, the support of the LSD can be determined by checking a necessary and sufficient condition, inside the support the derivative of the LSD is infinitely differentiable. Moreover, both the derivative and the condition on the support of the LSD are qualitatively tractable from an equation taking the form $z(m) = -1/m + c\int t/(1+tm)dH(t)$. These results are also useful for later developments on the spectral analysis of the matrix B_n.

One of the most significant results for B_n is on limiting behavior of its eigenvalues outside the support of its LSD. These are established in Bai and Silverstein (1998, 1999). Mainly, the earlier work proved that for any closed interval $[a, b]$ outside the support of its LSD, under appropriate conditions, with probability one there will be no eigenvalues of B_n appearing in this interval. The limiting behavior of the extreme eigenvalues of B_n can be followed from this result as a subsequence. Formally, if the largest eigenvalue of T_n converges to the largest number of H, then the largest eigenvalue of B_n converges to the largest number of the support of its LSD. Furthermore, if the smallest eigenvalue of T_n converges to the smallest number of the support of H, then in case of $c \leq 1$ the smallest eigenvalue of B_n converges to the smallest number of its LSD and in case of $c > 1$, the smallest eigenvalue of $\underline{B}_n = (1/N)X_n^*T_nX_n$ converges to the smallest number in the support of its LSD. Note the relation between B_n and $\underline{B}_n$. They have the same nonzero eigenvalues.

Bai and Silverstein (1999) went even farther. For any prescribed interval $[a, b]$, the result of Bai and Silverstein (1998) implies for all n large, $[a, b]$ is a gap in the spectrum of B_n. All eigenvalues of B_n must lie either to the left or to the right of this gap. Then a natural question is to inquire the number of eigenvalues to one side of the gap B_n put. Using the criterion given in Silverstein and Choi (1995) on how to determine the support of the LSD of B_n, a not so intuitive but definitely true fact can be shown which says that to such a gap $[a, b]$, there must be a interval $[a', b']$ which is the gap in the spectrum of T_n for all n large. Then the main result of Bai and Silverstein (1999) is to show with probability one for all n large the number of eigenvalues B_n put to one side of $[a, b]$ is equal to that of eigenvalues T_n put to the same side of $[a', b']$. There is only one exception with this beautiful accordance between the spectrum of B_n and T_n, which happens with the case when $c[1 - H\{0\}] > 1$ and $[a, b]$ lies in the intermediate segment between the origin and the first positive number in the support of B_n's LSD. But for any other cases when $c[1-H\{0\}] > 1$ but $[a, b]$ does not lie in this special segment, the result is still true. Here $H\{0\}$ denotes the point mass of H at zero. The reason of the exception is very intuitive, since it can be computed $F\{0\} = H\{0\}$ if and only if $c[1 - H\{0\}] \leq 1$. This result is called the exact separation of the eigenvalues of B_n.

Central limit theorems concerning certain functionals of the eigenvalues of B_n were first derived in Jonsson (1982) relying on the assumption that the entries of X_n are Gaussian random variables. In Bai and Silverstein (2004), a new way of establishing this type results was developed for a set of analytic functionals. Denote by F^{B_n} and $F^{c,H}$ respectively the ESD and the LSD of B_n. For any p.d.f. F, denote by $s_F(z)$ its Stieltjes transform. Define $G_n(x) = n[F^{B_n}(x) - F^{c,H}(x)]$ and $M_n(z) = n[s_{F^{B_n}}(z) - s_{F^{c,H}}(z)]$. Let $\mathcal{C}$ be a contour of the complex plane enclosing the interval

$$[\liminf_n \lambda_{min}(T_n)I_{(0,1)}(c)(1-\sqrt{c})^2, \limsup_n \lambda_{max}(T_n)(1+\sqrt{c})^2].$$

Assume that X_n is $N \times n$ consisting of i.i.d. entries with mean 0, variance 1 and finite fourth moment, T_n, independent of X_n, is $n \times n$ non-random nonnegative definite with uniformly bounded spectral norm whose ESD converges weakly to a p.d.f. H. Then it was shown, viewed as a random two dimensional process on the contour $\mathcal{C}$, $\{M_n(z)\}$ is tight. Furthermore, if the moments of the entries in X_n have the same fourth moment as the standard normal (real or complex), then $\{M_n(z)\}$ converges weakly to a two dimensional Gaussian process. For any integer k let $f_1, \cdots, f_k$ be functions analytic on an open interval containing the prescribed interval. Then central limit theorem on

$$\left(\int f_1(x)dG_n(x), \cdots, \int f_k(x)dG_n(x)\right)$$

follow immediately, provided that the moments of entries of X_n satisfy the previous condition.

The series of significant works of Bai and Silverstein (1998,1999, 2004) established two most important aspects of results on spectral analysis of random matrices of the form B_n. In view of the prominent value of the matrices, their works are even more marvellous. The achievements are obtained with the aids of the Stieltjes transforms of the ESDs of B_n, which offer a platform where classical probability theory on independent random variables can be applied to earn an optimal gain concerning the general assumptions underlying their random matrix model. This indicates that thorough spectral properties of random matrices can be analyzed effectively by systematically manipulating the Stieltjes transforms of ESDs.

There are many other random matrices studied in the random matrix theory in probability. For example, the convergence of the ESDs of the Toeplitz, Hankel and Markov matrices was shown in (Bryc, Dembo and Jiang (2006)) by using the moment method. However, due to the complexity of the problem, the LSDs are not known very much yet. The convergence of the ESD of the random matrix which is $n \times n$ consisting of i.i.d. complex entries to the circle law has been well known in the field. But the proof remains unknown until Girko (1984) provided a partial solution to the problem. The problem was later proved in Bai (1997) under the existence of the $(4+\varepsilon)$th moment of the matrix entries and some other smoothness conditions on their density function. In the monograph of Girko (1990), Girko defined a random matrix model which turns out to be very useful in applied areas. This random matrix is $n \times n$ Hermitian with independent entries on and above diagonal, all entries have mean 0 but variance σ_{ij}^2 for the (i, j)-th entry. It is assumed that the σ_{ij}^2 are uniformly bounded for all i, j and n and are such that the function defined by

$$w_n(x, y) = \sigma_{ij}^2, \quad \frac{i}{n} \le x \le \frac{i+1}{n}, \quad \frac{j}{n} \le y \le \frac{j+1}{n}$$

converges as $n \to \infty$ to a bounded limit function $w(x, y)$. It was shown as $n \to \infty$ the ESD of this random matrix converges weakly to a non-random p.d.f. whose Stieltjes transform is the unique solution to a certain equation. Other examples are random matrices with symmetry breaking structure $C =$

$I + (1/N)\sum_{k=1} T_k X_k^* X_k$ (Hoyle and Rattray (2003)) and information-plus-noise type matrices $D_n = (1/N)(R_n + \sigma X_n)^*(R_n + \sigma X_n)$ (Dozier and Silverstein (2004)). Due to the richness of the context and the interest of the present thesis, we shall not go deeper.

From the above review, it can be seen that random matrix theory in probability has gotten its success in mainly the following four aspects. The first aspect concerns the convergence of the ESDs of random matrices. The second aspect concerns the convergence rate of the ESDs. The third aspect concerns the limiting behavior of extreme eigenvalues of random matrices, or more generally, the limiting behavior of eigenvalues of random matrices outside the support of their LSDs. The fourth aspect concerns central limit theorems for analytic functionals of eigenvalues of random matrices. Indeed, there is also the fifth aspect in the field which concerns the limiting behavior of eigenvectors of random matrices. Results on the Wigner matrix and the sample covariance matrix can respectively be found in Girko, Kirsch and Kutzelnigg (1994) and Silverstein (1979,1981,1984,1989,1990). These five aspects only occupy a small portion of what is expected in the spectral analysis of large dimensional random matrices. As stated in Bai (1999), second order convergence problems need to be developed to deepen these five aspects of results whereas more other new random matrices need to be investigated.

For any random matrix, the first aspect of investigation summarized above is the fundamental part. This aspect shows the convergence of ESDs, finds the LSDs and studies important properties of the LSDs. The results can be used directly in applied areas such as in the wireless communications to give good predictions of performance measures of channel models or to achieve real-time implementation of the linear MMSE detector. Of course, spectral analysis of any random matrix must be started from this fundamental stage of work. In the meanwhile, this stage of work usually provides useful analytical tools to provide further investigation of other spectral properties of the random matrix. We thus come to the main purpose of the present thesis, to develop this fundamental stage of spectral analysis for the following three classes of random matrices:

- Large dimensional Wigner type random matrices.
- Large dimensional general sample covariance type random matrices.
- Large dimensional sparse random matrices.

1.1 Large Dimensional Wigner Type Random Matrices

1.1.1 The Problem

Generalizations of the Wigner matrices have attracted considerable interest since the inception of random matrix theory. In the theory, Wigner matrices have more important value than they have expressed in modelling complex nuclei. Almost every aspect of spectral analysis is carried out first to the Wigner matrices. Their foundational role has been recently further guaranteed by the establishment of the so-called free central limit theory, which proves that the semicircle law is the counter part of the normal distribution in free probability. Most extensively investigated asymptotic free random matrices are those relating to the Wigner matrices. In fact, in order to understand better spectral properties of Hermitian matrices which are not necessarily nonnegative definitive, generalizing the Wigner matrices in some appropriate sense is necessary.

Generalizations of the Wigner matrices have been done in various aspects. One important aspect is to show the semicircle law valid for some other random matrices rather than the Wigner ones. Examples of such random matrices include, for example, the normalized sample covariance matrices considered for the case when the vector dimension and the sample size both tend to infinity but their ratio tends to zero (Bai and Yin (1988)), or the sparse random matrices taking the form of the Hadamard products of a normalized sample covariance matrix and a sparsing matrix whose elements play the role of wearing down the correlation existing among the entries of the normalized sample covariance matrix (Kohrunzhy and Rodgers (1997, 1998)). These results revealed two important points to us. One point is that it is the lack of statistical correlation among the entries of the Wigner matrices that plays the essential role in the convergence to the semicircle law of their empirical spectral distributions.

The other point is if different limiting spectral distributions, especially some providing better predictions of the real system, are of interest, the generalization must be considered in another aspect, that is, to permit some statistical correlation structure to exist among the entries of the prescribed Hermitian matrix representing the Hamiltonian of the real system. This direction is indeed the motivation of the Wigner type random matrices dealt with in the present thesis.

1.1.2 The Objective

The Wigner type random matrices generalize the Wigner matrices by allowing the entries in the matrices to possess a statistical correlation structure. This can be conveniently achieved by expressing the Wigner type random matrices in the form of $(1/\sqrt{n})T_n^{1/2}W_nT_n^{1/2}$, where W_n is $n\times n$ Wigner matrix and T_n is $n\times n$ nonnegative definite, with W_n, T_n independent. Existing work on this class of random matrices is only confined to the special case where the entries of W_n are assumed Gaussian random variables (Monvel and Khorunzhy (1999)). In their work, Monvel and Khorunzhy proved as n tends to infinity, the empirical spectral distributions of the matrices will converge to a non-random limiting spectral distribution and under appropriate conditions the spectral norms of the matrices will converge to the upper endpoint of the support of the limiting spectral distribution. Although this limiting spectral distribution was given by an equation determining its Stieltjes transform, the result was indeed established through investigating a certain set of mixture moments relating to the empirical spectral distributions of the matrices. The main technical result in the paper is to construct appropriate bounds for those mixture moments by means of an equality on evaluating the mathematical expectation of a Gaussian random vector.

Concerning the universality property of random matrix theory, the Gaussian assumption in Monvel and Kohrunzhy (1999) is redundant. In our work, we propose to remove this redundant condition and to deal with the Wigner type matrices in a very general sense. The method of Monvel and Kohrunzhy strictly depends on the Gaussian assumption and so cannot be used anymore. We shall adopt both the Stieltjes transform method and the moment method to show the convergence of the ESD of the Wigner type random matrices, and based on the equation determining the Stieltjes transform of the LSD, we shall also derive properties of the LSD and their density functions.

1.1.3 Main Results

Let us introduce the following notation. Throughout the remainder of the present thesis,, for any probability distribution function G, $s_G(z)$ denotes its Stieltjes transform, S_G denotes its support set; for any set E, E^c denotes its complement and I_E denotes the indicator function on E; for any probability distribution function G and any Borel set E, $G(E)$ denotes the measure of E with respect to the measure generated by G on the real line $\mathbb{R}$; $\mathbb{R}^+$ and $\mathbb{R}^-$ denote respectively the positive half real line and the negative half real line; $\mathbb{C}$ denotes the set of complex numbers, $\mathbb{C}^+ = \{z \in \mathbb{C} : Im(z) > 0\}$ and $\mathbb{C}^- = \{z \in \mathbb{C} : Im(z) < 0\}$.

By imposing conditions on moments of the entries of the matrices instead of putting conditions on their distributions, we studied the Wigner type random matrices in the following general sense.

Definition 1.1.1. (Wigner type random matrix)

Let $A_n = \frac{1}{\sqrt{n}} T_n^{1/2} W_n T_n^{1/2}$*, where* $T_n^{1/2}$ *is any Hermitian square root of* T_n*. Then* A_n *is said to be a Wigner type random matrix if the following conditions are satisfied.*

(i) For $n = 1, 2, \cdots$, $W_n = (w_{ij})$ *is an* $n \times n$ *Hermitian matrix,* $w_{ij} \in \mathbb{C}$ *with* $Ew_{ij} = 0$ *and* $E|w_{ij}|^2 = 1$*, and* $\{w_{ij}, i \leq j\}$ *are independent satisfying condition (1.1).*

(ii) T_n *is a Hermitian nonnegative definite random matrix whose empirical spectral distribution function almost surely converges weakly to a non-random limiting distribution function* H *as* $n \to \infty$.

(iii) W_n *and* T_n *are independent.*

Remark 1. Although in Definition 1.1.1, we have assumed $Ew_{ij} = 0$ and $E|w_{ij}|^2 = 1$, the results in the next following can be easily extended over to the general case when $Ew_{ij} = \mu$ and $Var(w_{ij}) = \sigma^2$. In fact, as far as the w_{ij}'s have a common mean, by the second rank inequality of Lemma 2.1.1, we can show that $\|F^{A_n} - F^{\tilde{A}_n}\| \leq 1/n$, where $\tilde{A}_n = \frac{1}{\sqrt{n}} T_n^{1/2}(W_n - EW_n)T_n^{1/2}$ and F^{A_n}, $F^{\tilde{A}_n}$ are respectively the empirical spectral distributions of A_n, $\tilde{A}_n$. Thus F^{A_n} and $F^{\tilde{A}_n}$ converge simultaneously to the same limiting distribution function and hence μ does not show any effect on the results at all. However, the variance σ^2 of the w_{ij}'s does show effects. Nonetheless, $\sigma^{-1}A_n$ satisfies the assumptions in Definition 1.1.1 and so the next following results apply. Note that, supposing $F^{\sigma^{-1}A_n}$ denotes the empirical spectral distribution of $\sigma^{-1}A_n$, then $F^{A_n}(x) = F^{\sigma^{-1}A_n}(\sigma^{-1}x)$. This implies that, if $F^{\sigma^{-1}A_n}$ converges weakly to F and F^{A_n} to F_σ, then $F_\sigma(x) = F(\sigma^{-1}x)$ and their Stieltjes transforms satisfy $s_{F_\sigma}(z) = \sigma^{-1}s_F(\sigma^{-1}z)$. Thus from the results for $\sigma^{-1}A_n$ we can straightforwardly infer the corresponding ones for A_n.

We first prove the convergence of the empirical spectral distribution of A_n by using the Stieltjes transform method. So the limiting spectral distribution is naturally characterized by its Stieltjes transform satisfying a system of equations.

Theorem 1.1.1. (LSD of Wigner type random matrices)

Let A_n be the Wigner type random matrix defined in Definition 1.1.1. Then with probability 1*, as $n \to \infty$, the empirical spectral distribution function of A_n converges weakly to a non-random probability distribution function F for which if for any $z \in \mathbb{C}^+$*

$$\begin{cases} s(z) = -z^{-1} - z^{-1}\{p(z)\}^2 \\ p(z) = \int \frac{t}{-z-tp(z)} dH(t) \end{cases} \tag{1.1.1}$$

is viewed as a system of equations for the complex vector $(s(z), p(z))$, then the Stieltjes transform of F, $s_F(z)$, together with another function $g(z)$ analytic on $\mathbb{C}^+$, will satisfy that $(s_F(z), g(z))$ is the unique solution to (1.1.1) in the set $\{(s(z), p(z)) : Ims(z) > 0, Imp(z) \geq 0\}$.

Motivated by the work of Silverstein and Choi (1995) for the sample covariance type random matrices, we use the equations in (1.1.1) to derive some analytic properties of the limiting spectral distribution F for the Wigner type random matrices. We show that F is continuously differentiable at any point on the real line away from the origin. We give a necessary and sufficient condition of determining its support set.

We note that for any probability distribution H, if $H((-\infty, 0)) = 0$, then it can be constructed a sequence of diagonal matrices T_n with nonnegative diagonal entries such that F^{T_n} (the empirical spectral distribution of T_n) converges weakly to H as n tends to infinity. By Theorem 1.1.1, it follows that to H there must correspond a probability distribution function F whose Stieltjes transform $s_F(z)$ together with some function $g(z)$ satisfy the two equations in (1.1.1). Let us refer to F as the Wigner type limiting spectral distribution corresponding to H.

Theorem 1.1.2. (Fundamental properties of the LSD)

(1). *(i) Suppose $\{H_k\}$ is a sequence of probability distribution functions with $H_k((-\infty, 0)) = 0$ converging weakly to H. Let $\{F_k\}$ and F be the Wigner type limiting spectral distribution functions corresponding to $\{H_k\}$ and H. Then F_k converges weakly to F.*

Suppose, in the remainder of the present theorem, H and F are as defined in Definition 1.1.1 and Theorem 1.1.1.

(ii) $F(\{0\}) = H(\{0\})$, which implies that $F(x) = I_{\{0\}}(x)$ if and only if $H(t) = I_{\{0\}}(t)$, where $I_{\{0\}}(\cdot)$ is the indicator function on the singleton set $\{0\}$.

Suppose, in the remainder of the present theorem, $H(t) \neq I_{\{0\}}(t)$.

(2). *For any* $x \in \mathbb{R}^+ \cup \mathbb{R}^-$, *as* $z \in \mathbb{C}^+$ *tends to* x, $g(z)$ *and* $s_F(z)$ *converge. Let* $g(x)$ *and* $s_F(x)$ *denote their limits respectively. Then on* $\mathbb{R}^+ \cup \mathbb{R}^-$, $g(x)$ *and* $s_F(x)$ *are continuous and satisfy*

$$\begin{cases} s_F(x) = -x^{-1} - x^{-1}\{g(x)\}^2, \\ g(x) = \int t/\{-x - tg(x)\}dH(t), \end{cases} \tag{1.1.2}$$

with $Re(g(x))/x < 0$ *and* $\int t^2/|x + tg(x)|^2 dH(t) \leq 1$.

Consequently, on $\mathbb{R}^+ \cup \mathbb{R}^-$, $F(x)$ *is continuously differentiable with derivative*

$$f(x) = -2Re(g(x))Im(g(x))/(\pi x). \tag{1.1.3}$$

(3). (*i*) F *is symmetric, i.e.* $F(x) = 1 - F(-x)$ *for any* $x \in \mathbb{R}$.

(*ii*) *Let* $\tilde{S}_F = \{x \in \mathbb{R}^+ \cup \mathbb{R}^- : f(x) > 0\}$. *Then*

$$\tilde{S}_F = \{x \in \mathbb{R}^+ \cup \mathbb{R}^- : g = \textstyle\int t/\{-x - tg\}dH(t) \text{ has a solution in } \mathbb{C}^+\}$$

and $f(x)$ *is analytic in* $\tilde{S}_F$.

(*iii*) *For any* $x_0 \in \mathbb{R}^+ \cup \mathbb{R}^-$, $x_0 \in S_F^c$ *if and only if there exists some* $\delta_0 > 0$ *such that* $\int t^2/|x + tg(x)|^2 dH(t) < 1$ *for any* $x \in (x_0 - \delta_0, x_0 + \delta_0)$, *where* S_F *denotes the support of* F *and* $g(x)$ *is defined in* (2).

We also present a way to calculate the density function of the limiting spectral distribution for two important classes of the matrices T_n, the sample covariance matrices and the inverse matrices of the sample covariance matrices.

Theorem 1.1.3. (Density function of the LSD)

(1) *When H denotes the LSD of the inverse sample covariance matrices with ratio index $y' \in (0,1)$, then F has a density function*

$$f_1(x) = \begin{cases} -2g_1\sqrt{3g_1^2 - 2g_1/(xy') + (1-1/y')}/(\pi x), & 0 < x^2 < a_1, \\ \frac{1}{\pi}(\frac{4}{3}y' + 1)\sqrt{\frac{7}{3}}, & x = 0, \\ 0, & o.w., \end{cases}$$

where

$$a_1 = \frac{(2y'^2 + 5y' - 1/4) + \sqrt{32y'^3 + 12y'^2 + (3/2)y' + 1/16}}{2y'(1-y')^3},$$

and

$$g_1 = \frac{1}{3xy'} + (\frac{-t+\sqrt{\Delta}}{2})^{\frac{1}{3}} + (\frac{-t-\sqrt{\Delta}}{2})^{\frac{1}{3}},$$

with

$$t = \frac{1}{36x^3y'^3}(\frac{1}{3} + 3x^2y'(y' + \frac{1}{2})),$$

and

$$\Delta = \frac{1}{432x^4y'^4}[1 + (2y'^2 + 5y' - \frac{1}{4})x^2 - y'(1-y')^3x^4].$$

(2) *When H denotes the LSD of the sample covariance matrices with ratio index $y > 0$, then away from 0, F has a density function*

$$f_2(x) = \begin{cases} \frac{2}{\pi}\sqrt{\frac{g_1^2}{x^2}}\sqrt{g_1^2 + (\frac{1}{2} + \frac{1}{2y}) - \frac{1}{4y}\sqrt{\frac{x^2}{g_1^2}}}, & 0 < x^2 < a_2, \\ \frac{1}{\pi|1-y|}\sqrt{\frac{1}{2} + \frac{1}{2y} - \frac{1}{2y}|1-y|}, & x = 0, \\ 0, & o.w., \end{cases}$$

where

$$a_2 = \frac{-2(1+y)^3 + 72y(1+y) + 2(1+y^2+14y)^{3/2}}{27y},$$

and

$$g_1^2 = -\frac{1+y}{6y} + \frac{\sqrt{1+y^2+14y}}{6y}\cos\frac{\varphi}{3},$$

with $\varphi \in (0, \pi)$,

$$\cos\varphi = \frac{2(1+y)^3 - 72y(1+y) + 27x^2y}{2(1+y^2+14y)^{3/2}}.$$

When $y > 1$, F *has an additional point mass* $1 - 1/y$ *at the origin.*

Finally, we use the moment method to prove the convergence of the empirical spectral distribution of the Wigner type random matrix. Through a combinatorial argument, we get the explicit expression of the moments of the limiting spectral distribution.

Theorem 1.1.4. *Under the assumptions of Definition 1.1.1, with probability one, as* $n \to \infty$, *the empirical spectral distribution* F^{A_n} *converges weakly to a non-random limiting distribution function* F. *In the case when* H *possesses moments of all orders and for each positive integer* p, *the* p*-th moment of* F^{T_n} *almost surely converges to the* p*-th moment of* H, *the limiting spectral distribution* F *possesses moments of all orders and if* m_k *denotes the* k*-th moment of* F, *then* $m_0 = 1$ *and for* $l \geq 1$, $m_{2l-1} = 0$ *and*

$$m_{2l} = \sum_{s=0}^{l-1} g_{2s} g_{2(l-1-s)}, \tag{1.1.4}$$

where with α_p denoting the pth moment of $H(t)$, $g_0 = \alpha_1$, for $s \geq 1$, g_{2s} is given by

$$g_{2s} = \sum_{q=1}^{s} \sum_{\substack{j_1+j_2+\cdots+j_q=s+1-q \\ j_1+2j_2+\cdots+qj_q=s}} \frac{s!}{q!j_1!j_2!\cdots j_q!} \alpha_1^q \alpha_2^{j_1} \alpha_3^{j_2} \cdots \alpha_{q+1}^{j_q}. \tag{1.1.5}$$

Our results provide a theoretical foundation for further investigation of other spectral properties of this class of random matrices. Compared with the large dimensional sample covariance matrices, for which there have accumulated quite significant some results, the Wigner type random matrices need more works to reveal their many interesting spectral properties. It is of interest to consider, for example, the previously mentioned problem, which contains the convergence of the spectral norms of the matrices, that requires the limiting behavior of the eigenvalues of the matrices outside the support of the limiting spectral distribution, or the usually requested result which proves central limit theorems concerning some statistics of the matrices. Reading those derivations which have been developed for the purpose of establishing these type results for the large sample covariance matrices, we can expect that the derivations and analytic results we developed in our work would be helpful when such considerations are devoted to the large Wigner type random matrices.

1.2 Large Dimensional General Sample Covariance Matrices

1.2.1 The Problem and the Objective

The so-called general sample covariance matrices in the present thesis take the form $(1/N)T_{2n}^{1/2}X_n^*T_{1n}X_nT_{2n}^{1/2}$. Note that if T_{1n} is also assumed to be nonnegative definite, then they generalize the matrices $(1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$ to the case where statistical correlations are present for both the row vectors and the column vectors of the matrix X_n. Nonetheless, for the general sample covariance matrices, we propose to deal with the very general assumption that T_{1n} is Hermitian and T_{2n} is nonnegative definite. Thus they are the very much generalizations of the matrices $(1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$. This also explains why we call them the general sample covariance matrices. Applications of these general sample covariance matrices cover all those of the matrices $(1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$ and, more important, include many new matrices of considerable interest in statistical methods and applied areas. Therefore, it is of significant value to develop spectral analysis for these general sample covariance matrices.

Spectral analysis of this class of random matrices has been of interest for a long period. To applied areas, only results on the special cases of diagonal T_{1n} and T_{2n} have been accessible (Tulino and Verdú (2005)). A natural idea is to extend those spectral analysis arguments developed on $(1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$ to the general sample covariance matrices. However, this is not as so straightforward as a simple extension, especially when the fundamental stage of spectral analysis of showing convergence of ESD is concerned.

Examining those results reviewed for the matrices $(1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$, we can find that while the Stieltjes transform method plays an essentially important role, the idea of manipulating the Stieltjes transform method by taking the columns of the matrix X_n^* as perturbations to the resolvent matrices $((1/N)T_n^{1/2}X_n^*X_nT_n^{1/2} - zI)^{-1}$ has run through all derivations. Here we remark that the Stieltjes transforms of the empirical spectral distributions of the matrices are just given by the divided by n trace of the resolvent matrices. For the matrices $(1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$, this idea is rather sensible since it provides quite some convenience, concerning finding contributing terms in asymptotic discussions in order to catch promptly the limiting behavior of the main relations involved in the derivations. But for the general sample covariance matrices, it is not appropriate to follow the same way to manipulate the Stieltjes transform method, otherwise it may happen either the contributing terms are mixed together or the residual terms sum up to an uncontrollable level.

Therefore our objective is then to pursue a way suitable for the general sample covariance matrices in which the Stieltjes transform method can be manipulated for systematic investigations of the spectral properties of the general sample covariance matrices. The starting point for us is to seek the limiting spectral distribution of the general sample covariance matrices.

1.2.2 Main Result

The general sample covariance matrices we propose to deal with are defined as follows.

Definition 1.2.1. (General sample covariance matrix)

Let $B_n = (1/N)T_{2n}^{1/2}X_nT_{1n}X_n^*T_{2n}^{1/2}$. *Then* B_n *is said to be a general sample covariance matrix if the following conditions are satisfied.*

(i) $X_n = [x_{ij}]$ *is* $N \times n$ *consisting of independent complex random variables with* $Ex_{ij} = 0$, $E|x_{ij}|^2 = 1$ *satisfying for each* $\delta > 0$, *as* $n \to \infty$,

$$\frac{1}{\delta^2 nN}\sum_{ij} E\left(|x_{ij}|^2 I_{(|x_{ij}|>\delta\sqrt{n})}\right) \to 0. \tag{1.2.1}$$

(ii) T_{1n} *is* $n \times n$ *Hermitian and* T_{2n} *is* $N \times N$ *Hermitian nonnegative definite.*

(iii) With probability 1, *as* $n \to \infty$, *the empirical spectral distributions of* T_{1n} *and* T_{2n}, *denoted by* $F^{T_{1n}}$ *and* $F^{T_{2n}}$, *converge weakly to two probability functions* H_1 *and* H_2, *respectively.*

(iv) $N = N(n)$ *with* $n/N \to c > 0$.

(v) X_n, T_{1n}, T_{2n} *are independent.*

Under these assumptions, at the present stage, we proved as n tends to infinity, the empirical spectral distributions of the general sample covariance matrices converge weakly to a non-random limiting spectral distribution whose Stieljtes transform is given by a system of equations.

Theorem 1.2.1. (LSD of general sample covariance matrix)

Let B_n be the general sample covariance matrices defined in Definition 1.2.1. Then with probability 1*, as $n \to \infty$, the empirical spectral distribution of B_n converges weakly to some non-random probability distribution function $\underline{F}$ for which if $H_1 \equiv 1_{[0,\infty)}$ or $H_2 \equiv 1_{[0,\infty)}$, then $\underline{F} \equiv 1_{[0,\infty)}$; otherwise if for each $z \in \mathbb{C}^+$,*

$$\begin{cases} s(z) &= -z^{-1}(1-c) - z^{-1}c\int \frac{1}{1+q(z)x}dH_1(x) \\ s(z) &= -z^{-1}\int \frac{1}{1+p(z)y}dH_2(y) \\ s(z) &= -z^{-1} - p(z)q(z) \end{cases} \tag{1.2.2}$$

is viewed as a system of equations for the complex vector $(s(z), p(z), q(z))$, then the Stieltjes transform of $\underline{F}$, denoted by $s_{\underline{F}}(z)$, together with two other functions, denoted by $g_1(z)$ and $g_2(z)$, both of which are analytic on $\mathbb{C}^+$, will satisfy that $(s_{\underline{F}}(z), g_1(z), g_2(z))$ is the unique solution to (1.2.2) in the set

$$\begin{aligned} \tilde{U} &= \{(s(z), p(z), q(z)) : Ims(z) > 0, \\ &\qquad Im(zp(z)) > 0, Imq(z) > 0\}. \end{aligned} \tag{1.2.3}$$

Regarding the previously stated project to develop a new way to make feasible the application of the Stieltjes transform method, our accomplishment is to change by taking the (i,j)th element of X_n as perturbations to the resolvent matrices $(\frac{1}{N}T_{2n}^{\frac{1}{2}}X_nT_{1n}X_n^*T_{2n}^{\frac{1}{2}} - zI)^{-1}$. This change is rather natural but may sound trivial. However, it is not the case because there arises the need of solving the difficulty about how to develop methods in seeking the limiting spectral distribution. As a consequence, we presented a way to trace from simply the well known resolvent identity to the final desirable system of equations which determine the limiting spectral distribution but are not revealed to us in advance.

Furthermore, in the treatment of the general sample covariance matrices, we also need to face different computational difficulties. In fact, since there is only one nonnegative definite matrix T_n included, the matrix $(1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$ is also nonnegative definite, in which case it is rather simple to get the bounds for quantities involving its resolvent matrix. However, the general sample covariance matrix $\frac{1}{N}T_{2n}^{\frac{1}{2}}X_nT_{1n}X_n^*T_{2n}^{\frac{1}{2}}$ having the same eigenvalues as the matrix $\frac{1}{N}T_{2n}X_nT_{1n}X_n^*$, can only be regarded as the product of a nonnegative definite matrix and a Hermitian matrix. So in our treatment, we also have invented a new mathematical tool best suited to computations involving the resolvent matrices of products of a nonnegative definite matrix and a Hermitian matrix.

In conclusion, we presented an appropriate way of studying the Stieltjes transforms of the empirical spectral distributions of the general sample covariance matrices as well as an efficient approach of estimating quantities involving their resolvent matrices. With the aid of these methods, we proved the convergence of the empirical spectral distributions of the general sample covariance matrices and identified the limiting spectral distribution through a system of equations determining its Stieltjes transform. Based on the obtained results, further investigations of other spectral properties of the general sample covariance matrices can be carried out systematically by manipulating the Stieltjes transform method.

1.3 Large Dimensional Sparse Random Matrices

1.3.1 Literature Review

Large sparse random matrices have important applications in many fields. The motivations of using sparse random matrices are due to various aspects of considerations. Mainly, however, sparse random matrices are natural choice of models when the real systems cannot be observed completely or are not of full connectivity. Such situations occur very often in practical fields. For instances, in nuclear physics, since the particles move in a very high velocity in a small range, many exciting states in very short time cannot be observed; in neural network theory, the number of neurons in one person's brain is probably of several orders of magnitude larger than that of the dendrites connected with one individual neuron (Grenander and Silverstein (1977)). In general, whenever a real physical system cannot be observed completely or is not of full connectivity, the random matrix describing the states of the particles contained in the system or the interactions between the particles in the system will have a large proportion of zero elements taking the place of unobserved states or absent interactions.

Results on large sparse random matrices as well as their applications can be found in various areas. For example, see Barry and Pace (1999) and Boley and Goehring (2000) for results applied to linear algebra, Bekakos and Bartzi (1999) and Grenander and Silverstein (1977) to neural networks, Botta and Wubs (1999) and NeuB (2002) and Vassilevski (2002) to algorithms and computing, Mart, Glover and Campos (2001) to finance modelling, McKenzie and Bell (2001) to electrical engineering, Naulin (2002) to Bio-interactions, and Stariolo, Curado and Tamarit (1996) to theoretical physics.

Sparse random matrices can be properly expressed by the Hadamard products of two matrices. In each Hadamard product, one matrix is chosen to represent the physical nature of the system which the product is used to model, while the other matrix is called a sparsing matrix since the entries in this matrix indeed play the role of wearing down in some sense the magnitude of the entries in the other matrix prescribed.

The specialization of either of the two matrices is under determination. For the matrix other than the sparsing one, a common choice in literature is the sample covariance matrix suitably normalized. This is of course due to the highly attention received by the class of sample covariance matrices in a variety of fields. We shall therefore follow this convention. For the sparsing matrix, there are different definitions and assumptions appearing in the literature, whereas there can be found a common attribute shared by their interpretations of the key concept sparseness relating to sparse random matrices. This common attribute is just that the sparseness nature of sparse random matrices is typically embodied by letting the entries in the sparsing matrix be identically distributed random variables which take only values 0 and 1 but take value 1 with a so small probability that the sum of these probabilities per row or per column is of a smaller order of magnitude than the sample size relating to the sample covariance matrix.

Among these works, Kohrunzhy and Rodgers (1997, 1998)) developed comparatively systematic analytic results for such sparse random matrices, namely, the

Hadamard products of the normalized sample covariance matrix B_m (whose formal definition will be specified later) and a sparsing matrix $D_m = [d_{ij}]$ with the d_{ij}'s satisfying the above described conditions and further $p_{ij} = \frac{\alpha}{m^\beta}$, with $0 \leq \beta \leq 1$ and $0 < \frac{\alpha}{m^\beta} < 1$. They showed for such defined sparse random matrices, in probability their empirical spectral distributions converge weakly to the semicircle law. They also explained that this phenomenon of convergence to the semicircle law is due to the reason that the sparsing matrix (or dilute matrix in their terminology) has asymptotically gotten rid of the correlations between the entries of the normalized sample covariance matrix.

1.3.2 The Problem and the Objective

The sparseness described above is indeed a kind of zero-one and homogeneous sparseness. Its nature indeed confines the potential of sparse random matrices in applications. Thus we propose to extend the notion of sparseness to the case of non-zero-one and non-homogeneous sparseness, *i.e.* the entries of the sparsing matrices are not required to be identical distributed random variables anymore and are allowed to take non-zero-one values. We shall then seek the limiting spectral distribution for the sparse random matrices assumed such non-zero-one and non-homogeneous sparseness.

1.3.3 Main Result

We achieved our object by formulating the concept of sparseness in terms of the moments of the entries of the sparsing matrix. It is conceivable such formulation bypasses the problem of putting restrictions on the range of values taken by the entries of the sparsing matrix in an artificial way, but leaves the right to the sparsing factors themselves as long as their behavior does not violate the rule we put on their moments. As a consequence, under the non-zero-one and non-homogeneous sparseness assumption, the sparse random matrices provide many advantageous variabilities in modelling real systems. For example, it may happen, but obviously with small possibility, that some entries in the sparsing matrix take very big values which amplify the magnitude of the corresponding entries of the normalized sample covariance matrices. We have a detailed discussion on possible consequences of our result useful and interesting for practical considerations in Section 5.3.

In more details, our conditions on the sparsing matrix are as follows. We use the row (or column) sums of the second moments of the entries of the sparsing matrix to represent the level of sparseness, *i.e.* we require such row (or column) sums to be of similar order of magnitude with each other but to be of smaller order of magnitude than the sample size. Then we require all higher moments of order bigger than 2 of the entries of the sparsing matrix to be bounded uniformly by constant multiples of the second moments of them. This requirement is reasonable if we take into account of the basic rule concerning a sparsing matrix that its entries should take big values with small probabilities. Our condition on the first moments of the entries of the sparsing matrix is such that as a certain parameter varies in a closed interval, the restrictiveness level of the condition also varies. In the weakest case, the condition is a consequence of that on the second moments of the sparsing factors and so imposes nothing additional. In the strongest case, the condition requires that the row (or column) sums of the first moments of the sparsing factors are bounded by a constant multiple of the row (or column) sums of their second moments. For such defined sparse random matrices, we proved with probability one their empirical spectral distributions converge weakly to the semicircle law.

Definition 1.3.1. (Sparsing matrix)

Let $D_m = [d_{ij}]$ be $m \times m$ Hermitian matrix. Then D_m is said to be a sparsing matrix if the following conditions are satisfied.

$(D1)$ $\{d_{ij} : i \leq j\}$ *are independent complex random variables.*

$(D2)$ $\max_j |\sum_{i=1}^m p_{ij} - p| = o(p)$, *where* $p_{ij} = E|d_{ij}^2|$.

$(D3.1)$ *For some* $\delta \in [0, 1/2]$, *there exists a constant* $C_1 > 0$ *such that* $\max_j \sum_i E|d_{ij}| \leq C_1 m^\delta p^{1-\delta}$.

$(D3.2)$ *For each* $k > 2$ *there is a constant* C_k *such that* $E|d_{ij}|^k \leq C_k p_{ij}$.

Definition 1.3.2. (Normalized sample covariance matrix [1])

Let $B_m = (1/\sqrt{np})(X_{m,n}X_{m,n}^* - n\sigma^2 I_m)$, *where* $X_{m,n}$ *is* $m \times n$ *consisting of*

[1]For definition commonly used, see Definition 5.3.1.

independent complex random variables. Then B_m is said to a normalized sample covariance matrix, for which the following conditions are assumed to be satisfied.

$(X1)$ $Ex_{ij} = 0$, $E|x_{ij}|^2 = \sigma^2$.

$(X2.1)$ *For any* $\eta > 0$, $\frac{1}{mn}\sum_{ij} E|x_{ij}^2|I[|x_{ij}| > \eta\sqrt[4]{np}] \to 0$.

$(X2.2)$ *For any* $\eta > 0$, $\sum_{u=1}^{\infty} \frac{1}{mn}\sum_{ij} E|x_{ij}^2|I[|x_{ij}| > \eta\sqrt[4]{np}] < \infty$, *where u may be taken to be $[p]$, m, or n.*

$(X3)$ *For any* $\eta > 0$, $\frac{1}{m}\sum_{i=1}^{m} P(|\sum_{k=1}^{n}(|x_{ik}|^2 - \sigma^2)d_{ii}| > \eta\sqrt{np}) \to 0$.

Definition 1.3.3. (Large sparse random matrices)

Let $A_p = B_m \circ D_m$. Then A_p is said to be a sparse random matrix if the following conditions are satisfied.

(i) The matrix B_m and the matrix D_m are respectively defined in Definitions 1.3.1 and 1.3.2.

(ii) The entries of D_m are independent of those of $X_{m,n}$.

(iii) $p/n \to 0$ and $p \to \infty$.

(iv) Condition $(D3.1)$ holds for $\delta = 1/2$ and $m/n \to 0$; or condition $(D3.1)$ holds for some $\delta \in (0, 1/2)$ and $m \leq Kn$ for some constant K; or condition $(D3.1)$ holds for $\delta = 0$ and no restrictions between m and n.

Theorem 1.3.1. (LSD of large sparse random matrices)

Let A_p be the sparse random matrices defined in Definition 1.3.3. Then as $[p] \to \infty$, the empirical spectral distributions of A_p converge weakly to the semicircle *law* $F_{sc,\sigma^2}(x)$ *with scale parameter* σ^2*, which is given by*

$$\frac{d}{dx}F_{sc,\sigma^2}(x) = \begin{cases} \frac{1}{2\pi\sigma^4}\sqrt{4\sigma^4 - x^2}, & \text{if } |x| \leq 2\sigma^2, \\ 0, & \text{otherwise.} \end{cases} \tag{1.3.1}$$

The convergence is in probability if condition (X2.1) is assumed and the convergence is in the sense of almost surely for $[p] \to \infty$ *or* $m \to \infty$ *if condition (X2.2) is assumed for* $u = [p]$ *or* $u = m$ *respectively.*

In conclusion, the sparse random matrices discussed in our results form a very general class which has included many interesting random matrices which cannot be studied in the usual context of spectral analysis of large sparse random matrices. Nevertheless, we conjecture the class can still be generalized further to the Hadamard products of a normalized version of the sample covariance matrix $\frac{1}{N}XTX^*$ and a sparsing matrix with its conditions suitably adjusted.

Chapter 2

Methodologies

The present chapter is intended to elaborate two important methodologies in the spectral analysis of large dimensional random matrices. They are widely known as the moment method and the Stieltjes transform method. The many results we reviewed in the previous chapter are all derived by means of these two methods. In this chapter, we are aiming to discuss in more detail the manipulations of these two methods in the investigation of spectral properties of large random matrices, particularly in establishing limiting spectral distributions for random matrices which are closely related with the three classes of random matrices to be studied in the present thesis.

2.1 Preliminary Notions and Tools

The empirical spectral distribution of a random matrix is obviously a key concept in the spectral analysis of large random matrices. This notion is, in the present thesis, only defined for random matrices which possesses only real eigenvalues.

Definition 2.1.1. (Empirical spectral distribution)

Let A_n be an $n \times n$ matrix having real eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. Then the empirical spectral distribution of A_n is defined as

$$F^{A_n}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(\lambda_i \leq x)},$$

where $I_{(\cdot)}$ denotes the indicator function of the set $(\cdot)$.

Definition 2.1.2. (Limiting spectral distribution)

If for a class of random matrices A_n, it is proven with probability one (or in probability) as n tends to infinity the empirical spectral distribution F^{A_n} converges weakly to some distribution function F, then F is said to be the limiting spectral distribution of the matrices A_n in the strong sense (or in the weak sense).

In the present thesis, we shall use the convention that for any matrix A having real eigenvalues only, F^A denotes its empirical spectral distribution and that unless the opposite is stated, a limiting spectral distribution refers to a limiting spectral distribution in the strong sense.

When the empirical spectral distributions of two matrices are being compared to infer some information on the similarity of their limiting spectral distributions, we shall need the following inequalities concerning the distance or difference between two empirical spectral distributions (See proofs in Bai (1999)).

Lemma 2.1.1. (Rank inequality)

(1) *Let A and B be two $n \times n$ Hermitian matrix. Then*

$$\|F^A - F^B\| \leq \frac{1}{n} rank(A - B). \tag{2.1.1}$$

(2) *Let A and B be two $n \times N$ complex matrices. Then*

$$\|F^{AA^*} - F^{BB^*}\| \leq \frac{1}{n} rank(A - B). \tag{2.1.2}$$

Here $\| \cdot \|$ denotes the maximum norm which is defined for any function f to be $\|f\| = \sup_x |f(x)|$.

Lemma 2.1.2. (Difference inequality)

(1) *Let A and B be two $n \times n$ Hermitian matrix. Then*

$$L^3(F^A, F^B) \leq \frac{1}{n} tr(A - B)^2. \tag{2.1.3}$$

(2) *Let A and B be two $n \times N$ complex matrices. Then*

$$L(F^{AA^*}, F^{BB^*}) \leq \frac{2}{n^2} tr((A - B)(A - B)^*) tr(AA^* + BB^*). \tag{2.1.4}$$

Here $L(F, G)$ denotes the Levy distance between distribution functions F and G, which is defined to be

$$L(F, G) = \inf\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon, \text{ for all } x \in \mathbb{R}\}.$$

The next several results are contained in most textbooks on probability theory.

Lemma 2.1.3. (Bernstein's inequality)

Let $x_1, \cdots, x_n$ be independent random variables with $Ex_i = 0$, $Ex_i^2 = \sigma_i^2$, $|x_i| \leq b$. Then for any $\varepsilon > 0$,

$$P\left(|\sum_{i=1}^{n} x_i| \geq \varepsilon\right) \leq 2exp\left\{-\frac{\varepsilon^2}{2(\sum_i \sigma_i^2 + b\varepsilon)}\right\}.$$

Lemma 2.1.4. (Burkholder's inequality)

Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then for $p \geq 2$,

$$E|\sum X_k|^p \leq K_p \left(E\left(\sum E(|X_k|^2|\mathcal{F}_{k-1})\right)^{p/2} + E\sum |X_k|^p\right).$$

Lemma 2.1.5. (Borel-Cantelli's lemma)

If E_n is a sequence of events on probability space $(\Omega, \mathcal{F}, P)$ such that $\sum_n P(E_n)$ converges, then $P(\limsup_n E_n) = 0$.

Lemma 2.1.6. (Helly's selection theorem)

For every sequence $\{F_n\}$ of probability distribution functions there exists a subsequence $\{F_{n_k}\}$ and a nondecreasing, right-continuous function F such that $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ at all continuity points x of F.

Lemma 2.1.7. (Corollary to Helly's selection theorem)

If $\{\mu_n\}$ is a tight sequence of probability measures, and if each subsequence that converges weakly all converges weakly to the probability measure μ, then $\{\mu_n\}$ converges weakly to μ.

Suppose for any rectangular matrix A, $\sqrt{AA^*}$ denotes the matrix resulting from replacing the eigenvalues in the spectral decomposition of AA^* with their square roots.

Lemma 2.1.8. (Lemma 2.3 in Silverstein and Bai (1995))

Let x_1, x_2, x_3 be arbitrary non-negative numbers. For A, B, C square matrices of the same size,

$$\begin{aligned} F^{\sqrt{(ABC)(ABC)^*}}\{(x_1x_2x_3,\infty)\} &\leq F^{\sqrt{AA^*}}\{(x_1,\infty)\} + F^{\sqrt{BB^*}}\{(x_2,\infty)\} \\ &\quad + F^{\sqrt{CC^*}}\{(x_3,\infty)\}. \end{aligned}$$

Lemma 2.1.9. (The resolvent identity)

Let A and B be two nonsingular matrices of the same type. Then the following equality is called the resolvent identity:

$$A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}(or = B^{-1}(B-A)A^{-1}). \tag{2.1.5}$$

2.2 Moment Method

In the spectral analysis of large dimensional random matrices, the moment method appeared more than ten years earlier than the Stieltjes transform method and has played an important role in the field ever since Wigner firstly used this method to prove the famous semicircle law in his ground-breaking work (Wigner (1955)). This method studies the moments of empirical spectral distributions of a class of large random matrices to study their spectral properties.

2.2.1 Use of the Moment Method

In most situations, the moment method is used to seek limiting spectral distributions and to prove limiting theorems on extreme eigenvalues. In this latter aspect, by using the moment method, a necessary and sufficient condition to guarantee the almost sure convergence of the largest eigenvalues of Wigner random matrices to the largest number in the support of the semicircle law was proved in Bai and Yin (1988b) (See also Bai (1999)). For sample covariance matrices $\frac{1}{N}X_n^*X_n$, Yin, Bai and Krishnaiah (1988) and Bai, Silverstein and Yin (1988) proved respectively the existence of a finite fourth moment of the entries x_{ij} of the $N \times n$ matrix X_n, which are also assumed to be independent and identically distributed random variables, is sufficient and necessary for the largest eigenvalues of the sample covariance matrices to converge almost surely to the largest number in the support of its limiting spectral distribution, *i.e.* the Marcěnko-Pastur distribution. Previous or related works on this problem are Geman (1980, 1986) and Bai and Yin (1986). The almost sure convergence of the smallest eigenvalues of the sample covariance matrices was solved in Bai and Yin (1993). This result indeed proved altogether the almost sure convergence of the largest and the smallest eigenvalues of the sample covariance matrices through discussing moments of empirical spectral distributions of the centralized matrices $\frac{1}{N}X_n^*X_n - (1+c)I$, where $c = \lim \frac{n}{N} \in (0,1)$.

When the moment method is used to seek limiting spectral distributions, the underlying theory foundation is the moment convergence theorem in probability theory together with a condition known as Carleman's condition.

Lemma 2.2.1. (Moment convergence theorem)
Suppose that the probability distribution function F is determined by its moments, that the probability distribution function F_n has moments of all orders, and that for every positive integer k, the kth moment of F_n converges to the kth moment of F. Then F_n converges weakly to F.

There are various criteria in probability theory on deciding whether a probability distribution function is determined by its moments. One of these is given by Carleman which says that if a probability distribution function F has moments of all order, say $\{m_k\}_{k=1}^{\infty}$, then F is determined by its moments whenever

$$\sum_{k=1}^{\infty} m_{2k}^{-\frac{1}{2k}} = +\infty. \tag{2.2.1}$$

Suppose the limiting spectral distribution of random matrices A_n is pursued in the strong sense (or in the weak sense) by means of the moment method. Then with the moment convergence theorem, a feasible procedure is to show for each k, the kth moment of F^{A_n} converges almost surely (or in probability) to a non-random limit, which determines a non-random probability distribution function, say F. The main difficulty lying in this procedure is to find those limits of the moments of F^{A_n}. To conquer the difficulty, there needs usually a very complicated combinatoric argument. This accounts for, but not completely, why the moment method is not so preferable as the Stieltjes transform method.

2.2.2 Examples of Obtaining LSD's by Using the Moment Method

We shall introduce below some results on applying the moment method to seek the limiting spectral distributions of the Wigner matrices and the sample covariance matrices. Here we note that all these matrices are Hermitian matrices. Then we can use the fact that if A_n is $n \times n$ Hermitian, then the kth moment of F^{A_n} is equal to $\frac{1}{n}tr(A_n^k)$.

Theorem 2.2.1. (LSD of the Wigner matrices)

Let $W_n = [w_{ij}]$ be an $n \times n$ Hermitian random matrix. Assume that $\{w_{ij} : i \leq j\}$ are independent random variables with $Ew_{ij} = 0$, $E|w_{ij}|^2 = \sigma^2$ satisfying condition (1.1). Then with probability one, as $n \to \infty$, $F^{\frac{1}{\sqrt{n}}W_n}$ converges weakly to the *semicircle law $F_{sc,\sigma}(x)$ given by (1.3.1) with σ taking the place of σ^2 therein.*

Proof. Consider any sequence of numbers $\{a_m\}_{m=1}^{\infty}$ such that as $m \to \infty$, $a_m \downarrow 0$. Then for each fixed m, there exists n_m such that whenever $n \geq n_m$,

$$\frac{1}{a_m^2 n^2} \sum_{ij} E|w_{ij}|^2 1_{(|w_{ij}| > a_m \sqrt{n})} < a_m.$$

Let $\delta_n = a_m$ for $n \in [n_m, n_{m+1})$, $m = 1, 2, \cdots$. Then as $n \to \infty$, $\delta_n \downarrow 0$ and

$$\frac{1}{\delta_n^2 n^2} \sum_{ij} E|w_{ij}|^2 1_{(|w_{ij}| > \delta_n \sqrt{n})} \to 0. \tag{2.2.2}$$

This means one can select a sequence $\delta_n \downarrow 0$ such that condition (1.1) remains valid with δ replaced by δ_n.

The proof of the theorem can be done in the following procedure:

(a) Let $\tilde{W}_n = [\tilde{w}_{ij}]$ with (i,i)th entry $\tilde{w}_{ii} = 0$ and (i,j)th entry $\tilde{w}_{ij} = \hat{w}_{ij} - E\hat{w}_{ij}$, where $\hat{w}_{ij} = w_{ij} 1_{(|w_{ij}| \leq \delta_n \sqrt{n})}$.

(b) By Lemmas 2.1.1-2.1.3, almost surely, $L(F^{\frac{1}{\sqrt{n}}\tilde{W}_n}, F^{\frac{1}{\sqrt{n}}W_n}) \to 0$. Hence with probability one $F^{\frac{1}{\sqrt{n}}\tilde{W}_n}$ and $F^{\frac{1}{\sqrt{n}}W_n}$ converge weakly to the same limiting distribution.

(c) The moment method can be applied to the matrix $\frac{1}{\sqrt{n}}\tilde{W}_n$. Show that

$$E\frac{1}{n} tr(\frac{1}{\sqrt{n}}\tilde{W}_n)^k \to m_k = \begin{cases} 0, & k \text{ is odd }, \\ \frac{(2s)!\sigma^{2s}}{s!(s+1)!}, & k = 2s, \end{cases} \tag{2.2.3}$$

and that

$$E\left(\frac{1}{n} tr(\frac{1}{\sqrt{n}}\tilde{W}_n)^k - E\frac{1}{n} tr(\frac{1}{\sqrt{n}}\tilde{W}_n)^k\right)^4 = O(\frac{1}{n^2}).$$

By Lemma 2.1.5, almost surely,

$$\frac{1}{n}tr(\frac{1}{\sqrt{n}}\tilde{W}_n)^k \to m_k.$$

(d) Check $\{m_k\}$ satisfies Carleman's condition (2.2.1). $\square$

Theorem 2.2.2. (LSD of the sample covariance matrices)

Let $S_n = \frac{1}{N}X_n^*X_n$, where $X_n = [x_{ij}]$ is $N \times n$ consisting of independent random variables with $Ex_{ij} = 0$, $E|x_{ij}|^2 = \sigma^2$. Assume that as $n \to \infty$, $n/N \to c > 0$ and that for any $\delta > 0$,

$$\frac{1}{\delta^2 nN}\sum_{ij} E|x_{ij}|^2 I_{(|x_{ij}|>\delta\sqrt{n})} \to 0. \tag{2.2.4}$$

Then with probability one, as $n \to \infty$, F^{S_n} converges weakly to the Marčenko-Pastur distribution with ratio index c and scale parameter σ^2, denoted by $F_{M-P}^{c,\sigma^2}(x)$ which has a point mass of $1 - 1/c$ at the origin when $c > 1$ and a density function

$$\frac{d}{dx}F_{M-P}^{c,\sigma^2}(x) = \begin{cases} \frac{1}{2\pi x c\sigma^2}\sqrt{(b-x)(x-a)}, & \text{if } a \le x \le b, \\ 0, & \text{otherwise}, \end{cases} \tag{2.2.5}$$

where $a = \sigma^2(1-\sqrt{c})^2$ and $b = \sigma^2(1+\sqrt{c})^2$.

Proof. Select a sequence $\delta_n \downarrow 0$ such that condition (2.2.4) remains true with δ replaced by δ_n. By following the next steps, the theorem can be proven.

(a) Let $\tilde{X}_n = [\tilde{x}_{ij}]$ with (i,j)th entry $\tilde{x}_{ij} = \hat{x}_{ij} - E\hat{x}_{ij}$, where $\hat{x}_{ij} = x_{ij}I_{(|x_{ij}|\le\delta_n\sqrt{n})}$. Define $\tilde{S}_n = \frac{1}{N}\tilde{X}_n^*\tilde{X}_n$.

(b) By Lemmas 2.1.1-2.1.3, it can be shown with probability one, $L(F^{S_n}, F^{\tilde{S}_n}) \to 0$. This means $\tilde{S}_n$ has the same liming spectral distribution as S_n.

(c) Apply the moment method to the matrix $\tilde{S}_n$. Show that as $n \to \infty$,

$$E\frac{1}{n}tr(\tilde{S}_n)^k \to c_k = \sigma^{2k}\sum_{r=0}^{k-1}\frac{c^r}{r+1}\binom{k}{r}\binom{k-1}{r}, \tag{2.2.6}$$

and

$$E\left(\frac{1}{n}tr(\tilde{S}_n)^k - E\frac{1}{n}tr(\tilde{S}_n)^k\right)^4 = O(n^{-2}).$$

By Lemma 2.1.5, almost surely $\frac{1}{n}tr(\tilde{S}_n)^k \to c_k$.

(d) Check that $\sum_{k=1}^{\infty} c_{2k}^{-\frac{1}{2k}} = +\infty$. □

Theorem 2.2.3. (LSD of the sample covariance matrices)

Assume that X_n is as defined in Theorem 2.2.2, that T_n is Hermitian, independent of X_n and has a limiting spectral distribution H in the strong sense as $n \to \infty$, and that as $n \to \infty$, $n/N \to c > 0$. Then with probability one, as $n \to \infty$, the empirical spectral distribution of S_nT_n converges weakly to a non-random limiting distribution.

The proof of this theorem needs the following result.

Lemma 2.2.1. (Lemma 2.11 of Bai (1999))

Let G^0 be a connected graph with m vertices and h edges. To each vertex $v = (1, \cdots, m)$ there corresponds a positive integer n_v, and to each edge $e_j = (v_1, v_2)$ there corresponds a matrix $T_j = [t^{(j)}_{\eta,\zeta}]$ of order $n_{v_1} \times n_{v_2}$. Let E_c and E_{nc} denote the sets of cutting edges (those edges whose removal causes the graph disconnected) and non-cutting edges, respectively. Then there is a constant C, depending upon *m and h only, such that*

$$|\sum_{i_1,\cdots,i_m}\prod_{j=1}^{h} t^{(j)}_{i_{f_{ini}(e_j)}i_{f_{end}(e_j)}}| \leq Cn \prod_{e_j\in E_{nc}} \|T_j\| \prod_{e_j\in E_c} \|T_j\|_0,$$

where $n = \max(n_1, \cdots, n_m)$, $\|T_j\|$ *denotes the maximum singular value of* T_j*, and* $\|T_j\|_0$ *equals the product of the maximum dimension and the maximum absolute value of the entries of* T_j*; in the summation* i_v *runs over* $\{1, \cdots, n_v\}$, $f_{ini}(e_j)$ *and* $f_{end}(e_j)$ *denote the initial and end vertices of the edges* e_j*.*

Proof of Theorem 2.2.3. The proof of the theorem can be carried out by following the next steps.

(a) Let T_n^τ be the resulting matrix of replacing in the spectral decomposition of T_n those eigenvalues whose absolute values are bigger than τ with 0. When τ is a continuity point of H, $F^{T_n^\tau}$ converges weakly to a non-random limit, say H^τ, with probability one. Then the theorem will follow if for all τ sufficiently large (continuity points of H), the empirical spectral distribution of ST^τ converges weakly with probability one to a non-random limiting distribution function.

(b) Define $\tilde{X}_n$ and $\tilde{S}_n$ as previously. Then $\tilde{S}_n T_n^\tau$ and $S_n T_n^\tau$ have the same limiting spectral distribution.

(c) With the aid of Lemma 2.11 in Bai (1999), show that almost surely

$$\frac{1}{n} tr(\tilde{S}_n T_n^\tau)^k \to q_k = \sigma^{2k} \sum_{s=1}^{k} c^{k-s} \sum_{\substack{i_1+i_2+\cdots+i_s=k+1-s \\ i_1+2i_2+\cdots+si_s=k}} \frac{k!\alpha_1^{i_1}\alpha_2^{i_2}\cdots\alpha_s^{i_s}}{s!i_1!i_2!\cdots i_s!}, \qquad (2.2.7)$$

where $\alpha_m = \int t^m dH^\tau(t)$.

(d) Check $\{q_k\}$ satisfies Carleman's condition (2.2.1). $\square$

All proofs outlined above can be seen in their detailed derivations in Bai (1999). Concerning the result in Theorem 2.2.3, it is noteworthy to point out that Bai (1999) for the first time improved the result on the LSD of large sample covariance matrices $\frac{1}{N}X_n^*X_nT_n$ proven in Yin and Krishnaiah (1983) and Yin (1986) for the case of nonnegative definite matrices T_n to the case of Hermitian matrices T_n. Lemma 2.11 in Bai (1999) has a proof given in Bai and Silverstein (2005PEACH). The result in this lemma provides great help when the empirical spectral distributions of multiplications of random matrices are discussed through their moments. We shall make use of this lemma in Section 3.5 to prove result for the Wigner type random matrices by means of the moment method. There one can see how the lemma is interpreted and used.

Moreover, in all the proofs, the most difficult part is hidden in step (c), where the limits of moments of the empirical spectral distributions of the random matrices are pursued. Detailed derivations of these moments can be found in Bai (1999) or in for instance, Wigner (1955) and Yin (1986) etc. We shall carry out in detail a similiar type of argument for the Wigner type random matrices in Section 3.5. Further, although the treatment in (d) seems straightforward, the treatments in (a) and (b) have somewhat been developed into a customary technique to adopt before the moment method as well as the Stieltjes transform method are to be used. This technique, known as the truncation and centralization technique, was early invented in probability theory by Kolmogorov, and has been developed methodically by Bai in spectral analysis of large random matrices. The aim of using this technique is to yield appropriate conditions on the underlying random variables to work with.

Note that the limiting spectral distribution of both the Wigner matrices and the sample covariance matrices were originally obtained by using the moment method. Other such examples are the well known Toeplitz, Hankel and Markov matrices. Indeed, for these three classes of matrices which are of great interest in applications, results are only available on the existence of their limiting spectral distributions proven by the moment method, even basic properties of the limiting spectral distributions, such as whether it has a bounded support, are still unrevealed (Bryc, Dembo and Jiang (2004)).

Proofs of Theorems 2.2.1-2.2.3 can also be obtained by means of the Stieltjes transform method. This will be one part seen in the next section. In recent decades, to somewhat extent, the Stieltjes transform method has progressively taken the role played previously by the moment method in the spectral analysis of large random matrices. In the next section, our focus is on applying the Stieltjes transform method to seek limiting spectral distributions. We introduce the basic theory and concepts related to the method.

2.3 Stieltjes Transform Method

In the present section, we shall introduce the other important method in the spectral analysis of large dimensional random matrices: the Stieltjes transform method. Compared with the moment method introduced in the previous section, the Stieltjes transform method is usually more preferable to researchers. This is mainly because the moment method is typically accompanied with a sophisticated argument of combinatoric in nature, which hinders derivations as well as scrutiny, and results obtained using the moment method tend to be difficult for latent utilizations, whereas the Stieltjes transform method usually yields more transparent and far-reaching results which are advantageous for further investigations on other considerations.

2.3.1 Fundamental Facts

We firstly introduce in details the basic concepts and facts related to the Stieltjes transform method. The Stieltjes transform method studies the Stieltjes transforms of the empirical spectral distributions of a class of random matrices to investigate their spectral properties. Thus let us begin by introducing the definition of the Stieltjes transforms.

Definition 2.3.1. (The Stieltjes transform)

Let $F(x)$ be any function of bounded variation. Then the Stieltjes transform of $F(x)$ is defined as

$$s_F(z) = \int \frac{1}{x-z} dF(x), \quad (z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : Imz > 0\}). \tag{2.3.1}$$

We have several remarks about this definition.

Remark 1. Since the integrand in (2.3.1) is bounded in absolute value by $1/v$

($v = Imz$), the integral is always well defined. Due to the same reason, the definition applies equally well when z is considered in the set $\mathbb{C}^- \equiv \{z \in \mathbb{C} : Imz < 0\}$. As can be seen below, considering the Stieltjes transform on either $\mathbb{C}^+$ or $\mathbb{C}^-$ will be sufficient for discussing the corresponding function $F(x)$ (See Theorems 2.3.1 and 2.3.3). Thus in the present thesis, unless the opposite is stated, the Stieltjes transforms are defined on $\mathbb{C}^+$. However, in seldom cases, because of computational necessity, we do use Stieltjes transforms on $\mathbb{C}^-$. In those cases, we shall clarify the definition domain explicitly, thus no confusion will be caused.

Remark 2. Hereafter for any function $F(x)$ having bounded variation, $s_F(z)$ denotes its Stieltjes transform.

Remark 3. A straightforward derivation gives for any probability distribution function F,

$$\lim_{v\to\infty} ivs_F(iv) = -1, \quad \lim_{v\to\infty} -ivs_F(-iv) = -1,$$
$$\text{and} \quad F\{0\} = \lim_{v\downarrow 0}(-ivs_F(iv)),$$

where $F\{\cdot\}$ always denotes the measure of $\{\cdot\}$ talked with respect to the measure generated on $\mathbb{R}$ by F.

The terminology of Stieltjes transform is closely related with the classical moment problem, for example, in discussing the solubility and determinateness of the power moment problem. It has a variety of generalizations and hence various forms of inversion formulas correspondingly. For the case of the Stieltjes transforms of bounded variation functions, we introduce the following inversion formula.

Theorem 2.3.1. (Inversion formula)

Let $F(x)$ be any function of bounded variation. Then for any continuity points a, b of $F(x)$,

$$F(b) - F(a) = \lim_{v \downarrow 0} \frac{1}{\pi} \int_a^b Im s_F(u + iv) du. \tag{2.3.2}$$

Proof. Note that

$$\begin{aligned}
\frac{1}{\pi} \int_a^b Im s_F(u+iv)du &= \frac{1}{\pi} \int_a^b \int \frac{v}{(x-u)^2 + v^2} dF(x) du \\
&= \frac{1}{\pi} \int \int_a^b \frac{v}{(x-u)^2 + v^2} du dF(x) \\
&= \frac{1}{\pi} \int \left\{ \arctan\left(\frac{b-x}{v}\right) - \arctan\left(\frac{a-x}{v}\right) \right\} dF(x),
\end{aligned}$$

in which as $v \downarrow 0$,

$$\arctan\left(\frac{b-x}{v}\right) - \arctan\left(\frac{a-x}{v}\right) \to \frac{\pi}{2}\Big(1_{\{a\}}(x) + 1_{\{b\}}(x)\Big) + \pi 1_{\{(a,b)\}}(x).$$

It follows, by the dominated convergence theorem, we get

$$\lim_{v \downarrow 0} \frac{1}{\pi} \int_a^b Im s_F(u+iv)du = F(b) - F(a) - \frac{1}{2}\left(F\{b\} - F\{a\}\right).$$

In the case when a and b are continuity points of $F(x)$, we get (2.3.2) immediately.

□

As a consequence of this theorem, the following result was stated in Bai (1993a) and rigorously proved in Silverstein and Choi (1995).

Theorem 2.3.2. *Let $F(x)$ be any function of bounded variation. If at $x_0 \in \mathbb{R}$, $Ims_F(x_0) = \lim_{z\in\mathbb{C}^+\to x_0} Ims_F(z)$ exists, then $F(x)$ is differentiable at x_0 with derivative equal to $\frac{1}{\pi}\lim_{v\downarrow 0} Ims_F(x_0+iv)$.*

From Theorem 2.3.1, we also have the following result.

Theorem 2.3.3. *For any two functions of bounded variation, say $F(x)$ and $G(x)$, $F \equiv G$ if and only if $s_F(z) = s_G(z)$ for all $z \in \mathbb{C}^+$.*

This theorem tells explicitly that there is a one-to-one correspondence between a function of bounded variation and its Stieltjes transform.

Note that in probability theory, the characteristic function is a powerful means of studying weak convergence of distribution functions. This is due to the famously known continuity theorem regarding probability distribution functions and their characteristic functions. We shall next construct in parallel a continuity theorem regarding probability distribution functions and their Stieltjes transforms. This continuity theorem is the theory foundation of applying the Stieltjes transform method to seek limiting spectral distributions in the field of random matrix theory. We first present the continuity theorem regarding characteristic functions.

Theorem 2.3.4. (Theorem 26.3 in Billingsley (1995))

Let μ_n, μ be probability measures with characteristic functions ϕ_n, ϕ. A necessary *and sufficient condition for μ_n to converge weakly to μ is that $\phi_n(t) \to \phi(t)$.*

Theorem 2.3.5. (Continuity theorem for Stieltjes transforms)
Suppose that $\{F_n(x)\}$ is a sequence of probability distribution functions and that $F(x)$ is a probability distribution function. Then $F_n(x)$ converges weakly to $F(x)$ if and only if $s_{F_n}(z) \to s_F(z)$ for all $z \in \mathbb{C}^+$.

Proof. The necessity part follows directly from the equivalence of the weak convergence of probability distribution functions and the convergence of integrals of bounded continuous functions. The sufficiency part is established as follows. Write $z = u + iv$. Letting

$$p_v(x) = \frac{v}{\pi[x^2 + v^2]},$$

then $p_v(x)$ is the density function of the Cauchy distribution $C_v(x)$ with scale parameter v and by Definition 2.3.1,

$$\frac{1}{\pi} Ims_{F_n}(u + iv) = \int p_v(x - u)dF_n(x).$$

This implies $Ims_{F_n}(z)$, viewed as a function of the real part u of z, is indeed the density function of the convolution of $F_n(x)$ and $C_v(x)$.

We shall use the property of characteristic functions for the convolution of distribution functions. For that purpose, let $\psi_{nv}(t)$, $\phi_n(t)$ and $\phi_v(t)$ be respectively the characteristic functions of $\frac{1}{\pi}Ims_{F_n}(z)$, $F_n(x)$ and $C_v(x)$. Let $\psi_v(t)$ and $\phi(t)$ denote the characteristic functions of $\frac{1}{\pi}Ims_F(z)$ (function of argument u) and $F(x)$.

Then $\psi_{nv}(t) = \phi_n(t)\phi_v(t)$. Similarly, $\psi_v(t) = \phi(t)\phi_v(t)$. Note that $\phi_v(t) =$

$e^{-v|t|}$. Thus we get $\phi_n(t) = e^{v|t|}\psi_{nv}(t)$ and $\phi(t) = e^{v|t|}\psi(t)$. By hypothesis,

$$\frac{1}{\pi} Ims_{F_n}(z) \to \frac{1}{\pi} Ims_F(z),$$

and so $\psi_{nv}(t) \to \psi_v(t)$ and

$$\phi_n(t) \to \phi(t).$$

By the continuity theorem for characteristic functions, it follows immediately F_n converges weakly to F. We complete the proof of the theorem. □

In applying the theorem, it should be noted that the limit of the Stieltjes transforms of probability distribution functions may not be the Stieltjes transform of a probability distribution function. That is, supposing for the sequence of probability distribution functions $\{F_n(x)\}$, we know $s_{F_n}(z)$ converges to some limit $s(z)$, then we need to determine whether $s(z)$ is the Stieljtes transform of some probability distribution function F to conclude $\{F_n\}$ converges weakly, whose limit is F then.

This problem can be dismissed when $\{F_n\}$ is known to be a tight sequence. Since in this case by Theorem 2.3.3 and the necessity part of Theorem 2.3.5, the hypothesis that $s_{F_n}(z)$ converges will imply that all weak convergent subsequences of $\{F_n(x)\}$ have the same limiting distribution function. Thus by Lemma 2.1.7, it follows $\{F_n(x)\}$ converges weakly. Again by the necessity part of Theorem 2.3.5, it must follow also the weak limit of $\{F_n(x)\}$ takes the limit of $s_{F_n}(z)$ as its Stieltjes transform. Our assertion thus is true.

However, in some other cases, we may not be able to check for a sequence of convergent Stieltjes transforms, whether the corresponding sequence of probability distribution functions is tight. Thus we introduce the following result which will generally be useful in these cases.

Theorem 2.3.6. (Continuity theorem for Stieltjes transforms) *Suppose $\{F_n(x)\}$ is a sequence of probability distribution functions. If $\lim_{n\to\infty} s_{F_n}(z) = s(z)$ for all $z \in \mathbb{C}^+$, then there exists a probability distribution function F with Stieltjes transform $s(z)$ if and only if*

$$\lim_{v\to\infty}(ivs(iv)) = -1, \tag{2.3.3}$$

in which case F_n converges weakly to F.

This theorem is a result proven in Geronimo and Hill (2002). Its proof depends on mainly showing that $s(z) \equiv \lim_{n\to\infty} s_{F_n}(z)$ satisfies the following criterion which may be of interest on its own right. Note that for any sequence of probability distribution functions $\{F_n\}$, if $\lim_{n\to\infty} s_{F_n}(z)$ exists for a subset of $\mathbb{C}^+$ possessing a limit point in $\mathbb{C}^+$, then it must exist for every point in $\mathbb{C}^+$ and if we denote by $s(z)$ its limit then $s(z)$ must be analytic function on $\mathbb{C}^+$.

Theorem 2.3.7. (A criterion for Stieltjes transforms) *Let $s(z)$ be a function analytic on $\mathbb{C}^+$. Then there exists a probability distribution function F with Stieltjes transform $s(z)$ if and only is $s(z)$ satisfies that $Im s(z) > 0$ for each $z \in \mathbb{C}^+$ and that (2.3.3) holds.*

We shall not prove this result but refer the reader to Geronimo and Hill (2002) for a proof. We now introduce a convenient form of Theorem 2.3.6.

Theorem 2.3.8. (Continuity theorem for Stieltjes transforms) *Suppose $\{F_n(x)\}$ is a sequence of probability distribution functions. Let $K \subset \mathbb{C}^+$ be an infinite set with a limit point in $\mathbb{C}^+$. If $\lim_{n\to\infty} s_{F_n}(z) = s(z)$ for all $z \in K$, then there exists a probability distribution function F with Stieltjes transform $s(z)$ if and only if (2.3.3) holds, in which case F_n converges weakly to F.*

2.3.2 Use of the Stieltjes Transform Method

The Stieltjes transform method is very useful in the spectral analysis of large random matrices. Many works appearing in recent years show that thorough investigations of the spectral properties of random matrices can be carried out systematically by means of the Stieltjes transform method. The results we reviewed in section 1.2.1 on the sample covariance matrices $B_n \equiv \frac{1}{N} T_n^{1/2} X_n^* X_n T_n^{1/2}$ are relevant examples. Under certain conditions on the eigenvalues of T_n, based on showing that as the imaginary part of z converges to 0, the Stieltjes transforms of the empirical spectral distributions of B_n converge at an appropriate rate uniformly with respect to the real part of z over certain intervals, it was shown with probability one, for any closed interval outside the support of the limiting spectral distribution of B_n, there will be no eigenvalues of B_n appearing in this interval for all n sufficiently large (Bai and Silverstein (1998)).

The exact separation result proved that under appropriate conditions, for any interval $J \subset \mathbb{R}^+$ on which no eigenvalues of T_n appear for all n large, there exists some interval I exists such that with probability one for all n large the number of eigenvalues of B_n on one side of I matching up with those of T_n on the same side of J (Bai and Slilverstein (1999)). The proof of this result is easier to obtain when $c = \lim \frac{n}{N}$ is sufficient small. Then with the aid of the Stieltjes transforms to associate intervals in the complement of the support of the limiting spectral distribution of B_n with intervals in the complement of the support of the limiting spectral distribution of T_n, the proof of the exact separation for a general limiting ratio c was obtained through strategically increasing the numbers of columns of X_n^* while keeping track of the movements of the eigenvalues of the resulting new matrices B_n.

By using the Stieltjes transforms of $G_n(x) \equiv n(F^{B_n}(x) - F^{c_n,H_n}(x))$, viewed as a random two dimensional process defined on a contour $\mathcal{C}$ of the complex plane, to prove its tightness and investigate its limiting behavior, a central limit theorem was established for statistics $(\int f_1(x)dG_n(x), \cdots, \int f_k(x)dG_n(x))$, where the $f_j(x)$ are functions analytic on a certain open interval enclosing the support of $G_n(x)$ for all n large with probability one (Bai and Silverstein (2004)). Here if $F^{c,H}$ denotes the limiting spectral distribution of B_n, then F^{c_n,H_n} denotes its parallel with c, H replaced with c_n, $H_n \equiv F^{T_n}$; H being the limiting spectral distribution of T_n.

Nonetheless, the type of the foregoing results need to be preceded by an accomplishment of using the Stieltjes transform method to find the limiting spectral distribution. For the class of random matrices $B_n \equiv \frac{1}{N} T_n^{1/2} X_n^* X_n T_n^{1/2}$, this work was done in Silverstein (1995), to which an important related work was Silverstein and Bai (1995). Their works will be introduced later in this section as examples of applying the Stieltjes transform method to seek the limiting spectral distribution for a certain class of random matrices. At the present position, we shall discuss the main principles concerning this problem.

Consider $n \times n$ Hermitian random matrices A_n defined on the probability space $(\Omega, \mathcal{F}, P)$. Then we have the following basic result.

Theorem 2.3.9. (Basic rule)
If for each $z \in \mathbb{C}^+$ with probability one $s_{F^{A_n}}(z)$ converges to a non-random limit, then with probability one F^{A_n} converges vaguely to a non-random limit F , whose Stieltjes transform satisfies $s_F(z) = \lim_{n\to\infty} s_{F^{A_n}}(z)$. If it is further known F^{A_n} is tight with probability one or $s(z)$ satisfies the criterion in Theorem 2.3.7, then F^{A_n} converges weakly to F with probability one.

Proof. Choose a sequence $\{z_m\}_{m=1}^{\infty} \subset \mathbb{C}^+$ which possesses a limit point in $\mathbb{C}^+$. For each z_m, there exists a subspace Ω_{z_m} with $P(\Omega_{z_m}) = 1$ such that for any $\omega \in \Omega_{z_m}$,

$$\lim_{n\to\infty} s_{F^{A_n}}(z_m) = s(z_m). \tag{2.3.4}$$

Let $\Omega_0 = \cap_{m=1}^{\infty} \Omega_{z_m}$. Then $P(\Omega_0) = 1$ and for any $\omega \in \Omega_0$, (2.3.4) holds for all z_m.

For each $\omega \in \Omega_0$ fixed, consider any two subsequences of $F^{A_n(\omega)}$, say $F^{A_{n_i^{(1)}}(\omega)}$

and $F^{A_{n_i^{(2)}}(\omega)}$, which converge vaguely to limits $F_\omega^{(1)}$ and $F_\omega^{(2)}$ respectively. For ease of reference, let us write $F^\omega_{n_i^{(1)}} \equiv F^{A_{n_i^{(1)}}(\omega)}$ and $F^\omega_{n_i^{(2)}} \equiv F^{A_{n_i^{(2)}}(\omega)}$. Then by the equivalence between vague convergence of probability distribution functions and the convergence of functions $f(x)$, bounded continuous and vanishing as $|x| \to \infty$, we have for every $z \in \mathbb{C}^+$,

$$\lim_{i\to\infty} s_{F^\omega_{n_i^{(1)}}}(z) = s_{F_\omega^{(1)}}(z), \quad \lim_{i\to\infty} s_{F^\omega_{n_i^{(2)}}}(z) = s_{F_\omega^{(2)}}(z). \tag{2.3.5}$$

From (2.3.4), this means $s_{F_\omega^{(1)}}(z_m) = s_{F_\omega^{(2)}}(z_m) = s(z_m)$, for all z_m.

Note that $s_{F_\omega^{(1)}}(z)$ and $s_{F_\omega^{(2)}}(z)$ are analytic functions on $\mathbb{C}^+$. By elementary properties of analytic functions, it follows $s_{F_\omega^{(1)}}(z) = s_{F_\omega^{(2)}}(z)$ for all $z \in \mathbb{C}^+$. This implies by Theorem 2.3.3 $F_\omega^{(1)} = F_\omega^{(2)}$ and hence the vague convergence of the whole sequence $\{F^{A_n(\omega)}\}$. Such vague limit, for each $\omega \in \Omega_0$, can be denoted by F_ω. Then for any $\omega_1, \omega_2 \in \Omega_0$, (2.3.4) further implies $s_{F_{\omega_1}}(z_m) = s_{F_{\omega_2}}(z_m) = s(z_m)$, for all z_m. Hence, by the same reason, we get $s_{F_{\omega_1}}(z) = s_{F_{\omega_2}}(z)$ for all $z \in \mathbb{C}^+$ and hence $F_{\omega_1} = F_{\omega_2}$.

Thus we see for every $\omega \in \Omega_0$, F^{A_n} converges vaguely with the limit not dependent on ω. Therefore, we get with probability one, F^{A_n} converges vaguely to a non-random limit, which if we denote by F, then obviously for any $\omega \in \Omega_0$, $s_F(z) = \lim_{n\to\infty} s_{F^{A_n}}(z)$, for every $z \in \mathbb{C}^+$. The second part in the theorem follows trivially. This completes the proof of the theorem. $\square$

2.3.3 Examples of Obtaining LSD's by Using the Stieltjes Transform Method

We first note that if A_n is $n \times n$ Hermitian matrix, then the Stieltjes transform of the empirical spectral distribution of A_n is equal to $\frac{1}{n} tr(A_n - zI)^{-1}$, where $(A_n - zI)^{-1}$ is known as the resolvent of A_n. Also, hereafter we use the convention that on the complex plane, the square root function $\sqrt{(\cdot)}$ denotes the branch that has positive imaginary part on $\mathbb{C}^+$. We first introduce a proof given in Bai (1999) for the Wigner matrices.

Let A_k denote the resulting matrix by deleting from A_n the kth row and the kth column. Let β_k denote the kth column vector of A_n with its kth element removed. Let α_k denote the kth row vector of A_n with its kth element removed. Then if A_n and A_k are both nonsingular, it can be shown by using the inverse formula for partitioned matrices

$$A_n^{-1}(k,k) = \frac{1}{a_{kk} - \alpha_k{}' A_k^{-1} \beta_k}, \tag{2.3.6}$$

where $A_n^{-1}(k,k)$ denotes the (k,k)th element of A_n^{-1} and a_{kk} the (k,k)th element of A_n.

Theorem 2.3.10. (LSD of the Wigner matrices)

Let $W_n = [w_{ij}]$ be $n \times n$ Hermitian consisting of independent and identically distributed random variables with $Ew_{ij} = 0$, $E|w_{ij}|^2 = \sigma^2$. Then with probability one, $F^{\frac{1}{\sqrt{n}} W_n}$ converges weakly to the semicircle law $F_{sc,\sigma}$ with scale parameter σ whose

Stieltjes transform is given by

$$s_{F_{sc,\sigma}}(z) = \frac{1}{2\sigma^2}\left(-z + \sqrt{z^2 - 4\sigma^2}\right), \textit{ for every } z \in \mathbb{C}^+.$$

Proof. (a) With the aid of Lemmas 2.1.1-2.1.3 as well as the law of large numbers , it can be shown without loss of generality, we may assume in W_n, $w_{ii} = 0$ and $w_{ij} \le C$ for $i \neq j$, where C is some positive constant.

(b) Apply the Stieltjes transform method to the matrix $\frac{1}{\sqrt{n}}W_n$. Denote by $s_n(z)$ the Stieltjes transform of $F^{\frac{1}{\sqrt{n}}W_n}$. Then

$$\begin{aligned} s_n(z) &= \frac{1}{n}\sum_{k=1}^{n} \frac{1}{-z - \frac{1}{n}\alpha_k^*(n^{-\frac{1}{2}}W_k - zI_{n-1})^{-1}\alpha_k} \\ &= \frac{1}{n}\sum_{k=1}^{n} \frac{1}{-z - \sigma^2 s_n(z) + \varepsilon_k} \\ &\equiv -\frac{1}{z + \sigma^2 s_n(z)} + \delta_n(z), \end{aligned}$$

where W_k denotes the resulting matrix of removing from W_n its kth row and kth column and α_k the kth column of W_n with its kth element removed.

(c) Show that almost surely

$$\max_{1\le k\le n} |\varepsilon_k| \equiv \max_{1\le k\le n} |\sigma^2 s_n(z) - \frac{1}{n}\alpha_k^*(\frac{1}{\sqrt{n}}W_k - zI_{n-1})^{-1}\alpha_k| \to 0,$$

which implies

$$\begin{aligned} |\delta_n(z)| &= |\frac{1}{n}\sum_{k=1}^{n} \frac{-\varepsilon_k}{(-z - \sigma^2 s_n(z) + \varepsilon_k)(-z - \sigma^2 s_n(z))}| \\ &\le \frac{1}{v^2}\max_{1\le k\le n}|\varepsilon_k| \\ &\to 0. \end{aligned}$$

(d) By calculation,

$$s_n(z) = \frac{1}{2\sigma^2}\left(-z - \delta_n(z)\sigma^2 + \sqrt{(z-\delta_n(z)\sigma^2)^2 - 4\sigma^2}\right),$$

from which it follows almost surely,

$$s_n(z) \to s(z) = \frac{1}{2\sigma^2}\left(-z + \sqrt{z^2 - 4\sigma^2}\right),$$

which is the Stieltjes transform[1] of the semicircle law F_{sc,σ^2}. $\square$

Theorem 2.3.11. (LSD of the sample covariance type matrices)

Assume that

(i) $X_n = [x_{ij}]$ be $N \times n$ consisting of independent and identically distributed random variables with $Ex_{ij} = 0$, $Var(x_{ij}) = 1$.

(ii) $T_n = diag(\tau_1, \cdots, \tau_N)$ is real and with probability one F^{T_n} converges weakly to a probability distribution function H as $n \to \infty$.

(iii) A_n is $n \times n$ Hermitian and with probability one F^{A_n} converges vaguely to a non-random limit F_a as $n \to \infty$.

(iv) X_n, T_n and A_n are independent.

(v) As $n \to \infty$, $N = N(n) \to \infty$ with $n/N \to c > 0$.

Let $B_n = A_n + \frac{1}{N}X_n^* T_n X_n$. Then with probability one, F^{B_n} converges vaguely to a non-random distribution function F, whose Stieltjes transform $s_F(z)$ satisfies for each $z \in \mathbb{C}^+$,

$$s_F(z) = s_{F_a}\left(z - \int \frac{tdH(t)}{1+cts_F(z)}\right). \tag{2.3.7}$$

[1]For $z \in \mathbb{C}^-$, replace the sign '+' before the square root by '−'.

Proof. (a) Without loss of generality, one may assume further $|x_{ij}| \le \ln n$ and $\|T_n\| \le \tau$ (τ is some positive constant).

(b) Write $\mu_n = \frac{1}{N}\sum_{j=1}^{N} \frac{\tau_j}{1+\frac{n}{N}\tau_j s_{F^{B_n}}(z)}$. Then from the resolvent identity,

$$\begin{aligned} & s_{F^{B_n}}(z) \\ = \ & s_{F^{A_n}}(z-\mu_n) - \frac{1}{n} tr\{(A_n-(z-\mu_n)I)^{-1}(\frac{1}{N}X_n^* T_n X_n - \mu_n I)(B_n - zI)^{-1}\}. \end{aligned}$$

Thus it is conceivable, from Theorem 2.3.9, the result will follow once it is shown the second term on the right-hand side of the foregoing relation tends to zero.

(c) Write $X_n^* = [x_1, x_2, \cdots, x_N]$ and $B_{(j)} = B_n - \frac{1}{N}\tau_j x_j x_j^*$, where x_j is the jth column of X_n^*. Then $\frac{1}{N}X_n^* T_n X_n = \frac{1}{N}\sum_{j=1}^{N} \tau_j x_j x_j^*$,

$$\begin{aligned} & \frac{1}{n} tr\{\frac{1}{N}X_n^* T_n X_n (B_n - zI)^{-1}(A_n - (z-\mu_n)I)^{-1}\} \\ = \ & \frac{1}{N}\sum_{j=1}^{N} \frac{\frac{1}{n}\tau_j x_j^*(B_{(j)} - zI)^{-1}(A_n-(z-\mu_n)I)^{-1}x_j}{1+\frac{1}{N}\tau_j x_j^*(B_{(j)}-zI)^{-1}x_j} \\ = \ & \frac{1}{N}\sum_{j=1}^{N} \frac{\tau_j}{1+\frac{n}{N}\tau_j s_{F^{B_n}}(z)} \frac{(1+\frac{n}{N}\tau_j s_{F^{B_n}}(z))\frac{1}{n}x_j^*(B_{(j)}-zI)^{-1}(A_n-(z-\mu_n)I)^{-1}x_j}{1+\frac{1}{N}\tau_j x_j^*(B_{(j)}-zI)^{-1}x_j}. \end{aligned}$$

It follows that

$$\begin{aligned} & \frac{1}{n} tr\{(A_n-(z-\mu_n)I)^{-1}(\frac{1}{N}X_n^* T_n X_n - \mu_n I)(B_n-zI)^{-1}\} \\ = \ & \frac{1}{N}\sum_{j=1}^{N} \frac{\tau_j}{1+\frac{n}{N}\tau_j s_{F^{B_n}}(z)} \\ & \times \Bigg(\frac{(1+\frac{n}{N}\tau_j s_{F^{B_n}}(z))\frac{1}{n}x_j^*(B_{(j)}-zI)^{-1}(A_n-(z-\mu_n)I)^{-1}x_j}{1+\frac{1}{N}\tau_j x_j^*(B_{(j)}-zI)^{-1}x_j} \\ & -\frac{1}{n} tr\{(B_n-zI)^{-1}(A_n-(z-\mu_n)I)^{-1}\}\Bigg) \\ \equiv \ & \frac{1}{N}\sum_{j=1}^{N} \frac{\tau_j d_j}{1+\frac{n}{N}\tau_j s_{F^{B_n}}(z)}. \end{aligned}$$

(d) Show that $\max_{1\leq j\leq N}|d_j|\to 0$, almost surely. $\square$

Remark 4. If in Theorem 2.3.11, with the other conditions unchanged but we have $Var(x_{ij})=\sigma^2$, then the Stieltjes transform of the limiting spectral distribution of B_n will satisfy

$$s_F(z)=s_{F_a}\left(z-\sigma^2\int\frac{tdH(t)}{1+c\sigma^2ts_F(z)}\right). \tag{2.3.8}$$

In fact, for any Hermitian matrix A_n, we have $s_{F^{A_n}}(z)=as_{F^{aA_n}}(az)$, for any constant $a>0$. It follows $s_{F^{A_n/\sigma^2}}(z)\to\sigma^2 s_{F^a}(\sigma^2 z)$. Write $\tilde{B}_n=\frac{1}{\sigma^2}B_n$ and denote by $\tilde{F}$ its limiting spectral distribution. Then applying Theorem 2.3.11 to $\tilde{B}_n$ gives

$$s_{\tilde{F}}(z)=\sigma^2 s_{F_a}\left(\sigma^2 z-\sigma^2\int\frac{tdH(t)}{1+cts_{\tilde{F}}(z)}\right).$$

Again, $B_n=\sigma^2\tilde{B}_n$ implies $s_{\tilde{F}}(z)=\sigma^2 s_F(\sigma^2 z)$. It follows (2.3.8) immediately.

Remark 5. Suppose in Theorem 2.3.11, $Var(x_{ij})=\sigma^2$. If $A_n=O_n$ and $T_n=I_n$, then B_n reduces to the ordinary sample covariance matrix $S_n=\frac{1}{N}X_n^*X_n$, for which the limiting spectral distribution is known to be F_{M-P}^{c,σ^2}, the Marcĕnko-Pastur distribution with scale parameter σ and ratio parameter c. Then (2.3.8) implies the Stieltjes transform[2] of F_{M-P}^{c,σ^2}, denoted by $s(z)$ simply, is given by for $z\in\mathbb{C}^+$,

$$s(z)=\frac{-z+(1-c)\sigma^2+\sqrt{[z-(1+c)\sigma^2]^2-4c\sigma^4}}{2c\sigma^2 z}. \tag{2.3.9}$$

[2]For $z\in\mathbb{C}^-$, replace the sign '+' before the square root in the nominator by '$-$'.

To see this, we note that with $s_{F_a}(z) = -\frac{1}{z}$ and $H\{1\} = 1$, (2.3.8) gives us the following equation:

$$c\sigma^2 z[s(z)]^2 + [z - (1-c)\sigma^2]s(z) + 1 = 0.$$

This equation has two solutions, but for $z \in \mathbb{C}^+$, only one of them satisfies the condition in Theorem 2.3.7, which is just the one in (2.3.9).

Theorem 2.3.12. (LSD of the sample covariance matrices)

Assume that

(i) $X_n = [x_{ij}]$ *is* $N \times n$ *consisting of independent and identically distributed random variables with* $Ex_{ij} = 0$, $Var(x_{ij}) = 1$.

(ii) T_n *is* $n \times n$ *Hermitian nonnegative definite and with probability one* F^{T_n} *converges weakly to a probability distribution function* H *as* $n \to \infty$.

(iii) X_n, T_n *are independent.*

(iv) As $n \to \infty$, $N = N(n) \to \infty$ *with* $n/N \to c > 0$.

Let $B_n = \frac{1}{N} T_n^{1/2} X_n^* X_n T_n^{1/2}$. *Then with probability one,* F^{B_n} *converges weakly to a non-random distribution function* F, *whose Stieltjes transform* $s_F(z)$ *satisfies*

$$s(z) = \int \frac{1}{t(1-c-czs(z)) - z} dH(t) \tag{2.3.10}$$

in the sense that, for each $z \in \mathbb{C}^+$, $s(z) = s_F(z)$ *is the unique solution to (2.3.10) in* $\mathcal{D} \equiv \{s(z) \in \mathbb{C}^+ : -z^{-1}(1-c) + cs(z) \in \mathbb{C}^+\}$.

Proof. Note that we have the relations

$$F^{\underline{B}_n} = \left(1 - \frac{n}{N}\right) I_{[0,\infty)} + \frac{n}{N} F^{B_n},$$
$$s_{F^{\underline{B}_n}}(z) = -z^{-1}\left(1 - \frac{n}{N}\right) + \frac{n}{N} s_{F^{B_n}}(z),$$

where $\underline{B}_n = \frac{1}{N} X_n T_n X_n^*$. Then $F^{\underline{B}_n}$ and F^{B_n} converge simultaneously. Let $\underline{F}$ be the limiting spectral distribution of $F^{\underline{B}_n}$. It suffices to show

$$s_F(z) = \int \frac{1}{-z - z s_{\underline{F}}(z) t} dH(t). \tag{2.3.11}$$

For notational convenience, write $s_n(z) = s_{F^{B_n}}(z)$ and $\underline{s}_n(z) = s_{F^{\underline{B}_n}}(z)$. The proof can be obtained by following the next steps.

(a) Let $X_n^* = [x_1, x_2, \cdots, x_N]$, where x_j denotes the jth column of X_n^*. Let $r_j = \frac{1}{\sqrt{N}} T_n^{1/2} x_j$ and $B_{(j)} = B_n - r_j r_j^*$. Then $B_n = \sum_{j=1}^N r_j r_j^* = B_{(j)} + r_j r_j^*$ and by the resolvent identity of Lemma 2.1.9

$$\frac{1}{n} tr\{B_n(B_n - zI)^{-1}\} = \frac{1}{n} \sum_{j=1}^N \frac{r_j^*(B_{(j)} - zI)^{-1} r_j}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j},$$

which implies

$$\frac{1}{n} tr(B_n - zI)^{-1} = -z^{-1}\left(1 - \frac{N}{n}\right) - z^{-1} \frac{1}{n} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j},$$

and so

$$\underline{s}_n(z) = -z^{-1} \frac{1}{N} \sum_{j=1}^N \frac{1}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j}.$$

(b) Show that

$$\begin{aligned} &\frac{1}{n} tr\{B_n(B_n - zI)^{-1}(-zI - z\underline{s}_n(z) T_n)^{-1}\} \\ = \ &\frac{1}{n} \sum_{j=1}^N \frac{r_j^*(B_{(j)} - zI)^{-1}(-zI - z\underline{s}_n(z) T_n)^{-1} r_j}{1 + r_j^*(B_{(j)} - zI)^{-1} r_j}. \end{aligned}$$

(c) By the resolvent identity

$$\begin{aligned}
&\frac{1}{n}tr(-zI - z\underline{s}_n(z)T_n)^{-1} - \frac{1}{n}tr(B_n - zI)^{-1} \\
= \quad &\frac{1}{n}tr\{(-zI - z\underline{s}_n(z)T_n)^{-1}(B_n + z\underline{s}_n(z)T_n)(B_n - zI)^{-1}\} \\
= \quad &\frac{1}{n}tr\{B_n(B_n - zI)^{-1}(-zI - z\underline{s}_n(z)T_n)^{-1}\} \\
&+z\underline{s}_n(z)\frac{1}{n}tr\{T_n(B_n - zI)^{-1}(-zI - z\underline{s}_n(z)T_n)^{-1}\} \\
\equiv \quad &\frac{1}{n}\sum_{j=1}^{N}\frac{\varepsilon_j}{1 + r_j^*(B_{(j)} - zI)^{-1}r_j},
\end{aligned}$$

where

$$\begin{aligned}
\varepsilon_j \quad \equiv \quad &r_j^*(B_{(j)} - zI)^{-1}(-zI - z\underline{s}_n(z)T_n)^{-1}r_j \\
&-\frac{1}{N}tr\{T_n(B_n - zI)^{-1}(-zI - z\underline{s}_n(z)T_n)^{-1}\}.
\end{aligned}$$

(d) Show that $\max_{1\le j\le N}|\varepsilon_j| \to 0$ almost surely. Since $|\frac{1}{1+r_j^*(B_{(j)}-zI)^{-1}r_j}| \le \frac{|z|}{v}$, it follows then almost surely

$$\frac{1}{n}tr(-zI - z\underline{s}_n(z)T_n)^{-1} - \frac{1}{n}tr(B_n - zI)^{-1} \to 0.$$

(e) Based on the knowledge that the equation (2.3.10) has at most one solution in the set $\mathcal{D}$, by discussing any convergent subsequence of $s_n(z)$ must converge to a point in $\mathcal{D}$ satisfying equation (2.3.10), it follows the convergence of the whole sequence of $\{s_n(z)\}$ to a such limit. By Theorem 2.3.9, with a further result that $\{F^{B_n}\}$ is tight with probability one, the theorem is proven. $\square$

Now that we have proven the existence of the limiting spectral distribution, we can proceed to investigate its analytic properties. The following result was proven in Silverstein and Choi (1995).

Theorem 2.3.13. (Analytic properties of the LSD)

Let F be the limiting spectral distribution obtained in Theorem 2.3.12. Then for all $x \in \mathbb{R}$, $x \neq 0$,

$$\lim_{z \in \mathbb{C}^+ \to x} s_F(z) \equiv s_0(x) exists.$$

The function $s_0(x)$ is continuous on $\mathbb{R}/\{0\}$. Consequently, F has a continuous derivative on $\mathbb{R}/\{0\}$ given by $f(x) = \frac{1}{\pi} Im s_0(x)$. The density f is analytic (possesses a power series expansion) for every $x \neq 0$ for which $f(x) > 0$. Moreover, for these x, $\pi f(x)$ is the imaginary part of the unique $s \in \mathcal{D}$ satisfying

$$x = -\frac{1}{s} + c \int \frac{t dH(t)}{1 + ts}.$$

Now we consider the multivariate F-matrices as example to illustrate the use of Theorems 2.3.11-2.3.13. Let us firstly introduce the definition of multivariate F-matrices. Let $B_n = S_n V_n^{-1}$ with $S_n = \frac{1}{N} X_n^* X_n$ and $V_n = \frac{1}{N'} Y_n^* Y_n$. Assume that $X_n = [x_{ij}]$ is $N \times n$ consisting of independent standard normal random variables and that $Y_n = [y_{ij}]$ is $N' \times n$ consisting of independent standard normal random variables. Also assume that as $n \to \infty$, $N = N(n) \to \infty$ and $N' = N'(n) \to \infty$ with $n/N \to y > 0$ and $n/N' \to y' \in (0, 1)$. Then B_n is said to be a multivariate F-matrix.

By Theorem 2.3.11 for $A_n = O_n$ and $T_n = I_n$, with probability one F^{V_n} converge weakly to the Marcĕnko-Pastur distribution with ratio index y' and scale parameter 1, which is simply denoted here by $F_{y'}$. From (2.3.9) (see its footnote), for $z \in \mathbb{C}^-$, the Stieltjes transform of $F_{y'}$ is given by

$$s_{y'}(z) = \frac{-z + 1 - y' - \sqrt{(1 + y' - z)^2 - 4y'}}{2y'z}. \tag{2.3.12}$$

Let $T_n = V_n^{-1}$. Then we have with probability one $F^{T_n}(x) = F^{V_n}\{[\frac{1}{x}, \infty)\}1_{(0,\infty)}(x)$ and $s_{F^{T_n}}(z) = -z^{-1} - z^{-2}s_{F^{V_n}}(z^{-1})$ for all n large. It follows that with probability one, F^{T_n} converges weakly to $H(x) = \left(1 - F_{y'}\left(\frac{1}{x}\right)\right)1_{(0,\infty)}(x)$ and that for each $z \in \mathbb{C}^+$ the Stiejtjes transform of H is given by

$$\begin{aligned} s_H(z) &= \frac{z^{-2} - z^{-1}(1 + y') + z^{-1}\sqrt{(1 + y' - z^{-1})^2 - 4y'}}{2y'} \\ &= \frac{1 - z - zy' + \sqrt{(1 + z - zy')^2 - 4z}}{2y'z^2}. \end{aligned} \tag{2.3.13}$$

Noting that B_n has the same eigenvalues as $T_n^{1/2}S_nT_n^{1/2}$, by Theorem 2.3.12, we have with probability 1, the limiting spectral distribution of the multivariate F-matrices exists. More specifically, let $F_{y,y'}$ denote this limiting spectral distribution. Then for each $z \in \mathbb{C}^+$, it follows that $s_{F_{y,y'}}(z)$ satisfies equation (2.3.10), which can be equivalently written as

$$s(z) = \frac{1}{1 - y - yzs(z)} s_H\left(-\frac{1}{-z^{-1}(1 - y) + ys(z)}\right),$$

where $s_H(\cdot)$ is given by (2.3.13) and $-z^{-1}(1 - y) + ys_{F_{y,y'}}(z) \in \mathbb{C}^+$. Using the expression given by (2.3.13) yields

$$(y'z^2 + yz)s^2_{F_{y,y'}}(z) + [z(1 + y') - (1 - y)]s_{F_{y,y'}}(z) + 1 = 0,$$

which gives two analytic solutions

$$s_{1,2}(z) = \frac{(1-y) - z(1+y') \pm \sqrt{[z(1-y')+(1-y)]^2 - 4z}}{2z(y+y'z)}. \quad (2.3.14)$$

Compute that

$$\begin{aligned} ivs_1(iv) &= \frac{(1-y) - iv(1+y') + \sqrt{[iv(1-y')+(1-y)]^2 - 4iv}}{2(y+iy'v)} \\ &= \frac{\frac{(1-y)}{v} - i(1+y') + \sqrt{[i(1-y')+\frac{(1-y)}{v}]^2 - 4i\frac{1}{v}}}{2(\frac{y}{v}+iy')}. \end{aligned}$$

Letting $v \to \infty$, $[i(1-y')+\frac{(1-y)}{v}]^2 - 4i\frac{1}{v} \to -(1-y')^2$. By our convention on the square root function $\sqrt{(\cdot)}$ clarified at the beginning of this section, $\sqrt{[i(1-y')+\frac{(1-y)}{v}]^2 - 4i\frac{1}{v}} \to (1-y')i$. It follows $\lim_{v\to\infty} ivs_1(iv) = -1$. Similarly, we have $\lim_{v\to\infty} ivs_2(iv) = -\frac{1}{y'} < -1$. By Theorem 2.3.7, $s_2(z)$ cannot be the Stieltjes transform of any probability distribution function. Thus for each $z \in \mathbb{C}^+$, $s_{F_{y,y'}}(z)$ is given by $s_1(z)$. However, it can be shown for for each $z \in \mathbb{C}^-$, $s_{F_{y,y'}}(z)$ is given by $s_2(z)$.

From Theorem 2.3.13, we can further consider the density of $F_{y,y'}$. Then we first calculate its possible point mass at the origin, which by Remark 3 in the present section must be equal to

$$\lim_{v\downarrow 0}(-ivs_{F_{y,y'}}(iv)) = \frac{(1-y)+|1-y|}{2y}.$$

Now for each $x \in \mathbb{R}/\{0\}$, by Theorem 2.3.13,

$$\frac{d}{dx}F_{y,y'}(x) = \lim_{v\downarrow 0}\frac{1}{\pi}Im s_{F_{y,y'}}(x+iv). \quad (2.3.15)$$

Since

$$\lim_{v\downarrow 0} Im\left(\frac{(1-y)-z(1+y')}{2z(y+y'z)}\right)=0,$$
$$\lim_{v\downarrow 0} Im\left(\sqrt{[z(1-y')+(1-y)]^2-4z}\right)$$
$$= \begin{cases} \sqrt{4x-[x(1-y')+(1-y)]^2}, & \text{if } [x(1-y')+(1-y)]^2<4x, \\ 0, & \text{if } [x(1-y')+(1-y)]^2\geq 4x, \end{cases}$$

we get for each $x \in \mathbb{R}/\{0\}$,

$$\frac{d}{dx}F_{y,y'}(x)=\begin{cases} \frac{\sqrt{4x-[x(1-y')+(1-y)]^2}}{2\pi x(y+y'x)}, & \text{if } [x(1-y')+(1-y)]^2<4x, \\ 0, & \text{if } [x(1-y')+(1-y)]^2\geq 4x. \end{cases} \tag{2.3.16}$$

Thus we obtained the density function of the limiting spectral distribution of multivariate F-matrices.

Remark 6. Given any $z \in \mathbb{C}^+$, since in the expression of $s_1(z)$ in (2.3.14) if $y = 0$, then we get the Stieltjes transform of the limiting spectral distribution of the matrices V_n^{-1}, we may in this sense say that the limiting spectral distribution of the inverse matrices of the sample covariance matrices is a special case of the limiting spectral distribution of the multivariate F-matrices. Similarly, taking $y' = 0$ in the expression of $s_{F_{y,y'}}(z)$ yields the Stieltjes transform of the Marcĕnko-Pastur distribution with scale parameter 1 and ratio parameter y. So we are also allowed to say the limiting spectral distribution of the sample covariance matrices is a special case of the limiting spectral distribution of the multivariate F-matrices.

In the next chapter, we shall apply the Stieltjes transform method to the class of the Wigner type random matrices. We shall present in details our derivations of the results introduced in Section 1.1.4, in which Theorem 1.1.3 is concerned with the density function of the limiting spectral distribution in view that the results in Theorems 1.1.1.and 1.1.2 furnish us respectively the existence and differentiability of the limiting spectral distribution.

Chapter 3

Wigner Type Random Matrices

The present chapter is concerned with the limiting spectral distribution of the Wigner type random matrices defined in Definition 1.1.1. Mainly, we shall concern ourselves with the following results for the Wigner type random matrices:

- The almost sure existence of the limiting spectral distribution proven by means of the Stieltjes transform method.
- Some analytic properties of the limiting spectral distribution.
- Calculating the density function of the limiting spectral distribution for two important cases of given T_n.
- Moment method proof of the existence of the limiting spectral distribution.

The organization of this chapter is as follows. In Section 3.1, we prove some preliminary results necessary for proving Theorem 1.1.1. In Section 3.2, we prove Theorem 1.1.1 by means of the Stieltjes transform method. In Section 3.3, we derive the limiting spectral distribution possesses those analytic properties outlined in Theorem 1.1.2. In Section 3.4, we demonstrate a method of calculating the density of the limiting spectral distribution and compute the two density functions given in Theorem 1.1.3. In Section 3.5, we present an alternative proof of the existence of the limiting spectral distribution by using the moment method.

In the remainder of the thesis, K denotes a constant and may take on different values from one appearance to the next; for any complex number z, $Re(z)$ and $Im(z)$ denote respectively its real and imaginary part.

3.1 Preliminary Results.

This section is mainly devoted to deriving preliminary results necessary for proving Theorem 1.1.1.

3.1.1 Two Basic Lemmas: Tightness and Unique Solvability

Lemma 3.1.1. *With probability one, $\{F^{A_n}\}$ is a tight sequence.*

Proof. From Lemma 2.1.8, we have for any two positive numbers x_1 and x_2,

$$\begin{aligned} & F^{A_n}\{\lambda : |\lambda| > x_1x_2\} = F^{\sqrt{A_nA_n^*}}\{(x_1x_2, \infty)\} \\ \leq \; & F^{\sqrt{(\frac{1}{\sqrt{n}}W_n)^2}}\{(x_1, \infty)\} + 2F^{T_n^{1/2}}\{(\sqrt{x_2}, \infty)\} \\ \leq \; & F^{\frac{1}{\sqrt{n}}W_n}\{(-\infty, -x_1) \cup (x_1, \infty)\} + 2F^{T_n}\{(x_2, \infty)\}. \end{aligned} \tag{3.1.1}$$

Under assumption (i) of Theorem 1.1.1, it is known $F^{\frac{1}{\sqrt{n}}W_n}$ converges almost surely to the semicircle law (Bai (1999)). Thus with probability one, $\{F^{\frac{1}{\sqrt{n}}W_n}\}$ is tight. Assumption (ii) guarantees F^{T_n} is tight with probability one. Thus from (3.1.1), we get with probability one, $\{F^{A_n}\}$ is tight. $\square$

To implement the Stieltjes transform method, we need to show the unique solvability of the system of equations (1.1.1) in Theorem 1.1.1, which is obviously a consequence of the next result.

Lemma 3.1.2. *For each $z \in \mathbb{C}^+$, equation*

$$p(z) = \int \frac{t}{-z - tp(z)} dH(t), \tag{3.1.2}$$

has at most one solution in the set $\mathcal{B} \equiv \{p \in \mathbb{C} : Imp \geq 0\}$.

Proof. It is clear $p = 0$ is the only solution to equation (3.1.2) in the case $H(t) = 1_{\{0\}}(t)$. Now assume $H(t) \neq 1_{\{0\}}(t)$. For any $z = z_1 + iz_2$ with $z_2 > 0$, fixed. Let $p = p_1 + ip_2 \in \mathcal{B}$ satisfy equation (3.1.2). Then we have

$$p_2 = z_2 \int \frac{t}{|z + tp|^2} dH(t) + p_2 \int \frac{t^2}{|z + tp|^2} dH(t). \tag{3.1.3}$$

Since $z_2 > 0$ and $H\{0\} < 1$, the first term on the right hand side is positive. This implies $p_2 > 0$ and

$$\int \frac{t^2}{|z + tp|^2} dH(t) < 1. \tag{3.1.4}$$

Suppose for the fixed z, there is another $\tilde{p} = \tilde{p}_1 + i\tilde{p}_2 \in \mathcal{B}$ satisfying equation (3.1.2). Then

$$p - \tilde{p} = \int \frac{t^2}{(z+tp)(z+t\tilde{p})} dH(t)(p - \tilde{p}). \tag{3.1.5}$$

If $p \neq \tilde{p}$, by using Holder's inequality and (3.1.4), we have

$$\begin{aligned} & \int \frac{t^2}{(z+tp)(z+t\tilde{p})} dH(t) \\ < \ & \{\int \frac{t^2}{|z+tp|^2} dH(t)\}^{\frac{1}{2}} \{\int \frac{t^2}{|z+t\tilde{p}|^2} dH(t)\}^{\frac{1}{2}} \\ < \ & 1. \end{aligned} \tag{3.1.6}$$

It follows from (3.1.5) $p = \tilde{p}$. The proof is complete. □

Note that Theorem 1.1.1 asserts the existence of a such function $g(z) \in \mathcal{B}$ satisfying equation (3.1.2). The meaning of $g(z)$ will be clear later, whereas some of its basic properties can be derived from the equation.

Remark 1. If $H(t) \neq 1_{\{0\}}(t)$, then $g(z) \in \mathbb{C}^+$ for every $z \in \mathbb{C}^+$; otherwise, $g(z) \equiv 0$. This is a consequence of the fact that if there exists some $z^* \in \mathbb{C}^+$ such that $Img(z^*) = 0$, then $H(t) \equiv 1_{\{0\}}(t)$. To see this, let $g^* = g(z^*)$. If $Img(z^*) = 0$, then $\int t/|z^* + tg^*|^2 dH(t) = 0$, which implies $H(t) = 1_{\{0\}}(t)$ since $H\{(-\infty, 0)\} = 0$.

3.1.2 Simplification of Assumptions by Using Truncation and Centralization Technique

We shall see in the following, with the aid of the truncation and centralization technique, in proving Theorem 1.1.1, we can assume three more conditions hold without loss of generality. These three conditions are summarized as follows.

Note that it can be chosen a sequence of numbers δ_n such that $\delta_n \to 0$ as $n \to \infty$ while condition (1.1) remains true with δ replaced by δ_n.

Assumption 3.1.1.

(i) *There exists some constant* τ *such that* $\sup_n \|T_n\| \leq \tau$.

(ii) $Ew_{ij} = 0$, $|w_{ij}| \leq \delta_n \sqrt{n}$, $0 \leq \frac{1}{\delta_n^2 n^2} \sum_{ij}(1 - E|w_{ij}|^2) \to 0$ *as* $n \to \infty$.

(iii) *The matrices* T_n *are non-random.*

Consider condition (i) in Assumption 3.1.1. For any constant $\tau > 0$, denote by $\tilde{T}_n$ the resulting matrix of replacing in the spectral decomposition of T_n those eigenvalues bigger than τ with 0. Then $\|\tilde{T}_n\| \leq \tau$ and

$$F^{\tilde{T}_n}(t) = 1(t > \tau) + \{F^{T_n}(t) + 1 - F^{T_n}(\tau)\}1(0 \leq t \leq \tau).$$

As τ is a continuity point of $H(t)$, with probability one, $F^{\tilde{T}_n}(t)$ converges weakly to

$$H_\tau(t) = 1(t > \tau) + \{H(t) + 1 - H(\tau)\}1(0 \leq t \leq \tau). \tag{3.1.7}$$

For brevity, in the remainder, although we shall not clarify, the constant τ is always taken to be a continuity point of H.

Define $\tilde{A}_n = (1/\sqrt{n})\tilde{T}_n^{1/2} W_n \tilde{T}_n^{1/2}$. To have some intuition on the relationship between $F^{\tilde{A}_n}$ and F^{A_n}, we use Lemma 2.1.1 to see

$$\|F^{A_n} - F^{\tilde{A}_n}\| \ \leq \ 2\left(1 - F^{T_n}(\tau)\right) \to 2\left(1 - H(\tau)\right), \tag{3.1.8}$$

which tends to 0 as τ tends to infinity.

Suppose Theorem 1.1.1 holds for $\tilde{A}_n$. If H has a bounded support, then if τ is bigger than the largest value in the support of H, we have $H_\tau(t) = H(t)$ and $1 - H(\tau) = 0$. From Lemma 3.1.2, $g_\tau(z)$ are the same for all τ large. Hence $s_\tau(z)$ are the same for all τ large. From the inversion formula in Theorem 2.3.1, this implies F_τ are the same for all τ large and hence are properly denoted by F. On the other hand, from (3.1.8) we have for all τ large, with probability one $\limsup_{n\to\infty} \|F^{A_n} - F^{\tilde{A}_n}\| = 0$, which implies F^{A_n} and $F^{\tilde{A}_n}$ converges simultaneously to the same limit, *i.e.* F. Therefore, it follows when H has a bounded support, Theorem 1.1.1 must also hold for A_n.

We proceed to consider the case when H has an unbounded support. Denote by F_τ the limiting spectral distribution of $F^{\tilde{A}_n}$ and $s_\tau(z)$ the stieltjes transform of F_τ. From Theorem 1.1.1, for each $z \in \mathbb{C}^+$, there is an associating function say $g_\tau(z)$ such that $(s_\tau(z), g_\tau(z))$ is the unique point in the set $\{(s(z), p(z)) : Ims(z) > 0, Imp(z) \geq 0\}$ satisfying the relations

$$\begin{cases} s_\tau(z) = -z^{-1} - z^{-1}[g_\tau(z)]^2, \\ g_\tau(z) = \int \frac{t}{-z - t g_\tau(z)} dH_\tau(t). \end{cases} \tag{3.1.9}$$

In this case, we firstly prove as τ tends to infinity, for every $z \in \mathbb{C}^+$, $s_\tau(z)$ and

$g_\tau(z)$ converge, the vector of their limits satisfying (1.1.1).

Write $g_\tau(z) = g_{1\tau}(z) + ig_{2\tau}(z)$ and $z = z_1 + iz_2$ in (3.1.9). Then we have

$$g_{2\tau}(z) = z_2 \int \frac{t}{|z + tg_\tau(z)|^2} dH_\tau(t) + g_{2\tau}(z) \int \frac{t^2}{|z + tg_\tau(z)|^2} dH_\tau(t). \quad (3.1.10)$$

Since in this case we must have $H \neq 1_{[0,\infty)}$, as in the proof of Lemma 3.1.2, we deduce that $g_{2\tau}(z) > 0$ and

$$\int t^2/|z + tg_\tau(z)|^2 dH_\tau(t) < 1. \quad (3.1.11)$$

By Hölder's inequality and the second equation in (3.1.9), it follows that $|g_\tau(z)| < 1$.

For any convergent subsequence $\{g_{\tau_m}(z)\}$, denote its limit by $q(z)$. Then $Im(q(z)) \geq 0$. If $Im(q(z)) = 0$, since $z_2 > 0$ is fixed, from (3.1.10) and (3.1.11), as $m \to \infty$,

$$\int \frac{t}{|z + tg_{\tau_m}(z)|^2} dH_{\tau_m}(t) \to 0. \quad (3.1.12)$$

Further, by (3.1.9), it follows

$$g_{1\tau}(z) = -z_1 \int \frac{t}{|z + tg_\tau(z)|^2} dH_\tau(t) - g_{1\tau}(z) \int \frac{t^2}{|z + tg_\tau(z)|^2} dH_\tau(t),$$

which implies

$$[1 + \int \frac{t^2}{|z + tg_\tau(z)|^2} dH_\tau(t)] g_{1\tau}(z) = -z_1 \int \frac{t}{|z + tg_\tau(z)|^2} dH_\tau(t).$$

In view of (3.1.12), we get $\lim_{m\to\infty} g_{1\tau_m}(z) = 0$. This means if $Im(q(z)) = 0$, then $q(z) = 0$. Turn back to consider (3.1.12). Noting the relationship between $H(t)$

and $H_\tau(t)$, we have for all τ_m large

$$\begin{aligned}\int_{0\le t\le M}\frac{t}{|z+tg_{\tau_m}(z)|^2}dH(t) &\le \int_{0\le t\le \tau_m}\frac{t}{|z+tg_{\tau_m}(z)|^2}dH(t)\\ &= \int\frac{t}{|z+tg_{\tau_m}(z)|^2}dH_{\tau_m}(t)\to 0. \quad (3.1.13)\end{aligned}$$

On the other hand, when $0\le t\le M$, $0\le t/|z+tg_{\tau_m}(z)|^2\le M/z_2^2$ so that the dominated convergence theorem (d.c.t.) is applicable. It follows

$$\int_{0\le t\le M}\frac{t}{|z+tg_{\tau_m}(z)|^2}dH(t)\to\frac{1}{|z|^2}\int_{0\le t\le M}tdH(t).$$

Combining this with (3.1.13), we get $\int_{0\le t\le M}tdH(t)=0$. Therefore, $H(0,M]=0$. Since M is arbitrary, we get $H(0,\infty)=0$, or equivalently, $H\{0\}=1$. This obviously contradicts the assumption that H has an unbounded support. Therefore we assert that $Im(q(z))>0$.

Now we let $\tau=\tau_m\to\infty$ in (3.1.9). Since $|\frac{t}{-z-tg_{\tau_m}(z)}|\le\frac{1}{g_{2\tau_m}(z)}\to\frac{1}{Im(q(z))}$ and thus is bounded, the dominated convergence theorem is applicable. Thus from the second equation of (3.1.9), we get $q(z)$ satisfies (3.1.2) for each $z\in\mathbb{C}^+$. Note that $\{\tau_m\}$ is arbitrarily chosen. From Lemma 3.1.2, it follows that $\{g_\tau(z)\}$ converges, the limit satisfying equation (3.1.2), or equivalently, the second equation in (1.1.1). From the first equation of (3.1.9), $\{s_\tau(z)\}$ also converges. Denote the limits of $\{s_\tau(z)\}$ and $\{g_\tau(z)\}$ by $s(z)$ and $g(z)$, respectively. Then $(s(z),g(z))$ satisfies (1.1.1).

Let $s_n(z)$ and $\tilde{s}_n(z)$ denote respectively the Stieltjes transforms of F^{A_n} and $F^{\tilde{A}_n}$. To show Theorem 1.1.1 must also hold for A_n, by Theorem 2.3.9 and Lemma

3.1.1, we only need show further $s_n(z)$ converges to $s(z)$. We firstly show

$$\limsup_{n\to\infty} |s_n(z) - \tilde{s}_n(z)| \leq (\frac{4}{z_2} + \frac{2M}{z_2^2})(1 - H(\tau)). \tag{3.1.14}$$

From the proof of Lemma 3.1.1, we can choose M such that for all n,

$$\max(F^{A_n}\{\lambda : |\lambda| > M\}, F^{\tilde{A}_n}\{\lambda : |\lambda| > M\}) < 1 - H(\tau).$$

It follows then

$$\int_{|\lambda|>M} \frac{1}{|\lambda - z|} d\{F^{A_n}(\lambda) + F^{\tilde{A}_n}(\lambda)\} \leq \frac{2}{z_2}\{1 - H(\tau)\}.$$

Then, by integration by parts,

$$\begin{aligned}
& \int_{|\lambda|\leq M} \frac{1}{\lambda - z} d\{F^{A_n}(\lambda) - F^{\tilde{A}_n}(\lambda)\} \\
= \; & |\frac{1}{M - z}\{F^{A_n}(M) - F^{\tilde{A}_n}(M)\} - \frac{1}{-M - z}\{F^{A_n}(-M) - F^{\tilde{A}_n}(-M)\} \\
& + \int_{|\lambda|\leq M} \frac{1}{(\lambda - z)^2}\{F^{A_n}(\lambda) - F^{\tilde{A}_n}(\lambda)\}d\lambda| \\
\leq \; & (\frac{2}{z_2} + \frac{2M}{z_2^2})\|F^{A_n} - F^{\tilde{A}_n}\|.
\end{aligned}$$

Therefore,

$$|s_n(z) - \tilde{s}_n(z)| \leq (\frac{2}{z_2} + \frac{2M}{z_2^2})\|F^{A_n} - F^{\tilde{A}_n}\| + \frac{2}{z_2}\{1 - H(\tau)\}.$$

From (3.1.8), (3.1.14) follows.

Using the triangular inequality, (3.1.14) and the fact $\lim \tilde{s}_n(z) = s_\tau(z)$, we get for any τ, $\limsup_{n\to\infty} |s_n(z) - s(z)| \leq (\frac{4}{z_2} + \frac{2M}{z_2^2})(1 - H(\tau)) + |s_\tau(z) - s(z)|$. Let $\tau \to \infty$. Since $s_\tau(z) - s(z) \to 0$, it follows $s_n(z) \to s(z)$ almost surely. Hence we arrive at the conclusion that in proving Theorem 1.1.1 we may assume condition (i) in Assumption 3.1.1 holds without generality.

Now we consider condition (ii) in Assumption 3.1.1. Assume that now A_n satisfies condition (i) in Assumption 3.1.1. Find sequence $\delta_n \downarrow 0$ so that condition (1.1) still holds when δ is replaced by δ_n. Let $\hat{w}_{ij} = w_{ij} I[|w_{ij}| \leq \delta_n \sqrt{n}]$ and $\tilde{w}_{ij} = \hat{w}_{ij} - E\hat{w}_{ij}$. Define $\hat{A}_n$ and $\tilde{A}_n$ in parallel to A_n but with w_{ij} respectively replaced by $\hat{w}_{ij}$ and $\tilde{w}_{ij}$. Let $E|\tilde{w}_{ij}|^2 = \sigma_{ij}^2$. Then it is straightforward to show $\sigma_{ij}^2 \leq 1$ and $\frac{1}{\delta_n^2 n^2} \sum_{ij}(1 - \sigma_{ij}^2) \to 0$. Thus $\tilde{A}_n$ satisfies conditions (i), (ii) in Assumption 3.1.1. Suppose Theorem 1.1.1 holds for $\tilde{A}_n$.

By the rank inequality in Lemma 2.1.1, we have

$$\|F^{A_n} - F^{\hat{A}_n}\| \leq \frac{1}{n} rank(W_n - \hat{W}_n) \leq \frac{1}{n} \sum_{ij} I[|w_{ij}| > \delta_n \sqrt{n}].$$

Using Bernstein's inequality in Lemma 2.1.3, noting that

$$\frac{1}{n} \sum_{ij} P(|w_{ij}| > \delta_n \sqrt{n}) \leq \frac{1}{\delta_n^2 n^2} \sum_{ij} E|w_{ij}|^2 I[|w_{ij}| > \delta_n \sqrt{n}] \to 0,$$

we have

$$\begin{aligned}
& P(\frac{1}{n} \sum_{ij} I[|w_{ij}| > \delta_n \sqrt{n}] > \varepsilon) \\
\leq \;& P(\sum_{ij} \{I[|w_{ij}| > \delta_n \sqrt{n}] - P(|w_{ij}| > \delta_n \sqrt{n})\} > \frac{n\varepsilon}{2}) \\
\leq \;& P(\sum_{i \leq j} \{I[|w_{ij}| > \delta_n \sqrt{n}] - P(|w_{ij}| > \delta_n \sqrt{n})\} > \frac{n\varepsilon}{4}) \\
\leq \;& 2exp\{-\frac{(\frac{n\varepsilon}{4})^2}{2(\frac{n\varepsilon}{4} + \frac{n\varepsilon}{2})}\} \\
\leq \;& 2exp\{-(\frac{\varepsilon}{24})n\}.
\end{aligned}$$

By Borel-Cantelli's lemma in Lemma 2.1.5,, this implies with probability 1, $\frac{1}{n} \sum_{ij} I[|w_{ij}| > \delta_n \sqrt{n}] \to 0$ and so $\|F^{A_n} - F^{\hat{A}_n}\| \to 0$.

On the other hand, from Lemma 2.1.2 (the difference inequality), we have

$$
\begin{aligned}
& L^3(F^{\hat{A}_n}, F^{\tilde{A}_n}) \leq \frac{1}{n} tr(\hat{A}_n - \tilde{A}_n)^2 \\
= & \frac{1}{n^2} tr(T_n^{1/2}(\hat{W}_n - \tilde{W}_n) T_n (\hat{W}_n - \tilde{W}_n) T_n^{1/2}) \\
\leq & \frac{\tau^2}{n^2} \sum_{ij} |E\hat{w}_{ij}|^2 = \frac{\tau^2}{n^2} \sum_{ij} |Ew_{ij} I[|w_{ij}| > \delta_n \sqrt{n}]|^2 \\
\leq & \frac{\tau^2}{n^2} \sum_{ij} E|w_{ij}|^2 I[|w_{ij}| > \delta_n \sqrt{n}] \to 0.
\end{aligned}
$$

The above results assert that the empirical spectral distributions of $\tilde{A}_n$ and A_n converge simultaneously to the same limit. Therefore Theorem 1.1.1 must also be true for A_n. This guarantees in proving Theorem 1.1.1, without loss of generality, we may assume further condition (ii) of Assumption 3.1.1 holds.

Consider condition (iii). Given any $\omega \in \Omega$, $T_n \equiv T_n(\omega)$ is non-random. Define random matrix $A_n^\omega \equiv \frac{1}{\sqrt{n}} T_n^{1/2}(\omega) W_n T_n^{1/2}(\omega)$ in which the matrix W_n is random. Assumption (ii) in Theorem 1.1.1 guarantees there is a subspace Ω_0 with $P(\Omega_0) = 1$ such that for every $\omega \in \Omega_0$, F^{T_n} converges weakly to H. Suppose, for each $\omega \in \Omega_0$, Theorem 1.1.1 holds for A_n^ω. Since H, the limiting spectral distribution of $T_n(\omega)$, does not depend on the given ω, the limiting spectral distribution of A_n^ω does not depend on ω either and hence can be denoted by F. Then for each $z \in \mathbb{C}^+$,

$$
Es_{F^{A_n^\omega}}(z) \to s_F(z), \tag{3.1.15}
$$

and by Lemma 3.2.2 of the present chapter,

$$
E|s_{F^{A_n^\omega}}(z) - Es_{F^{A_n^\omega}}(z)|^4 \leq Kn^{-2},
$$

where K is a constant depending only on τ and $v = Im(z)$.

Note that since W_n is independent of T_n, we have for any $\omega \in \Omega$

$$E(s_{F^{A_n}}(z)|T_n = T_n(\omega)) = Es_{F^{A_n^\omega}}(z) \tag{3.1.16}$$

and

$$\begin{aligned} & E|s_{F^{A_n}}(z) - E(s_{F^{A_n}}(z)|T_n = T_n(\omega))|^4 \\ = \; & E\left\{E\left(|s_{F^{A_n}}(z) - E(s_{F^{A_n}}(z)|T_n = T_n(\omega))|^4|T_n = T_n(\omega)\right)\right\} \\ = \; & E\left(E|s_{F^{A_n^\omega}}(z) - Es_{F^{A_n^\omega}}(z)|^4\right) \\ = \; & \int_{\Omega_0} E|s_{F^{A_n^\omega}}(z) - Es_{F^{A_n^\omega}}(z)|^4 dP(\omega) \\ \leq \; & Kn^{-2}. \end{aligned} \tag{3.1.17}$$

By (3.1.15), (3.1.16), and $P(\Omega_0) = 1$, $E(s_{F^{A_n}}(z)|T_n = T_n(\omega)) \to s_F(z)$ almost surely. From (3.1.17), $s_{F^{A_n}}(z) - E(s_{F^{A_n}}(z)|T_n = T_n(\omega)) \to 0$ almost surely. Thus, $s_{F^{A_n}}(z) \to s_F(z)$ almost surely. That is, Theorem 1.1.1 must also hold for A_n. Therefore, in proving Theorem 1.1.1, adding condition (iii) in Assumption 3.1.1 does not reduce the generality either.

3.2 Existence of the LSD: Proof of Theorem 1.1.1 by Using the Stieltjes Transform Method

As we have mentioned, by Theorem 2.3.9 and Lemma 3.1.1, to finish the proof of Theorem 1.1.1, it suffices to show for each $z \in \mathbb{C}^+$, $s_n(z)$ converges almost surely to a non-random limit satisfying the system of equations (1.1.1). The main target of the present section is then with the aid of the Stieltjes transform method, to accomplish this task.

In this section, some basic properties of matrices will be used. These include for any matrix B and vectors a and b such that a^*Bb is well defined, the inequality

$$|a^*Bb| \leq \|B\|(a^*a)^{1/2}(b^*b)^{1/2}, \tag{3.2.1}$$

and for any invertible matrix A, the identity

$$(A + \alpha\beta^*)^{-1} = A^{-1} - \frac{A^{-1}\alpha\beta^*A^{-1}}{1 + \beta^*A^{-1}\alpha}. \tag{3.2.2}$$

Lemma 3.2.1. *Let* $G_n(z) = (A_n - zI)^{-1}$, $s_n(z) = (1/n)trG_n(z)$ *and* $g_n(z) = (1/n)tr\{T_nG_n(z)\}$. *Then for all* n *and every* $z \in \mathbb{C}^+$, $|s_n(z)| \leq 1/v$, $|g_n(z)| \leq \tau/v$ *and* $Img_n(z) \geq 0$.

Proof. Note that for any Hermitian matrix B, we have $\|(B - zI)^{-1}\| \leq 1/v$. Thus obviously $|s_n(z)| \leq 1/v$. It can also be seen $|g_n(z)| \leq \|T_n\|\|G_n(z)\| \leq \tau/v$. We now show $Img_n(z) \geq 0$. Indeed, let $\Lambda = diag((\lambda_1 - z)^{-1}, \cdots, (\lambda_n - z)^{-1})$, $\Delta = diag(\mu_1, \cdots, \mu_n)$, where λ_i and μ_i denote the eigenvalues of A_n and T_n, respectively. Then there exist unitary matrices P and Q such that $G_n(z) = P^*\Lambda P$ and $T_n = Q\Delta Q^*$. Let $U = PQ$. Then

$$g_n(z) = \frac{1}{n}tr\{\Delta U^*\Lambda U\} = \frac{1}{n}\sum_{i,j=1}^{n} |u_{ij}|^2(\lambda_i - z)^{-1}\mu_j.$$

Thus

$$Img_n(z) = \frac{1}{n}\sum_{i,j=1}^{n} |u_{ij}|^2\mu_j\frac{v}{|\lambda_i - z|^2} \geq 0.$$

□

We next prove the following result, which is a consequence of Burkholder's inequality in Lemma 2.1.4.

Now we introduce some notations. Throughout the remainder of this section, we shall use v to denote the positive imaginary part of a complex number $z \in \mathbb{C}^+$. We also denote $T_n^{1/2} = [\xi_1, \cdots, \xi_n]$ with ξ_i the ith column of $T_n^{1/2}$. Then we have

$$A_n = \sum_{i,j} (1/\sqrt{n}) w_{ij} \xi_i \xi_j^*. \tag{3.2.3}$$

For any $i \neq j$, define A_{ij} and B_{ij} by

$$\begin{aligned} A_n &= B_{ij} + (1/\sqrt{n}) w_{ij} \xi_i \xi_j^* & (3.2.4) \\ &= A_{ij} + (1/\sqrt{n}) w_{ij} \xi_i \xi_j^* + (1/\sqrt{n}) \overline{w}_{ij} \xi_j \xi_i^*. & (3.2.5) \end{aligned}$$

For $1 \leq i \leq n$, define $A_{ii} = A_n - (1/\sqrt{n}) w_{ii} \xi_i \xi_i^*$.

Then A_{ij} is the matrix obtained by taking out from A_n the component involving w_{ij} and $\overline{w}_{ij}$. Obviously, A_{ij} is Hermitian so that, for any $z \in \mathbb{C}^+$, $(A_{ij} - zI)^{-1}$ always exists and $\|(A_{ij} - zI)^{-1}\| \leq 1/v$. While, B_{ij} is not Hermitian, but we can show that $(B_{ij} - zI)^{-1}$ also exists and $\|(B_{ij} - zI)^{-1}\|$ is also bounded. Indeed, from the identity that for any nonsingular matrix A, $\det(A + ab^*) = det(A)(1 + b^* A^{-1} a)$, we have $\det(B_{ij} - zI) = (1 - c_{ij}) \det(A_n - zI)$, where $c_{ij} = (w_{ij}/\sqrt{n}) \xi_j^* (A_n - zI)^{-1} \xi_i$. Since $|c_{ij}| \leq \delta_n (\tau/v) \to 0$, $(A_n - zI)^{-1}$ exists implying $\det(A_n - zI) \neq 0$, we get for all n large, $\det(B_{ij} - zI) \neq 0$ and so $(B_{ij} - zI)^{-1}$ exists. Further, from

$(B_{ij} - zI)^{-1} = (A_{ij} - zI)^{-1} - n^{-1/2}\overline{w}_{ij}(A_{ij} - zI)^{-1}\xi_j\xi_i^*(B_{ij} - zI)^{-1}$, it follows that $\|(B_{ij} - zI)^{-1}\| \leq 1/(v - \delta_n\tau)$ and thus is uniformly bounded for all n large.

Lemma 3.2.2. *Under the assumptions in Theorem 1.1.1 and Assumption 3.1.1, with probability one, as $n \to \infty$,*

$$s_n(z) - Es_n(z) \to 0, \tag{3.2.6}$$

$$g_n(z) - Eg_n(z) \to 0. \tag{3.2.7}$$

Proof. Define the increasing σ-fields generated by $\{w_{ij}\}$ as follows. For any $i \leq j$, write $k = \sum_{l=1}^{i-1}(n - l + 1) + (j - i + 1)$ and

$$\mathcal{F}_k = \sigma\{w_{11}, \cdots, w_{1n}, w_{22}, \cdots, w_{2n}, \cdots, w_{ii}, \cdots, w_{ij}\}.$$

Further define $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then $\mathcal{F}_k$ is a sequence of increasing σ-field, $0 \leq k \leq n(n+1)/2$. Write $m = n(n+1)/2$.

For each k (which is naturally related with a pair (i, j), define

$$Y_k = (1/n)tr\{(A_{ij} - zI)^{-1} - G_n(z)\},$$

where A_{ij} and $G_n(z)$ are respectively as defined in (3.2.5) and Lemma 3.2.1. Then

$$s_n(z) - Es_n(z) = -\sum_{k=1}^{m}(E_k - E_{k-1})Y_k$$

Further, by the resolvent identity, when $i \neq j$,

$$Y_k = \frac{1}{n\sqrt{n}}w_{ij}\xi_j^*(A_{ij} - zI)^{-1}G_n(z)\xi_i + \frac{1}{n\sqrt{n}}\bar{w}_{ij}\xi_i^*(A_{ij} - zI)^{-1}G_n(z)\xi_j,$$

and, when $i = j$, $Y_k = \frac{1}{n\sqrt{n}} w_{ii}\xi_i^*(A_{ii} - zI)^{-1}G_n(z)\xi_i$. From (3.2.1), in either case, it holds $|Y_k| \leq K|w_{ij}|/(n\sqrt{n})$.

Denote by $E_k(\cdot)$ the conditional expectation with respect to the σ-field $\mathcal{F}_k$. Then we get

$$\begin{aligned} &E[|(E_k - E_{k-1})Y_k|^2|\mathcal{F}_{k-1}] \leq E_{k-1}|Y_k|^2 \leq \frac{K}{n^3}, \\ &E|(E_k - E_{k-1})Y_k|^p \leq 2^p E|Y_k|^p \leq \frac{K}{n^{p+1}}. \end{aligned} \tag{3.2.8}$$

From Lemma 2.1.4, we get for any $p \geq 2$,

$$\begin{aligned} & E|s_n(z) - Es_n(z)|^p \\ \leq\ & K\{E(\sum_{k=1}^{m} E[|(E_k - E_{k-1})Y_k|^2|\mathcal{F}_{k-1}])^{p/2} \\ & + \sum_{k=1}^{m} E|(E_k - E_{k-1})Y_k|^p\} \\ \leq\ & Kn^{-p/2}, \end{aligned}$$

from which (3.2.6) follows.

In parallel, define, when $i \neq j$,

$$\tilde{Y}_k = \frac{1}{n\sqrt{n}} w_{ij}\xi_j^*(A_{ij} - zI)^{-1}T_nG_n(z)\xi_i + \frac{1}{n\sqrt{n}}\bar{w}_{ij}\xi_i^*(A_{ij} - zI)^{-1}T_nG_n(z)\xi_j,$$

and, when $i = j$, $Y_k = \frac{1}{n\sqrt{n}} w_{ii}\xi_i^*(A_{ii} - zI)^{-1}T_nG_n(z)\xi_i$. Then we have $|\tilde{Y}_k| \leq K|w_{ij}|/(n\sqrt{n})$, and $g_n(z) - Eg_n(z) = -\sum_{k=1}^{m}(E_k - E_{k-1})\tilde{Y}_k$. Following similar argument, we can get (3.2.7). This completes the proof of Lemma 3.2.2. □

It can be seen that $s_n(z)$ is the Stieltjes transform of the empirical spectral distribution of A_n, F^{A_n}. Thus Lemma 3.2.2 ensures that we may achieve our forestated target by showing for each $z \in \mathbb{C}^+$, the convergence of $(Es_n(z), Eg_n(z))$ to a limit which satisfies the system of equations (1.1.1) in Theorem 1.1.1.

Lemma 3.2.3. *Let $R_n(z) = (-zI - g_n(z)T_n)^{-1}$. Then under the assumptions in Theorem 1.1.1 and Assumption 3.1.1, as $n \to \infty$,*

$$E(\frac{1}{n}tr\{\Phi_n A_n G_n(z)\}) + E(g_n(z)\frac{1}{n}tr\{\Phi_n T_n G_n(z)\}) \to 0, \tag{3.2.9}$$

where Φ_n may take I_n or $T_n R_n(z)$.

Proof of Lemma 3.2.3. Define $G_{ij}(z) = (A_{ij} - zI)^{-1}$, $g_{ij}(z) = (1/n)tr\{T_n G_{ij}(z)\}$, $R_{ij}(z) = (-zI - g_{ij}(z)T_n)^{-1}$. For $\Phi_n = T_n R_n(z)$, define $\Phi_{ij} = T_n R_{ij}(z)$.

We have $Img_n(z) \geq 0$ and $Img_{ij}(z) \geq 0$. It is thus easy to see $\|\Phi_n\|$ and $\|\Phi_{ij}\|$ are both bounded by $1/v$. We have

$$|g_n(z) - g_{ij}(z)| \leq K/(n\sqrt{n})|w_{ij}|. \tag{3.2.10}$$

Note that when $\Phi_n = T_n R_n(z)$, $\Phi_{ij} - \Phi_n = (g_{ij}(z) - g_n(z))\Phi_{ij}\Phi_n$. It follows that

$$\|\Phi_{ij} - \Phi_n\| \leq K/(n\sqrt{n})|w_{ij}|. \tag{3.2.11}$$

By use of (3.2.3), we have

$$\begin{aligned} & \frac{1}{n}tr[\Phi_n A_n G_n(z)] \\ = & \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}w_{ii}\xi_i^* G_n(z)\Phi_n\xi_i \\ & +\frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}\xi_j^* G_n(z)\Phi_n\xi_i, \end{aligned}$$

in which the first term is bounded in absolute value by $K\delta_n$, hence, its expectation converges to 0. Note that for the case when $\Phi_n = I_n$, this means in latter proof we only need to consider the expectation of the following term

$$\frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}\xi_j^* G_n(z)\xi_i. \tag{3.2.12}$$

When $\Phi_n = T_n R_n(z)$, we write

$$\begin{aligned} & \frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}\xi_j^* G_n(z)\Phi_n\xi_i \\ = & \frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}\xi_j^* G_n(z)\Phi_{ij}\xi_i \\ & +\frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}\xi_j^* G_n(z)(\Phi_n-\Phi_{ij})\xi_i. \end{aligned} \tag{3.2.13}$$

From (3.2.1) and (3.2.11), the second term on the right hand side of (3.2.13) is bounded in absolute value by $K\delta_n^2$, hence, its expectation converges to 0. Thus for the case when $\Phi_n = T_n R_n(z)$, we only need to consider the expectation of the first term in (3.2.13).

From (3.2.2), we have

$$\begin{aligned} & \frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}\xi_j^* G_n(z)\Phi_{ij}\xi_i \\ = & \frac{1}{n}\sum_{i\neq j}\Big\{\frac{\frac{1}{\sqrt{n}}w_{ij}\xi_j^*(A_{ij}-zI)^{-1}\Phi_{ij}\xi_i}{1+\frac{1}{\sqrt{n}}w_{ij}\xi_j^*(B_{ij}-zI)^{-1}\xi_i} \\ - & \frac{\frac{1}{n}|w_{ij}|^2\xi_j^*(A_{ij}-zI)^{-1}\xi_j\xi_i^*(A_{ij}-zI)^{-1}\Phi_{ij}\xi_i}{[1+\frac{1}{\sqrt{n}}w_{ij}\xi_j^*(B_{ij}-zI)^{-1}\xi_i][1+\frac{1}{\sqrt{n}}\overline{w}_{ij}\xi_i^*(A_{ij}-zI)^{-1}\xi_j]}\Big\}. \end{aligned}$$

For notational convenience, in the following we denote for each fixed pair $i \neq j$,

$$\begin{aligned} & p_{kl} = \xi_l^*(A_{ij}-zI)^{-1}\xi_k, \tilde{p}_{kl} = \xi_l^* G_n(z)\xi_k, \\ & b_{kl} = \xi_l^*(B_{ij}-zI)^{-1}\xi_k, \\ & r_{kl} = \xi_l^*(A_{ij}-zI)^{-1}\Phi_{ij}\xi_k, \tilde{r}_{kl} = \xi_l^* G_n(z)\Phi_n\xi_k. \end{aligned} \tag{3.2.14}$$

Here for brevity, also because i, j is fixed, we only use subscripts k and l to indicate the kth and lth column of $T_n^{1/2}$ involved in each term, the indices i and j omitted.

Also, k and l may take values of i and j. By (3.2.1) and $\xi_k^*\xi_k \le \tau$ for any k, we have $|p_{kl}|$, $|\tilde{p}_{kl}|$, $|b_{kl}|$, $|r_{kl}|$ and $|\tilde{r}_{kl}|$ are all bounded.

Therefore, for the case when $\Phi_n = T_n R_n(z)$, with notations (3.2.14), we get

$$\begin{aligned}
&\frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}\xi_j^* G_n(z)\Phi_{ij}\xi_i \\
=\ & \frac{1}{n}\sum_{i\neq j}\{\frac{\frac{1}{\sqrt{n}}w_{ij}r_{ij}}{1+\frac{1}{\sqrt{n}}w_{ij}b_{ij}} - \frac{\frac{1}{n}|w_{ij}|^2 p_{jj}r_{ii}}{[1+\frac{1}{\sqrt{n}}w_{ij}b_{ij}][1+\frac{1}{\sqrt{n}}\overline{w}_{ij}p_{ji}]}\} \\
=\ & -\frac{1}{n^2}\sum_{i\neq j}|w_{ij}|^2 p_{jj}r_{ii} + \frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}r_{ij} \\
& -\frac{1}{n^2}\sum_{i\neq j}{w_{ij}}^2 r_{ij}b_{ij} + \frac{1}{n^2}\sum_{i\neq j}\frac{\frac{1}{\sqrt{n}}{w_{ij}}^3 r_{ij}b_{ij}^2}{1+\frac{1}{\sqrt{n}}w_{ij}b_{ij}} \\
& +\frac{1}{n^2}\sum_{i\neq j}\frac{|w_{ij}|^2 p_{jj}r_{ii}[\frac{1}{\sqrt{n}}w_{ij}b_{ij} + \frac{1}{\sqrt{n}}\overline{w}_{ij}p_{ji} + \frac{1}{n}|w_{ij}|^2 b_{ij}p_{ji}]}{[1+\frac{1}{\sqrt{n}}w_{ij}b_{ij}][1+\frac{1}{\sqrt{n}}\overline{w}_{ij}p_{ji}]}. \qquad (3.2.15)
\end{aligned}$$

For the case when $\Phi_n = I_n$, by replacing the r_{kl}'s by p_{kl}'s in (3.2.15), we obtain the expression for the term (3.2.12).

Now we calculate the expectation of the terms involved for both cases. Since w_{ij} is independent of p_{ij} and r_{ij}, it is easy to see

$$E(\frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}r_{ij}) = E(\frac{1}{n}\sum_{i\neq j}\frac{1}{\sqrt{n}}w_{ij}p_{ij}) = 0. \qquad (3.2.16)$$

From (3.2.2),

$$b_{ij} = p_{ij} - \frac{\frac{1}{\sqrt{n}}\overline{w}_{ij}p_{jj}p_{ii}}{1+\frac{1}{\sqrt{n}}\overline{w}_{ij}p_{ji}}. \qquad (3.2.17)$$

Substitute this identity into the third term on the right-hand side of (3.2.15). We get (excluding its menus sign)

$$\frac{1}{n^2}\sum_{i\neq j}{w_{ij}}^2 r_{ij}b_{ij} = \frac{1}{n^2}\sum_{i\neq j}{w_{ij}}^2 r_{ij}p_{ij} - \frac{1}{n^2}\sum_{i\neq j}{w_{ij}}^2\frac{\frac{1}{\sqrt{n}}\overline{w}_{ij}r_{ij}p_{jj}p_{ii}}{1+\frac{1}{\sqrt{n}}\overline{w}_{ij}p_{ji}}. \qquad (3.2.18)$$

Since $|p_{ij}|$, $|b_{ij}|$, $|r_{ij}|$ are bounded, $|w_{ij}| \le \delta_n\sqrt{n}$, $1/|1+(1/\sqrt{n})w_{ij}b_{ij}|$ and $1/|1+(1/\sqrt{n})\overline{w}_{ij}p_{ji}|$ are also bounded. It follows that the absolute values of the last two terms on the right-hand side of (3.2.15) and the second term on the right-hand side of (3.2.18) are bounded by $K(\delta_n/n^2)\sum_{i\neq j}|w_{ij}|^2$. Hence the expectations of these terms are bounded by $K\delta_n$ and thus converge to 0. Obviously, this argument is still valid after the r_{kl}'s are replaced by p_{kl}'s in these terms. Then for the case $\Phi_n = T_nR_n(z)$ we only need to consider the first terms in (3.2.15) and (3.2.18). We get

$$\begin{aligned} & E\frac{1}{n}tr\{\Phi_n A_n G_n(z)\} \\ = \;& -E(\frac{1}{n^2}\sum_{i\neq j}|w_{ij}|^2 p_{jj}r_{ii}) - E(\frac{1}{n^2}\sum_{i\neq j}{w_{ij}}^2 r_{ij}p_{ij}) + o(1). \end{aligned} \tag{3.2.19}$$

Similarly, for the case when $\Phi_n = I_n$, we get

$$\begin{aligned} & E\frac{1}{n}tr\{\Phi_n A_n G_n(z)\} = E\frac{1}{n}tr\{A_n G_n(z)\} \\ = \;& -E(\frac{1}{n^2}\sum_{i\neq j}|w_{ij}|^2 p_{jj}p_{ii}) - E(\frac{1}{n^2}\sum_{i\neq j}{w_{ij}}^2 p_{ij}^2) + o(1). \end{aligned} \tag{3.2.20}$$

We need the following result to finish our proof.

Lemma 3.2.4. *With the notations defined in (3.2.14), we have*

$$E\sum_{kl}|r_{kl}|^2 \le Kn \text{ and } E\sum_{kl}|p_{kl}|^2 \le Kn. \tag{3.2.21}$$

Proof. From (3.2.1) and (3.2.11), we have

$$|\xi_l^*(A_{ij}-zI)^{-1}(\Phi_{ij}-\Phi_n)\xi_k| \le \frac{K}{n\sqrt{n}}|w_{ij}|.$$

From the resolvent identity (2.1.5), with notations (3.2.14), we have

$$\begin{aligned}&|\xi_l^*\{(A_{ij}-zI)^{-1}-G_n(z)\}\Phi_n\xi_k|\\&=|\frac{1}{\sqrt{n}}w_{ij}p_{il}\tilde{r}_{kj}+\frac{1}{\sqrt{n}}\bar{w}_{ij}p_{jl}\tilde{r}_{ki}|\leq\frac{K}{\sqrt{n}}|w_{ij}|.\end{aligned}$$

It follows that $|r_{kl}-\tilde{r}_{kl}|\leq K|w_{ij}|/\sqrt{n}$ and hence $E|r_{kl}-\tilde{r}_{kl}|^2\leq K/n$. On the other hand, we have $\sum_{kl}|\tilde{r}_{kl}|^2=tr\{T_n^{1/2}G_n(z)\Phi_nT_n\Phi_n^*(A_n-\bar{z}I)^{-1}T_n^{1/2}\}\leq Kn$. Therefore, we get $E\sum_{kl}|r_{kl}|^2\leq Kn$.

Similarly, we have

$$\begin{aligned}&|p_{kl}-\tilde{p}_{kl}|=|\frac{1}{\sqrt{n}}w_{ij}p_{il}\tilde{p}_{kj}+\frac{1}{\sqrt{n}}\bar{w}_{ij}p_{jl}\tilde{p}_{ki}|\leq\frac{K}{\sqrt{n}}|w_{ij}|, &(3.2.22)\\&\sum_{kl}|\tilde{p}_{kl}|^2=tr\{T_n^{1/2}G_n(z)T_n(A_n-\bar{z}I)^{-1}T_n^{1/2}\}\leq Kn,&\end{aligned}$$

which gives $E\sum_{kl}|p_{kl}|^2\leq Kn$. The proof of Lemma 3.2.4 is finished.$\square$

Now let us continue with our proof of Lemma 3.2.3. Note that w_{ij} is independent of p_{kl} and r_{kl}. From lemma 3.2.4, we have

$$|E(\frac{1}{n^2}\sum_{i\neq j}w_{ij}^2p_{ij}^2)|\leq\frac{1}{n^2}E\sum_{i\neq j}|p_{ij}|^2\leq K/n\to 0,$$

and

$$|E(\frac{1}{n^2}\sum_{i\neq j}w_{ij}^2p_{ij}r_{ij})|\leq\frac{1}{n^2}(E\sum_{i\neq j}|p_{ij}|^2)^{1/2}(E\sum_{i\neq j}|r_{ij}|^2)^{1/2}\leq K/n\to 0.$$

From (3.2.22), we have $|p_{kl}-\tilde{p}_{kl}|\leq K\delta_n$. Then we get

$$|\frac{1}{n^2}\sum_{i\neq j}(p_{jj}p_{ii}-\tilde{p}_{jj}\tilde{p}_{ii})|\leq K\delta_n.$$

Note that

$$\frac{1}{n^2}\sum_{i\neq j}\tilde{p}_{jj}\tilde{p}_{ii} = \{g_n(z)\}^2 - \frac{1}{n^2}\sum_{i=1}^{n}\tilde{p}_{ii}^2,$$

in which the second term on the right-hand side is bounded in absolute value by K/n. Therefore

$$\begin{aligned}
& |E\frac{1}{n^2}\sum_{i\neq j}|w_{ij}|^2 p_{jj}p_{ii} - E\{g_n(z)\}^2| \\
\leq\ & |E\frac{1}{n^2}\sum_{i\neq j}p_{jj}p_{ii} - E\{g_n(z)\}^2| + K\frac{1}{n^2}\sum_{ij}(1 - E|w_{ij}|^2) \\
\leq\ & K\delta_n \to 0.
\end{aligned}$$

Thus we get, from (3.2.20),

$$E(\frac{1}{n}tr\{A_n G_n(z)\} + \{g_n(z)\}^2) \to 0.$$

This implies (3.2.9) holds for the case $\Phi_n = I_n$.

Similarly, we have $|r_{kl} - \tilde{r}_{kl}| \leq K\delta_n$ and

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{r}_{ii} = \frac{1}{n}tr\{\Phi_n T_n G_n(z)\}.$$

Then we have

$$\begin{aligned}
& |\frac{1}{n^2}\sum_{i\neq j}(p_{jj}r_{ii} - \tilde{p}_{jj}\tilde{r}_{ii})| \leq K\delta_n, \\
& |\frac{1}{n^2}\sum_{i\neq j}(\tilde{p}_{jj}\tilde{r}_{ii}) - g_n(z)\frac{1}{n}tr\{\Phi_n T_n G_n(z)\}| \leq \frac{K}{n}.
\end{aligned}$$

From (3.2.19), we get

$$|E(\frac{1}{n}tr\{\Phi_n A_n G_n(z)\} + g_n(z)\frac{1}{n}tr\{\Phi_n T_n G_n(z)\})| \to 0.$$

This shows (3.2.9) holds for the case when $\Phi_n = T_n R_n(z)$. We complete the proof of Lemma 3.2.3 now. $\square$

Lemma 3.2.5. *For each $z \in \mathbb{C}^+$, as $n \to \infty$, $(Es_n(z), Eg_n(z))$ converges to a limit which satisfies the system of equations (1.1.1) in Theorem 1.1.1.*

Proof. By Lemmas 3.1.2 and 3.2.1, we need only show that any convergent subsequence $\{(Es_{n_i}(z), Eg_{n_i}(z))\}$ converges to a limit satisfying the system of equations (1.1.1). With the resolvent identity (2.1.5), when $\Phi_n = I_n$, (3.2.9) gives us

$$E\left(zs_n(z) + 1 + \{g_n(z)\}^2\right) \to 0, \tag{3.2.23}$$

as $n \to \infty$. When $\Phi_n = T_n R_n(z)$, (3.2.9) gives us as $n \to \infty$.

$$\begin{aligned} & E\left(\frac{1}{n} tr\{T_n R_n(z)\} - g_n(z)\right) \\ = \quad & E\frac{1}{n} tr\{T_n R_n(z)(A_n + g_n(z)T_n)G_n(z)\} \\ = \quad & E\left(\frac{1}{n} tr\{\Phi_n A_n G_n(z)\} + g_n(z)\frac{1}{n} tr\{\Phi_n T_n G_n(z)\}\right) \\ \to \quad & 0. \end{aligned} \tag{3.2.24}$$

Let $(s(z), g(z))$ denote the limit of $\{(Es_{n_i}(z), Eg_{n_i}(z))\}$. Then note that $|g_n(z)| \leq 2\tau/v$. From (3.2.7) and the dominated convergence theorem, we have $E\{g_{n_i}(z)\}^2 \to g(z)^2$. By (3.2.23), this implies that $zs(z)+1+\{g(z)\}^2 = 0$ and so we see $(s(z), g(z))$ satisfies the first equation in (1.1.1).

We now prove

$$E\frac{1}{n_i}tr\{T_{n_i}R_{n_i}(z)\} \to \int \frac{t}{-z-tg(z)}dH(t). \tag{3.2.25}$$

Define $S_n(z) = (-zI - Eg_n(z)T_n)^{-1}$. Then

$$|E\frac{1}{n}tr\{T_nR_n(z)\} - \frac{1}{n}tr\{T_nS_n(z)\}| \leq \left(\frac{\tau}{v}\right)^2 E|g_n(z) - Eg_n(z)| \to 0.$$

Let $f_n(t) = t/(-z - tEg_n(z))$ and $f(t) = t/(-z - tg(z))$. Then $ImEg_n(z) \geq 0$, $Img(z) \geq 0$, and $|g(z)| \leq \tau/v$ imply $f(t)$ is bounded continuous and

$$|f_{n_i}(t) - f(t)| \leq (\tau^2/v^2)|Eg_{n_i}(z) - g(z)| \to 0.$$

It follows

$$\begin{aligned} & \frac{1}{n_i}tr\{T_{n_i}S_{n_i}(z)\} = \int f_{n_i}(t)dF^{T_{n_i}}(t) \\ = & \int (f_{n_i}(t) - f(t))dF^{T_{n_i}}(t) + \int f(t)dF^{T_{n_i}}(t) \\ \to & \int f(t)dH(t) = \int \frac{t}{-z-tg(z)}dH(t). \end{aligned}$$

Hence (3.2.25) is proved. By (3.2.24), it follows $g(z)$ satisfies the second equation in (1.1.1). As we have claimed at the beginning, this completes the proof. $\square$

Up to this point, we have indeed finished the proof of Theorem 1.1.1, since Lemmas 3.2.2 and 3.2.5 together tell us for each $z \in \mathbb{C}^+$, with probability one, $(s_n(z), g_n(z))$ converges to a non-random limit $(s(z), g(z))$ satisfying the system of equations (1.1.1) in Theorem 1.1.1. By Lemma 3.2.1, $s(z)$ must be the Stieltjes transform of the limiting spectral distribution. While the meaning of the function $g(z)$ associating with the Stieltjes transform of the limiting spectral distribution is also clear now.

3.3 Analytic Properties of the LSD: Proof of Theorem 1.1.2

This section is devoted to the proof of Theorem 1.1.2. In the following, F denotes the limiting spectral distribution of the Wigner type random matrices and $s_F(z)$ denotes the Stieltjes transform of F. Then we have from the previous section

$$\begin{cases} s_F(z) = -z^{-1} - z^{-1}\{g(z)\}^2, \\ g(z) = \int t/\{-z - tg(z)\}dH(t), \end{cases} \tag{3.3.1}$$

Proof of Theorem 1.1.2–(1).

(i) By Theorem 2.3.5 and the first equation of (3.3.1), we only need to show for any $z \in \mathbb{C}^+$, $g_k(z)$ converges to $g(z)$. Here, $g_k(z)$, with $Im(g_k(z)) \geq 0$, is related to the Stieltjes transform $s_{F_k}(z)$ of F_k through $s_{F_k}(z) = -z^{-1} - z^{-1}\{g_k(z)\}^2$ and satisfies the equation

$$g_k(z) = \int t/\{-z - tg_k(z)\}dH_k(t). \tag{3.3.2}$$

By computing the imaginary part of the second equation of (3.3.1), one gets

$$\int t^2/\{|z + tg(z)|^2\}dH(t) < 1.$$

From Hölder's inequality, this implies $|g(z)| < 1$. Similarly, $|g_k(z)| < 1$.

If $H(t) = I_{\{0\}}(t)$, then $g(z) = 0$. We show $g_k(z) \to 0$. Suppose not. Then, since $|g_k(z)| < 1$, there must exist some subsequence of $\{g_k(z)\}$ which converges to some nonzero limit $g_0(z)$. Without loss of generality, we suppose $g_k(z) \to g_0(z)$.

This implies that we can find $M > 0$ such that for all large k and $t \geq 0$

$$
\begin{aligned}
& |t/\{z + tg_k(z)\}| \\
\leq\ & (t/v)I_{(0,M]}(t) + |\frac{1}{z}| \cdot |\frac{z}{g_k(z)}| \cdot \frac{t}{t - |z/g_k(z)|} I_{(M,\infty)}(t) \\
\leq\ & (t/v)I_{[0,M]}(t) + (M/v)I_{(M,\infty)}(t),
\end{aligned}
$$

where $v = Im(z)$. Thus, as H_k converges weakly to $I_{\{0\}}(t)$,

$$
\begin{aligned}
|g_k(z)| &\leq \int t^2/|z + tg_k(z)| dH_k(t) \\
&\leq \int \{(t/v)I_{[0,M]}(t) + (M/v)I_{(M,\infty)}(t)\} dH_k(t) \\
&\rightarrow \int \{(t/v)I_{[0,M]}(t) + (M/v)I_{(M,\infty)}(t)\} dH(t) \\
&= 0.
\end{aligned}
$$

It follows then, $g_k(z) \rightarrow 0$. Thus, $g_k(z) \rightarrow g(z)$ holds for the case of $H(t) = I_{\{0\}}(t)$.

Now suppose $H(t) \neq I_{\{0\}}(t)$. We first prove that any convergent subsequence of $\{g_k(z)\}$ does not converge to 0. For simplicity, by the way of contradiction, suppose $g_k(z) \rightarrow 0$. From equation (3.3.2),

$$
Im(g_k(z)) \left(1 - \int t^2/|z + tg_k(z)|^2 dH_k(t)\right) = v \int t/|z + tg_k(z)|^2 dH_k(t).
$$

It follows then, $\int_{(o,M]} t/|z+tg_k(z)|^2 dH_k(t) \rightarrow 0$, for any $M > 0$. By the dominated convergence theorem, however, $\int_{(o,M]} t/|z + tg_k(z)|^2 dH_k(t) \rightarrow \int_{(o,M]} t^2 dH(t)/|z|^2$. Thus, $H\{(0, M]\} = 0$ for any $M > 0$. This means $H(t) = I_{\{0\}}(t)$, a contradiction. Therefore, any convergent subsequence of $\{g_k(z)\}$ does not converge to 0.

Consider any subsequence $\{k_j\}_{j=1}^{\infty}$ such that $\{g_{k_j}(z)\}_{j=1}^{\infty} \rightarrow \tilde{g}(z)$. From the preceding argument, $\tilde{g}(z) \neq 0$. Thus, there is $M > 0$ such that for all large j,

$|z/g_{k_j}(z)| < M/2$ and so $|z/\tilde{g}(z)| < M/2$. It follows that

$$|t/\{-z - tg_{k_j}(z)\} - t/\{-z - t\tilde{g}(z)\}| \leq (M/v)^2 |g_{k_j}(z) - \tilde{g}(z)|$$

and that $t/\{-z - t\tilde{g}(z)\}$ is bounded and continuous. Consequently,

$$\int t/\{-z - tg_{k_j}(z)\} dH_{k_j}(t) - \int t\{-z - t\tilde{g}(z)\} dH_{k_j}(t) \to 0$$
$$\int t/\{-z - t\tilde{g}(z)\} dH_{k_j}(t) \to \int t/\{-z - t\tilde{g}(z)\} dH(t)$$

so that $\tilde{g}(z) = \int t/\{-z - t\tilde{g}(z)\} dH(t)$. Since, obviously, $Im(\tilde{g}(z)) \geq 0$, by Lemma 3.1.2, we get $\tilde{g}(z) = g(z)$. Thus, for the case when $H(t) \neq I_{[0,\infty)}(t)$, $g_k(z)$ also converges to $g(z)$. This completes the proof of part (i) of Theorem 1.1.2 (1).

(ii) Recall the basic property of the Stieltjes transform that

$$F\{0\} = \lim_{a \downarrow 0} \{-ias_F(ia)\} = 1 + \lim_{a \downarrow 0} \{g(ia)\}^2.$$

Write $g(ia) = g_1(ia) + ig_2(ia)$. From (3.3.1),

$$g_1(ia) = -g_1(ia) \int t^2 / |ia + tg(ia)|^2 dH(t).$$

This implies $g_1(ia) = 0$. Thus $\{g(ia)\}^2 = -\{g_2(ia)\}^2$ and

$$g_2(ia) = \int t/\{a + tg_2(ia)\} dH(t). \tag{3.3.3}$$

We only need to show $\lim_{a \downarrow 0} \{g_2(ia)\}^2 = 1 - H\{0\}$.

Consider first the case of $F\{0\} = 1$. Then $\lim_{a \downarrow 0} g_2(ia) = 0$. From (3.3.3), since $t \geq 0$, $a > 0$ and $g_2(ia) > 0$, we have

$$g_2(ia) \geq \int_{[0<t\leq M]} t/\{a + Mg_2(ia)\} dH(t).$$

Then as $a \downarrow 0$,

$$\int_{[0<t\leq M]} t dH(t) \leq g_2(ia)\{a + M g_2(ia)\} \to 0.$$

This implies, for any $M > 0$, $H(0, M] = 0$. Then $H(0, \infty) = 0$ and $H\{0\} = 1$. So in this case we get $F\{0\} = H\{0\}$.

Now consider the case of $F\{0\} < 1$. Then $\lim_{a\downarrow 0} g_2(ia) > 0$. From (3.3.3) and the dominated convergence theorem, we have

$$\lim_{a\downarrow 0}\{g_2(ia)\}^2 = \int_{t>0} \lim_{a\downarrow 0} \frac{t}{t + a/g_2(ia)} dH(t) = 1 - H\{0\}.$$

Thus we also get $F\{0\} = H\{0\}$ in this case. Therefore, it always holds $F\{0\} = H\{0\}$.

Proof of Theorem 1.1.2–(2).

From the above proof, we have $|g(z)| < 1$, for all $\forall z \in \mathbb{C}^+$. Thus for any $x \in \mathbb{R}^+ \cup \mathbb{R}^-$, we need only consider an arbitrarily chosen subsequence $\{z_n\} \subset \mathbb{C}^+ \to x$ such that $\{g(z_n)\}$ converges. Denote $g(x) = \lim g(z_n)$. Write $z_n = z_{1n} + i z_{2n}$, $g(z_n) = g_1(z_n) + i g_2(z_n)$ and $g(x) = g_1(x) + i g_2(x)$.

From the second equation of (3.3.1),

$$g(z_n) = \int \frac{t}{-z_n - t g(z_n)} dH(t). \tag{3.3.4}$$

By calculating the imaginary part for both sides of (3.3.4), we get

$$\int \frac{t^2}{|z_n + t g(z_n)|^2} dH(t) < 1. \tag{3.3.5}$$

By Fatou's lemma, it then follows

$$\int \liminf_{n\to\infty} \frac{t^2}{|z_n + t g(z_n)|^2} dH(t) \leq \liminf_{n\to\infty} \int \frac{t^2}{|z_n + t g(z_n)|^2} dH(t) \leq 1.$$

Note

$$\liminf_{n\to\infty} \frac{t^2}{|z_n + tg(z_n)|^2} = \frac{t^2}{|x + tg(x)|^2} I_{(|x+tg(x)|>0)} + \infty \cdot I_{(|x+tg(x)|=0)}.$$

This implies $H\{t : x + tg(x) = 0\} = 0$ and hence

$$\begin{aligned}
& \int \frac{t^2}{|x + tg(x)|^2} dH(t) \\
= & \int \frac{t^2}{|x + tg(x)|^2} I_{(|x+tg(x)|>0)} dH(t) \\
\leq & \liminf_{n\to\infty} \int \frac{t^2}{|z_n + tg(z_n)|^2} dH(t) \\
\leq & \; 1.
\end{aligned} \tag{3.3.6}$$

On the other hand, by computing the real part for both sides of (3.3.4), we get

$$\begin{aligned}
& \int \frac{t}{|z_n + tg(z_n)|^2} dH(t) \\
= & -\frac{g_1(z_n)}{z_{1n}} \left(1 + \int \frac{t^2}{|z_n + tg(z_n)|^2} dH(t)\right) \\
\leq & -2\frac{g_1(z_n)}{z_{1n}},
\end{aligned} \tag{3.3.7}$$

so that

$$\limsup_{n\to\infty} \int \frac{t}{|z_n + tg(z_n)|^2} dH(t) \leq -2\frac{g_1(x)}{x}. \tag{3.3.8}$$

By Fatou's lemma, it implies

$$\int \frac{t}{|x + tg(x)|^2} dH(t) \leq \liminf_{n\to\infty} \int \frac{t}{|z_n + tg(z_n)|^2} dH(t) \leq -2\frac{g_1(x)}{x}. \tag{3.3.9}$$

Note

$$\begin{aligned}
& \int \left(\frac{t}{-z_n - tg(z_n)} - \frac{t}{-x - tg(x)}\right) dH(t) \\
= & \int \frac{t}{(z_n + tg(z_n))(x + tg(x))} dH(t)(z_n - x) \\
& + \int \frac{t^2}{(z_n + tg(z_n))(x + tg(x))} dH(t)(g(z_n) - g(x)),
\end{aligned} \tag{3.3.10}$$

in which by Hölder's inequality and (3.3.5), (3.3.6),

$$\begin{aligned}
& |\int \frac{t^2}{(z_n + tg(z_n))(x + tg(x))} dH(t)| \\
\leq & \left(\int \frac{t^2}{|z_n + tg(z_n)|^2} dH(t)\right)^{1/2} \left(\int \frac{t^2}{|x + tg(x)|^2} dH(t)\right)^{1/2} \\
\leq & \ 1,
\end{aligned}$$

by Hölder's inequality and (3.3.8), (3.3.9)

$$\begin{aligned}
& \limsup_{n\to\infty} |\int \frac{t}{(z_n + tg(z_n))(x + tg(x))} dH(t)| \\
\leq & \left(\limsup_{n\to\infty} \int \frac{t}{|z_n + tg(z_n)|^2} dH(t)\right)^{1/2} \left(\int \frac{t}{|x + tg(x)|^2} dH(t)\right)^{1/2} \\
\leq & -2\frac{g_1(x)}{x}.
\end{aligned}$$

Let $n \to \infty$ in (3.3.10). Since $z_n \to x$ and $g(z_n) \to g(x)$, it follows

$$\lim_{n\to\infty} \int \frac{t}{-z_n - tg(z_n)} dH(t) = \int \frac{t}{-x - tg(x)} dH(t).$$

This shows $g(x)$ satisfies the second equation of (1.1.2) of Theorem 1.1.2, *i.e.* the following equation

$$g(x) = \int \frac{t}{-x - tg(x)} dH(t). \tag{3.3.11}$$

Thus far it has been seen $g(x)$ satisfies simultaneously equations (3.3.6), (3.3.9) and (3.3.11). Further, we show for any $x \in \mathbb{R}^+ \cup \mathbb{R}^-$, there is at most one solution $g(x)$ satisfying simultaneously equations (3.3.6) and (3.3.11). We use the way of contradiction. Let $g(x) \neq \tilde{g}(x)$ both satisfy the two equations simultaneously. It follows

$$g(x) - \tilde{g}(x) = (g(x) - \tilde{g}(x)) \int \frac{t^2}{(x + tg(x))(x + t\tilde{g}(x))} dH(t). \tag{3.3.12}$$

But from Hölder's inequality, $g(x) \neq \tilde{g}(x)$ implies

$$\begin{aligned} & |\int \frac{t^2}{(x+tg(x))(x+t\tilde{g}(x))} dH(t)| \\ < \; & \{\int \frac{t^2}{|x+tg(x)|^2} dH(t)\}^{1/2} \{\int \frac{t^2}{|x+t\tilde{g}(x)|^2} dH(t)\}^{1/2} \\ \leq \; & 1. \end{aligned}$$

This, by (3.3.12) however, implies $g(x) = \tilde{g}(x)$. Thus we arrive at a contradiction. The contradiction asserts that any convergent subsequence $\{g(z_n)\}$ must converge to the same limit, that is, the unique solution to equation (3.3.11) satisfying condition (3.3.6).

Therefore, as $z \in \mathbb{C}^+ \to x$, $g(z)$ converges, and if its limit is denoted by $g(x)$, then $g(x)$ is the unique solution to equation (3.3.11) satisfying condition (3.3.6). Computing the real part for both sides of equation (3.3.11) gives immediately

$$g_1(x) \left(1 + \int \frac{t^2}{|x+tg(x)|^2} dH(t)\right) = -x \int \frac{t}{|x+tg(x)|^2} dH(t),$$

which implies $g_1(x)/x < 0$, that is, $Re(g(x))/x < 0$. Further, by the first equation of (3.3.1), it is trivial to see $s_F(z)$ also converges with the limit $s_F(x)$ satisfying the first equation of (1.1.2) of Theorem 1.1.2.

We now show $s_F(x)$, $g(x)$ are continuous on $\mathbb{R}^+ \cup \mathbb{R}^-$. It is sufficient to show the result for $g(x)$. Consider any $x_0 \in \mathbb{R}$, $x_0 \neq 0$. For any given $\varepsilon > 0$, from $g(x_0) = \lim_{z \in \mathbb{C}^+ \to x_0} g(z)$, there exists $\delta > 0$ such that when $z \in \mathbb{C}^+$ and $|z - x_0| \leq \delta$, $|g(z) - g(x_0)| < \varepsilon$. Obviously, we can choose δ such that $0 \notin (x_0 - \delta, x_0 + \delta)$. Then for any $x \in (x_0 - \delta, x_0 + \delta)$, $g(x) = \lim_{z \in \mathbb{C}^+ \to x} g(z)$. For the given ε, there exists δ_x such that when $z \in \mathbb{C}^+$ and $|z - x| \leq \delta_x$, $|g(z) - g(x)| < \varepsilon$. Choose

$z \in \mathbb{C}^+$ satisfying $|z - x_0| \le \delta$ and $|z - x| \le \delta_x$ simultaneously, then we get $|g(x) - g(x_0)| < 2\varepsilon$, for all $x \in (x_0 - \delta, x_0 + \delta)$. This means $g(x)$ is continuous at the arbitrarily chosen x_0. Thus $g(x)$ is continuous on $\mathbb{R}^+ \cup \mathbb{R}^-$. By Theorem 2.3.2, the limiting spectral distribution F of the Wigner type random matrices possesses a continuous derivative on $\mathbb{R}^+ \cup \mathbb{R}^-$ given by $f(x) = -2Re(g(x))Im(g(x))/(\pi x)$.

Proof of Theorem 1.1.2–(3).

(i) To show F is symmetric, we only need to show for any $z \in \mathbb{C}^+$, $s_F(-\bar{z}) = -\overline{s_F(z)}$. This is because the inversion formula then implies for any $0 < a < b$,

$$\begin{aligned} F(b) - F(a) &= \frac{1}{\pi} \lim_{v \downarrow 0} \int_a^b Im s_F(u + iv) du \\ &= \frac{1}{\pi} \lim_{v \downarrow 0} \int_{-b}^{-a} Im s_F(-u + iv) du \\ &= \frac{1}{\pi} \lim_{v \downarrow 0} \int_{-b}^{-a} Im\{-\overline{s_F(u + iv)}\} du \\ &= \frac{1}{\pi} \lim_{v \downarrow 0} \int_{-b}^{-a} Im s_F(u + iv) du \\ &= F(-a) - F(-b). \end{aligned}$$

Let $b \to \infty$ on both sides of $F(b) - F(a) = F(-a) - F(-b)$. Since $F(b) \to 1$ and $F(-b) \to 0$, it then follows $F(a) = 1 - F(-a)$ for any $a \in \mathbb{R}$, *i.e.* F is symmetric.

Now we show $s_F(-\bar{z}) = -\overline{s_F(z)}$. Again, we first show $g(-\bar{z}) = -\overline{g(z)}$. For any $z \in \mathbb{C}^+$, we have

$$g(z) = \int \frac{t}{-z - tg(z)} dH(t).$$

Simple operations of complex conjugate and multiplication by (-1) give

$$-\overline{g(z)} = \int \frac{t}{-(-\bar{z}) - t(-\overline{g(z)})} dH(t).$$

Since $-\bar{z} \in \mathbb{C}^+$ and $Im(-\overline{g(z)}) \geq 0$, by Lemma 3.1.2, it then follows $g(-\bar{z}) = -\overline{g(z)}$. Thus,

$$s_F(-\bar{z}) = \bar{z}^{-1} + \bar{z}^{-1}\{g(-\bar{z})\}^2 = \bar{z}^{-1} + \bar{z}^{-1}\{-\overline{g(z)}\}^2 = \overline{z^{-1} + z^{-1}\{g(z)\}^2} = -\overline{s_F(z)}.$$

(ii) Note that $f(x) > 0$ $(x \neq 0)$ if and only if $Im(g(x)) > 0$. Thus

$$\tilde{S}_F = \{x \in \mathbb{R}^+ \cup \mathbb{R}^- : Im(g(x)) > 0\}.$$

Further note that $g(x)$ satisfies the equation $g = \int t/\{-x - tg\}dH(t)$. It thus follows

$$\tilde{S}_F \subseteq \{x \in \mathbb{R}^+ \cup \mathbb{R}^- : g = \textstyle\int t/\{-x - tg\}dH(t) \text{ has a solution in } \mathbb{C}^+\}.$$

On the other hand, for any $x \in \mathbb{R}^+ \cup \mathbb{R}^-$, if $g = \int t/\{-x - tg\}dH(t)$ has a solution in $\mathbb{C}^+$, by Lemma 3.1.2 and the conclusion on $g(x)$ shown in (2), this solution must be $g(x)$. So, $Im(g(x)) > 0$. Thus the converse relation also holds, *i.e.*

$$\tilde{S}_F \supseteq \{x \in \mathbb{R}^+ \cup \mathbb{R}^- : g = \textstyle\int t/\{-x - tg\}dH(t) \text{ has a solution in } \mathbb{C}^+\}.$$

Now let us show the analyticity of $f(x)$ on $\tilde{S}_F$. That is, given any $x_o \in \tilde{S}_F$, $f(x)$ has a formal power series expansion near x_0. Without loss of generality, assume that $x_0 > 0$. Note that $\tilde{S}_F$ is an open set due to the continuity of $g(x)$. So it can further be assumed that $0 \notin (x_0 - \delta_0, x_0 + \delta_0) \subset \tilde{S}_F$, for some $\delta_0 > 0$. For clarity of our following argument, recall that $\sqrt{(\cdot)}$ denotes the analytic branch of the complex square root function that has positive imaginary part in the upper complex plane.

Following the same idea as Silverstein and Choi (1995), let us first construct an function that plays the role of the inverse function of $g(z)$ but not exactly equal to it. Observe that for any $z \in \mathbb{C}^+$,

$$g(z) = \int t/\{-z - tg(z)\}dH(t) \Leftrightarrow z^2 = -\{z/g(z)\}^2 \int t/\{t + z/g(z)\}dH(t).$$

Write $\phi(z) = -z/g(z)$. Then

$$z^2 = -\phi^2(z) - \phi^3(z) \int 1/\{t - \phi(z)\}dH(t). \tag{3.3.13}$$

Therefore, we define the following function:

$$q(\phi) = -\phi^2 - \phi^3 \int \frac{1}{t - \phi} dH(t).$$

We can see that since $q(\phi)$ is analytic in $\mathbb{C}^+$, the composition function $\sqrt{q(\phi)}$ is analytic in $\mathbb{C}^+$. Furthermore, as $z \in \mathcal{D} = \{z = z_1 + iz_2 : z_1 > 0, z_2 > 0\}$, $z^2 \in \mathbb{C}^+$ implying that $\sqrt{z^2} = z$ and so from (3.3.13)

$$z = \sqrt{q(\phi(z))}, \quad \text{for any } z \in \mathcal{D}. \tag{3.3.14}$$

Let $\phi_0 = -x_0/g(x_0)$. Then $\phi_0 \in \mathbb{C}^+$ and so $\lim_{\phi \to \phi_0} \sqrt{q(\phi)} = \sqrt{q(\phi_0)}$. However, since $\phi(z) \to \phi_0$ (as $z \in \mathbb{C}^+ \to x_0$), we have $\lim_{z \to x_0} \sqrt{q(\phi(z))} = \lim_{\phi \to \phi_0} \sqrt{q(\phi)}$. It follows that $x_0 = \sqrt{q(\phi_0)}$. Similarly, we get

$$x = \sqrt{q(\phi(x))}, \quad \text{for any } x \in (x_0 - \delta_0, x_0 + \delta_0). \tag{3.3.15}$$

Note that from $g(x_0) = \int t/\{-x_0 - tg(x_0)\}dH(t)$ and $Im(g(x_0)) > 0$, we have

$$\int t/(t - \phi_0)dH(t) = -x_0^2/\phi_0^2$$

and, since $\int t^2/|x_0 + tg(x_0)|^2 dH(t) = 1$,

$$\int t^2/(x_0 + tg(x_0))^2 dH(t) \neq 1 \Leftrightarrow \int t^2/(t - \phi_0)^2 dH(t) \neq x_0^2/\phi_0^2.$$

These give

$$q'(\phi_0) = \int t^2/(t - \phi_0)^2 dH(t) + \int t/(t - \phi_0) dH(t) \neq 0.$$

Thus,

$$\left.\frac{d}{d\phi}\sqrt{q(\phi)}\right|_{\phi=\phi_0} \neq 0.$$

Thus we can now use the inverse function theorem of complex analysis to assert that there is a neighborhood U of ϕ_0 and a neighborhood V of x_0 such that $\sqrt{q(\phi)} : U \to V$ has an analytic inverse function. Let us denote this inverse function by $\hat{\phi}(z)$. Then (3.3.14) and (3.3.15) imply that

$$\hat{\phi}(z) = \phi(z)(= -z/g(z)), \quad \text{for any } z \in (\mathcal{D} \cup (x_0 - \delta_0, x_0 + \delta_0))) \cap V.$$

It follows that $g(z)$ has an analytic extension onto V. Thus near x_0, $g(z)$ can be expanded into $g(z) = \sum_{n=0}^{\infty} a_n (z - x_0)^n$ and hence

$$f(x) = -\frac{2}{\pi} \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} Re(a_k) Im(a_{n-k}) \right) (x - x_0)^n,$$

i.e. $f(x)$ is analytic on $\tilde{S}_F$. This completes the proof of $(3)(ii)$.

Before we prove $(3)(iii)$, let us show that $g(z)$ is analytic in $\mathbb{C}^+$.

This is a consequence of that $g(z)$ is continuous in $\mathbb{C}^+$. To see this, we note that $s_F(z)$ is analytic in $\mathbb{C}^+$ implying $\{g(z)\}^2$ is analytic in $\mathbb{C}^+$. Recall that the definition of an analytic function in an open set is only saying that the function is differentiable in the set. From the following relation

$$\frac{g(z)-g(z_0)}{z-z_0}\{g(z)+g(z_0)\} = \frac{\{g(z)\}^2-\{g(z_0)\}^2}{z-z_0},$$

provided that the continuity of $g(z)$ in $\mathbb{C}^+$ is shown, the analyticity of $\{g(z)\}^2$ in $\mathbb{C}^+$ then implies that of $g(z)$ in $\mathbb{C}^+$. Here we also remind that the assumption $H \neq I_{[0,\infty)}$ guarantees $g(z_0) \neq 0$.

Therefore, we need only prove $g(z)$ is continuous in $\mathbb{C}^+$. Given any $z_0 \in \mathbb{C}^+$, choose $\delta > 0$ such that $B(z_0,\delta) = \{z \in \mathbb{C} : |z-z_0| < \delta\} \subset \mathbb{C}^+$. Then there exists $\varepsilon > 0$ such that $Im(z) > \varepsilon$ for any $z \in B(z_0,\delta)$. Further note that previously we have shown for any $z \in \mathbb{C}^+$, $\int t^2/|z+tg(z)|^2 dH(t) < 1$. Denote $r_0 = \{\int t^2/|z_0+tg(z_0)|^2 dH(t)\}^{1/2}$. By Hölder's inequality, we then get $|\int t/[\{z+tg(z)\}\{z_0+tg(z_0)\}]dH(t) < 1/\varepsilon$ and $|\int t^2/[\{z+tg(z)\}\{z_0+tg(z_0)\}]dH(t) < r_0 < 1$. From

$$\begin{aligned} g(z)-g(z_0) &= \int \frac{t}{\{z+tg(z)\}\{z_0+tg(z_0)\}} dH(t)(z-z_0) \\ &\quad + \int \frac{t^2}{\{z+tg(z)\}\{z_0+tg(z_0)\}} dH(t)(g(z)-g(z_0)), \end{aligned}$$

it follows $|g(z)-g(z_0)| \leq \frac{1}{\varepsilon(1-r_0)}|z-z_0| \to 0$. Thus $g(z)$ is continuous at z_0. We conclude $g(z)$ is analytic in $\mathbb{C}^+$.

Now let us extend the definition of $g(z)$ to $\mathbb{C}^-$. This can be done by straightforwardly reproving Theorem 1.1.1 for $z \in \mathbb{C}^-$. Examining the proof we can see that with suitable adjustment all our arguments remain valid. In that case, we in parallel get $s_F(z)$ associated with $g(z)$ satisfy the system of equations in Theorem 1.1.1

but only that $Im(s_F(z)) < 0$ and $Im(g(z)) \leq 0$ in accordance with $Im(z) < 0$. Then for $z \in \mathbb{C}^-$, $g(z)$ with $Im(g(z)) \leq 0$ satisfies $g(z) = \int t/\{-z - tg(z)\}dH(t)$. Thus, $\overline{g(z)} = \int t/\{-\bar{z} - t\overline{g(z)}\}dH(t)$. Note that $\bar{z} \in \mathbb{C}^+$ and $Im(\overline{g(z)}) \geq 0$. By Lemma 3.1.2, we then get $g(\bar{z}) = \overline{g(z)}$, for $z \in \mathbb{C}^-$.

In view of the analyticity of $g(z)$ in $\mathbb{C}^+$, we assert that after extension $g(z)$ is analytic in $\mathbb{C}^+ \cup \mathbb{C}^-$.

Further, the relation $\overline{g(z)} = g(\bar{z})$ implies that for any $x \in \mathbb{R}^+ \cup \mathbb{R}^-$,

$$\lim g(z) = \overline{g(x)}, \text{ as } z \in \mathbb{C}^- \to x,$$

where $g(x)$ is defined in (2), *i.e.*

$$g(x) = \lim g(z), \text{ as } z \in \mathbb{C}^+ \to x.$$

(iii) The sufficiency part is obvious. In fact, for any $x \in (x_0 - \delta_0, x_0 + \delta_0)$, $g(x) = \int t/\{-x - tg(x)\}dH(t)$ implies $Im(g(x)) = Im(g(x)) \int t^2/|x + tg(x)|^2 dH(t)$. Thus, by the hypothesis of the sufficiency part, it follows that for any $x \in (x_0 - \delta_0, x_0 + \delta_0)$, $Im(g(x)) = 0$ and so $f(x) = 0$, where $f(x)$ is the derivative of $F(x)$. This implies that for any $\delta < \delta_0$, $F((x_0 - \delta, x_0 + \delta)) = 0$ and so $(x_0 - \delta, x_0 + \delta) \subset S_F^c$, where $F((x_0 - \delta, x_0 + \delta))$ denotes the measure of the interval $(x_0 - \delta, x_0 + \delta)$ with respect to the measure specified by the distribution function F on the real line. Thus, $x_0 \in S_F^c$.

Now let us prove the necessity part. Suppose $x_0 \neq 0$ and $x_0 \in S_F^c$. Then, since S_F^c is open, there exists $\delta > 0$ such that $0 \notin (x_0 - \delta, x_0 + \delta) \subset S_F^c$. Define $B(x_0, \delta) = \{z \in \mathbb{C} : |z - x_0| < \delta\}$. Then $S_F(z)$ is analytic in $B(x_0, \delta)$ and, for any $x \in (x_0 - \delta, x_0 + \delta)$, $s_F(x) = \int 1/(\lambda - x) dF(\lambda)$, $s'_F(x) = \int 1/(\lambda - x)^2 dF(\lambda)$. Thus, for any $x \in (x_0 - \delta, x_0 + \delta)$, $g(x) \in \mathbb{R}$ so that

$$\lim_{z \in \mathbb{C}^-} g(z) = \lim_{z \in \mathbb{C}^+} g(z) = g(x).$$

This implies that $g(z)$ is continuous in $B(x_0, \delta)$. Remember that since $H \neq I_{[0,\infty)}$, $g(z) \neq 0$, for any $z \neq 0$. Similar to our previous argument, by using the analyticity of $\{g(z)\}^2$, we then get $g(z)$ is analytic in $B(x_0, \delta)$. It is proper to define $\phi(z) = -z/g(z)$. Then $\phi(z)$ is also analytic in $B(x_0, \delta)$.

Let $\phi_0 = \phi(x_0)$. We first show H is differentiable at ϕ_0 with derivative zero. By Lemma 2.3.2, this will follow once it is proven $\lim_{\phi \in \mathbb{C}^+ \to \phi_0} s_H(\phi) \in \mathbb{R}$. Note that from $g(z) = \int t/\{-z - tg(z)\} dH(t)$, we have $s_H(\phi(z)) = z^{-1}[g(z) + \{g(z)\}^3]$. In view of the fact that $\lim_{z \to x_0} s_H(\phi(z)) = -x_0^{-1}[g(x_0) + \{g(x_0)\}^3] \in \mathbb{R}$, we only need to show $\lim_{\phi \in \mathbb{C}^+ \to \phi_0} s_H(\phi) \in \mathbb{R} = \lim_{z \to x_0} s_H(\phi(z))$. This in turn follows once it is proven $\phi'(x_0) \neq 0$. In fact, $\phi'(x_0) \neq 0$ guarantees we can use the inverse function theorem of complex analysis to see that for any ε sufficiently small, there is η such that for any $\phi \in B(\phi_0, \eta)$, there exists $z \in B(x_0, \varepsilon)$ such that $\phi = \phi(z)$. Hence we only need to show that $\phi'(x_0) \neq 0$.

Since $g(z)$ is analytic in $B(x_0, \delta_0)$, $\phi'(x) = \{g'(x) - g(x)/x\}/[x\{g(x)\}^2]$. From $\{g(x)\}^2 = -1 - xs_F(x)$, we get $-2g(x)g'(x) = s_F(x) + xs'_F(x)$. Since F is symmetric, it follows

$$\begin{aligned} s_F(x) + x s_F'(x) &= \int \frac{\lambda}{(\lambda - x)^2} dF(\lambda) \\ &= \int_0^\infty \left\{ \frac{\lambda}{(\lambda - x)^2} - \frac{\lambda}{(\lambda + x)^2} \right\} dF(\lambda). \end{aligned}$$

For $\lambda > 0$, if $x > 0$, $|\lambda - x| < |\lambda + x|$ hence the integrand of the above last relation is positive; similarly, if $x < 0$, the integrand is negative. Thus $s_F(x) + x s_F'(x)$, or equivalently, $-2g(x)g'(x)$ has the same sign as x. Note that since it always holds that $Re(g(x))/x < 0$, now that $g(x) \in \mathbb{R}$, it holds that $g(x)/x < 0$. Thus, $g'(x) > 0$. It follows that $\phi'(x)$ has the same sign as x for any $x \in (x_0 - \delta_0, x_0 + \delta_0)$. So, $\phi'(x) \neq 0$.

As claimed previously, this implies that H is differentiable at ϕ_0 with derivative zero. Note that the above argument applies equally well to any other point $x \in (x_0 - \delta_0, x_0 + \delta_0)$. Thus, H has derivative 0 at $\phi(x)$, for any $x \in (x_0 - \delta_0, x_0 + \delta_0)$. Note that $\phi(z)$ is analytic in $B(x_0, \delta_0)$. By the open mapping theorem, $\{\phi(x) : x \in (x_0 - \delta_0, x_0 + \delta_0)\}$ is an open set and so contains the interval $(\phi_0 - \eta_0, \phi_0 + \eta_0)$, for any η_0 sufficiently small. Hence we get $(\phi_0 - \eta_0, \phi_0 + \eta_0) \subset S_H^c$. Without loss of generality, let η_0 be small enough such that for any $\phi \in (\phi_0 - \eta_0, \phi_0 + \eta_0)$, $|t - \phi| > d_0 > 0$, for any $t \in S_H$. Further, from the continuity of the function $\phi(x)$ in $(x_0 - \delta_0, x_0 + \delta_0)$, there is ε_0 such that for any $x \in (x_0 - \varepsilon_0, x_0 + \varepsilon_0)$, $\phi(x) \in (\phi_0 - \eta_0, \phi_0 + \eta_0)$.

It follows that for any $x \in (x_0 - \varepsilon_0, x_0 + \varepsilon_0)$

$$\begin{aligned}
|\frac{1}{x+tg(x)}| &\leq \frac{1}{d_0|g(x)|}, \\
|\frac{t}{x+tg(x)}| &= \left|\frac{1}{g(x)}\left[1 - \frac{x}{g(x)}\frac{1}{t+x/g(x)}\right]\right| \\
&\leq \frac{1}{|g(x)|}\left[1 + \frac{|x|}{d_0|g(x)|}\right].
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{g(x)-g(x_0)}{x-x_0} &= \int \frac{t}{\{x+tg(x)\}\{x_0+tg(x_0)\}}dH(t) \\
&+ \int \frac{t^2}{\{x+tg(x)\}\{x_0+tg(x_0)\}}dH(t)\frac{g(x)-g(x_0)}{x-x_0}.
\end{aligned}$$

Therefore, by the dominated convergence theorem,

$$g'(x_0) = \int t/\{x_0+tg(x_0)\}^2 dH(t) + \int t/\{x_0+tg(x_0)\}^2 dH(t) g'(x_0),$$

which implies, due to $H \neq I_{[0,\infty)}(t)$, $\int t^2/\{x_0+tg(x_0)\}^2 dH(t) < 1$. Similarly, the above argument applies to any $x \in (x_0 - \delta_0, x_0 + \delta_0)$. Thus,

$$\int \frac{t^2}{|x+tg(x)|^2}dH(t) < 1, \quad \text{for any } x \in (x_0 - \delta_0, x_0 + \delta_0).$$

This completes the proof of Theorem 1.1.2 (3)(iii).

3.4 Density Function of the LSD: Proof of Theorem 1.1.3

We shall in the present section mainly focus on calculating the density of the limiting spectral distribution when the matrices T_n are given or more precisely when their limiting spectral distribution is known.

As is stated in the previous section, the density $f(x)$ is positive at a point x ($x \neq 0$) if and only if at x, the equation $g = \int t/(-x - tg)dH(t)$ admits a unique solution $g = g_1 + ig_2$ such that $g_2 > 0$ and $g_1/x < 0$, in which case $f(x) = -\frac{2g_1g_2}{\pi x}$. This indicates we can calculate the density function for given H through solving the equation for a such solution.

Obviously, we cannot get an explicitly expressed density function for any arbitrarily given H. We are interested in the case when H denotes the limiting spectral distribution of the multivariate F-matrices. Let $B_n = S_n V_n^{-1}$ be the multivariate F-matrices we defined in Section 2.3.3, where we obtained their limiting spectral distribution $F_{y,y'}$ and its Stieltjes transform. Recall that for each $z \in \mathbb{C}^+$,

$$s_{F_{y,y'}}(z) = \frac{(1-y) - z(1+y') + \sqrt{[z(1-y') + (1-y)]^2 - 4z}}{2z(y + y'z)}. \tag{3.4.1}$$

Setting $y' = 0$, we get the Stieltjes transform of the limiting spectral distribution of the sample covariance matrices S_n, which is written here as

$$s_y(z) = \frac{-z + 1 - y + \sqrt{(1 + y - z)^2 - 4y}}{2yz}; \tag{3.4.2}$$

and setting $y = 0$, we get the Stieltjes transform of the limiting spectral distribution of the inverse matrices of the sample covariance matrices V_n^{-1}, which is written here as

$$s_{y'}(z) = \frac{1 - z - zy' + \sqrt{(1 + z - zy')^2 - 4z}}{2y'z^2}. \tag{3.4.3}$$

Thus the limiting spectral distribution of the sample covariance matrices (their inverse matrices) correspond to the special case of $y' = 0$ ($y = 0$) of $F_{y,y'}$.

We first see when $H = F_{y,y'}$, what form the equation $g = \int t/(-x - tg)dH(t)$ will take. If $H = F_{y,y'}$, then the equation implies

$$
\begin{aligned}
-g^2 - 1 &= -\frac{x}{g} s_{y,y'}(-\frac{x}{g}) \\
&= \frac{(1-y) + \frac{x}{g}(1+y') \pm \sqrt{[(1-y) - \frac{x}{g}(1-y')]^2 + 4\frac{x}{g})}}{2(y - \frac{x}{g}y')},
\end{aligned}
$$

and then

$$
\begin{aligned}
&2[-yg^2 + y'xg + (y'-1)\frac{x}{g} - y] - [(1-y) - \frac{x}{g}(1-y')] \\
&= \pm\sqrt{[(1-y) - \frac{x}{g}(1-y')]^2 + 4\frac{x}{g}},
\end{aligned}
$$

which can be organized into the following equation:

$$
\begin{aligned}
&y^2g^5 - 2yy'xg^4 + [y'^2x^2 + y(1+y)]g^3 \\
&+x[y - y' - 2yy']g^2 + [y + y'(y'-1)x^2]g - y'x = 0. \qquad (3.4.4)
\end{aligned}
$$

Thus when the matrices T_n are taken to be the general multivariate F-matrices, we are facing a quintic equation (degree 5 polynomial equation). However, if in the equation we take $y = 0$, that is we take T_n to be the inverse matrices of the sample covariance matrices, we get

$$
y'xg^3 - g^2 + (y'-1)xg - 1 = 0; \qquad (3.4.5)
$$

If we let $y' = 0$, we get when T_n are taken to be the sample covariance matrices, the equation becomes

$$
yg^4 + (1+y)g^2 + xg + 1 = 0. \qquad (3.4.6)
$$

As is shown above, when the two special cases of the multivariate F-matrices are considered instead, the equation $g = \int t/(-x-tg)dH(t)$ reduces from a quintic equation to a cubic and a quartic equation. As is known now, the general degree 5 and above polynomial equations are not solvable by hand, whereas the cubic and quartic equations can. Actually, we can see below, by discussing the solutions to two cubic equations with real coefficients, Theorem 1.1.3 follows. That is, we also do not need to solve a quartic equation. This is a result by separately setting the real and imaginary part of the left hand side of (3.4.5) and (3.4.6) zero and then reorganizing the obtained equations. Now we introduce the Cardano method for solving a cubic equation.

Let $x^3 + px^2 + qx + r = 0$ be a cubic equation for x, where p, q, and r are real. Replace x by $y - p/3$. Then we get $y^3 + sy + t = 0$, where $s = q - p/3$ and $t = r - pq/3 + 2p^3/27$. Further replace y by $u + v$. Then we have a system of equations for (u, v):

$$\begin{cases} 3uv = -s, \\ u^3 + v^3 = -t. \end{cases}$$

Solve this system for (u^3, v^3) to have $u^3, v^3 = (-t \pm \sqrt{t^2 + 4s^3/27})/2$. Denote by $\Delta = t^2 + 4s^3/27$. Then $u^3, v^3 = (-t \pm \sqrt{\Delta})/2$. If $\Delta \geq 0$, $(-t \pm \sqrt{\Delta})/2$ are real and by letting $[(-t \pm \sqrt{\Delta})/2]^{1/3}$ denote the real cubic root of a real number, we have

$$\begin{aligned} u &= \left(\frac{-t + \sqrt{\Delta}}{2}\right)^{1/3}, \qquad \text{or } \left(\frac{-t + \sqrt{\Delta}}{2}\right)^{1/3}\left(-\frac{1}{2} \pm i\frac{\sqrt{3}}{2}\right), \\ v &= \left(\frac{-t - \sqrt{\Delta}}{2}\right)^{1/3}, \qquad \text{or } \left(\frac{-t - \sqrt{\Delta}}{2}\right)^{1/3}\left(-\frac{1}{2} \pm i\frac{\sqrt{3}}{2}\right). \end{aligned}$$

Note that a solution to the system of equation for (u, v) needs to satisfy $uv \in \mathbb{R}$.

It follows that the three cubic roots for the original cubic equation are

$$\begin{aligned}
x_1 &= -p/3 + (\frac{-t+\sqrt{\Delta}}{2})^{1/3} + (\frac{-t-\sqrt{\Delta}}{2})^{1/3}, \\
x_2 &= -p/3 - \frac{1}{2}[(\frac{-t+\sqrt{\Delta}}{2})^{1/3} + (\frac{-t-\sqrt{\Delta}}{2})^{1/3}] \\
&+ i\frac{\sqrt{3}}{2}[(\frac{-t+\sqrt{\Delta}}{2})^{1/3} - (\frac{-t-\sqrt{\Delta}}{2})^{1/3}], \\
x_3 &= -p/3 - \frac{1}{2}[(\frac{-t+\sqrt{\Delta}}{2})^{1/3} + (\frac{-t-\sqrt{\Delta}}{2})^{1/3}] \\
&- i\frac{\sqrt{3}}{2}[(\frac{-t+\sqrt{\Delta}}{2})^{1/3} - (\frac{-t-\sqrt{\Delta}}{2})^{1/3}]. \qquad (3.4.7)
\end{aligned}$$

Here we define that for any real number c, $c^{1/3}$ denotes its real cubic root. If $\Delta < 0$, then $|\Delta| = -\Delta$ and

$$(-t \pm \sqrt{\Delta})/2 = -t/2 \pm i\sqrt{|\Delta|}/2 \equiv \rho(\cos\varphi \pm i\sin\varphi),$$

where $\rho = \sqrt{t^2 - \Delta}/2 = \sqrt{-s^3/27}$, and $\cos\varphi = -t/(2\rho)$. Then by letting $\rho^{1/3}$ denote the real cubic root of ρ, we have

$$u = \rho^{\frac{1}{3}}e^{i\frac{\varphi}{3}}, \rho^{\frac{1}{3}}e^{i(\frac{\varphi}{3}+\frac{2\pi}{3})}, \rho^{\frac{1}{3}}e^{i(\frac{\varphi}{3}+\frac{4\pi}{3})},$$

$$v = \rho^{\frac{1}{3}}e^{i(-\frac{\varphi}{3})}, \rho^{\frac{1}{3}}e^{i(-\frac{\varphi}{3}+\frac{4\pi}{3})}, \rho^{\frac{1}{3}}e^{i(-\frac{\varphi}{3}+\frac{2\pi}{3})}.$$

Then we get the three cubic roots to the original equation are

$$\begin{aligned}
x_1 &= -p/3 + 2\rho^{1/3}\cos\frac{\varphi}{3}, \\
x_2 &= -p/3 + 2\rho^{1/3}\cos(\frac{\varphi}{3} + \frac{2\pi}{3}), \\
x_3 &= -p/3 + 2\rho^{1/3}\cos(\frac{\varphi}{3} + \frac{4\pi}{3}). \qquad (3.4.8)
\end{aligned}$$

Note that in this case all three roots are real numbers. Also, it is a simple matter to examine (3.4.7) and (3.4.8) coincide in the case $\Delta = 0$.

Proof of Theorem 1.1.3. (1) Our task is to determine S and for each $x \in S$, find the solution $g = g_1 + ig_2$ to (3.4.5) satisfying $g_2 > 0$ and $g_1/x < 0$. In (3.4.5), write $g = g_1 + ig_2$. Then we get two equations:

$$y'(g_1^3 - 3g_1g_2^2) - x^{-1}(g_1^2 - g_2^2) + (y' - 1)g_1 - x^{-1} = 0, \tag{3.4.9}$$

$$y'(3g_1^2g_2 - g_2^3) - x^{-1}2g_1g_2 + (y' - 1)g_2 = 0. \tag{3.4.10}$$

From (3.4.10),

$$g_2^2 = 3g_1^2 - \frac{2}{xy'}g_1 + (1 - \frac{1}{y'}). \tag{3.4.11}$$

Substitute (3.4.11) into (3.4.9). We get

$$g_1^3 - \frac{1}{xy'}g_1^2 + (\frac{y' - 1}{4y'} + \frac{1}{4x^2y'^2}) + \frac{1}{8xy'^2} = 0. \tag{3.4.12}$$

Let $g_2^2 > 0$. From (3.4.11), we have

$$3g_1^2 - \frac{2}{xy'}g_1 + (1 - \frac{1}{y'}) > 0,$$

which gives

$$\frac{g_1}{x} < \frac{1 - \sqrt{1 + 3x^2y'(1 - y')}}{3x^2y'} \text{ or } \frac{g_1}{x} > \frac{1 + \sqrt{1 + 3x^2y'(1 - y')}}{3x^2y'}.$$

Note that we need $g_1/x < 0$. It follows that $x \in S$ if and only if (3.4.12) has a solution such that

$$\frac{g_1}{x} < \frac{1 - \sqrt{1 + 3x^2y'(1 - y')}}{3x^2y'}. \tag{3.4.13}$$

Using the Cardano method to solve (3.4.12), we have in this case,

$$p = -\frac{1}{xy'}, q = \frac{y' - 1}{4y'} + \frac{1}{4x^2y'^2}, r = \frac{1}{8xy'^2}.$$

Then

$$s = q - \frac{p^2}{3} = -\frac{1}{12x^2y'^2}[1 + 3x^2y'(1-y')],$$

$$t = \frac{2}{27}p^3 - \frac{1}{3}pq + r = \frac{1}{36x^3y'^3}(\frac{1}{3} + 3x^2y'(y' + \frac{1}{2})). \tag{3.4.14}$$

We can see that $s < 0$ since $0 < y' < 1$. Let $\Delta = t^2 + 4s^3/27$. Then

$$t^2 - \Delta = \frac{4}{27}\{\frac{1}{12x^2y'^2}[1 + 3x^2y'(1-y')]\}^3. \tag{3.4.15}$$

If at x, $\Delta \leq 0$, then the three roots of (3.4.12) are

$$-p/3 + 2\rho^{1/3}\cos\frac{\varphi}{3}, -p/3 + 2\rho^{1/3}\cos(\frac{\varphi}{3} + \frac{2\pi}{3}), -p/3 + 2\rho^{1/3}\cos(\frac{\varphi}{3} + \frac{4\pi}{3}),$$

where

$$\rho^{1/3} = \sqrt{-s/3} = \frac{\sqrt{1 + 3x^2y'(1-y')}}{6|x|y'}. \tag{3.4.16}$$

Divided by x, the above three roots become

$$\frac{1}{3x^2y'} + \frac{\sqrt{1 + 3x^2y'(1-y')}}{3|x|xy'}\cos(\alpha), \tag{3.4.17}$$

where $\alpha = \varphi/3$, $\varphi/3 + 2\pi/3$, and $\varphi/3 + 4\pi/3$. Since $|\cos(\alpha)| \leq 1$, we have

$$\frac{1}{3x^2y'} + \frac{\sqrt{1 + 3x^2y'(1-y')}}{3|x|xy'}\cos(\alpha) \geq \frac{1 - \sqrt{1 + 3x^2y'(1-y')}}{3x^2y'}. \tag{3.4.18}$$

This means if $x \neq 0$ such that $\Delta \leq 0$, then (3.4.12) does not have a solution satisfying (3.4.13). Indeed, as can be seen below, the solution exists for the case when at x, $\Delta > 0$. This means the support set of the density function consists of those x's such that $\Delta > 0$. Note that

$$\Delta = \frac{1}{432x^4y'^4}[1 + (2y'^2 + 5y' - \frac{1}{4})x^2 - y'(1-y')^3x^4]. \tag{3.4.19}$$

Thus $\Delta > 0$ means

$$y'(1-y')^3x^4 - (2y'^2 + 5y' - \frac{1}{4})x^2 - 1 < 0.$$

It follows that

$$0 < x^2 < \frac{(2y'^2 + 5y' - \frac{1}{4}) + \sqrt{32y'^3 + 12y'^2 + \frac{3}{2}y' + \frac{1}{16}}}{2y'(1-y')^3}. \tag{3.4.20}$$

When $\Delta > 0$, (3.4.12) has only one real root:

$$g_1 = \frac{1}{3xy'} + (\frac{-t+\sqrt{\Delta}}{2})^{1/3} + (\frac{-t-\sqrt{\Delta}}{2})^{1/3}, \tag{3.4.21}$$

t and Δ given by (3.4.14) and (3.4.19). We can check when $\Delta > 0$, g_1 satisfies (3.4.13), or equivalently, when $x > 0$,

$$g_1 < [1 - \sqrt{1 + 3x^2y'(1-y')}]/(3xy');$$

when $x < 0$,

$$g_1 > [1 - \sqrt{1 + 3x^2y'(1-y')}]/(3xy').$$

Note that $t^2 - \Delta = -4s^3/27 > 0$. Thus $|t| > \sqrt{\Delta}$. Also, note that t and x have the same sign. Thus when $x > 0$, $(-t + \sqrt{\Delta})/2$ and $(-t - \sqrt{\Delta})/2$ are both negative. By Hölder's inequality, when $\Delta > 0$,

$$\begin{aligned} & -(\frac{-t+\sqrt{\Delta}}{2})^{1/3} - (\frac{-t-\sqrt{\Delta}}{2})^{1/3} > 2\sqrt{-\frac{s}{3}} \\ = \quad & \frac{\sqrt{1+3x^2y'(1-y')}}{3|x|y'} = \frac{\sqrt{1+3x^2y'(1-y')}}{3xy'}, \end{aligned}$$

and thus

$$g_1 < \frac{1}{3xy'} - \frac{\sqrt{1+3x^2y'(1-y')}}{3xy'} = \frac{1-\sqrt{1+3x^2y'(1-y')}}{3xy'}.$$

When $x < 0$, $t < 0$, and thus $(-t + \sqrt{\Delta})/2$ and $(-t - \sqrt{\Delta})/2$ are both positive. Similarly, by Hölder's inequality, when $\Delta > 0$,

$$
\begin{aligned}
& (\frac{-t+\sqrt{\Delta}}{2})^{1/3} + (\frac{-t-\sqrt{\Delta}}{2})^{1/3} \\
> \; & \frac{\sqrt{1+3x^2y'(1-y')}}{3|x|y'} = -\frac{\sqrt{1+3x^2y'(1-y')}}{3xy'}.
\end{aligned}
$$

Thus

$$
g_1 > \frac{1}{3xy'} - \frac{\sqrt{1+3x^2y'(1-y')}}{3xy'} = \frac{1-\sqrt{1+3x^2y'(1-y')}}{3xy'}.
$$

That is, when at x, $\Delta > 0$, g_1 given by (3.4.21), is the solution to (3.4.12) satisfying (3.4.13). Therefore by substituting g_1 given by (3.4.21) into (3.4.11), we obtain the expression of the density function for points in the support set given by (3.4.20).

By using Taylor's expansion, it can further be shown that

$$
\lim_{x\to 0} g_1(x)/x = -(2/3)y' - 1/2, \quad \lim_{x\to 0} g_2(x) = \sqrt{7/3}
$$

and so

$$
\lim_{x\to 0} f(x) = \frac{1}{\pi}(\frac{4}{3}y' + 1)\sqrt{\frac{7}{3}}.
$$

Thus we obtained the density function for the case when T_n are known to be the inverse matrices of the sample covariance matrices.

To prove the next part in Theorem 1.1.3, we again start by writing $g = g_1 + ig_2$ in (3.4.6). Then we get

$$
y([g_1^2 - g_2^2]^2 - 4g_1^2g_2^2) + (1+y)[g_1^2 - g_2^2] + xg_1 + 1 = 0, \tag{3.4.22}
$$

$$
y4g_1g_2[g_1^2 - g_2^2] + 2(1+y)g_1g_2 + xg_2 = 0. \tag{3.4.23}
$$

From (3.4.23),

$$g_2^2 = g_1^2 + (\frac{1}{2} + \frac{1}{2y}) + \frac{x}{4yg_1}. \tag{3.4.24}$$

Substitute (3.4.24) into (3.4.22). We have

$$g_1^6 + \frac{1+y}{2y} g_1^4 + \frac{(1-y)^2}{16y^2} g_1^2 - \frac{x^2}{64y^2} = 0. \tag{3.4.25}$$

From (3.4.24), we have

$$g_2 = \sqrt{g_1^2 + (\frac{1}{2} + \frac{1}{2y}) + \frac{x}{4yg_1}}. \tag{3.4.26}$$

Thus $g_2 > 0$ implies

$$g_1^2 + (\frac{1}{2} + \frac{1}{2y}) > -\frac{x}{4yg_1}. \tag{3.4.27}$$

Note that $-x/g_1 > 0$. It follows

$$[g_1^2 + (\frac{1}{2} + \frac{1}{2y})]^2 > \frac{x^2}{16y^2 g_1^2},$$

or equivalently,

$$g_1^6 + (1 + \frac{1}{y}) g_1^4 + \frac{1}{4}(1 + \frac{1}{y})^2 g_1^2 - \frac{x^2}{16y^2} > 0. \tag{3.4.28}$$

From (3.4.25),

$$\frac{x^2}{16y^2} = 4g_1^6 + \frac{2(1+y)}{y} g_1^4 + \frac{(1-y)^2}{4y^2} g_1^2.$$

Substitute it into (3.4.28). We get

$$-3g_1^6 - \frac{1+y}{y} g_1^4 + \frac{1}{y} g_1^2 > 0, \tag{3.4.29}$$

and thus

$$g_1^4 + \frac{1+y}{3y} g_1^2 - \frac{1}{3y} < 0. \tag{3.4.30}$$

It follows that

$$0 < g_1^2 < \frac{-(1+y) + \sqrt{1+y^2+14y}}{6y}. \tag{3.4.31}$$

Thus we only need to determine for whichever x's, (3.4.25) viewed as a cubic equation for g_1^2 has a solution satisfying (3.4.31) and find the solution for these x's. Then with $p = \frac{1+y}{2y}$, $q = \frac{(1-y)^2}{16y^2}$, and $r = -\frac{x^2}{64y^2}$, we have

$$\begin{aligned} s &= q - \frac{p^2}{3} = -\frac{1+y^2+14y}{48y^2}, \\ t &= \frac{2}{27}p^3 - \frac{1}{3}pq + r = \frac{1}{(12y)^3}[-2(1+y)^3 + 72y(1+y) - 27x^2 y]. \end{aligned}$$

Let $\Delta = t^2 + 4s^3/27$. Then

$$t^2 - \Delta = -\frac{4s^3}{27} = \frac{4}{27}(\frac{1+y^2+14y}{48y^2})^3 > 0.$$

Let's first consider the case $\Delta \geq 0$. Then $t^2 - \Delta > 0$ implies $|t| > \sqrt{\Delta}$. Thus, $(-t+\sqrt{\Delta})/2$ and $(-t-\sqrt{\Delta})/2$ have the same sign. From Hölder's inequality, it follows that $|[(-t+\sqrt{\Delta})/2]^{1/3} + [(-t-\sqrt{\Delta})/2]^{1/3}| \geq \sqrt{1+y^2+14y}/(6y)$. Note that in this case, $-(1+y)/(6y) + [(-t+\sqrt{\Delta})/2]^{1/3} + [(-t-\sqrt{\Delta})/2]^{1/3}$ is the only real solution to the equation (3.4.25) and as is shown above, it does not satisfy the condition (3.4.31). That is, if at x, $\Delta \geq 0$, then (3.4.22) does not have a pair of solution (g_1, g_2) with $g_2 > 0$ and $g_1/x < 0$. However, as can be seen below, such a pair of solution exists if and only if at x, $\Delta < 0$. Then by solving $\Delta < 0$, we get the support set of the density function $f(x)$. Since $\Delta = t^2 + 4s^3/27$, $\Delta < 0$ gives

$-2\sqrt{-s^3/27} < t < 2\sqrt{-s^3/27}$. Compute that $2\sqrt{-s^3/27} = \frac{2(1+y^2+14y)^{3/2}}{(12y)^3}$. Thus we get

$$\frac{-2(1+y)^3 + 72y(1+y) - 2(1+y^2+14y)^{3/2}}{27y} < x^2$$
$$< \frac{-2(1+y)^3 + 72y(1+y) + 2(1+y^2+14y)^{3/2}}{27y}.$$

It can be computed

$$[-2(1+y)^3 + 72y(1+y)]^2 = 4(1+y^2+14y)^3 - 27 \times 16y(1-y)^4.$$

Since $x^2 > 0$, we have

$$0 < x^2 < \frac{-2(1+y)^3 + 72y(1+y) + 2(1+y^2+14y)^{3/2}}{27y}. \tag{3.4.32}$$

This gives the support set of the density function. Now for these x, we find the solution to (3.4.25) satisfying the condition (3.4.31). For $\Delta < 0$, $\rho^{1/3} = \sqrt{-s/3} = \sqrt{1+y^2+14y}/(12y)$. Thus the three solutions to (3.4.25) (viewed as a cubic equation for g_1^2) are

$$-\frac{1+y}{6y} + 2\rho^{\frac{1}{3}}\cos\alpha = -\frac{1+y}{6y} + \frac{\sqrt{1+y^2+14y}}{6y}\cos\alpha, \tag{3.4.33}$$

where α takes $\varphi/3$, $\varphi/3 + 2\pi/3$, and $\varphi/3 + 4\pi/3$. Here, φ is defined by

$$\cos\varphi = \frac{-t}{2\rho} = \frac{-t}{2\sqrt{-s^3/27}} = \frac{2(1+y)^3 - 72y(1+y) + 27x^2y}{2(1+y^2+14y)^{3/2}}, \tag{3.4.34}$$

and $\sin\varphi = \sqrt{|\Delta|}/(2\rho)$. Since $\sin\varphi > 0$, we define $\varphi \in (0,\pi)$. Noting that $|\cos\alpha| < 1$, to find the one in the three solutions which satisfies (3.4.31), we only need to consider whichever is positive. From (3.4.33), this is equivalent to deciding which of $\varphi/3$, $\varphi/3+2\pi/3$, and $\varphi/3+4\pi/3$, satisfies $\cos\alpha > (1+y)/\sqrt{1+y^2+14y}$.

Note that $\varphi/3 + 2\pi/3 \in (2\pi/3, \pi)$. Then $\cos(\varphi/3 + 2\pi/3)$ is < 0 and obviously does not satisfy the condition. We only need to consider $\varphi/3$ and $\varphi/3 + 4\pi/3$. In the sequel, we will prove only $\varphi/3$ satisfies the condition. For that purpose, consider the function $c(y) = (1 + y)/\sqrt{1 + y^2 + 14y}$ for $y \in (0, \infty)$. It is easy to see $c'(y) = 6(y - 1)\{\sqrt{1 + y^2 + 14y}\}^{-3/2}$. Thus $c(y)$ achieves a minimum of $1/2$ at $y = 1$ and we get

$$(1 + y)/\sqrt{1 + y^2 + 14y} \in (1/2, 1) = \{\cos(\frac{\theta}{3}) : \theta \in (0, \pi)\}$$

for $y \in (0, \infty)$. Therefore we can choose a unique $\theta \in (0, \pi)$ so that $\cos(\theta/3) = (1 + y)/\sqrt{1 + y^2 + 14y}$. Then $\cos\theta = 4\cos^3(\theta/3) - 3\cos(\theta/3) = (1 + y)(1 + y^2 - 34y)/(1 + y^2 + 14y)^{3/2}$. Then from (3.4.34), since $x^2 > 0$,

$$\cos\varphi > \frac{2(1 + y)^3 - 72y(1 + y)}{2(1 + y^2 + 14y)^{3/2}} = \cos\theta.$$

Since $\varphi, \theta \in (0, \pi)$ and $\cos(\cdot)$ is decreasing on $(0, \pi)$, we have $\varphi < \theta$. Then $0 < \varphi/3 < \theta/3 < \pi$ and $\pi < \varphi/3 + 4\pi/3 < \theta/3 + 4\pi/3 < 2\pi$. Due to the monotonicity of $\cos(\cdot)$ on $(0, \pi)$ and $(\pi, 2\pi)$, respectively, we have $\cos(\varphi/3) > \cos(\theta/3)$ and $\cos(\varphi/3 + 4\pi/3) < \cos(\theta/3 + 4\pi/3)$. Thus $\cos(\varphi/3) > (1 + y)/\sqrt{1 + y^2 + 14y}$ is proved. To show $\cos(\varphi + 4\pi/3) < (1 + y)/\sqrt{1 + y^2 + 14y}$, we only need to show $\cos(\theta/3 + 4\pi/3) < \cos(\theta/3)$. However, this is a consequence of the fact that $\cos(\theta/3) = \cos(2\pi - \theta/3)$ and $\pi < \theta/3 + 4\pi/3 < 2\pi - \theta/3 < 2\pi$. Therefore, we have proved that only $\varphi/3$ satisfies $\cos(\varphi/3) > (1 + y)/\sqrt{1 + y^2 + 14y}$ and thus

$$g_1^2 = -\frac{1 + y}{6y} + \frac{\sqrt{1 + y^2 + 14y}}{6y}\cos(\frac{\varphi}{3}), \tag{3.4.35}$$

with $\varphi \in (0, \pi)$ and $\cos\varphi$ given by (3.4.34) is the only solution to (3.4.25) satisfying

(3.4.31). By substituting g_1^2 into (3.4.26), we get g_2. Since $-g_1/x > 0$, we can write $-g_1/x = \sqrt{g_1^2/x^2}$,

$$-\frac{2g_1g_2}{\pi x} = \frac{2}{\pi}\sqrt{\frac{g_1^2}{x^2}}g_2 = \frac{2}{\pi}\sqrt{\frac{g_1^2}{x^2}}\sqrt{g_1^2 + (\frac{1}{2}+\frac{1}{2y}) - \frac{1}{4y}\sqrt{\frac{x^2}{g_1^2}}}.$$

Thus we get the expression of the density function (1.1.4) for points in the support set defined by (3.4.32).

We then calculate the limit of $\frac{-2g_1(x)g_2(x)}{\pi x}$ as $x \to 0$. Note that as $x \to 0$, $\cos\varphi \to \cos\theta$ and hence

$$\cos\frac{\varphi}{3} \to \cos\frac{\theta}{3} = \frac{1+y}{\sqrt{1+y^2+14y}}.$$

It follows that

$$4(\cos^2\frac{\varphi}{3} + \cos\frac{\varphi}{3}\cos\frac{\theta}{3} + \cos^2\frac{\theta}{3}) - 3 \to \frac{9(1-y)^2}{1+y^2+14y}.$$

Note that

$$\frac{1}{x^2}(\cos\varphi - \cos\theta) = \frac{27y}{2(1+y^2+14y)^{3/2}},$$

$$\begin{aligned}
\frac{1}{x^2}(\cos\frac{\varphi}{3} - \cos\frac{\theta}{3}) &= \frac{\frac{1}{x^2}(\cos\varphi - \cos\theta)}{4(\cos^2\frac{\varphi}{3} + \cos\frac{\varphi}{3}\cos\frac{\theta}{3} + \cos^2\frac{\theta}{3}) - 3} \\
&\to \frac{3y}{2(1-y)^2\sqrt{1+y^2+14y}}.
\end{aligned}$$

Then we get

$$\begin{aligned}
\frac{g_1^2(x)}{x^2} &= \frac{1}{x^2}(-\frac{1+y}{6y} + \frac{\sqrt{1+y^2+14y}}{6y}\cos\theta \\
&\quad + \frac{\sqrt{1+y^2+14y}}{6y}(\cos\varphi - \cos\theta)) \\
&= \frac{\sqrt{1+y^2+14y}}{6y}\frac{1}{x^2}(\cos\varphi - \cos\theta) \\
&\to \frac{1}{4(1-y)^2}.
\end{aligned}$$

It follows $g_1(x) \to 0$, $g_1(x)/x \to -1/(2|1-y|)$ and hence

$$g_2(x) \to \sqrt{\frac{1}{2} + \frac{1}{2y} - \frac{1}{2y}|1-y|}.$$

Thus we also obtained the density function given in (1.1.4) for the case when T_n are known to be the sample covariance matrices. The proof of Theorem 1.1.3 is complete. □

3.5 Existence of the LSD: Proof of Theorem 1.1.4 by Using the Moment Method

In this section, we present a proof of using the moment method to establish the almost sure weak convergence of the empirical spectral distributions of the Wigner type random matrices. We prove Theorem 1.1.4 for the matrices $A_n = \frac{1}{\sqrt{n}} T_n^{1/2} W_n T_n^{1/2}$ defined in Definition 1.1.1.

3.5.1 Truncation and Centralization Treatment

Let us begin by applying the truncation and centralization technique to the matrices W_n and T_n in order to do the proof under the additional conditions in Assumption 3.1.1.

We first prove the condition that $\|T_n\| \le \tau$ can be added without reducing the generality of the result. Define T_n^τ to be the resulting matrix of replacing in the spectral decomposition of T_n those eigenvalues bigger than τ with 0 and A_n^τ the analog of A_n with the matrix T_n replaced by T_n^τ. Then we have if τ is a continuity point of $H(t)$, with probability one, $F^{\tilde{T}_n}(t)$ converges weakly to $H^\tau(t)$ given by (3.1.7).

Suppose that Theorem 1.1.4 is true for the matrix A_n^τ. Then let F^τ denote the limiting spectral distribution of A_n^τ. By Lemma 2.1.1, whenever τ is a continuity point of H, with probability one

$$\|F^{A_n} - F^{A_n^\tau}\| \;\leq\; \frac{2}{n} F^{T_n}\{(\tau, \infty)\} \to 1 - H(\tau), \tag{3.5.1}$$

which tends to 0 as τ tends to infinity.

By the Helly selection theorem[1], there exists a subsequence $\{F^{\tau_m}\}$ and a non-decreasing, right-continuous function F such that $\lim_m F^{\tau_m}(x) = F(x)$ at all continuity points x of F. Let D be the set of numbers which are continuity points of F and all functions F^{τ_m}. Since the discontinuity points of all these functions must be countable, the set D is dense in $\mathbb{R}$. By hypothesis and triangular inequality, we have with probability one for every point $x \in D$,

$$\begin{aligned}
&\limsup_{n\to\infty} |F^{A_n}(x) - F(x)| \\
\leq\;& \limsup_{n\to\infty} |F^{A_n}(x) - F^{A_n^{\tau_m}}(x)| + \limsup_{n\to\infty} |F^{A_n^{\tau_m}}(x) - F^{\tau_m}(x)| + |F^{\tau_m}(x) - F(x)| \\
=\;& (1 - H(\tau_m)) + |F^{\tau_m}(x) - F(x)|,
\end{aligned}$$

which tends to 0 as $m \to \infty$. Thus we get with probability one $\lim_n F^{A_n}(x) = F(x)$ holds for every $x \in D$, *i.e.* there exists a subspace Ω^* with $P(\Omega^*) = 1$ such that for any $\omega \in \Omega^*$, $\lim_n F^{A_n}(x) = F(x)$, for every $x \in D$. This implies, indeed, for any $\omega \in \Omega^*$, $\lim_n F^{A_n}(x) = F(x)$, for every continuity point x of F. To see

[1]See Billingsley(1995) p.336 Theorem 25.9.

this, we note that D is dense in the set of continuity points of F so that for each continuity point x of F and each $\varepsilon > 0$, we can find two numbers in D, say x_1 and x_2, such that $x_1 < x < x_2$, $F(x) - \varepsilon \leq F(x_1) \leq F(x) \leq F(x_2) \leq F(x) + \varepsilon$. Since, for any $\omega \in \Omega^*$, $F(x_1) = \lim_{n\to\infty} F^{A_n}(x_1) \leq \liminf_{n\to\infty} F^{A_n}(x)$ and $F(x_2) = \lim_{n\to\infty} F^{A_n}(x_2) \geq \limsup_{n\to\infty} F^{A_n}(x)$, we get $F(x) - \varepsilon \leq \liminf_{n\to\infty} F^{A_n}(x) \leq \limsup_{n\to\infty} F^{A_n}(x) \leq F(x) + \varepsilon$. Thus, for each $\omega \in \Omega^*$, $F^{A_n}(x)$ converges to $F(x)$ for all continuity points of F. This implies with probability one $\{F^{A_n}\}$ converges vaguely to F. However, $\{F^{A_n}\}$ is tight, hence indeed with probability one, $\{F^{A_n}\}$ converges weakly to $F(x)$.

We can also show that F^τ converges weakly to F. Indeed, the above proof simultaneously shows every subsequence of $\{F^\tau\}$ contains a further weak convergent subsequence and all weak convergent subsequences of $\{F^\tau\}$ converge to the same limit, hence by Theorem 25.10 and its corollary on p.336-337 of Billingsley (1995), F^τ converges weakly to F.

Note that it remains to show as H possesses moments of all orders and for each positive integer p the pth moment of F^{T_n} converges weakly to the pth moment of H, the moments of F are given by (1.1.4) and (1.1.5) of Theorem 1.1.4. By hypothesis, in view that H^τ possesses all moments and for each positive integer p the pth moment of $F^{T_n^\tau}$ converges weakly to the pth moment of H^τ, the moments of F^τ are given by the analogues of (1.1.4) and (1.1.5) with the function H replaced by H^τ. Here we use the corollary to Theorem 25.12 on p.338 of Billingsley (1995) which states that for any positive integer r, if probability distribution functions G_n converge weakly to G and $\sup_n \int |x|^{r+\varepsilon} dG_n(x) < \infty$ where $\varepsilon > 0$, then $\int |x|^r dG(x) < \infty$ and $\int x^r dG_n(x) \to \int x^r dG(x)$. We apply this result to the probability distribution functions F^τ and their weak limit F.

Let us denote by m_p^τ the pth moment of F^τ. We have since H possesses moments of all orders, for any positive integer m, $\int t^m dH(t) < \infty$ and so $\int t^m dH(t) - \int t^m dH^\tau(t) = \int_{t>\tau} t^m dH(t) \to 0$, from which it follows for every order p, $\lim_\tau m_p^\tau$ exists and is given by (1.1.4) and (1.1.5). However, for every even p the existence of $\lim_\tau m_p^\tau$ implies that $\sup_\tau m_p^\tau < \infty$. Thus we obtain from the just stated result that for every order p, the pth moment of F exists and is given by the limit $\lim_\tau m_p^\tau$. Therefore, we proved Theorem 1.1.4 must be true for the matrix A_n provided that it is true for the matrix A_n^τ. We conclude in proving Theorem 1.1.4, without loss of generality, we may assume condition (i) of Assumption 3.1.1 to be true.

The truncation and centralization of the random variables in the matrix W_n can be carried out the same way as in Section 3.1.2. By defining the matrices $\hat{A}_n$ and $\tilde{A}_n$, we get F^{A_n} and $F^{\tilde{A}_n}$ converge to the same limit. Thus if Theorem 1.1.4 is true for the latter it must be equally true for the former. Thus it only remains to show in proving Theorem 1.1.4, without loss of generality, we may assume T_n is non-random.

For each $\omega \in \Omega$, similarly define the matrix $A_n^\omega = \frac{1}{\sqrt{n}} T_n^{1/2}(\omega) W_n T_n^{1/2}(\omega)$ as in Section 3.1.2. By assumption (ii) of Theorem 1.1.4, let Ω_0 still denote the subspace with $P(\Omega_0) = 1$ such that for every $\omega \in \Omega_0$, $F^{T_n(\omega)}$ converges weakly to H. Then, for any $\omega \in \Omega_0$, by Theorem 3.5.1 of the present chapter, $F^{A_n^\omega}$ converges weakly to a non-random limit F with

$$En^{-1}tr(A_n^{\omega})^k \to m_k, \tag{3.5.2}$$

where m_k is the k-th moment of F and satisfies (3.5.5), (3.5.6), and by Theorem 3.5.11,

$$E(n^{-1}tr(A_n^{\omega})^k - En^{-1}tr(A_n^{\omega})^k)^4 \le Kn^{-2},$$

where K depends only on τ and k.

Note that since W_n is independent of T_n, we have for any $\omega \in \Omega$

$$E(n^{-1}tr(A_n)^k | T_n = T_n(\omega)) = En^{-1}tr(A_n^{\omega})^k \tag{3.5.3}$$

and

$$\begin{aligned}
& E|n^{-1}tr(A_n)^k - E(n^{-1}tr(A_n)^k|T_n = T_n(\omega))|^4 \\
= \ & E\left\{E\left(|n^{-1}tr(A_n)^k - E(n^{-1}tr(A_n)^k|T_n = T_n(\omega))|^4|T_n = T_n(\omega)\right)\right\} \\
= \ & E\left(E|n^{-1}tr(A_n^{\omega})^k - En^{-1}tr(A_n^{\omega})^k|^4\right) \\
= \ & \int_{\Omega_0} E|n^{-1}tr(A_n^{\omega})^k - En^{-1}tr(A_n^{\omega})^k|^4 dP(\omega) \\
\le \ & Kn^{-2}. && (3.5.4)
\end{aligned}$$

Note that, because of $P(\Omega_0) = 1$, (3.5.2) and (3.5.3) imply $E(n^{-1}tr(A_n)^k|T_n = T_n(\omega)) \to m_k$, almost surely. By (3.5.4), $n^{-1}tr(A_n)^k - E(n^{-1}tr(A_n)^k|T_n = T_n(\omega)) \to 0$, almost surely. Thus, $n^{-1}tr(A_n)^k \to m_k$, almost surely. That is, Theorem 1.2.1 must also hold for A_n. Therefore, in proving Theorem 1.1.4, without loss of generality, we may also assume T_n to be non-random. In the sequel, we assume all assumptions of Definition 1.1.1 and all conditions of Assumption 3.1.1 hold.

3.5.2 Moment Method Proof: Preliminary Derivations

The preceding truncation and centralization treatment guarantees that Theorem 1.1.4 will follow once the following theorem is proved.

Theorem 3.5.1. *Under assumptions $(i)-(iii)$ of Definition 1.1.1 and conditions $(i)-(iii)$ of Assumption 3.1.1, with probability one, as $n \to \infty$, the empirical spectral distribution F^{A_n} converges weakly to a non-random limiting distribution function F which is determined by its moments. Moreover, if m_k denotes the k-th moment of F, then $m_0 = 1$ and*

$$m_k = \begin{cases} 0, & k \text{ is odd,} \\ \sum_{s=0}^{l-1} g_{2s} g_{2(l-1-s)}, & k \text{ is even} = 2l, \end{cases} \tag{3.5.5}$$

where with α_p denoting the pth moment of $H(t)$, $g_0 = \alpha_1$, for $s \geq 1$, g_{2s} is given by

$$g_{2s} = \sum_{q=1}^{s} \sum_{\substack{j_1+j_2+\cdots+j_q=s+1-q \\ j_1+2j_2+\cdots+qj_q=s}} \frac{s!}{q! j_1! j_2! \cdots j_q!} \alpha_1^q \alpha_2^{j_1} \alpha_3^{j_2} \cdots \alpha_{q+1}^{j_q}. \tag{3.5.6}$$

Carleman's Condition

To see F is determined by its moments, we show $\{m_k\}$ satisfies Carleman's condition. Note that by condition (ii) of Assumption 3.1.1, the moments of H satisfies Carleman's condition since $(\alpha_{2k})^{-\frac{1}{2k}} \geq \tau^{-1}$. Next we show whenever the moments of H satisfy Carleman's condition, the moments m_k given by (3.5.5) and (3.5.6) also satisfy Carleman's condition. By using the calculations of Yin and Krishnaiah

(1983) p.502, we have

$$\begin{aligned}
&\sum_{\substack{j_1+j_2+\cdots+j_q=s+1-q \\ j_1+2j_2+\cdots+qj_q=s}} \frac{(s-q+1)!}{j_1!j_2!\cdots j_q!} \\
= &\text{ the coefficient of } x^s \text{ term of } (x+x^2+\cdots+x^q)^{s-q+1} \\
\le &\text{ the coefficient of } x^s \text{ term of } (x+x^2+\cdots)^{s-q+1} \\
= &\text{ the coefficient of } x^s \text{ term of } x^{s-q+1}(1-x)^{-s+q-1} \\
= &\frac{(s-1)!}{(s-q)!(q-1)!}.
\end{aligned}$$

By Hölder's inequality, since $\alpha_m \le (\alpha_{2s+1})^{\frac{m}{2s+1}}$ and $q+2j_1+\cdots+(q+1)j_q = 2s+1$, it follows

$$\begin{aligned}
g_{2s} &\le \alpha_{2s+1} \sum_{q=1}^{s} \frac{s!}{q!(s-q+1)!} \sum_{\substack{j_1+j_2+\cdots+j_q=s+1-q \\ j_1+2j_2+\cdots+qj_q=s}} \frac{(s-q+1)!}{j_1!j_2!\cdots j_q!} \\
&\le \alpha_{2s+1} \sum_{q=1}^{s} \frac{s!}{q!(s-q+1)!} \frac{(s-1)!}{(s-q)!(q-1)!} \\
&= \alpha_{2s+1} \frac{1}{s+1} \sum_{v=0}^{s-1} C_{s+1}^{v+1} C_{s-1}^{v} \\
&\le \alpha_{2s+1} 2^{2s}.
\end{aligned}$$

Thus

$$m_{2k} \le \sum_{s=0}^{k-1} \alpha_{2s+1}\alpha_{2(k-s-1)+1} 2^{2(k-1)} \le k\alpha_{2k} 2^{2(k-1)}.$$

We get with $C = 2^{-1}3^{-1/6}$,

$$\sum_{k=1}^{\infty} m_{2k}^{-\frac{1}{2k}} \ge C \sum_{k=1} (\alpha_{2k})^{-\frac{1}{2k}} = \infty.$$

Now that $\{m_k\}$ satisfies Carleman's condition, Theorem 3.5.1 can be proven using the moment method. We only need to show for each positive integer k, as $n \to \infty$, the kth moment of F^{A_n} converges almost surely to m_k.

Construction of Graphs

For every positive integer k, write

$$M_k = \frac{1}{n} tr(A_n^k) = n^{-\frac{k}{2}-1} \sum_{i_1,i_2,\cdots,i_{2k}} (w_{i_1i_2}w_{i_3i_4}\cdots w_{i_{2k-1}i_{2k}})(t_{i_2i_3}t_{i_4i_5}\cdots t_{i_{2k}i_1}). \tag{3.5.7}$$

Then M_k is the kth moment of F^{A_n}.

For each sequence $i_1, i_2, \cdots, i_{2k}$, we construct a graph G as follows. We first partition $E = \{1, 2, \cdots, 2k\}$ into several disjoint subsets, say, E_1, E_2, $\cdots$, E_m according to the rule that any two indices u and v in the set E are assigned to the same subset if and only if $i_u = i_v$. By this rule, if u and v belong to different subsets in the partition, then $i_u \neq i_v$. Thus, the total number of subsets contained in the partition, *i.e.* m, should be the number of distinct values appearing in $i_1, i_2, \cdots, i_{2k}$. We can also order E_j by requiring the smallest number of E_j is increasing in j. For ease of reference, a such partition will be written as $\Delta(i_1, i_2, \cdots, i_{2k})$. To construct the graph describing the sequence, we first draw a line and put m different vertices on it. For each $1 \leq j \leq m$ put at the jth vertex on the line all of those i_u with $u \in E_j$. Now for each $1 \leq v \leq k$, draw an edge pointing from i_{2v-1} to i_{2v} corresponding to the variable $w_{i_{2v-1}i_{2v}}$ and we refer to this edge as the vth w-edge; draw an edge pointing from i_{2v} to i_{2v+1} corresponding to the variable $t_{i_{2v}i_{2v+1}}$ and we refer to this edge as the vth t-edge.

Let $\mathcal{L}$ be the set of all graphs G constructed in the above way. For each graph $G \in \mathcal{L}$, let us denote by G_B its subgraph consisting of all its w-edges and all its vertices and denote by G_H its subgraph consisting of all its t-edges and all its vertices. Let $w_{G_B} = w_{i_1i_2}w_{i_3i_4}\cdots w_{i_{2k-1}i_{2k}}$ and $t_{G_H} = t_{i_2i_3}t_{i_4i_5}\cdots t_{i_{2k}i_1}$. With these notation, we have

$$M_k = n^{-\frac{k}{2}-1} \sum_{G \in \mathcal{L}} (w_{G_B})(t_{G_H}). \tag{3.5.8}$$

Let $\pounds$ be the set of all graphs which contain no single w-edges. Since whenever a graph G contains a single w-edge, the expectation of w_{G_B} is equal to zero, we have

$$EM_k = n^{-\frac{k}{2}-1} \sum_{G \in \pounds} (Ew_{G_B})(t_{G_H}). \tag{3.5.9}$$

Two Characteristic Numbers

For each graph G, let l be the number of non-coincident w-edges it contains. Let $\mathcal{Z}(G)$ be the resulting graph obtained from gluing coincident w-edges in the graph G into one edge. Then $\mathcal{Z}(G)$ contains l glued w-edges. If one glued w-edge satisfies that once it is removed from $\mathcal{Z}(G)$, the resulting graph will not be connected any more, then this glued w-edge is said to be a cutting edge. Note that the edges in the original graph G form a closed path starting from i_1 and terminating at i_1 and so removing any edge from G does not cause the graph connected. This guarantees us removing any t-edge from $\mathcal{Z}(G)$ will not cause the graph disconnected. Thus the number of cutting glued w-edges is just the total number of cutting edges in $\mathcal{Z}(G)$ whose removal will cause the graph disconnected. Let r denote the number of cutting w-edges contained in $\mathcal{Z}(G)$. We shall also refer to r as the number of cutting non-coincident w-edges contained in G. Thus we have defined two characteristic numbers for each graph G.

Isomorphic Classes of Graphs

Suppose G' and G'' are two graphs constructed as above respectively for sequences $i'_1, i'_2, \cdots, i'_{2k}$ and $i''_1, i''_2, \cdots, i''_{2k}$. Then G' and G'' are said to be isomorphic with each other if and only if $\Delta(i'_1, i'_2, \cdots, i'_{2k}) = \Delta(i''_1, i''_2, \cdots, i''_{2k})$. It can be seen if G' and G'' are isomorphic graphs, then except that the specific values taken by the m vertices are different between the two graphs all the other properties of the two graphs are the same. Therefore, the previously defined two characteristic numbers l and r are indeed the same for all graphs in one isomorphic class.

With this definition, we may separate the graphs in $\pounds$ into isomorphic classes such that each class contains only isomorphic graphs and graphs in different classes are not isomorphic with each other. One isomorphic class of graphs will be denoted by $\mathcal{G}$. Note that different isomorphic classes may possess the same characteristic numbers l, r. On the other hand, it is easy to see for each isomorphic class obtained from $\pounds$, it must hold $r \leq l \leq k/2$. Thus in the following, we denote by $\mathbb{G}_1$ the set of isomorphic classes whose characteristic numbers satisfy $l < k/2$, $\mathbb{G}_2$ the set of isomorphic classes whose characteristic numbers satisfy $r < l = k/2$, and $\mathbb{G}_3$ the set of isomorphic classes whose characteristic numbers satisfy $r = l = k/2$.

The definition of isomorphic graphs also enables us to obtain a representative graph for each isomorphic class. Recall that all graphs in one isomorphic class possess the same partition E_1, E_2, $\cdots$, E_m of the set $\{i_1, i_2, \cdots, i_{2k}\}$. When we define a graph corresponding to a given sequence, we put at each of the m vertices all those i_u for $u \in E_j$ and thus the value of the vertex is implicitly determined to be the common value taken by those i_u. And as we have claimed, it is these values assigned to the m-vertices that make the graphs differ from each other. To define a representative graph for an isomorphic class, we express the m vertices by I_1, I_2, $\cdots$, I_m whose values are not determined and only satisfy the restriction that they should be m different integers taken from 1 to n.

In the following, let us use $G(\mathcal{G})$ to denote the representative graph of an isomorphic class $\mathcal{G}$. We similarly define $\mathcal{Z}(\mathcal{G})$ to be the resulting graph obtained by gluing coincident w-edges contained in $G(\mathcal{G})$ into one edge. The vertices of $\mathcal{Z}(\mathcal{G})$ are the same as $G(\mathcal{G})$, *i.e.* I_1, I_2, $\cdots$, I_m. The meaning of $\sum_{I_1,\cdots,I_m}$ refers to taking summation over all possibilities of the values taken by the vertices I_1, I_2, $\cdots$, I_m of $G(\mathcal{G})$ satisfying the restriction mentioned in the preceding paragraph.

Two Preliminary Theorems

Theorem 3.5.2. *For each $\mathcal{G} \in \mathbb{G}_1$ or $\mathbb{G}_2$, as $n \to \infty$,*

$$n^{-\frac{k}{2}-1} \sum_{G \in \mathcal{G}} (Ew_{G_B})(t_{G_H}) \to 0.$$

Proof. We apply Lemma 2.2.1 (Lemma 2.11 of Bai (1999)) to $\mathcal{Z}(\mathcal{G})$. It contains totally k t-edges, each of which corresponds to the matrix T_n. It contains l glued w-edges among which there are r are cutting edges. Suppose the l glued w-edges are respectively composed of for $1 \leq i \leq l$, μ_i w-edges of one direction and ν_i w-edges of the other direction. Then the ith glued w-edge corresponds to matrix B with (a,b) entry equal to $E(w_{ab}^{\mu_i} \bar{w}_{ab}^{\nu_i})$. It is easy to see $\|B\|_0 \equiv n \max |B(a,b)| \leq$

$(\delta_n\sqrt{n})^{\mu_i+\nu_i-2}n$. At the same time, noting that B is Hermitian, with n dimensional vector u such that $u^*u = 1$ we have by Hölder's inequality, $\|B\| = \max_u |u^*Bu| \leq \max|B(a,b)|\sum_{a,b}|u_a||u_b| \leq (\delta_n\sqrt{n})^{\mu_i+\nu_i-2}n$.

It is legal to denote the totally $l+k$ edges of $\mathcal{Z}(\mathcal{G})$ by $e_1, \cdots, e_r$ the r cutting glued w-edges, $e_{r+1}, \cdots, e_l$ the other $l-r$ non-cutting glued w-edges, and $e_{l+1}, \cdots, e_{l+k}$ the k t-edges. We adopt the functions $f_{ini}(\cdot)$ and $f_{end}(\cdot)$ of any edge e to represent respectively the initial vertex and the ending vertex of e. Then for each e_i with $1 \leq i \leq l+k$, $f_{ini}(e_i)$ and $f_{end}(e_i)$ should be from the vertices $I_1, I_2, \cdots, I_m$ of $\mathcal{Z}(\mathcal{G})$.

By using Lemma 2.2.1, for $\mathcal{G} \in \mathbb{G}_1$,

$$\begin{aligned}
& n^{-\frac{k}{2}-1}|\sum_{G\in\mathcal{G}}(Ew_{G_B})(t_{G_H})| \\
= \;& n^{-\frac{k}{2}-1}|\sum_{I_1,\cdots,I_m}\prod_{i=1}^{l}E(w^{\mu_i}_{f_{ini}(e_i),f_{end}(e_i)}\bar{w}^{\nu_i}_{f_{ini}(e_i),f_{end}(e_i)})\prod_{i=l+1}^{l+k}t_{f_{ini}(e_i),f_{end}(e_i)}| \\
\leq \;& n^{-\frac{k}{2}-1}Cn\prod_{i=1}^{r}\{(\delta_n\sqrt{n})^{\mu_i+\nu_i-2}n\}\prod_{i=r+1}^{l}\{(\delta_n\sqrt{n})^{\mu_i+\nu_i-2}n\} \\
= \;& C\delta_n^{k-2l} \\
\to \;& 0,
\end{aligned} \tag{3.5.10}$$

in which C is a constant which depends only on τ, m, l and k. Here we have also used the fact $\|T_n\| \leq \tau$ and $\sum_{i=1}^{l}(\mu_i+\nu_i) = k$.

For $\mathcal{G} \in \mathbb{G}_2$, since $2l = k$ and there is no single w-edge in each $G \in \mathcal{G}$, each non-coincident w-edge in G consists of exactly 2 w-edges. This results in for each edge e_i in $\mathcal{Z}(\mathcal{G})$, $1 \leq i \leq l$, $\mu_i + \nu_i = 2$. Since $r < l$, there must be at least one non-cutting w-edge contained in $\mathcal{Z}(\mathcal{G})$. Note that one cutting w-edge cannot be loop, but one non-cutting w-edge can. Therefore, we consider two cases in the following.

If there is some non-cutting w-edge in $\mathcal{Z}(\mathcal{G})$ which is degenerate into a loop, without loss of generality, we suppose e_l is one such edge. Then the matrix corresponding to e_l becomes diagonal with (a,a)th entry $B(a,a) = E(w_{aa}^{\mu_l}\bar{w}_{aa}^{\nu_l})$ and its spectral norm is bounded by $(\delta_n\sqrt{n})^{\mu_l+\nu_l-2}(=1)$. There may be other degenerate non-cutting w-edges, but since considering one of them is already enough for our proof, we still use $(\delta_n\sqrt{n})^{\mu_i+\nu_i-2}n(=n)$ as the bound of the spectral norm of the matrices corresponding to them. The validity of this treatment is obvious. Applying Lemma 2.2.1 in the same way as above, then besides that all $(\delta_n\sqrt{n})^{\mu_i+\nu_i-2}n$ is replaced by n for $1 \le i \le l-1$, $(\delta_n\sqrt{n})^{\mu_l+\nu_l-2}n$ will be replaced by 1. Thus there should be one n factor disappear from the second inequality of (3.5.10). It follows in this case, noting $2l = k$,

$$n^{-\frac{k}{2}-1}|\sum_{G\in\mathcal{G}}(Ew_{G_B})(t_{G_H})| \le Cn^{-1} \to 0.$$

Otherwise, if in $\mathcal{Z}(\mathcal{G})$ there is no degenerate w-edge we arbitrarily select one non-cutting w-edge. For simplicity and without loss of generality, we suppose the selected edge is e_l. Then by Hölder's inequality, we have

$$\begin{aligned}
& n^{-\frac{k}{2}-1}|\sum_{G\in\mathcal{G}}(Ew_{G_B})(t_{G_H})| \\
= \; & n^{-\frac{k}{2}-1}|\sum_{f_{ini}(e_l),f_{end}(e_l)} E(w^{\mu_l}_{f_{ini}(e_l),f_{end}(e_l)}\bar{w}^{\nu_l}_{f_{ini}(e_l),f_{end}(e_l)}) \\
& \sum_{\{I_1,\cdots,I_m\}\backslash\{f_{ini}(e_l),f_{end}(e_l)\}}\prod_{i=1}^{l-1}E(w^{\mu_i}_{f_{ini}(e_i),f_{end}(e_i)}\bar{w}^{\nu_i}_{f_{ini}(e_i),f_{end}(e_i)})\prod_{i=l+1}^{l+k}t_{f_{ini}(e_i),f_{end}(e_i)}|
\end{aligned}$$

$$\leq n^{-\frac{k}{2}-1}\left\{\sum_{f_{ini}(e_l),f_{end}(e_l)}|E(w^{\mu_l}_{f_{ini}(e_l),f_{end}(e_l)}\bar{w}^{\nu_l}_{f_{ini}(e_l),f_{end}(e_l)})|^2\right\}^{1/2}$$
$$\left\{\sum_{f_{ini}(e_l),f_{end}(e_l)}|\sum_{\{I_1,\cdots,I_m\}\backslash\{f_{ini}(e_l),f_{end}(e_l)\}}\prod_{i=1}^{l-1}E(w^{\mu_i}_{f_{ini}(e_i),f_{end}(e_i)}\bar{w}^{\nu_i}_{f_{ini}(e_i),f_{end}(e_i)})\right.$$
$$\left.\prod_{i=l+1}^{l+k}t_{f_{ini}(e_i),f_{end}(e_i)}|^2\right\}^{1/2}. \tag{3.5.11}$$

Note that since $\mu_l+\nu_l=2$, $|E(w^{\mu_l}_{f_{ini}(e_l),f_{end}(e_l)}\bar{w}^{\nu_l}_{f_{ini}(e_l),f_{end}(e_l)})|\leq 1$. Thus

$$\left\{\sum_{f_{ini}(e_l),f_{end}(e_l)}|E(w^{\mu_i}_{f_{ini}(e_l),f_{end}(e_l)}\bar{w}^{\nu_i}_{f_{ini}(e_l),f_{end}(e_l)})|^2\right\}^{1/2}\leq n.$$

For the other term appearing in the last inequality of (3.5.11), we introduce the following definitions.

Denote by $\mathcal{Z}(G)\backslash e_l$ the resulting graph of deleting from $\mathcal{Z}(G)$ the edge e_l. Note that $\mathcal{Z}(G)\backslash e_l$ contains all the m vertices of $\mathcal{Z}(G)$ including the two vertices $f_{ini}(e_l)$, $f_{end}(e_l)$ of e_l. Now we make a copy of $\mathcal{Z}(G)\backslash e_l$ but we keep using only the two vertices $f_{ini}(e_l)$, $f_{end}(e_l)$ and change all other vertices I_u's into J_u's. However, the copies of the edges e_i's will all be changed their notations into $\tilde{e}_i$'s. Then we glue the original $\mathcal{Z}(G)\backslash e_l$ and its copy at their common vertices $f_{ini}(e_l)$, $f_{end}(e_l)$ and denote the resulting graph by $\mathcal{Z}^0(\mathcal{G})$.

Let us consider the restriction on the values of the vertices of $\mathcal{Z}^0(\mathcal{G})$. To make explanations clear, let us now simply assume $f_{ini}(e_l)=I_1$ and $f_{end}(e_l)=I_2$. Then the vertices of $\mathcal{Z}^0(\mathcal{G})$ consist of I_1, I_2, I_3, $\cdots$, I_m, J_3, $\cdots$, J_m. These vertices should take values satisfying the following restriction: vertices I_1 and I_2 never take common values with each other and with all the left I_u's and J_u's, vertices I_3, $\cdots$, I_m cannot take common values among themselves, vertices J_3, $\cdots$, J_m cannot take common values among themselves, but there maybe some I-vertex taking a common value with some J-vertex or vice versa.

According to the rule governing the values taken by vertices I_1, I_2, I_3, $\cdots$, I_m, J_3, $\cdots$, J_m, we see in $\mathcal{Z}^0(\mathcal{G})$ there are at least m and at most $2m-2$ non-coincident vertices. For any integer $m \le m' \le 2m-2$, denote by $\aleph(m')$ all possible coincidence way between I_3, $\cdots$, I_m and J_3, $\cdots$, J_m such that $\mathcal{Z}^0(\mathcal{G})$ contains m' non-coincident vertices and denote by L_1, $\cdots$, $L_{m'}$ these vertices.

With these definitions, we get from (3.5.11),

$$\begin{aligned}
& n^{-\frac{k}{2}-1} \Big| \sum_{G \in \mathcal{G}} (E w_{G_B})(t_{G_H}) \Big| \\
\le\ & n^{-\frac{k}{2}-1} \times n \times \Bigg\{ \sum_{I_1,\cdots,I_m,J_3,\cdots,J_m} \prod_{i=1}^{l-1} E(w^{\mu_i}_{f_{ini}(e_i),f_{end}(e_i)} \bar{w}^{\nu_i}_{f_{ini}(e_i),f_{end}(e_i)}) \\
& \prod_{i=l+1}^{l+k} t_{f_{ini}(e_i),f_{end}(e_i)} \prod_{i=1}^{l-1} E(w^{\mu_i}_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)} \bar{w}^{\nu_i}_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)}) \\
& \prod_{i=l+1}^{l+k} t_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)} \Bigg\}^{1/2} \\
=\ & n^{-\frac{k}{2}-1} \times n \times \Bigg\{ \sum_{m'=m}^{2m-2} \sum_{\aleph(m')} \sum_{L_1,\cdots,L_{m'}} \prod_{i=1}^{l-1} E(w^{\mu_i}_{f_{ini}(e_i),f_{end}(e_i)} \bar{w}^{\nu_i}_{f_{ini}(e_i),f_{end}(e_i)}) \\
& \prod_{i=l+1}^{l+k} t_{f_{ini}(e_i),f_{end}(e_i)} \prod_{i=1}^{l-1} E(w^{\mu_i}_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)} \bar{w}^{\nu_i}_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)}) \\
& \prod_{i=l+1}^{l+k} t_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)} \Bigg\}^{1/2}
\end{aligned} \tag{3.5.12}$$

It is clear $\mathcal{Z}^0(\mathcal{G})$ contains $2(l+k-1)$ edges. They are simply e_i and $\tilde{e}_i$ for $i = 1, \cdots, l-1, l+1, \cdots, k+l$. We consider the case when all of I_1, I_2, I_3, $\cdots$, I_m, J_3, $\cdots$, J_m take distinct value from each other. In this case, it is clear since except the two vertices I_1 and I_2, the original part and the copy part in $\mathcal{Z}^0(\mathcal{G})$ have no other coincident vertices, each e_i and each $\tilde{e}_i$ for $1 \le i \le r$ are cutting edges of

$\mathcal{Z}^0(\mathcal{G})$. Further for e_i, $(r+1 \le i \le k+l$ and $i \ne l)$, since before the edge e_l is removal, e_i is non-cutting in $\mathcal{Z}(\mathcal{G})$, now that we can find a path only through edges in the copy part which connecting the vertices I_1 and I_2, e_i is still non-cutting in $\mathcal{Z}^0(\mathcal{G})$. Due to the symmetry of the graph, each $\tilde{e}_i$ $(r+1 \le i \le k+l$ and $i \ne l)$ is also non-cutting in $\mathcal{Z}^0(\mathcal{G})$. In the case where some I-vertex and J-vertex take common values, we just note that no cutting edges may be added.

Concerning the matrix corresponding to each edge of $\mathcal{Z}^0(\mathcal{G})$, we simply use the same matrix for an original edge and its copy and such matrix has been defined previously. Therefore, we are ready to use Lemma 2.2.1. Note that due to $\mu_i+\nu_i = 2$, $\|\cdot\|_0$ and $\|\cdot\|$ are both bounded by n. Thus when we applying Lemma 2.2.1 to $\mathcal{Z}^0(\mathcal{G})$, noting $l = k/2$, the $2(l-1)$ w-edge in it totally generate one factor n^{k-2}, while all the other t-edges generate constant factor τ^{2k}. Thus for each m' and each possible way in $\aleph(m')$, from Lemma 2.2.1, we get

$$\begin{aligned}
&|\sum_{L_1,\cdots,L_{m'}} \prod_{i=1}^{l-1} E(w^{\mu_i}_{f_{ini}(e_i),f_{end}(e_i)} \bar{w}^{\nu_i}_{f_{ini}(e_i),f_{end}(e_i)}) \prod_{i=l+1}^{l+k} t_{f_{ini}(e_i),f_{end}(e_i)} \\
&\prod_{i=1}^{l-1} E(w^{\mu_i}_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)} \bar{w}^{\nu_i}_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)}) \prod_{i=l+1}^{l+k} t_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)}| \\
\le\ & Cnn^{k-2} \\
=\ & Cn^{k-1}.
\end{aligned}$$

By (3.5.12), it follows

$$\begin{aligned}
n^{-\frac{k}{2}-1}|\sum_{G\in\mathcal{G}}(Ew_{G_B})(t_{G_H})| &\le Cn^{-\frac{k}{2}-1} \times n \times (n^{k-1})^{1/2} \\
&\le Cn^{-\frac{1}{2}}.
\end{aligned}$$

This completes the proof of the theorem. □

Theorem 3.5.3. *For each $\mathcal{G} \in \mathbb{G}_3$, as $n \to \infty$,*

$$n^{-\frac{k}{2}-1} \sum_{G \in \mathcal{G}} (1 - Ew_{G_B})(t_{G_H}) \to 0.$$

Proof. For the representative graph $G(\mathcal{G})$ of an isomorphic class $\mathcal{G} \in \mathbb{G}_3$, we first assert that each of its non-coincident w-edge must consist of two w-edges with opposite directions. If there is one non-coincident w-edge consisting of two w-edges of the same direction, there must be a path which connects the two vertices of this edge in the graph without need of passing this edge. This means removing the two w-edges included in a such non-coincident w-edge from $G(\mathcal{G})$ will not cause the graph disconnected and hence after gluing all the coincident w-edges in $G(\mathcal{G})$, the resulting graph $\mathcal{Z}(\mathcal{G})$ contains one non-cutting w-edge. This contradicts the hypothesis $\mathcal{G} \in \mathbb{G}_3$.

We still denote by e_i for $1 \leq i \leq l$ (here $l = k/2$) the l cutting w-edge in $\mathcal{Z}(\mathcal{G})$ and e_i for $l+1 \leq i \leq k+l$ the k t-edges in $\mathcal{Z}(\mathcal{G})$. Then

$$\begin{aligned}
& 1 - Ew_{G_B} \\
= \; & 1 - \prod_{i=1}^{l} E|w_{f_{ini}(e_i), f_{end}(e_i)}|^2 \\
= \; & \sum_{j=1}^{l} \left(\prod_{i=1}^{j-1} E|w_{f_{ini}(e_i), f_{end}(e_i)}|^2 \right) (1 - E|w_{f_{ini}(e_j), f_{end}(e_j)}|^2).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
& n^{-\frac{k}{2}-1}\sum_{G\in\mathcal{G}}(1-Ew_{G_B})(t_{G_H}) \\
= \quad & \sum_{j=1}^{l} n^{-\frac{k}{2}-1}\sum_{I_1,\cdots,I_m}\left(\prod_{i=1}^{j-1}E|w_{f_{ini}(e_i),f_{end}(e_i)}|^2\right)(1-E|w_{f_{ini}(e_j),f_{end}(e_j)}|^2) \\
& \prod_{i=l+1}^{k+l} t_{f_{ini}(e_i),f_{end}(e_i)}. \qquad (3.5.13)
\end{aligned}
$$

Fix any j. We may want to apply Lemma 2.2.1 to the inner sum of the above expression. According to the form of the summand, we define new matrices corresponding to the edges of $\mathcal{Z}(\mathcal{G})$. Let each edge e_i with $1 \le i < j$ correspond to the matrix B with (a,b)th entry $B(a,b) = E|w_{ab}|^2$, the edge e_j correspond to the matrix B with (a,b)th entry $Q(a,b) = 1 - E|w_{ab}|^2$, each edge e_i with $j < i \le l$ correspond to the matrix R with (a,b)th entry $R(a,b) = 1$. Then it is easy to see $\|B\|_0 \le n$, $\|R\|_0 \le n$. However, $\|Q\|_0 \le n\max(1 - E|w_{ab}|^2)$ is not appropriate to yield the result. Therefore, as we have done for the case of $\mathcal{G} \in \mathbb{G}_2$, we again make use of Hölder's inequality. We obtain

$$
\begin{aligned}
& |\sum_{I_1,\cdots,I_m}\left(\prod_{i=1}^{j-1}E|w_{f_{ini}(e_i),f_{end}(e_i)}|^2\right)(1-E|w_{f_{ini}(e_j),f_{end}(e_j)}|^2)\prod_{i=l+1}^{k+l}t_{f_{ini}(e_i),f_{end}(e_i)}| \\
\le \quad & \left\{\sum_{f_{ini}(e_j),f_{end}(e_j)}(1-E|w_{f_{ini}(e_j),f_{end}(e_j)}|^2)^2\right\}^{1/2} \\
& \left\{\sum_{f_{ini}(e_j),f_{end}(e_j)}|\sum_{\{I_1,\cdots,I_m\}\backslash\{f_{ini}(e_j),f_{end}(e_j)\}}\left(\prod_{i=1}^{j-1}E|w_{f_{ini}(e_i),f_{end}(e_i)}|^2\right)\right. \\
& \left.\prod_{i=l+1}^{k+l}t_{f_{ini}(e_i),f_{end}(e_i)}|^2\right\}^{1/2}. \qquad (3.5.14)
\end{aligned}
$$

The following calculations will be in parallel to those of (3.5.11) and (3.5.12). It is easy to see

$$\begin{aligned} & \left\{ \sum_{f_{ini}(e_j), f_{end}(e_j)} (1 - E|w_{f_{ini}(e_j), f_{end}(e_j)}|^2)^2 \right\}^{1/2} \\ \leq & \left(\sum_{a,b} (1 - E|w_{a,b}|^2) \right)^{1/2} \\ = & \; o(\delta_n n). \end{aligned} \tag{3.5.15}$$

For the second square root term appearing in (3.5.14), define similarly $\mathcal{Z}(\mathcal{G})\backslash e_j$ and its copy. Then similarly glue the original and copy parts together at their two common vertices $f_{ini}(e_j)$, $f_{end}(e_j)$ to get $\mathcal{Z}^0(\mathcal{G})$. For simplicity and without loss of generality, we now assume the two vertices $f_{ini}(e_j)$, $f_{end}(e_j)$ of e_j are just I_1, I_2 in $\mathcal{Z}(\mathcal{G})$.

For any $1 \leq j \leq l$ fixed, we have defined for each edge e_i with $i \neq j$ and $1 \leq i \leq l + k$ of $\mathcal{Z}(\mathcal{G})$ the matrix corresponding to it. Therefore, for each edge of $\mathcal{Z}^0(\mathcal{G})$, its corresponding matrix is known by the rule that a copy edge $\tilde{e}_i$ corresponds to a matrix equal to the matrix corresponding to the original edge e_i.

Similarly denote by I_1, I_2, I_3, $\cdots$, I_m, J_3, $\cdots$, J_m the vertices of $\mathcal{Z}^0(\mathcal{G})$. Similarly, for $m \leq m' \leq 2m - 2$, let $\aleph(m')$ be all the possible coincidence way between I_3, $\cdots$, I_m, and J_3, $\cdots$, J_m such that $\mathcal{Z}^0(\mathcal{G})$ possesses exactly m' non-coincident vertices L_1, $\cdots$, $L_{m'}$.

However, the connectivity property of $\mathcal{Z}^0(\mathcal{G})$ is not the same as the previous case. In this case, due to the cutting nature of the removed edge e_j, $\mathcal{Z}^0(\mathcal{G})$ is composed of two disjoint connected subgraphs. We may conveniently add one edge to connect the two vertices I_1, I_2 left by e_j and let this edge correspond to the matrix R. The resulting graph will possess the same vertices as $\mathcal{Z}^0(\mathcal{G})$ and one more new edge which is cutting corresponding to the matrix R.

We now yield

$$\begin{aligned}
& \sum_{f_{ini}(e_j),f_{end}(e_j)} | \sum_{\{I_1,\cdots,I_m\}\backslash\{f_{ini}(e_j),f_{end}(e_j)\}} \left(\prod_{i=1}^{j-1} E|w_{f_{ini}(e_i),f_{end}(e_i)}|^2\right) \\
& \prod_{i=l+1}^{k+l} t_{f_{ini}(e_i),f_{end}(e_i)}|^2 \\
= & \sum_{I_1,I_2} | \sum_{I_3,\cdots,I_m} \left(\prod_{i=1}^{j-1} E|w_{f_{ini}(e_i),f_{end}(e_i)}|^2\right) \prod_{i=l+1}^{k+l} t_{f_{ini}(e_i),f_{end}(e_i)}|^2 \\
= & \sum_{m'=m}^{2m-2} \sum_{\aleph(m')} \sum_{L_1,\cdots,L_{m'}} \left(\prod_{i=1}^{j-1} E|w_{f_{ini}(e_i),f_{end}(e_i)}|^2\right) \\
& \prod_{i=l+1}^{k+l} t_{f_{ini}(e_i),f_{end}(e_i)} \left(\prod_{i=1}^{j-1} E|w_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)}|^2\right) \prod_{i=l+1}^{k+l} t_{f_{ini}(\tilde{e}_i),f_{end}(\tilde{e}_i)}.
\end{aligned}$$

To get estimate of the inner sum in the above relation, we apply Lemma 2.2.1 to the resulting graph of adding one edge to $\mathscr{Z}^0(\mathcal{G})$. Then the inner sum can be considered as a summation taken over all of the vertices of this graph while the matrix corresponding to each edge of this graph depending on the form of the summand has been defined in preceding illustrations on $\mathscr{Z}^0(\mathcal{G})$ and the adding edge. There are totally $2(l-1)+1$ cutting edges each of which corresponds to a matrix whose $\|\cdot\|_0$ does not exceeding n and hence they totally generates factor $n^{2l-1} = n^{k-1}$. Therefore, the absolute value of the inner sum is bounded by $Cnn^{k-1} = Cn^k$. Combining this result with (3.5.15), in view of (3.5.14), we then get for each fixed j, the absolute value of the inner sum of (3.5.13) is bounded by $C \cdot o(\delta_n) n^{\frac{k}{2}+1}$. It follows then

$$|n^{-\frac{k}{2}-1} \sum_{G\in\mathcal{G}} (1 - Ew_{G_B})(t_{G_H})| = o(\delta_n) \to 0.$$

The proof is complete. $\square$

Theorems 3.5.2 and 3.5.3 imply that

$$EM_k = n^{-\frac{k}{2}-1} \sum_{\mathcal{G}\in\mathbb{G}_3} \sum_{G\in\mathcal{G}} (t_{G_H}) + o(1). \tag{3.5.16}$$

Since when k is odd, $\mathbb{G}_3$ is empty, we get in this case $EM_k \to 0$. Now assume k is even. To calculate the leading term on the right-hand, we need to give a reclassifications of the isomorphic classes involved in $\mathbb{G}_3$.

Property of $G(\mathcal{G})$ for $\mathcal{G} \in \mathbb{G}_3$

We consider some basic properties of the representative graph $G(\mathcal{G})$ of an isomorphic class $\mathcal{G}$ belonging to the set $\mathbb{G}_3$. first, we recall that in this case due to $r = l = k/2$, every non-coincident w-edge of $G(\mathcal{G})$ is cutting and so every non-coincident w-edge of $G(\mathcal{G})$ consists of exactly two w-edges of opposite directions.

For each $\mathcal{G} \in \mathbb{G}_3$, denote by $G_H(\mathcal{G})$ the resulting graph of removing from $G(\mathcal{G})$ all its w-edges. Then $G_H(\mathcal{G})$ consists of exactly $l+1$ disjoint subgraphs each of which consists of only t-edges and does not contain cutting edge. In the sequel, these $l+1$ subgraphs will be said to be $l+1$ blocks and denoted by $B_0(\mathcal{G}), \cdots, B_l(\mathcal{G})$. For clarity, we give them a such order that if v_j is the smallest interger such that the v_jth t-edge (i_{2v_j}, i_{2v_j+1}) belongs to $B_j(\mathcal{G})$ for $0 \le j \le l$, then $1 = v_0 < v_1 < \cdots < v_l$. Consequently, for each $B_j(\mathcal{G})$ with $1 \le j \le l$, the w-edge (i_{2v_j-1}, i_{2v_j}) appears before all those edges in $B_j(\mathcal{G})$ if we draw the edges of the graph $G(\mathcal{G})$ one by one begining at (i_1, i_2). Thus the w-edge (i_{2v_j-1}, i_{2v_j}) is also the one among all those w-edges connected with $B_j(\mathcal{G})$ that appears the first. Therefore, for each $1 \leq j \leq l$ we call the w-edge (i_{2v_j-1}, i_{2v_j}) as the initiating w-edge of the block $B_j(\mathcal{G})$.

For each block $B_j(\mathcal{G})$ with $0 \leq j \leq l$, there is associated a subset U_j of the set $\{1, \cdots, k\}$ such that v belongs to U_j if and only if the vth t-edge (i_{2v}, i_{2v+1}) belongs to $B_j(\mathcal{G})$. Then U_0, U_1, $\cdots$, U_l form a separation of $\{1, \cdots, k\}$ which for brevity we denote by $\Upsilon(\mathcal{G})$. It can be seen the prescribed v_j is the smallest number of U_j. It can also be seen U_j describes the composition of each block $B_j(\mathcal{G})$, *i.e.* which many t-edges constitute the block.

For each of the l w-edges (i_{2v_j-1}, i_{2v_j}) which initiate l blocks, there is one and only one w-edge coincident with it. Thus let us suppose the coincident w-edge of (i_{2v_j-1}, i_{2v_j}) is (i_{2u_j-1}, i_{2u_j}) with $i_{2u_j} = i_{2v_j-1}$,$i_{2u_j-1} = i_{2v_j}$ for $1 \leq j \leq l$. Then $\{v_1, u_1\}$, $\cdots$, $\{v_l, u_l\}$ form another separation of $\{1, \cdots, k\}$. Let us write this separation as $\Delta(\mathcal{G})$. We can indeed obtain $\Delta(\mathcal{G})$ in another way. That is, partitioning $\{1, \cdots, k\}$ into l disjoint subsets, any pair of integers u and v with $1 \leq u, v \leq k$ belong to the same subset if and only if the uth w-edge and the vth w-edge of $G(\mathcal{G})$ are coincident with each other. Thus separation $\Delta(\mathcal{G})$ describes the coincident way of the w-edges of $G(\mathcal{G})$.

We now prove for each $\mathcal{G} \in \mathbb{G}_3$, $\Delta(\mathcal{G})$ determines $\Upsilon(\mathcal{G})$. We first note that if we view each block as a vertex, then the l w-edges and the $l+1$ blocks form a tree. This property dertermines that if one person walks along the edges of $G(\mathcal{G})$ then once he leaves a block through one w-edge, he can only reenter the block through the w-edge coincident with the prescribed w-edge and he leaves the block ultimately only through the w-edge coincident with the iniatiating w-edge of the block. With this understanding, we consider an arbitrary $B_j(\mathcal{G})$ with $1 \leq j \leq l$. Its initiating w-edge is (i_{2v_j-1}, i_{2v_j}) coincident with (i_{2u_j-1}, i_{2u_j}), so the first and last t-edge appearing in the block are respectrively (i_{2v_j}, i_{2v_j+1}) and (i_{2u_j-2}, i_{2u_j-1}). If one person enters this block and walks along the first edge of (i_{2v_j}, i_{2v_j+1}), he will leave the block through the w-edge (i_{2v_j+1}, i_{2v_j+2}). As stated, he reenter the block only through the w-edge which is coincident with (i_{2v_j+1}, i_{2v_j+2}). Since we have $\Delta(\mathcal{G})$, this coincident edge can be known. Then sequentially we know the next t-edge in the block and the next w-edge along which he again leaves the block. Based on $\Delta(\mathcal{G})$, the third t-edge in this block is then known. Continue this process until we get the last t-edge (i_{2u_j-2}, i_{2u_j-1}) of the block. Since each t-edge in the block must initiate one w-edge which leaves the block and each t-edge of the block must at the same time be initiated by one w-edge which enters the block, by this procedure we obtain the compostion U_j of $B_j(\mathcal{G})$ for all $1 \leq j \leq l$. As for the block $B_0(\mathcal{G})$, once we know the composition of all the others, its composition U_0 is naturally known.

As another consequence of the preceding proof, we can see each block is indeed an Euler circuit, *i.e.* a person can walks along the edges of each block one by one and return to his start point in the block. If the end vertices of all non-coincident w-edges are not coincident with each other in $G(\mathcal{G})$ (so there are totally $2l$ non-coincideent vertices in the graph), then the number of non-coincident vertices of each block is equal to that of the non-coincident w-edges connected with the block and hence that of the t-edges constituting the block. In this case, the $l+1$ Euler circuits become $l+1$ cycles. Or more precisely, in this case, the graph $G_H(\mathcal{G})$

consists of exactly $l+1$ cycles in this case. Here a cycle refers to a connected graph which possesses the same number of edges and vertices. Intuitively, a cycle is a circuit or a closed path touching every vertex exactly once.

Reclassification of Dominating Isomorphic Classes

Using $\Delta(\mathcal{G})$, we can separate all the isomorphic classes included in $\mathbb{G}_3$ into disjoint subsets like this: any two isomorphic classes $\mathcal{G}_1$ and $\mathcal{G}_2$ are classified into the same subset if and only if $\Delta(\mathcal{G}_1) = \Delta(\mathcal{G}_2)$. In the sequel, a subset of $\mathbb{G}_3$ obtained from this classification will be denoted by $\mathcal{C}$.

Consider any two isomorphic classes $\mathcal{G}_1$ and $\mathcal{G}_2$ classified into a same subset $\mathcal{C}$. Then by definition, $\Delta(\mathcal{G}_1) = \Delta(\mathcal{G}_2)$. From the previously shown result, it follows then $\Upsilon(\mathcal{G}_1) = \Upsilon(\mathcal{G}_2)$. Hence $B_j(\mathcal{G}_1)$ and $B_j(\mathcal{G}_2)$ have the same composition U_j for each $0 \leq j \leq l$. Therefore, the difference between $B_j(\mathcal{G}_1)$ and $B_j(\mathcal{G}_2)$ lies in the vertices of them. Suppose we define the length of a block as the number of the t-edges constituting it. Let y_j be the length of $B_j(\mathcal{G}_1)$ for $0 \leq j \leq l$. Then there are exactly y_j non-coincident w-edges connected with $B_j(\mathcal{G}_1)$, each of which has exactly one end vertex which is the end vertices of two t-edges belonging to the block. Denote these y_j vertices of the y_j w-edges by $I_1^{(j)}, \cdots, I_{y_j}^{(j)}$. Then for each $0 \leq j \leq l$, the difference between $B_j(\mathcal{G})$ for $\mathcal{G} \in \mathcal{C}$ lies in the coincidence way among the y_j vertices.

Among all the isomorphic classes $\mathcal{G}$ in one set $\mathcal{C}$, there is one and only one isomorphic class $\mathcal{G}$ for which $G(\mathcal{G})$ contains exactly $2l$ non-coincident vertices and the $l+1$ blocks of $G_H(\mathcal{G})$ are $l+1$ cycles. In the following, we shall call the representative graph of such an isomorphic class as the representative graph of the whole set $\mathcal{C}$. For clarity, we shall change to denote by $G(\mathcal{C})$ this representative graph in the future. Similarly, we change to use notations $G_H(\mathcal{C})$, $B_j(\mathcal{C})$ instead of $G_H(\mathcal{G})$, $B_j(\mathcal{G})$.

Characteristics of any set $\mathcal{C}$ can be obtained from its representative graph. We first find the non-coincident w-edge in $G(\mathcal{C})$ containing the first w-edge (i_1, i_2) and then we remove it from the graph (remove both (i_1, i_2) and the w-edge coincident with (i_1, i_2)). The resulting graph are disjoint two parts now. Let s be the number of non-coincident w-edges included in the part which contains the cycle $B_1(\mathcal{C})$. Then $l-1-s$ is the number of non-coincident w-edges contained in the other part. The ordered pair of characteristic numbers $(s, l-1-s)$ enable us to investigate the property of $G(\mathcal{C})$ through separately considering the prescribed two parts. These two parts have similar characteristics obviously. We focus on the part containing the cycle $B_1(\mathcal{C})$. This part contains totally s non-coincident w-edges as claimed and hence $s+1$ cycles the total length of which are $2s+1$. Denote by q the number of cycles of length 1. Recall that by viewing each cycle as a vertex, this part is a tree containing s edges. This implies the maximum length of all cycles contained in this part does not exceed $q+1$. For $1 \leq v \leq q$, let j_v be the number of cycles with length $v+1$. Then we have $q + j_1 + \cdots + j_q = s+1$ and $q+2j_1+3j_2+\cdots+(q+1)j_q = 2s+1$. Or equivalently, the sequence of characteristic numbers $(s; q, j_1, \cdots, j_q)$ satisfies the restiction that $j_1 + \cdots + j_q = s+1-q$ and $j_1 + 2j_2 + \cdots + qj_q = s$ where $1 \leq q \leq s$. In parallel, we may define a sequence of numbers $(l-1-s; q', j_1', \cdots, j_{q'}')$ for the other part which satisfies a similar restriction that $j_1' + j_2' + \cdots + j_{q'}' = l - s - q'$ and $j_1' + 2j_2' + \cdots + q' j_{q'}' = l - 1 - s$.

In the following, let $\mathcal{C}\begin{pmatrix} s;q,j_1,\cdots,j_q \\ l-1-s;q',j_1',\cdots,j_{q'}' \end{pmatrix}$ be the collection of all sets $\mathcal{C}$ which corresponds to the two sequences of numbers $(s; q, j_1, \cdots, j_q)$ and $(l-1-s; q', j_1', \cdots, j_{q'}')$ as a consequence of the preceding procedure. Define summations

$$\sum_{(s;q,j_1,\cdots,j_q)} = \sum_{q=1}^{s} \sum_{\substack{j_1+j_2+\cdots+j_q=s+1-q \\ j_1+2j_2+\cdots+qj_q=s}},$$

and

$$\sum_{(l-1-s;q',j_1',\cdots,j_{q'}')} = \sum_{q'=1}^{l-1-s} \sum_{\substack{j_1'+j_2'+\cdots+j_{q'}'=l-s-q' \\ j_1'+2j_2'+\cdots+q'j_{q'}'=l-1-s}}.$$

It follows

$$\begin{aligned} & n^{-\frac{k}{2}-1} \sum_{\mathcal{G}\in\mathbb{G}_3} \sum_{G\in\mathcal{G}} (t_{G_H}) \\ = \; & n^{-k/2-1} \sum_{s=0}^{l-1} \sum_{(s;q,j_1,\cdots,j_q)} \sum_{(l-1-s;q',j_1',\cdots,j_{q'}')} \sum_{\mathcal{C}\in\mathcal{C}\begin{pmatrix} s;q,j_1,\cdots,j_q \\ l-1-s;q',j_1',\cdots,j_{q'}' \end{pmatrix}} \sum_{\mathcal{G}\in\mathcal{C}} \sum_{G\in\mathcal{G}} (t_{G_H}). \end{aligned} \tag{3.5.17}$$

Preliminary Theorem

Theorem 3.5.4. *For each* $\mathcal{C} \in \mathcal{C}\begin{pmatrix} s;q,j_1,\cdots,j_q \\ l-1-s;q',j_1',\cdots,j_{q'}' \end{pmatrix}$, *as* $n \to \infty$,

$$n^{-\frac{k}{2}-1} \sum_{\mathcal{G}\in\mathcal{C}} \sum_{G\in\mathcal{G}} (t_{G_H}) = \left((\alpha_1)^q \prod_{v=1}^{q} (\alpha_{v+1})^{j_v} \right) \left((\alpha_1)^{q'} \prod_{v=1}^{q'} (\alpha_{v+1})^{j_v'} \right) + o(1).$$

Proof. If we reuse $I_1^{(j)}, \cdots, I_{y_j}^{(j)}$ to denote the vertices of $B_j(\mathcal{C})$, then with the understanding $I_{y_j+1}^{(j)} \equiv I_1^{(j)}$, we can write

$$t_{G_H} = \prod_{j=0}^{l} \prod_{v=1}^{y_j} t_{I_v^{(j)} I_{v+1}^{(j)}}.$$

The total set of vertices of $G(\mathcal{C})$ are given by

$$\begin{gathered} I_1^{(0)}, I_2^{(0)}, \cdots, I_{y_0}^{(0)}, \\ I_1^{(1)}, I_2^{(1)}, \cdots, I_{y_1}^{(1)}, \\ \cdots, \cdots, \cdots, \cdots, \\ I_1^{(l)}, I_2^{(l)}, \cdots, I_{y_l}^{(l)}. \end{gathered}$$

Note that $y_0 + y_1 + \cdots + y_l = k$ is the total number of t-edges in $G(\mathcal{C})$. These k vertices can be expressed conformly by $I_1, I_2, \cdots, I_k$, where for each $1 \le a \le k$, since a can be expressed by $a = \sum_{u=0}^{j-1} y_u + v$ for some $0 \le j \le l$ and $1 \le v \le y_j$, we define $I_a = I_v^{(j)}$. With these notations, for each $0 \le j \le l$, letting $b_{-1} = 0$, $b_j = \sum_{u=0}^{j} y_u$, we have $\prod_{v=1}^{y_j} t_{I_v^{(j)} I_{v+1}^{(j)}} = \left(\prod_{a=b_{j-1}+1}^{b_j - 1} t_{I_a I_{a+1}}\right) t_{I_{b_j} I_{b_{j-1}+1}}$ so that

$$n^{-\frac{k}{2}-1} \sum_{\mathcal{G}\in\mathcal{C}} \sum_{G\in\mathcal{G}} t_{G_H} = n^{-\frac{k}{2}-1} \sum_{I_1,\cdots,I_k}^{res} \prod_{j=0}^{l} \left\{ \left(\prod_{a=b_{j-1}+1}^{b_j-1} t_{I_a I_{a+1}} \right) t_{I_{b_j} I_{b_{j-1}+1}} \right\}, \quad (3.5.18)$$

where the summation is taken over all possible values of the vertices $I_1, I_2, \cdots, I_{2k}$ satisfying the restriction that if $V_j \equiv \{I_{b_{j-1}+1}, \cdots, I_{b_j}\}$ for $0 \le j \le l$, then for any $j_1 \ne j_2$, $V_{j_1} \cap V_{j_2} = \emptyset$. It is easy to see if the summation is taken over all possible values of the vertices $I_1, I_2, \cdots, I_k$ without imposing any restriction, then the value of the right-hand side of (3.5.18) is equal to $((\alpha_1)^q \prod_{v=1}^{q} (\alpha_{v+1})^{j_v}) \left((\alpha_1)^{q'} \prod_{v=1}^{q'} (\alpha_{v+1})^{j'_v}\right)$.

For any given values of the vertices $I_1, I_2, \cdots, I_k$, correponding to all the t-variables involved in the summand of the summation on the right-hand side of

(3.5.18), we may draw a k-edge graph. In the case where the given values of the vertices satisfy the above stated restriction, the resulting graph is obviously consisting of $l+1$ disjoint blocks each of which is connected and does not contain cutting edges. However, for other cases of given values, the resulting graph may not have this property. Nonetheless, the set of the graphs drawn for all possible given values of the vertices can be analyzed by defining isomorphic classes.

We now define these isomorphic classes. Given values of the vertices I_1, I_2, $\cdots$, I_k, or equivalently, given any graph defined above, define a partation of the set $\{1, 2, \cdots, k\}$ into m subsets F_1, F_2, $\cdots$, F_m according to the rule that for any $1 \leq u, v \leq k$, u and v belong to the same subset if and only if I_u and I_v take the same value. Without loss of generality, suppose the order of these m subsets is given such that $\min F_i$ is increasing in i. Any two graphs are defined to be isomorphic with each other if the partitions so defined for them are the same. An isomorphic class is defined to be a set of isomorphic graphs which satisfies that any graph outside the set cannot be isomorphic with the graphs included in the set.

Due to the definition of isomorphic graphs, since the partitions corresponding to the graphs in one isomorphic class must be the same, each isomorphic class corresponds to a unique partition. Also, we should note that if in an isomorphic class there is one graph satisfies the restriction specified by the summation on the right-hand side of (3.5.18), then all graphs included in the class satisfy the restriction. This suggests us putting a criterion on the partitions of the isomorphic classes to characterize whichever classes of graphs satisfy the restriction and whichever classes do not satisfy. For that purpose, let us denote $E_0 = \{1, \cdots, b_0\}$, $E_1 = \{b_0 + 1, \cdots, b_1\}$, $\cdots$, $E_l = \{b_{l-1} + 1, \cdots, b_l\}$, where the numbers b_i for $0 \le i \le l$ were as defined above (3.5.18).

All those isomorphic classes of graphs satisfing that restriction can be identified by using the following criterion on their separation: $m \ge l + 1$ and there exists integers $0 \le i_0 < i_1 < \cdots < i_l = m$ such that $\{F_1, \cdots, F_{i_0}\}$ forms a partition of E_0, $\{F_{i_0+1}, F_{i_0+2}, \cdots, F_{i_1}\}$ forms a partition of E_1, $\cdots$, $\{F_{i_{l-1}+1}, \cdots, F_{i_l}\}$ forms a partition of E_l. The meaning of $i_v - i_{v-1}$ for each $1 \le v \le l$ with $i_{-1} \equiv 1$ is then the number of non-coincident values taken by the vertices of the vth block. Therefore, let $\mathbb{F}_1$ be the collection of all those isomorphic classes whose partitions satisfy the restriction and $\mathbb{F}_2$ be the collection of all those isomorphic classes whose partitions do not satisfy the restriction.

Denote by F and $\mathcal{F}$ respectively an arbitrary graph and an arbitrary isomorphic class defined above. For each isomorphic class $\mathcal{F}$, define its representative graph by replacing the m vertices in an arbitrary graph selected from the class with J_1, $\cdots$, J_m whose values are not determined and are only restricted to be m different integers from the set $\{1, \cdots, n\}$. Denote the representative graph of $\mathcal{F}$ by $F(\mathcal{F})$.

Therefore,

$$\begin{aligned} &\left((\alpha_1)^q \prod_{v=1}^{q} (\alpha_{v+1})^{j_v}\right) \left((\alpha_1)^{q'} \prod_{v=1}^{q'} (\alpha_{v+1})^{j'_v}\right) - n^{-\frac{k}{2}-1} \sum_{\mathcal{G} \in \mathcal{C}} \sum_{G \in \mathcal{G}} t_{G_H} \\ = \; & n^{-\frac{k}{2}-1} \sum_{\mathcal{F} \in \mathbb{F}_2} \sum_{J_1, \cdots, J_m} \prod_{j=1}^{l} \left\{ \left(\prod_{a=b_{j-1}+1}^{b_j - 1} t_{I_a I_{a+1}} \right) t_{I_{b_j} I_{b_{j-1}+1}} \right\}, \end{aligned} \tag{3.5.19}$$

where the summation $\sum_{J_1, \cdots, J_m}$ is taken over all possible values of J_1, $\cdots$, J_m which should be m different integers from $\{1, \cdots, n\}$, and for each a, I_a takes values from the set of values $\{J_1, \cdots, J_m\}$. Note that replacing the class $\mathbb{F}_2$ with $\mathbb{F}_1$ in (3.5.19) yields the restricted summation on the right-hand side of (3.5.18). We next show the value of (3.5.19) varnishes as n tends to infinity.

For each $\mathcal{F} \in \mathbb{F}_2$, due to the property of the graphs defined above for the summand on the right-hand side of (3.5.18), also due to the definition of $\mathbb{F}_2$, the representative graph $\mathcal{F}$, $F(\mathcal{F})$, must be composed of at most l disjoint blocks each of which is connected and does not contain any cutting edge. Let each edge of $F(\mathcal{F})$ correspond to the matrix T_n and then we add the least number of edges to $F(\mathcal{F})$ so that the disjoint parts of the graph are connected altogether and let each of the added edge correspond to the matrix whose entries are all equal to one. Note that since this treatment is equivalent with adding to the product of the summand of (3.5.19) another several factors of 1, the summation value does not change. The number of such added edges cannot exceed $l-1$. Each added edge is cutting in the resulting graph and the $\|\cdot\|_0$ of its corresponding matrix is n. Thus by Lemma 2.2.1, for each $\mathcal{F} \in \mathbb{F}_2$,

$$|\sum_{J_1,\cdots,J_m} \prod_{j=1}^{l} \left\{ \left(\prod_{a=b_{j-1}+1}^{b_j-1} t_{I_a I_{a+1}} \right) t_{I_{b_j} I_{b_{j-1}+1}} \right\} | \leq C n^l.$$

Noting that $l = k/2$, in view of (3.5.19), we then get our result. □

Thus we now face the problem of counting the number of $\mathcal{C}$, which are subsets of isomorphic classes belonging to $\mathbb{G}_3$, included in $\mathcal{C}\begin{pmatrix} s;q,j_1,\cdots,j_q \\ l-1-s;q',j'_1,\cdots,j'_{q'} \end{pmatrix}$. This can be done by counting the number of their representative graphs. Given any such representative graph $G(\mathcal{C})$, by deleting from it the w-edge (i_1, i_2) and the w-edge coincident with (i_1, i_2), we can obtain two disjoint parts: one part contains s non-coincident w-edges and $s+1$ cycles which include q cycles of degree 1, j_v cycles of length $v+1$ for $1 \le v \le q$; one part contains $l-1-s$ non-coincident w-edges and $l-s$ cycles which include q' cycles of degree 1, j'_v cycles of length $v+1$ for $1 \le v \le q'$. Let $N_{s;q,j_1,\cdots,j_q}$ be the number of all possible graphs of the part of s non-coincident w-edges. Due to the symmetry of the two parts involved in one respective graph $G(\mathcal{C})$, the number of all possible graphs of the other part of $l-1-s$ non-coincident w-edges is $N_{s';q',j'_1,\cdots,j'_{q'}}$ where $s' = l-1-s$. By the rule of product in enumerative combinatorics, the number of all possible representative graphs we are concerning is given by $N_{s;q,j_1,\cdots,j_q} \times N_{s';q',j'_1,\cdots,j'_{q'}}$. In another word, the number of all possible subsets $\mathcal{C}$ in $\mathcal{C}\binom{s;q,j_1,\cdots,j_q}{l-1-s;q',j'_1,\cdots,j'_{q'}}$ is given by $N_{s;q,j_1,\cdots,j_q} \times N_{s';q',j'_1,\cdots,j'_{q'}}$. In the following, we need only calculate $N_{s;q,j_1,\cdots,j_q}$ and so in parallel we obtain also $N_{s';q',j'_1,\cdots,j'_{q'}}$.

3.5.3 Count of the Number of Graphs

The previous arguments guarantee us to finish the proof of Theorem 3.5.1, we need only show the following theorem.

Theorem 3.5.5. *For $s \ge 1$,*

$$N_{s;q,j_1,\cdots,j_q} = \frac{s!}{q!j_1!j_2!\cdots j_q!}.$$

Characterization of the Graphs Involved

Since the part of graph we now concern is preceded by the w-edge (i_1, i_2) and contains totally s non-coincident w-edges, it is a circuit of $4s+1$ edges as shown below:

$$i_2 \xrightarrow{t} i_3 \xrightarrow{w} i_4 \xrightarrow{t} i_5 \xrightarrow{w} \cdots \xrightarrow{t} i_{4s+1} \xrightarrow{w} i_{4s+2} \xrightarrow{t} i_{4s+3} (= i_2). \quad (3.5.20)$$

In our graph, this circuit is characterized by the following three conditions:

(i) The $2s$ w-edges in the circuit form s non-coincident w-edges each of which consists of two w-edges of opposite directions and is cutting in the graph.

(ii) There are exactly $2s+1$ non-coincident vertices in the graph.

(iii) The resulting graph of removing all the s non-coincident w-edges consists of $s+1$ cycles including q cycles of length 1, j_v cycles of length $v+1$ for $1 \le v \le q$.

The statement that a non-coincident w-edge is cutting means removing the two w-edges included in this non-coincident w-edge will cause the graph disconnected. Note that due to the property of a circuit, removing any single w-edge or any single t-edge cannot make the graph disconnected. Thus any t-edge is not cutting in the graph. This implies that the resulting graph of removing all of the s non-coincident w-edges from the graph consists of $s+1$ disjoint blocks each of which is connected and does not contain any cutting edge. Condition (ii) guarantees then each of the $s+1$ blocks is a cycle which contains the same number of t-edges and vertices. Let us prove this assertion. The condition implies that each non-coincident w-edge possesses two vertices which are distinct from the vertices of any other non-coincident w-edge and that the vertex composed of $i_2 = i_{4s+3}$, which is not connected with any w-edge at all by the definition of the circuit, is also distinct from all the other $2s$ vertices in the graph. Denote by C_0 the block containing this special vertex. Consider any block other than C_0. Suppose the block is connected with v non-coincident w-edges. Then the number of the vertices possessed by the block must be v now. On the other hand, since there are totally $2v$ w-edges connected with the block while each t-edge in the block must be connected with two w-edges connected with the block, we get there are exactly v t-edges in the block. Since the block is connected, now that it contains the same number of edges and vertices, it must be a cycle. Applying the same argument to C_0 and noting that the vertex composed of $i_2 = i_{4s+3}$ is also distinct from all other vertices in C_0 and is naturally connected with two t-edges, we can see C_0 is also a cycle. Define the length of a cycle to be the number of edges included in it. Thus conditions (i) and (ii) guarantees the validity of condition (iii).

Transformation Procedure

We now define a function $f(\cdot)$ on the cycles. For the heart cycle C_0, we simply define $f(C_0) = 0$. Consider the other s cycles. Let the s non-coincident w-edges in the graph are respectively composed of the w-edge (i_{2u_a+1}, i_{2u_a+2}) and the w-edge (i_{2v_a+1}, i_{2v_a+2}) with $u_a < v_a$ for $1 \leq a \leq s$. Also let $u_1 < u_2 < \cdots < u_s$. Then since each non-coincident w-edge is cutting, for each a there is a unique cycle among the s cycles which contains the t-edge (i_{2u_a+2}, i_{2u_a+3}). We say the w-edge (i_{2u_a+1}, i_{2u_a+2}) is the initiating w-edge of this cycle and give the value a to this cycle as the value of the function $f(\cdot)$ on this cycle. Then we give notation J_1, J_2, $\cdots$, J_q to the q loop cycles (cycles of length 1) according to $f(J_1) < f(J_2) < \cdots < f(J_q)$ and notation C_1, C_2, $\cdots$, C_{s-q} to the left $s-q$ cycles according to $f(C_1) < f(C_2) < \cdots < f(C_{s-q})$. Thus the meaning of the value of $f(\cdot)$ on a cycle is the occurence order of the cycle and we say the cycle C_{b_1} occurs earlier than the cycle C_{b_2} whenever $1 \le b_1 < b_2 \le s-q$.

If in the above for each $1 \le a \le s$, we further define the vertex composed of $i_{2u_a+2} = i_{2v_a+1}$ to be the end vertex of the non-coincident w-edge composed of the w-edge (i_{2u_a+1}, i_{2u_a+2}) and the w-edge (i_{2v_a+1}, i_{2v_a+2}), then now we proceed to define a function $g(\cdot)$ on the end vertices of the s non-coinicident w-edges. Due to the nature of the graph, supposing one person starts a walk from the end vertex composed of $i_{2u_a+2} = i_{2v_a+1}$ with his first step passing the t-edge (i_{2u_a+2}, i_{2u_a+3}), then he can never return to this end vertex unless he meets a loop cycle at somewhere during his walk. But there is the possibility that he meets more than one loop cycles during his walk. Now that we have defined notation J_v with $1 \le v \le q$ for the loop cycles, for each $1 \le a \le s$, we define the value of the function $g(\cdot)$ at the end vertex composed of $i_{2u_a+2} = i_{2v_a+1}$ to be J_v if J_v is the last loop cycle he meets during the prescribed walk before he returns to this end vertex for the first time.

We apply the following procedure, called the transformation procedure, to the graph of any given circuit of (3.5.20) satisfying conditions $(i) - (iii)$. We first shrink the q loop cycles into q vertices and still denote the resulting vertices by J_1, $\cdots$, J_q. Note that among the total s non-coincident w-edges, there are exactly q non-coincident w-edges containing the initiating w-edges of the q loop cycles. After the prescribed treatment to the q loop cycles, by our definition of end vertex, these q non-coincident w-edges take $J_1, \cdots, J_q$ as their end vertices. By definition, each of the cycles $C_1, \cdots, C_{s-q}$ possesses one end vertex of one of the other $s-q$ non-coincident w-edges. Cut at each of these $s-q$ vertices and each time leave the vertex only to the non-coincident w-edge. Glue all of those vertices in these $s-q$ vertices (now belonging only to the w-edges) on which the function $g(\cdot)$ takes the same value J_v with the vertex J_v (resulting from the original loop cycle indexed J_v). Carry out this treatment for every $1 \le v \le q$. If for some v, there is no vertices in these $s-q$ vertices on which the function $g(\cdot)$ takes the value J_v, then we just pass on to $v+1$. For each cycle of $C_1, C_2, \cdots, C_{s-q}$, join up the pair of t-edges which have been cut off the middle end vertex between them to get one t-edge possessing the same direction as the original two ones. Then every cycle of $C_1, C_2, \cdots, C_{s-q}$ is again complete but with its length reduced by 1. As a final step of the procedure, replace the two t-edges (i_2, i_3) and (i_{4s+2}, i_{4s+3}) in the heart cycle C_0 by one t-edge possessing the same direction. The procedure is finished.

Property of the Resulting Graph

The resulting graph of the prescribed procedure will possess the following characteristics. Each w-edge takes one end vertex from $J_1, \cdots, J_q$ and is still cutting in the resulting graph. There are totally $s-q+1$ cycles including j_v cycles of length v for $1 \le v \le q$. The total length of these cycles is simply $j_1 + 2j_2 + \cdots + vj_v = s$. Thus the resulting graph contains totally $2s$ w-edges and s t-edges. Moreover, the procedure includes indeed $s-q$ times of the same type of treatment as this: cut at one end vertex, then immediatedly complete the broken cycle and glue the end vertex with its corresponding J_v vertex. Since each cut followed by a completion of the broken cycle results in two circuits, the act of gluing the end vertex of the w-edge with the J_v vertex is just to glue two circuits at one vertex. Since gluing two circuits at one vertex still yields a circuit, the resulting graph of our produre is also a circuit. This means no matter one person starts from which point of the graph, by following the edges and their directions in the graph one by one, he can return to his start point by passing every edge in the graph once and only once.

Determination of Resulting Circuit

Let us now obtain the following circuit from the resulting graph as a correspondence to the original given circuit of (3.5.20) satisfying conditions $(i)-(iii)$. We achieve our purpose by supposing one person starts a walk passing every edge in the graph once and only once and then using the order he finishes each edge as the order of this edge appearing in the circuit. We let the w-edge resulting from (i_3, i_4) be the first edge of his walk and also the first edge of the circuit. Then he comes to one J_v vertex. If this J_v vertex is connected with only one non-coincident w-edge in the graph, then his next step is clear and must be following the w-edge coincident with the prescribed w-edge resulting from (i_3, i_4) to turn back to the present cycle, which must be the cycle resulting from C_0. Suppose the J_v vertex is connected with, except the non-coincident w-edge containing the w-edge resulting from (i_3, i_4), also d other non-coincident w-edges. Then these d non-coincident w-edges are connected with d cycles. Suppose these d cycles are resulting from cycles C_{b_1}, C_{b_2}, $\cdots$, C_{b_d} with $b_1 < b_2 < \cdots < b_d$. For each $1 \le j \le d$, in view that by removing from the graph the non-coincident w-edge connected with J_v and the cycle resulting from C_{b_j} we get two disjoint components, we define the component containing the cycle resulting from C_{b_j} as the jth outer branch of J_v. Due to the cutting nature of each non-coincident w-edge, once this person takes one w-edge into one of these d cycles, say the one resulting from C_{b_j}, then unless he completes all the edges included in the jth outer branch of J_v, he cannot return to J_v. Also, unless the person finishes all the edges included in these d outer branches of J_v as well as the d non-coincident w-edges connected with J_v, he cannot ultimately leave J_v. Recall that he can only ultimately leave J_v through taking the w-edge coincident with the one along which he first reaches J_v. Thus by requiring the person finishes the d outer branches following the rule that the earlier occurent be finished also earlier, *i.e.* he finishes in turn the 1st, 2nd, $\cdots$, dth outer branch connected with J_v, it is now clear each time he returns to the vertex J_v, whichever next w-edge he should take. Note that when he walks into a particular outer branch, say the jth outer branch, if the cycle resulting from C_{b_j} contains more than one edge, then after he walkes along one edge on the cycle, he will leave that cycle to a new $J_{v'}$ vertex. Then the previously defined rule for the preceding J_v vertex is well to be used here for $J_{v'}$.

We now see by defining at each J_v vertex the previous rule on choice of the next w-edge to take, the walking path of the person along the resulting graph is determined. Due to the nature of the graph, once the person walks into a cycle, his next step is naturally clear, *i.e.* walks one step along the present t-edge and then leaves the cycle along one w-edge to some J_v vertex. During these two steps happening after he arrives at a cycle, he need only follow the directions of the two edges consecutively oriented from his standing point on the cycle. This means once he walks into a cycle, his walking path is determined until he arrives at one J_v vertex. Thus once we determine the rule on his choice of taking whichever next w-edge at a J_v vertex, his walking path along the whole graph is determined. Therefore, it is now proper to use this determined walking path as the correspondent circuit to the given circuit of (3.5.20) satisfying conditions $(i) - (iii)$.

Characterization of the Resulting Circuit

To describe the obtained circuit, we renew the index in the graph by following the order the person passed each edge in the graph. So in the sequel, the vth edge in the obtained circuit is simply the vth edge in the walk path of the person. With this understanding, noting that the walk is a sequence of edges satisfying every two w-edges are followed by one t-edge, for every $1 \leq v \leq s$, we index the $(2v-1)$th w-edge (x_{2v-1}, y_v), the $2v$th w-edge (x_{2v}, y_v), and the vth t-edge (x_{2v}, x_{2v+1}) with $i_{2v+1} \equiv i_1$. Using the new indices, we can write the circuit obtained from the resulting graph as below:

$$x_1 \xrightarrow{w} y_1 \xrightarrow{w} x_2 \xrightarrow{t} x_3 \xrightarrow{w} \cdots \xrightarrow{t} x_{2s-1} \xrightarrow{w} y_s \xrightarrow{w} x_{2s} \xrightarrow{t} x_1. \quad (3.5.21)$$

Reviving the property of the resulting graph, the circuit is characterized by the following three conditions:

$(i)'$ The $2s$ w-edges in the circuit form s non-coincident w-edges each of which consists of two w-edges of opposite directions and is cutting in the graph.

$(ii)'$ There are s non-coincident vertices included in $\{x_1, x_2, \cdots, x_{2s}\}$ and q non-coincident vertices included in $\{y_1, y_2, \cdots, y_s\}$.

$(iii)'$ The s t-edges form $s-q+1$ cycles including j_v cycles of length v for $1 \le v \le q$.

Basic Facts of the Transformation Procedure

We first investigate some basic facts relating to the transformation procedure. Suppose that U is an arbitrarily given circuit of (3.5.20) satisfying conditions $(i)-(iii)$ and $G(U)$ is the graph of U and that after applying the transformation procedure to $G(U)$, the resulting circuit and its graph are respectively V and $G(V)$.

Consider the heart cycle C_0 in $G(U)$ and the cycle Q_0 in $G(V)$ resulting from C_0. Suppose C_0 is connected with totally d_0 non-coincident w-edges in $G(U)$. Then there exist integers $1 = u_1 < u_2 < \cdots < u_{d_0} < 2s$ such that the d_0 non-coincident w-edges connected with C_0 are respectively composed of (i_{2u_a+1}, i_{2u_a+2}) and (i_{2v_a+1}, i_{2v_a+2}) with $u_a < v_a$ and $i_{2u_a+2} = i_{2v_a+1}$, for $1 \le a \le d_0$. Here for each $1 \le a \le d_0 - 1$, $v_a = u_{a+1} - 1$, and $v_{d_0} = 2s$. For ease of reference, we speak the non-coincident w-edge composed of (i_{2u_a+1}, i_{2u_a+2}) and (i_{2v_a+1}, i_{2v_a+2}) to be the ath stem edge of C_0 and the graph of the edges (i_{2u_a+1}, i_{2u_a+2}), (i_{2u_a+2}, i_{2u_a+3}), $\cdots$, (i_{2v_a+1}, i_{2v_a+2}) to be the ath branch of C_0. Then the $d_0 + 1$ t-edges included in C_0 are respectively, (i_2, i_3), (i_{4s+2}, i_{4s+3}), and (i_{2v_a+2}, i_{2v_a+3}) for $1 \le a \le d_0 - 1$.

The first fact on the effect of the transformation procedure is: Let $c_m = \frac{3}{2}(v_m - u_m) + \frac{1}{2}$ for $1 \le m \le d_0$. Then for each $1 \le a \le d_0$, the ath branch of C_0 is transformed into from the $\left(\sum_{m=1}^{a-1}(c_m+1)+1\right)$th to the $\left(\sum_{m=1}^{a-1}(c_m+1)+c_a\right)$th edges in the resulting circuit V. The second fact is: For each $1 \le a \le d_0 - 1$, the t-edge (i_{2v_a+2}, i_{2v_a+3}) of C_0 is transformed into the $\sum_{m=1}^{a}(c_m+1)$th edge of the resulting circuit V. Remember that by the procedure, the left two t-edges of C_0, (i_2, i_3) and (i_{4s+2}, i_{4s+3}), have been combined into the last edge in the resulting circuit V.

The proof of the above two facts is a consequence of our definition of the J_v vertex corresponding to the end vertex of a non-coincident w-edge as well as the choice of the edge resulting from the w-edge (i_3, i_4) to be the first edge in the resulting circuit V. By our definition of the value J_v taken by the function $g(\cdot)$ on the end vertex of a non-coincident w-edge, say composed of (i_{2u_a+1}, i_{2u_a+2}) and (i_{2v_a+1}, i_{2v_a+2}) with $u_a < v_a$, $i_{2u_a+2} = i_{2v_a+1}$, J_v is the last loop cycle one person meets during his walk long the $(v_a - u_a - 1)$ edges (i_{2u_a+2}, i_{2u_a+3}), (i_{2u_a+3}, i_{2u_a+4}), $\cdots$, (i_{2v_a}, i_{2v_a+1}). Based on this definition, we see the end vertices of all the non-coincident w-edges included in the same branch of C_0 correspond to J_v vertices resulting from loop cycles only occurrent in this branch. Thus the images of the d_0 branches of C_0 in the graph of the resulting circuit V should be connected only by the cycle resulting from C_0, *i.e.* Q_0 in $G(V)$, and constitute respectively d_0 branches of Q_0 if we extend the above definition of branch to the graph $G(V)$. We can compute for each $1 \le a \le d_0$, the number of w-edges included in the ath branch of C_0 is $(v_a - u_a + 1)$ so that the number of non-coincident w-edge included in this branch is $\frac{1}{2}(v_a - u_a + 1)$ and hence that of the cycles included in this branch is $\frac{1}{2}(v_a - u_a + 1)$. Thus after the transformation procedure, the image of the ath branch of C_0 will totally contain $\frac{3}{2}(v_a - u_a) + \frac{1}{2}$ edges. Since (i_3, i_4), or equivalently

(i_{2u_1+1}, i_{2u_1+2}), is chosen to be the first edge of V, the results stated in the two facts follow.

Before we state our main result, we introduce another useful fact concerning the effect of the transformation procedure. Let us define further the graph of the edges (i_{2u_a+2}, i_{2u_a+3}), (i_{2u_a+3}, i_{2u_a+4}), $\cdots$, (i_{2v_a}, i_{2v_a+1}) to be the ath outer branch of C_0. Note that the ath branch of C_0 is then composed of the ath outer branch as well as the ath stem edge of it. Then as a consequence of the previous two facts, we have for each $1 \leq a \leq d_0$, the ath outer branch of C_0 is transformed into from the $\left(\sum_{m=1}^{a-1}(c_m+1)+2\right)$th to the $\left(\sum_{m=1}^{a-1}(c_m+1)+c_a-1\right)$th edges in the resulting circuit V. Let us denote this part of the circuit V by $\tilde{V}^{(a)}$.

Note that when an outer branch of the heart cycle is viewed as an independent circuit, then this circuit is also of the type of circuit of (3.5.20) satisfying conditions $(i)-(iii)$ with suitable changes of the sequence of characteristic numbers and so the transformation procedure can be applied to this circuit. Specifically, consider the ath outer branch of C_0 as a circuit and denote it by $U^{(a)}$. Suppose after applying the transformation procedure to this circuit, the resulting circuit is denoted by $V^{(a)}$. Then the relation between $\tilde{V}^{(a)}$ and $V^{(a)}$ is for $3 \leq v \leq (c_a - 2)$, the vth edge of $\tilde{V}^{(a)}$ is the $(v-2)$th edge of $V^{(a)}$ while the first two edges of $\tilde{V}^{(a)}$ become the last two edges of $V^{(a)}$. This is the third fact we introduce.

To prove the asserted result, we need utilize further the rule concerning at each J_v vertex in the resulting graph the occurence order of each non-coincident w-edge connected with the vertex, which is proposed by us in the process of determining the resulting circuit from the resulting graph of applying the transformation procedure. By our defnition of the ath branch of C_0, there is one cycle in this branch which contains the two t-edges (i_{2u_a+2}, i_{2u_a+3}) and (i_{2v_a}, i_{2v_a+1}) with $i_{2u_a+2} = i_{2v_a+1}$. Extend our definition of stem edges and branches to this cycle. Then the end vertex of the ath stem edge of C_0 possess the same value of J_v as the end vertex of the last stem edge of this later mentioned cycle, say denoted by C_b for convenience of reference. Then during the transformation procedure, the two t-edges (i_{2u_a+2}, i_{2u_a+3}) and (i_{2v_a}, i_{2v_a+1}) contained in C_b should be combined into one edge. Regarding the position of this resulting edge in the resulting graph $G(V)$, our rule on the choice of the next w-edge at each vertex J_v guarantees that the resulting edge is the $\left(\sum_{m=1}^{a-1}(c_m+1)+3\right)$th edge of the resulting circuit V, or in another word, is the second edge of $\tilde{V}^{(a)}$. But this edge is the last edge of $V^{(a)}$, the assertion follows.

We now proceed to prove for any given two different circuits of (3.5.20) satisfying conditions $(i)-(iii)$, the obtained two circuits of (3.5.21) satisfying conditions $(i)'-(iii)'$ are also different. For ease of reference, we say the two given circuits are U_1 and U_2, while the circuits obtained from them are correspondingly V_1 and V_2. For the four circuits, we say the graphs of them are respectively $G(U_1)$, $G(U_2)$, and $G(V_1)$, $G(V_2)$.

Theorem 3.5.6. *Suppose the heart cycle in $G(U_i)$ is denoted by $C_0^{(i)}$ for $i=1$, 2.*

(i) If the number of branches possessed by $C_0^{(1)}$ is not equal to that of the branches possessed by $C_0^{(2)}$ or if there is some a such that the number of edges

possessed by the ath branch of $C_0^{(1)}$ is not equal to that of the edges possessed by the ath branch of $C_0^{(2)}$, then $V_1 \neq V_2$.

(2) *If there is some a such that when the ath outer branch of $C_0^{(1)}$ and the ath outer branch of $C_0^{(2)}$ are considered to be two circuits $U_1^{(a)}$ and $U_2^{(a)}$, the resulting two circuits of applying the transformation procedure to $U_1^{(a)}$ and $U_2^{(a)}$ are not equal, then $V_1 \neq V_2$.*

Proof. The first conclusion follows straightforwardly from the first two facts previously shown. In terms of the conditions there, those two facts tell us if the heart cycle in the graph of a given circuit U possesses those branches specified there, then the $\sum_{m=1}^{a}(c_m + 1)$th (for $1 \leq a \leq d_0 - 1$) edges and the last edge of the resulting circuit V form a circle in its graph. Thus (i) follows. The second conclusion follows from the third fact. Denote by $\tilde{V}_i^{(a)}$ the part of edges in the circuit V_i which are resulting from the ath outer branch of $C_0^{(i)}$, $i = 1$, 2. Therefore, according to the third fact proven previously, $V_1^{(a)} \neq V_2^{(a)}$ results that $\tilde{V}_1^{(a)} \neq \tilde{V}_2^{(a)}$ and hence obviously $V_1 \neq V_2$. □

Theorem 3.5.7. *If $U_1 \neq U_2$, then $V_1 \neq V_2$.*

Proof. Based on (i) of Theorem 3.5.6, we first compare the number of branches possessed by the heart cycles $C_0^{(1)}$ and $C_0^{(2)}$. If they do not have the same number of branches, then we get our result, otherwise if there is some a such the ath branch of $C_0^{(1)}$ does not have the same number of edges as the ath branch of $C_0^{(2)}$, then we still get our result. Now suppose $C_0^{(1)}$ and $C_0^{(2)}$ have the same number of branches and each branch of $C_0^{(1)}$ has the same number of edges as the corresponding branch of $C_0^{(2)}$. But since $U_1 \neq U_2$, there must be some a such that the ath branch of $C_0^{(1)}$ is different from the ath branch of $C_0^{(2)}$. Then by (ii) of Theorem 3.5.6, we need only show that $V_1^{(a)} \neq V_2^{(a)}$. Recall that $V_1^{(a)}$ and $V_2^{(a)}$ are the resulting circuits of applying the transformation procedure to the circuits formed by the ath outer branch of $C_0^{(1)}$ and the ath outer branch of $C_0^{(2)}$, *i.e.* $U_1^{(a)}$ and $U_2^{(a)}$. Then again we compare the heart cycles in the graph of $U_1^{(a)}$ and $U_2^{(a)}$. We compare the number of branches and the number of edges each branch possesses. If these two steps of comparations do not give any difference, then since $U_1^{(a)} \neq U_2^{(a)}$ due to choice of a, we similarly proceed to find an integer a such that the ath branch of the heart cycle in the graph of $U_1^{(a)}$ is different from the ath branch of the heart cycle in the graph of $U_2^{(a)}$. Again from (ii) of Theorem 3.5.6, to get our result, we need only continue the preceding argument for the ath outer branch of the heart cycle in the graph of $U_1^{(a)}$ and the ath outer branch of the heart cycle in the graph of $U_2^{(a)}$. Note that since the number of cycles in the graphs is finite, the conclusion on the existence of two different branches in the previous treatment cannot hold throughout the way. Therefore, there must be some where during the process, the heart cycles being compared either possess different number of branches or possess branches having different number of edges. Thus from (i) of Theorem 3.5.6, we get our result. $\square$

Recovery Procedure

We now prove for any given circuit of (3.5.21) satisfying conditions $(i)' - (iii)'$, there also corresponds a unique circuit of (3.5.20) satisfying conditions $(i) - (iii)$. This is achieved with the aid of a procedure which plays an inverse effect compared with the previous transformation procedure. We shall then in the sequel refer to this procedure as the recovery procedure.

Suppose V is an arbitrarily given circuit of (3.5.21) satisfying conditions $(i)' - (iii)'$ and $G(V)$ is its graph. Let us denote by $J_1,\ J_2,\ \cdots,\ J_q$ the q non-coincident vertices included in $\{y_1, y_2, \cdots, y_s\}$. For each $1 \le m \le q$, suppose the vertex J_m is connected with $(d_m + 1)$ non-coincident w-edges in the graph $G(V)$. Then we can find integers $1 \le u_1 < u_2 < \cdots < u_{d_m+1} \le s$ for which $y_{u_1} = y_{u_2} = \cdots = y_{u_{d_m+1}}$, and the $(d_m + 1)$ non-coincident w-edges connected with J_m are respectively resulting from (x_{2u_1-1}, y_{u_1}) coincident with $(x_{2u_{d_m+1}}, y_{u_{d_m+1}})$, and (x_{2u_a}, y_{u_a}) coincides with $(x_{2u_{a+1}-1}, y_{u_{a+1}})$ for $1 \le a \le d_m$. We similarly define these $(d_m + 1)$ non-coincident w-edges as in turn the 0th, the 1st, $\cdots$, the d_mth stem edge of J_m. For $1 \le a \le d_m$, we define the graph of the edges (x_{2u_a}, y_{u_a}), (x_{2u_a}, x_{2u_a+1}), $\cdots$, $(x_{2u_{a+1}-1}, y_{u_{a+1}})$ as the ath branch of J_m. We also define the 0th branch of J_m to be the resulting graph obtained by removing from $G(V)$ all the other d_m branches of J_m. For each branch of J_m, we correspondingly define the resulting subgraph of removing from the branch the stem edge of J_m to be an outer branch of J_m.

The recovery procedure is applied to the graph $G(V)$ as follows. We consecutively carry out similar treatment to the q vertices $J_1,\ \cdots,\ J_q$. Thus suppose we are dealing with J_m for $1 \leq m \leq q$, for which the above concepts of stem edges, branches and outer branches are well defined. If $d_m = 0$, we simply replace the vertex J_m with a loop cycle of t-edge with a direction consistent with the two w-edges connected with the original J_m vertex. If $d_m \geq 1$, then the first step of us is to cut the graph at the vertex J_m into two components. We let one component consist of only the 0th branch of J_m and be denoted by P_1. We let the other component consist of all the other d_m branches of J_m and be denoted by P_2. There needs a consideration on which component pertains the vertex J_m. We let P_2 keep the vertex J_m but add one vertex to P_1 to take the place which is previously taken by the vertex J_m, *i.e.* to be one end vertex of the 0th stem edge of J_m in the component P_1. However, we shall not denote the added vertex by J_m anymore. Now we split the first t-edge (x_{2u_1}, x_{2u_1+1}) of P_2 into two ones possessing the same direction as the original one edge. Glue the added vertex of component P_1 with the splitting point of the split t-edge in component P_2.

The resulting graph will possess the following properties: Each non-coincident w-edge is still cutting in the graph; Since both the component P_1 with one vertex added and the component P_2 are circuits, the resulting graph is still a circuit; The 0th branch of J_m has been combined into part of the 1st branch of J_m and there remain d_m branches we originally defined for J_m unchanged. For simplicity, let us still use the terms of the 1st branch, $\cdots$, the d_mth branch to refer to these changed and unchanged branches of J_m.

Continue our procedure in a similar way to the resulting graph. Cut the (resulting) graph at J_m into two components. Let the component consisting of only the 1st branch of J_m be denoted by P_1 and let the other component consisting of the other $(d_m - 1)$ branches of J_m be denoted by P_2. Similaryly, let P_2 keep the vertex J_m but add one vertex to P_1. Split the first t-edge (x_{2u_2}, x_{2u_2+1}) of P_2 into two ones with the same direction as the original one edge. Glue the added vertex of P_1 with the splitting point of P_2. The resulting graph obviously pertains the first two properties prescribed and it is not hard to see the third property for it becomes: The 0th and the 1st branches of J_m have been combined into part of the 2nd branch of J_m and there remain $(d_m - 1)$ branches we originally defined for J_m unchanged. Let us still speak of the 2nd branch, $\cdots$, the d_mth branch of J_m. Then similar treatment is applied to the resulting graph.

After such treatment is carried out consecutively for $(d_m - 1)$ times, the resulting graph will contain only the $(d_m - 1)$th and the d_mth branches of J_m, where the $(d_m - 1)$th branch of J_m of course refers to the combination of the component P_1 yielded in the $(d_m - 2)$th time of treatment and the originally defined $(d_m - 1)$th branch of J_m. Cut the resulting graph at J_m into two components, one component P_1 containing the $(d_m - 1)$th branch of J_m and one component P_2 containing the d_mth branch of J_m. Let P_2 pertain J_m but add one vertex to P_1 to take the place of the vertex J_m. Split the t-edge $(x_{2u_{d_m}}, x_{2u_{d_m}+1})$ of P_2 into two edges of the same direction as the original one edge. Glue the added vertex in P_1 with the split point in P_2. Finally, we replace the vertex J_m in the resulting graph with a loop cycle of t-edge possessing a direction consistent with the directions of the two w-edges connected with the original vertex J_m. This completes our treatment to the vertex J_m.

Once the above treatment is carried out at each of the vertices $J_1, \cdots, J_q$, we finish the recovery procedure by replacing the t-edge (x_{2s}, x_1) of the graph with two t-edges with the same direction as (x_{2s}, x_1). This is because every cycle except the one containing this t-edge must belong to the ath branch of J_m for some m and some $a \geq 1$, so every cycle except this particular cycle has experienced splitting of its first t-edge into two edges when the above treatment is applied to all J_m vertices.

We determine a circuit from the resulting graph by defining the t-edge, which is the one of the two edges obtained by splitting (x_{2s}, x_1) pointing to x_1, to be the first edge. Then we let this circuit be denoted by U and be the correspondent circuit to the given circuit V. To show U is a circuit of (3.5.20) satisfying conditions $(i) - (iii)$, we need only consider conditions (ii) and (iii). We added d_m vertices in the treatment of each J_m as well as one additional vertex to the cycle containing the t-edge (x_{2s}, x_1) in the final step, so totally we added $\sum_{m=1}^{q} d_m + 1 = \sum_{m=1}^{q}(d_m + 1) - q + 1 = s - q + 1$ vertices during the procedure. Hence taking into account of the existed $s+q$ vertices in the original graph, the resulting graph totally possesses $2s+1$ non-coincident vertices. Condition (ii) holds. Condition (iii) is also obvious. Due to the procedure, each J_v vertex produces exactly one loop cycle and each of the $s - q + 1$ cycles in the original graph is added one edge. Thus in the resulting graph, there are q loop cycles and j_v cycles of length $v+1$ for $1 \leq v \leq q$. Therefore from the given circuit V of (3.5.21) satisfying conditions $(i)' - (iii)'$, by applying the recovery procedure, we obtained one circuit U of (3.5.20) satisfying conditions $(i) - (iii)$.

We now prove some basic facts on the effect of the recovery procedure. Denote by Q_0 the cycle in $G(V)$ which contains the t-edge (x_{2s}, x_1) and C_0 the cycle in $G(U)$ which contains the two t-edges resulting from splitting the t-edge (x_{2s}, x_1). Then C_0 is indeed the image of Q_0 resulting from applying the recovery procedure.

Suppose there are b_0 non-coincident w-edges connected with Q_0. Then we can find integers $1 = u_1 < u_2 < \cdots < u_{b_0} < s$ for which the b_0 non-coincident w-edges connected with Q_0 are respectively composed of (x_{2u_a-1}, y_{u_a}) coincident with (x_{2v_a}, y_{v_a}) with $y_{u_a} = y_{v_a}$ for $1 \leq a \leq b_0$, where $u_{a+1} = v_a + 1$ for $1 \leq a \leq b_0 - 1$ and $v_{b_0} = s$. For each $1 \leq a \leq b_0$, define the non-coincident w-edge consisting of (x_{2u_a-1}, y_{u_a}) and (x_{2v_a}, y_{v_a}) to be the ath stem edge of Q_0, the graph of (x_{2u_a-1}, y_{u_a}), (x_{2u_a}, y_{u_a}), (x_{2u_a}, x_{2u_a+1}), $\cdots$, (x_{2v_a-1}, y_{v_a}), (x_{2v_a}, y_{v_a}) to be the ath branch of Q_0, and the graph of (x_{2u_a}, y_{u_a}), (x_{2u_a}, x_{2u_a+1}), (x_{2u_a+1}, y_{u_a+1}), $\cdots$, $(x_{2(v_a-1)}, x_{2v_a-1})$, (x_{2v_a-1}, y_{v_a}) to be the ath outer branch of Q_0. Thus again, the ath outer branch is the resulting graph of removing from the ath branch the a stem edge. Note that the b_0 t-edges of Q_0 are respectively (x_{2v_a}, x_{2v_a+1}) for $1 \leq a \leq (b_0 - 1)$ and (x_{2s}, x_1).

Note that if we view the ath outer branch of Q_0 as a circuit $V^{(a)}$ starting from the w-edge (x_{2u_a+1}, y_{u_a+1}) and ending at the t-edge (x_{2u_a}, x_{2u_a+1}), then this circuit is still of type (3.5.21) with a suitable changement of parameters. Thus the recovery procedure can be applied to this circuit and we shall define the resulting circuit by $U^{(a)}$.

We summarize the basic facts regarding the effect of the recovery procedure as follows: Let $c_m = 4(v_m - u_m) + 2$. Then

(1) For any $1 \leq a \leq b_0$, after the recovery procedure the ath branch of Q_0 is

transformed into from the $\left(\sum_{m=1}^{a-1}(1+c_m)+2\right)$th to the $(\sum_{m=1}^{a}(1+c_m))$th edges in the graph $G(U)$ of the resulting circuit U;

(2) For any $1 \leq a \leq (b_0-1)$, after the recovery procedure the t-edges (x_{2v_a}, x_{2v_a+1}) of Q_0 is transformed into the $(\sum_{m=1}^{a}(1+c_m)+1)$th edge of U, while the t-edge (x_{2s}, x_1) of Q_0 is split into repectively the first and the last edge of U;

(3) The ath outer branch of Q_0 is transformed into the circuit $U^{(a)}$.

We take a look at the proof of these results. From the recovery procedure, we can see at each J_m vertex, the ath stem edge of it is glued with the cycle connected with J_m through the $(a+1)$th stem edge of it, while the last stem edge of it is installed one loop cycle at one end. Hence the images of the b_0 branches of Q_0 must be disjoint components in $G(U)$ forming b_0 branches of the cycle C_0 in $G(U)$. Due to the cutting nature of the non-coincident w-edges as well as the previous choice of the first edge of the resulting circuit U, we then get the the first two results outlined above. More specifically, it can be computed that the ath branch of Q_0 contains $3(v_a-u_a)+2$ edges including (v_a-u_a) non-coincident w-edges so that after the recovery procedure the image of the ath branch of Q_0 contains totally $4(v_a-u_a)+2$ edges. Noting that in the resulting graph, the image of the 1st branch of Q_0 is preceded by one t-edge, the first two results follow. The third result follows from the fact that by choosing the w-edge (x_{2u_a+1}, y_{u_a+1}) and the t-edge (x_{2u_a}, x_{2u_a+1}) to be respectively the first and last edge of the circuit of the ath outer branch, the resulting circuit of applying the recovery procedure to the circuit of the ath outer branch of Q_0 is exactly the same as the part of circuit of U

containing its from the $\left(\sum_{m=1}^{a-1}(1+c_m)+3\right)$th to the $(\sum_{m=1}^{a}(1+c_m)-1)$th edges.

Now we are ready to prove for any two different given circuits V_1 and V_2 of (3.5.21), the resulting two circuits of applying the recovery procedure are also different.

Theorem 3.5.8 *If $V_1 \neq V_2$, then $U_1 \neq U_2$.*

Proof. Denote by $Q_0^{(i)}$ the parallel of Q_0 with the given circuit replaced by V_i, $i = 1, 2$. By the first two results above, we can see if $Q_0^{(1)}$ and $Q_0^{(2)}$ have different number of branches or one pair of their branches, for instance say the ath branch of $Q_0^{(1)}$ and the ath branch of $Q_0^{(2)}$, have different number of edges, then we must have $U_1 \neq U_2$. Otherwise, since $V_1 \neq V_2$ there must exist integer a such that the ath branch of $Q_0^{(1)}$ is different from the ath branch of $Q_0^{(2)}$. Based on the third result above, $U_1 \neq U_2$ follows if $U_1^{(a)} \neq U_2^{(a)}$. Here $U_i^{(a)}$ is the parallel of $U^{(a)}$ above with the given circuit V replaced by V_i, $i = 1, 2$. However, we then look for in the graphs of $U_1^{(a)}$ and $U_2^{(a)}$ the two cycles which play the role in parallel to the cycles $Q_0^{(1)}$ and $Q_0^{(2)}$. Then we again compare the number of branches as well as the number of edges contained in each pair of corresponding branches of these two cycles. If there is some difference found, then we arrive at our conclusion. Otherwise, there must be another integer a such that the pair of the ath branches of the two cycles being compared are different. Thus the above argument can be developed similarly to the ath outer branches of these two cycles. Since the graph has only finite cycles, the conclusion of the existence of an integer a such that the

ath branches of the two cycles being compared are not the same cannot always be addressed. Therefore, there must meet two cycles in comparation possessing either different number of branches or branches containing different number of edges. Hence the conclusion is proven. □

Count of the Number of Circuits

Theorems 3.5.7 and 3.5.8 guarantee that the number of the circuits of (3.5.20) satisfying conditions $(i)-(iii)$ is exactly the same as that of the circuits of (3.5.21) satisfying conditions $(i)'-(iii)'$. The number of the latter set of circuits has been shown in Yin and Krishnaiah (1983) to be $\frac{s!}{q!j_1!\cdots j_q!}$. Their method is to define a set of sequences of numbers corresponding to the set of circuits of (3.5.21) satisfying conditions $(i)'-(iii)'$. Specifically, given any circuit, let it correspond to such a sequence of numbers:

(1) The 1st number is 0;

(2) The 2nd, 4th, $\cdots$, 2sth numbers are 1.

(3) If the t-edge (x_{2r}, x_{2r+1}) just completes a cycle of length v, then the $(2r+1)$th number is $-v$, otherwise the $(2r+1)$th number is 0.

Such a sequence of numbers can be expressed as

$$(0, 1, a_1, 1, a_2, \cdots, 1, a_s), \tag{3.5.22}$$

where among the s numbers a_i's, there are j_v numbers equal to $-v$ for $1 \le v \le q$ while $(q-1)$ numbers equal to 0.

There is a one to one correspondence between the set of circuits of (3.5.21) satisfying conditions $(i)' - (iii)'$ and the set of all sequences of (3.5.22) each of which satisfies that all partial sums of it are nonnegative. We only need to see for an arbitrarily given sequence of numbers defined above whose partial sums are all nonnegative, there corresponds a unique circuit of (3.5.21) satisfying conditions $(i)' - (iii)'$.

Suppose the given sequence is (3.5.22). If an integer m_1 is such that $a_1 = \cdots = a_{m_1-1} = 0$ and $a_{m_1} = -v_1$, then we let $x_{2u-1} = x_{2u}$ for $m_1 - v_1 + 2 \leq u \leq m_1$ while $x_{2m_1+1} = x_{2(m_1+1-v_1)}$ and $y_{m_1+1} = y_{m_1+1-v_1}$. Let us give an intuitive explanation of this.

Draw two lines in parallel, say the X-line and the Y-line, and we put all x-vertices on the X-line and all y-vertices on the Y-line. Suppose for each $1 \leq u \leq s$, we draw one up vertical edge pointing from y_u to x_{2u}, one horizontal edge pointing from x_{2u} to x_{2u+1}, one down vertical edge pointing from x_{2u+1} to y_{u+1} and call these three edges a unit of the graph. Then for the m_1 1's appearing before $a_{m_1} = -v_1$, we consecutively draw out the first m_1 units of the graph and we let the down vertical edge of the m_1th unit coincide with the up vertical edge of the $(m_1 - v_1 + 1)$th unit and let all those adjacent vertical down and up edges appearing between the prescribed two vertical edges coincide with each other.

Note that in the above, due to the restriction of nonnegative partial sums, we have $m_1 - v_1 \geq 0$. We totally drew out m_1 units of the graph and used v_1 of them to form a circle of length v_1 connected with v_1 non-coincident w-edges each of which consists of exactly two w-edges of the opposite directions. Hence there are still $(m_1 - v_1)$ free units in the graph.

For the given sequence there must be another integer m_2 such that $a_{m_1+1} = \cdots = a_{m_2-1} = 0$ and $a_{m_2} = -v_2$. We proceed to draw out the next $(m_2 - m_1)$ units of the graph. Plus the previous $(m_1 - v_1)$ free units, there are totally $(m_2 - v_1)$ free units in the graph now. We shall never use those units which have been used to form cycle before. Thus we just order the existed free units as from the 1st to the $(m_2 - v_1)$th units, *i.e.* the later drawn $(m_2 - m_1)$ units are arranged to follow the $(m_1 - v_1)$th unit left in the previous step. Then we let the down vertical edge of the $(m_2 - v_1)$th free unit coincide with the up vertical edge of the $(m_2 - v_1 - v_2 + 1)$th free unit while let all adjacent vertical down and up edges appearing between them coincide with each other. Note that in this process, due to the restriction of nonnegative definite partial sums, we have $m_2 - v_1 - v_2 \geq 0$. After this step, there are $(m_2 - v_1 - v_2)$ free units in the graph and there has appeared a new cycle of length v_2 which is connected with v_2 non-coincident w-edges. Also, there can be seen one non-coincident w-edge of this cycle has the same vertex on the Y-line as the non-coincident w-edge consisting of (x_{2m_1+1}, y_{m_1+1}) and $(x_{2(m_1+1-v_1)}, y_{m_1+1-v_1})$ which belongs to the previous cycle.

Continue the above dealings until we come to the last negative integer which must be a_s. Note that due to the restriction $\sum_{m=1}^{s} a_m = -\sum_{v=1}(vj_v) = -s$, the number of free units at the moment, *i.e.* $s + \sum_{m=1}^{s-1} a_m$, is equal to $-a_s$. We may simply complete the procedure by letting the down vertical edge of the last free unit coincide with the up vertical edge of the first free unit as well as letting adjacent vertical and up edges between them coincide with each other.

From the procedure, it can be seen with the occurence of each of first occurred

$s-q$ cycles, say a cycle with length v, there occurred $(v-1)$ non-coincident J-vertices in the graph while with the occurrence of the last cycle, say of length v', there occurred v' non-concident J-vertices, so there are totally $\sum_{v=1}^{q}(v-1)j_v+1=q$ non-coincident J-vertices in the drawn graph. Thus, taking into account of the number of the cycles as well as their length, we can see the drawn graph is indeed the graph of a circuit of (3.5.21) satisfying conditions $(i)'-(iii)'$.

Therefore, the number of circuits of (3.5.21) satisfying conditions $(i)'-(iii)'$ is equal to the number of sequences of (3.5.22) with nonnegative partial sums. However, to obtain the number of these sequences, we need the following result.

Theorem 3.5.9. (LEMMA of Yin and Krishnaiah (1983) p.503) *Suppose $a_1, \cdots, a_k$ are nonnegative integers such that $\sum_{i=1}^{k} a_i = -(k-1)$. Then*

(1) *There exists a unique integer r, $1 \le r \le k$, such that all partial sums of*

$$(a_r, 1, a_{r+1}, 1, \cdots, 1, a_{r+k-1})$$

are nonnegative.

(2) *For any $1 \le r_1 < r_2 \le k$,*

$$(a_{r_1}, 1, a_{r_1+1}, 1, \cdots, 1, a_{r_1+k-1}) \ne (a_{r_2}, 1, a_{r_2+1}, 1, \cdots, 1, a_{r_2+k-1}).$$

Here the indices are the residue classes (mod k).

Proof. The proof of (1) can be found in their paper, while (2) is a consequence of (1). We only give a proof of (2) then. We need only show that for any $1 < r \le k$,

$$(a_1, 1, a_2, 1, \cdots, 1, a_k) \ne (a_r, 1, a_{r+1}, 1, \cdots, 1, a_{r+k-1}).$$

Suppose not. Then there is some $0 < r_1 \leq k$ such that

$$(a_1, 1, a_2, 1, \cdots, 1, a_k) = (a_{r_1}, 1, a_{r_1+1}, 1, \cdots, 1, a_{r_1+k-1}).$$

Hence for each $1 \leq i \leq k$, $a_i = a_{r_1+i-1}$. From (1), there is a unique integer r_0 such that the sequence

$$(a_{r_0}, 1, a_{r_0+1}, 1, \cdots, 1, a_{r_0+k-1})$$

have nonnegative partial sums. However, by hypothesis, it follows that

$$(a_{r_0}, 1, a_{r_0+1}, 1, \cdots, 1, a_{r_0+k-1}) = (a_{r_1+r_0-1}, 1, a_{r_1+r_0}, 1, \cdots, 1, a_{r_1+r_0+k-2}).$$

Here the indices are the residue classes (mod k). Hence by the result of (1), there must be some nonnegative integer m such that $r_1 + r_0 - 1 = r_0 + mk$ so that $m = \frac{r_1-1}{k}$. But $0 < \frac{r_1-1}{k} < 1$. Thus we reach a contradiction. $\square$

Now we count the number of sequences of (3.5.22) possessing nonnegative partial sums. Note that whenever a sequence of

$$(a_0, 1, a_1, 1, a_2, \cdots, 1, a_s), \tag{3.5.23}$$

where among the $(s+1)$ numbers a_i's, there are j_v numbers equal to $-v$ for $1 \leq v \leq q$ while q numbers equal to 0, has nonnegative partial sums, the first number a_0 must be equal to 0. Hence the number of sequences of (3.5.22) possessing nonnegative partial sums is equal to the number of sequences of (3.5.23) possessing nonnegative partial sums. Note that the total number of sequences of (3.5.23) is obviously equal to $\frac{(s+1)!}{q! j_1! \cdots j_q!}$. By Theorem 3.5.9 (1) and (2), every $(s+1)$ sequences produce one and only one sequence with nonnegative partial sums, thus the number of sequences of (3.5.23) and that of (3.5.22), with nonnegative partial sums is $\frac{s!}{q!j_1!\cdots j_q!}$. Hence we also get the number of circuits of (3.5.21) satisfying conditions $(i)' - (iii)'$ is $\frac{s!}{q!j_1!\cdots j_q!}$. Up to this point, the proof of Theorem 3.5.5 is complete.

Theorems 3.5.4 and 3.5.5 together with relations (3.5.16) and (3.5.17) further guarantee us to conclude the following result.

Theorem 3.5.10. *For any positive integer k,*

$$\lim_{n\to\infty} EM_k = m_k,$$

where m_k is given by (3.5.5) and (3.5.6) of Theorem 3.5.1.

Estimation of the Fourth Moment

Theorem 3.5.11. *For any positive integer k,*

$$E(M_k - EM_k)^4 = O(n^{-2}).$$

Proof. Note that $EM_k \in \mathbb{R}$. We may write by using the terminologies of graphs,

$$E(M_k - EM_k)^4 = n^{-2k-4} \sum_{\cup_{\ell=1}^4 G^{(\ell)}} \prod_{\ell=1}^4 (t_{G_H^{(\ell)}}) E \prod_{\ell=1}^4 (w_{G_B^{(\ell)}} - Ew_{G_B^{(\ell)}}). \quad (3.5.24)$$

Here for $\ell = 1, \cdots, 4$, $G^{(\ell)}$'s are four graphs drawn independently 4 times for the summand appearing in the expression (3.5.7) of M_k.

It is easy to see from the independency among the variables, if among the four graphs there is one graph having no vertical edge coincident with some vertical edge of the other three graphs or if there is single w-edges in $\cup_{\ell=1}^{4} G^{(\ell)}$, then the summand of (3.5.24) is equal to zero. Hence in the sequel, we let $\mathcal{G}_I$ denote the set of graphs $\cup_{\ell=1}^{4} G^{(\ell)}$ where the four graphs are connected together and there is no single w-edge existing in their union and let $\mathcal{G}_{II}$ denote the set of graphs $\cup_{\ell=1}^{4} G^{(\ell)}$ where the four graphs form two disjoint components each of which contains two graphs connected together and contains no single w-edge. Denote by l the number of non-coincident w-edges contained in $\cup_{\ell=1}^{4} G^{(\ell)}$ and r the number of cutting non-coincident w-edges contained in $\cup_{\ell=1}^{4} G^{(\ell)}$. Then $2l \le 4k$.

For either case of $\mathcal{G}_I$ and $\mathcal{G}_{II}$, we use Lemma 2.2.1. For every non-coincident w-edge consisting of μ_i edges of one direction and ν_i edgs of the other direction, we let it correspond to the matrix B, whose (a, b)th entry is $E(w_{ab}^{\mu_i} \bar{w}_{ab}^{\nu_i})$. For every t-edge, we let it correspond to the matrix T_n. In the case of $\mathcal{G}_{II}$, we add one edge to connect the two components and let the edge correspond to the matrix whose entries are all equal to 1. Note that we still have any t-edge must be non-cutting. Since $\|B\|_0$ and $\|B\|$ have the same bound $\{(\delta_n \sqrt{n})^{\mu_i+\nu_i-2} n\}$ while $\|T_n\|$ is bounded by constant, similar to our previous calculations in proving Theorem 3.5.2, we get by Lemma 2.2.1,

$$
\begin{aligned}
& |n^{-2k-4} \sum_{\cup_{\ell=1}^{4} G^{(\ell)} \in \mathcal{G}_I} \prod_{\ell=1}^{4} (t_{G_H^{(\ell)}}) E \prod_{\ell=1}^{4} (w_{G_B^{(\ell)}} - E w_{G_B^{(\ell)}})| \\
\leq \quad & n^{-2k-4} C n \prod_{i=1}^{l} \left[(\delta_n \sqrt{n})^{\mu_i + \nu_i - 2} n \right] \\
\leq \quad & C \delta_n^{4k-2l} n^{-3},
\end{aligned}
$$

and

$$
\begin{aligned}
& |n^{-2k-4} \sum_{\cup_{\ell=1}^{4} G^{(\ell)} \in \mathcal{G}_{II}} \prod_{\ell=1}^{4} (t_{G_H^{(\ell)}}) E \prod_{\ell=1}^{4} (w_{G_B^{(\ell)}} - E w_{G_B^{(\ell)}})| \\
\leq \quad & n^{-2k-4} C n \prod_{i=1}^{l} \left[(\delta_n \sqrt{n})^{\mu_i + \nu_i - 2} n \right] \times n \\
\leq \quad & C \delta_n^{4k-2l} n^{-2}.
\end{aligned}
$$

The proof of the theorem is complete. □

By Borel-Cantelli's lemma in Lemma 2.1.5, Theorems 3.5.10 and 3.5.11 together imply as $n \to \infty$, M_k converges almost surely to m_k. Hence it follows Theorem 3.5.1 is proven. We finished the presentation of using the moment method to find the limiting spectral distribution of the Wigner type random matrices.

Chapter 4

General Sample Covariance Matrices

The present chapter is intended to develop a way of applying the Stieltjes transform method appropriate for spectral analysis of the class of general sample covariance matrices $B_n = \frac{1}{N} T_{2n}^{1/2} X_n T_{1n} X_n^* T_{2n}^{1/2}$. We exemplified our procedure through strategically finding the limiting spectral distribution for the class of matrices under the general assumption that T_{1n} is Hermitian and T_{2n} is nonnegative definite, specific definition of B_n formulated in Definition 1.2.1.

The organization of the chapter is as follows. In Section 4.1, we introduce in details the way we develop the Stieltjes transform method for the matrices of our concern. In Section 4.2, we prove the main mathematical tool used in the present chapter, which is developed directly for dealing with the resolvent of products of a Hermitian matrix and a nonnegative definite matrix. In Section 4.3, we prove preliminary results needed for proving Theorem 1.2.1. In Section 4.4, we finish the proof of Theorem 1.2.1.

4.1 Manipulation of the Stieltjes Transform Method

In this section, after a brief discussion on the matrices we shall deal with, we shall introduce in detail the way we apply the Stieltjes transform method to the matrices.

4.1.1 A Brief Introduction on the Matrices

Let us introduce firstly the following two matrices: $\forall_n = (1/N)T_{2n}X_nT_{1n}X_n^*$ and $A_n = (1/N)T_{1n}X_n^*T_{2n}X_n$. The relationship between the three classes of matrices B_n, $\forall_n$ and A_n is as follows.

The matrices B_n and $\forall_n$ have the same eigenvalues, so all the spectral properties of the two matrices are the same. In particular, their empirical spectral distributions are the same, *i.e.* $F^{B_n} = F^{\forall_n}$. Therefore, the proof of Theorem 1.2.1 can be obtained by proving the result stated in the theorem holds for the matrix $\forall_n$.

The nonzero eigenvalues of A_n are the same as those of $\forall_n$, while the number of zero eigenvalues of A_n equals the number of zero eigenvalues of $\forall_n$ plus $n - N$. This implies

$$F^{A_n}(x) = \frac{N}{n}F^{\forall_n}(x) + \left(1 - \frac{N}{n}\right)1_{[0,\infty)}(x), \tag{4.1.1}$$

and hence

$$s_{F^{A_n}}(z) = \frac{N}{n} s_{F^{\forall_n}}(z) - z^{-1}\left(1 - \frac{N}{n}\right). \tag{4.1.2}$$

Therefore $F^{A_n}(x)$ and $F^{\forall_n}(x)$ must converge simultaneously. Denote by F the limiting distribution of F^{A_n}. Then F and $\underline{F}$ must satisfy

$$F(x) = \frac{1}{c}\underline{F}(x) + \left(1 - \frac{1}{c}\right) 1_{[0,\infty)}(x),$$

and

$$s_F(z) = \frac{1}{c} s_{\underline{F}}(z) - z^{-1}\left(1 - \frac{1}{c}\right).$$

Moreover, we can formulate a limit theorem on F^{A_n} as follows.

Theorem 4.1.1. *Let $A_n = (1/N)T_{1n}X_n^* T_{2n}X_n$. Then under the assumptions of Theorem 1.2.1, with probability 1, the empirical spectral distribution of A_n, denoted by F^{A_n}, converges weakly to some non-random probability distribution function F for which if $H_1 \equiv 1_{[0,\infty)}$ or $H_2 \equiv 1_{[0,\infty)}$, then $F \equiv 1_{[0,\infty)}$; otherwise if for each $z \in \mathbb{C}^+$,*

$$\begin{cases} s(z) &=& -z^{-1}(1-c^{-1}) - z^{-1}c^{-1}\int \frac{1}{1+p(z)y} dH_2(y) \\ s(z) &=& -z^{-1}\int \frac{1}{1+q(z)x} dH_1(x) \\ s(z) &=& -z^{-1} - c^{-1}p(z)q(z) \end{cases} \tag{4.1.3}$$

is viewed as a system of equations for the complex vector $(s(z), p(z), q(z))$, then the Stieltjes transform of F, denoted by $s_F(z)$, together with the two functions in Theorem 1.2.1, $g_1(z)$ and $g_2(z)$, will satisfy that $(s_F(z), g_1(z), g_2(z))$ is the unique *solution to (4.1.3) in the set*

$$U = \{(s(z), p(z), q(z)) \;\; : \;\; Im(-z^{-1}(1-c) + cs(z)) > 0,$$
$$Im(zp(z)) > 0, Imq(z) > 0\}. \qquad (4.1.4)$$

With the aid of relations (4.1.1) and (4.1.2), Theorem 4.1.1 is a direct consequence of Theorem 1.2.1 and vice versa. The main task of the following sections in this chapter is then to prove Theorem 1.2.1 holds for the matrices $\forall_n$. However, during the process of seeking the limiting spectral distribution, no matter for $\forall_n$ or for A_n, both of them will show their effects. We then introduce our procedure to use the Stieltjes transform method on these two matrices.

4.1.2 The Stieltjes Transform Method

The basic rule relating to the Stieltjes transform method, *i.e.* Theorem 2.3.9, says that provided the sequence of the empirical spectral distributions is tight with probability one, to have the result asserted in Theorem 1.2.1, we only need to show for each $z \in \mathbb{C}^+$, the Stieltjes transforms of the empirical spectral distributions of the matrices converge almost surely to some non-random limit, which is determined by the system of equations in the theorem.

It can be verified[1] the Stieltjes transforms of $F^{\forall_n}$ and F^{A_n} are respectively given by $(1/N)tr(\forall_n - zI)^{-1}$ and $(1/n)tr(A_n - zI)^{-1}$. Then the matrices $(\forall_n - zI)^{-1}$ and $(A_n - zI)^{-1}$ are called the resolvent matrices. By using the resolvent identity

[1] A proof can be found in Section 4.2.

in Lemma 2.1.9, we can obtain

$$\frac{1}{N}tr(\forall_n - zI)^{-1} = -z^{-1} + z^{-1}\frac{1}{N}tr\{\forall_n(\forall_n - zI)^{-1}\},$$
$$\frac{1}{n}tr(A_n - zI)^{-1} = -z^{-1} + z^{-1}\frac{1}{n}tr\{A_n(A_n - zI)^{-1}\}.$$

We can develop the Stieltjes transform method in this way. We firstly expand the quantities, $\frac{1}{N}tr\{\forall_n(\forall_n - zI)^{-1}\}$ and $\frac{1}{n}tr\{A_n(A_n - zI)^{-1}\}$, with respect to some perturbation terms, whose presence and absence in these quantities do not affect the properties of the resolvent matrices but will reveal some latent rule governing the values of these quantities. The choice of such perturbations is quite important in finding the limit of the Stieltjes transforms of interest. Different choices will induce different limiting processes to occur, although the final limit should be the same. For those sample covariance matrices $(1/N)T_n^{1/2}X_n^*X_nT_n^{1/2}$, the columns of the matrix X_n^* have been an ideal choice to develop systematic investigations on large spectral properties of the matrices. However, as we have indicated in Section 1.2, the same choice is not so appropriate for the matrices B_n, or $\forall_n$ and A_n. Therefore, we change to use the entries of the matrix X_n as perturbations. So our procedure is to expand the preceding two quantities with respect to the entries x_{ij} and apply the resolvent identity to each resolvent matrix involved in the expansion with the comparing matrices chosen to be different from $\forall_n$ and A_n only by those components dependent on x_{ij}.

Due to the complexity of the matrices, we need still to carry out the above procedure for more generally evolved quantities. These are

$$\Phi_n^{(k)}(z) = \frac{1}{N}tr\{A_n^{(k)}(A_n - zI)^{-1}\}, \quad \Psi_n^{(k)}(z) = \frac{1}{N}tr\{\forall_n^{(k)}(\forall_n - zI)^{-1}\}, \quad (4.1.5)$$

where k is any positive integer for which

$$A_n^{(k)} \equiv \frac{1}{N} T_{1n} X_n^* T_{2n}^k X_n, \quad \forall_n^{(k)} \equiv \frac{1}{N} T_{2n} X_n T_{1n}^k X_n^*.$$

Thus we can see the quantities $\Phi_n^{(k)}(z)$ and $\Psi_n^{(k)}(z)$ are closely related with the Stieltjes transforms of $F^{\forall_n}$ and F^{A_n}.

We then carry out the preceding procedure for $\Phi_n^{(k)}(z)$ and $\Psi_n^{(k)}(z)$ as follows. Let us use e_i to denote the ith column of an $N \times N$ identity matrix and f_j the jth column of an $n \times n$ identity matrix. Then we can express the matrix X_n in the form

$$X_n = \sum_{ij} x_{ij} e_i f_j' \equiv X_{ij} + x_{ij} e_i f_j', \tag{4.1.6}$$

where in the summation $i = 1, \cdots, N$, $j = 1, \cdots, n$.

By using (4.1.6), we get

$$A_n^{(k)} = \frac{1}{N} \sum_{ij} x_{ij} T_{1n} X_{ij}^* T_{2n}^k e_i f_j' + \frac{1}{N} \sum_{ij} |x_{ij}|^2 \xi_{ii}^{(k)} T_{1n} f_j f_j', \tag{4.1.7}$$

where $\xi_{ii}^{(k)} = T_{2n}^k[i,i]$, *i.e.* the ith diagonal element of T_{2n}^k.

For each fixed pair (i, j), write for any $m = 1, \cdots, n$, $l = 1, \cdots, N$,

$$\tilde{a}_{ml}^{(k)} = f_m'(A_n - zI)^{-1} T_{1n} X_{ij}^* T_{2n}^k e_l,$$

$$\tilde{p}_{ml} = f_m'(A_n - zI)^{-1} T_{1n} f_l.$$

Then by using (4.1.7), we get

$$\Phi_n^{(k)}(z) = \frac{1}{N^2} \sum_{ij} x_{ij} \tilde{a}_{ji}^{(k)} + \frac{1}{N^2} \sum_{ij} |x_{ij}|^2 \xi_{ii}^{(k)} \tilde{p}_{jj}. \tag{4.1.8}$$

In the above we finished expressing $\Phi_n^{(k)}(z)$ into an expansion with respect to the perturbations x_{ij}.

Now we need to use the resolvent identity to the resolvent matrix $(A_n - zI)^{-1}$ involved in (4.1.8). For that purpose, we obtain the comparing matrix with A_n in this way. Let $A_{ij} \equiv (1/N)T_{1n}X_{ij}^* T_{2n} X_{ij}$. Then A_{ij} is the resulting matrix of eliminating from A_n the component dependent on the perturbation term x_{ij}. The difference between A_{ij} from A_n is given by

$$\begin{aligned} A_n - A_{ij} &= \frac{1}{N} x_{ij} T_{1n} X_{ij}^* T_{2n} e_i f_j' \\ &\quad + \frac{1}{N} \bar{x}_{ij} T_{1n} f_j e_i' T_{2n} X_{ij} + \frac{1}{N} |x_{ij}|^2 \xi_{ii}^{(1)} T_{1n} f_j f_j'. \end{aligned} \tag{4.1.9}$$

Use the resolvent identity to $(A_n - zI)^{-1} - (A_{ij} - zI)^{-1}$. Write further

$$\begin{aligned} p_{ml} &= f_m'(A_{ij} - zI)^{-1} T_{1n} f_l, \\ \hat{a}_{ml}^{(1)} &= e_m' T_{2n} X_{ij} (A_{ij} - zI)^{-1} T_{1n} f_l, \\ a_{ml}^{(k)} &= f_m' (A_{ij} - zI)^{-1} T_{1n} X_{ij}^* T_{2n}^k e_l, \\ d_{ii}^{(k)} &= e_i' T_{2n} X_{ij} (A_{ij} - zI)^{-1} T_{1n} X_{ij}^* T_{2n}^k e_i. \end{aligned}$$

It then follows

$$\tilde{p}_{jj} = p_{jj} - \frac{1}{N} x_{ij} \tilde{a}_{ji}^{(1)} p_{jj} - \frac{1}{N} \bar{x}_{ij} \tilde{p}_{jj} \hat{a}_{ij}^{(1)} - \frac{1}{N} |x_{ij}|^2 \xi_{ii}^{(1)} \tilde{p}_{jj} p_{jj}, \tag{4.1.10}$$

and

$$\tilde{a}_{ji}^{(k)} = a_{ji}^{(k)} - \frac{1}{N} x_{ij} \tilde{a}_{ji}^{(1)} a_{ji}^{(k)} - \frac{1}{N} \bar{x}_{ij} \tilde{p}_{jj} d_{ii}^{(k)} - \frac{1}{N} |x_{ij}|^2 \xi_{ii}^{(1)} \tilde{p}_{jj} a_{ji}^{(k)}. \tag{4.1.11}$$

Relations (4.1.8), (4.1.10) and (4.1.11) constitute the foundation of our later arguments which are to be developed in aiming to find the asymptotic behavior of $\Phi_n^{(k)}(z)$ and of course as a special case that of $s_{F^{A_n}}(z)$.

Let us now develop a similar treatment for the matrix $\forall_n$. In this case, for each positive integer k,

$$\forall_n^{(k)} \equiv \frac{1}{N}\sum_{ij}\bar{x}_{ij}T_{2n}X_{ij}T_{1n}^k f_j e_i' + \frac{1}{N}\sum_{ij}|x_{ij}|^2\zeta_{jj}^{(k)}T_{2n}e_ie_i',$$

where $\zeta_{jj}^{(k)} = T_{1n}^k[j,j]$. It follows then, by writing

$$\tilde{\sigma}_{ml}^{(k)} = e_m'(\forall_n - zI)^{-1}T_{2n}X_{ij}T_{1n}^k f_l,$$

$$\tilde{q}_{ml} = e_m'(\forall_n - zI)^{-1}T_{2n}e_l,$$

we have

$$\frac{1}{N}tr\{\forall_n^{(k)}(\forall_n - zI)^{-1}\} = \frac{1}{N^2}\sum_{ij}\bar{x}_{ij}\tilde{\sigma}_{ij}^{(k)} + \frac{1}{N^2}\sum_{ij}|x_{ij}|^2\zeta_{jj}^{(k)}\tilde{q}_{ii}. \quad (4.1.12)$$

Let $\forall_{ij} \equiv (1/N)T_{2n}X_{ij}T_{1n}^kX_{ij}^*$. Then

$$\begin{aligned}\forall_n - \forall_{ij} &= \frac{1}{N}x_{ij}T_{2n}e_if_j'T_{1n}X_{ij}^* \\ &\quad + \frac{1}{N}\bar{x}_{ij}T_{2n}X_{ij}T_{1n}f_je_i' + \frac{1}{N}|x_{ij}|^2\zeta_{jj}^{(1)}T_{2n}e_ie_i'. \qquad (4.1.13)\end{aligned}$$

Write

$$\sigma_{ml}^{(k)} = e_m'(\forall_{ij} - zI)^{-1}T_{2n}X_{ij}T_{1n}^k f_l,$$

$$q_{ml} = e_m'(\forall_{ij} - zI)^{-1}T_{2n}e_l,$$

$$b_{ml}^{(k)} = f_m'T_{1n}X_{ij}^*(\forall_{ij} - zI)^{-1}T_{2n}X_{ij}T_{1n}^k f_l,$$

$$\hat{\sigma}_{ml}^{(1)} = f_m'T_{1n}X_{ij}^*(\forall_{ij} - zI)^{-1}T_{2n}e_l.$$

By the resolvent identity we get

$$\tilde{q}_{ii} = q_{ii} - \frac{1}{N}x_{ij}\tilde{q}_{ii}\hat{\sigma}_{ji}^{(1)} - \frac{1}{N}\bar{x}_{ij}\tilde{\sigma}_{ij}^{(1)}q_{ii} - \frac{1}{N}|x_{ij}|^2\zeta_{jj}^{(1)}\tilde{q}_{ii}q_{ii}, \quad (4.1.14)$$

and

$$\tilde{\sigma}_{ij}^{(k)} = \sigma_{ij}^{(k)} - \frac{1}{N} x_{ij} \tilde{q}_{ii} b_{jj}^{(k)} - \frac{1}{N} \bar{x}_{ij} \tilde{\sigma}_{ij}^{(1)} \sigma_{ij}^{(k)} - \frac{1}{N} |x_{ij}|^2 \zeta_{jj}^{(1)} \tilde{q}_{ii} \sigma_{ij}^{(k)}. \quad (4.1.15)$$

The derivations in later sections are all devoted to establishing results to attain the asymptotic behavior of relation (4.1.8) equipped with (4.1.10) and (4.1.11) and that of relation (4.1.12) equipped with (4.1.14) and (4.1.15).

4.2 Mathematical Tools

In this section, we shall develop the main mathematical tools used in the remainder of the present chapter. Let us begin by listing a collection of useful inequalities on trace of matrices, which are well defined multiplications of two or more matrices.

Lemma 4.2.1.

(1) For rectangular matrices A, B,

$$|tr(AB)| \le \{tr(AA^*)\}^{1/2} \{tr(BB^*)\}^{1/2}.$$

(2) For Hermitian matrix A and nonnegative definite matrix B,

$$|tr(AB)| \le \|A\| tr(B).$$

(3) For rectangular matrices A, B, C, D,

$$|tr(ABCD)| \le \|A\| \|C\| \{tr(BB^*)\}^{1/2} \{tr(DD^*)\}^{1/2}.$$

(4) For Hermitian matrices A, B,

$$tr(AB)^2 \le \|A\|^2 tr(B^2).$$

(5) For rectangular matrix A and Hermitian matrix B,

$$tr(ABA^*)^2 \le \|B\|^2 tr(AA^*)^2.$$

(6) For rectangular matrix A, complex vectors a and b,

$$|a^*Ab| \le \|A\|(a^*a)^{1/2}(b^*b)^{1/2}.$$

Most of the derivations in the present chapter are based on the following theorem and its several consequences. For any m numbers $a_1, \cdots, a_m$, denote by $diag(a_1, \cdots, a_m)$ the diagonal matrix with diagonal elements $a_1, \cdots, a_m$. Moreover, if $a_i \ge 0$, for all i, and $\Delta \equiv diag(a_1, \cdots, a_m)$, then $\Delta^{1/2} \equiv diag(\sqrt{a_1}, \cdots, \sqrt{a_m})$.

Theorem 4.2.1. *Let B_n be $n \times n$ Hermitian nonnegative definite whose spectral decomposition is $B_n = U\Delta U^*$, where $\Delta = diag(\mu_1, \mu_2, \cdots, \mu_n)$ with $\mu_1 \ge \mu_2 \ge \cdots \ge \mu_n$, and U is $n \times n$ unitary matrix. Let $r = rank(B_n)$, $\tilde{\Delta} = diag(\mu_1, \mu_2, \cdots, \mu_r)$, and $U = (U_1, U_2)$ with U_1 consisting of the first r columns of U. Let H_n be $n \times n$ Hermitian and $z \in \mathbb{C}^+$ with $v = Imz > 0$. Then for any positive integer k,*

$$(B_n H_n - zI_n)^{-k} B_n^{1/2} = B_n^{1/2}[G(z)]^k, \tag{4.2.1}$$

where $B_n^{1/2} = U\Delta^{1/2}U^*$,

$$G(z) = U \begin{pmatrix} G_{11}^{-1}(z) & O \\ O & O \end{pmatrix} U^*,$$

with $G_{11}(z) = \tilde{\Delta}^{1/2} U_1^* H_n U_1 \tilde{\Delta}^{1/2} - zI_r$.

Proof. We first show (4.2.1) for $k = 1$. By using the formula

$$\begin{pmatrix} A & B \\ O & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}BD^{-1} \\ O & D^{-1} \end{pmatrix}$$

valid in case of nonsingular sub-matrices A and D, we have

$$\begin{aligned}
& (\Delta U^* H_n U - zI_n)^{-1} \\
= & \begin{pmatrix} \tilde{\Delta} U_1^* H_n U_1 - zI_r & \tilde{\Delta} U_1^* H_n U_2 \\ O & -zI_{n-r} \end{pmatrix}^{-1} \\
= & \begin{pmatrix} (\tilde{\Delta} U_1^* H_n U_1 - zI_r)^{-1} & z^{-1}(\tilde{\Delta} U_1^* H_n U_1 - zI_r)^{-1} \tilde{\Delta} U_1^* H_n U_2 \\ O & -z^{-1} I_{n-r} \end{pmatrix} \\
= & \begin{pmatrix} \tilde{\Delta}^{1/2} G_{11}^{-1}(z) \tilde{\Delta}^{-1/2} & z^{-1} \tilde{\Delta}^{1/2} G_{11}^{-1}(z) \tilde{\Delta}^{1/2} U_1^* H_n U_2 \\ O & -z^{-1} I_{n-r} \end{pmatrix}.
\end{aligned} \tag{4.2.2}$$

It follows that

$$\begin{aligned}
& (B_n H_n - zI_n)^{-1} B_n^{1/2} \\
= \; & U (\Delta U^* H_n U - zI_n)^{-1} \Delta^{1/2} U^* \\
= \; & U \begin{pmatrix} \tilde{\Delta}^{1/2} G_{11}^{-1}(z) & O \\ O & O \end{pmatrix} U^* = U \Delta^{1/2} \begin{pmatrix} G_{11}^{-1}(z) & O \\ O & O \end{pmatrix} U^* \\
= \; & B_n^{1/2} G(z).
\end{aligned}$$

Thus (4.2.1) holds for the case $k = 1$. For $k \geq 2$, we use the induction method. We have

$$\begin{aligned}
(B_n H_n - zI_n)^{-k} B_n^{1/2} &= (B_n H_n - zI_n)^{-1} (B_n H_n - zI_n)^{-k+1} B_n^{1/2} \\
&= (B_n H_n - zI_n)^{-1} B_n^{1/2} [G(z)]^{k-1} \\
&= B_n^{1/2} [G(z)]^k.
\end{aligned}$$

Thus (4.2.1) is proved. □

Corollary 4.2.1. *Under the assumptions of Theorem 4.2.1, we have for any* a, $b \in \mathbb{C}^n$ *and every* $z \in \mathbb{C}^+$ *with* $v \equiv Imz$,

(1) $$|a^*(B_nH_n - zI_n)^{-k}B_nb| \leq \frac{1}{v^k}(a^*B_na)^{1/2}(b^*B_nb)^{1/2}.$$

(2) $$|a^*(B_nH_n - zI_n)^{-k}b| \leq \frac{1}{v^k}|a^*b| + \frac{k}{v^{k+1}}(a^*B_na)^{1/2}(b^*H_nB_nH_nb)^{1/2}.$$

(3) $$|a^*B_n(H_nB_n - zI_n)^{-k}b| \leq \frac{1}{v^k}(a^*B_na)^{1/2}(b^*B_nb)^{1/2}.$$

(4) $$|a^*(H_nB_n - zI_n)^{-k}b| \leq \frac{1}{v^k}|a^*b| + \frac{k}{v^{k+1}}(a^*H_nB_nH_na)^{1/2}(b^*B_nb)^{1/2}.$$

Proof. Note that $\|G(z)\| \leq \|G_{11}^{-1}(z)\| \leq 1/v$. By using (6) of Lemma 4.2.1 and (4.2.1) of Theorem 4.2.1, we get

$$\begin{aligned} |a^*(B_nH_n - zI_n)^{-k}B_nb| &= |a^*B_n^{1/2}[G(z)]^kB_n^{1/2}b| \\ &\leq \frac{1}{v^k}(a^*B_na)^{1/2}(b^*B_nb)^{1/2}. \end{aligned}$$

Thus we proved (1). We prove (2) by induction. When $k = 1$, from the resolvent identity (2.1.5),

$$(B_nH_n - zI_n)^{-1} = -z^{-1}I_n + z^{-1}(B_nH_n - zI_n)^{-1}B_nH_n,$$

we get

$$\begin{aligned} |a^*(B_nH_n - zI_n)^{-1}b| &= |-z^{-1}a^*b + z^{-1}a^*(B_nH_n - zI_n)^{-1}B_nH_nb| \\ &\leq \frac{1}{v}|a^*b| + \frac{1}{v^2}(a^*B_na)^{1/2}(b^*H_nB_nH_nb)^{1/2}, \end{aligned}$$

where we have used the result of (1). Thus (2) is true for $k = 1$. When $k \geq 2$,

$$\begin{aligned}
& |a^*(B_nH_n - zI_n)^{-k}b| \\
\leq\ & |a^*(B_nH_n - zI_n)^{-k+1}[-z^{-1}I_n + z^{-1}(B_nH_n - zI_n)^{-1}B_nH_n]b| \\
\leq\ & \frac{1}{v}|a^*(B_nH_n - zI_n)^{-k+1}b| + \frac{1}{v}|a^*(B_nH_n - zI_n)^{-k}B_nH_nb| \\
\leq\ & \frac{1}{v}\left(\frac{1}{v^{k-1}}|a^*b| + \frac{k-1}{v^k}(a^*B_na)^{1/2}(b^*H_nB_nH_nb)^{1/2}\right) \\
& + \frac{1}{v^{k+1}}(a^*B_na)^{1/2}(b^*H_nB_nH_nb)^{1/2} \\
=\ & \frac{1}{v^k}|a^*b| + \frac{k}{v^{k+1}}(a^*B_na)^{1/2}(b^*H_nB_nH_nb)^{1/2}.
\end{aligned}$$

Therefore, (2) is proved. From (1) and (2), it is straightforward to derive (3) and (4). □

Corollary 4.2.2. *Under the assumptions of Theorem 4.2.1, we have*

$$tr(B_nH_n - zI_n)^{-1} = tr(B_n^{1/2}H_nB_n^{1/2} - zI_n)^{-1} = tr(H_nB_n - zI_n)^{-1}.$$

Proof. From (4.2.2),

$$\begin{aligned}
tr(B_nH_n - zI_n)^{-1} &= tr(\Delta U^*H_nU - zI_n)^{-1} \\
&= tr(\tilde{\Delta}U_1^*H_nU_1 - zI_r)^{-1} - z^{-1}(n-r) \\
&= tr\{\tilde{\Delta}^{1/2}G_{11}^{-1}(z)\tilde{\Delta}^{-1/2}\} - z^{-1}(n-r) \\
&= trG_{11}^{-1}(z) - z^{-1}(n-r).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
tr(B_n^{1/2}H_nB_n^{1/2} - zI_n)^{-1} &= tr(\Delta^{1/2}U^*H_nU\Delta^{1/2} - zI_n)^{-1} \\
&= tr\begin{pmatrix} G_{11}^{-1}(z) & O \\ & \\ O & -z^{-1}I_{n-r} \end{pmatrix} \\
&= trG_{11}^{-1}(z) - z^{-1}(n-r).
\end{aligned}$$

This completes the proof. □

Now we apply the above results to the matrices A_n and $\forall_n$. For A_n, take $B_n = (1/N)X_n^*T_{2n}X_n$ and $H_n = T_{1n}$. Then Corollary 4.2.2 gives us

$$(1/n)tr(A_n - zI_n)^{-1} = (1/n)tr(B_n^{1/2}T_{1n}B_n^{1/2} - zI_n)^{-1}.$$

But since $C_n \equiv B_n^{1/2}T_{1n}B_n^{1/2}$ is Hermitian having the same eigenvalues as A_n,

$$\frac{1}{n}tr(B_n^{1/2}T_{1n}B_n^{1/2} - zI_n)^{-1} = s_{F^{C_n}}(z) = s_{F^{A_n}}(z).$$

It therefore follows

$$s_{F^{A_n}}(z) = (1/n)tr(A_n - zI_n)^{-1}.$$

Similar argument gives

$$s_{F^{\forall_n}}(z) = (1/N)tr(\forall_n - zI_N)^{-1}.$$

From the resolvent identity,

$$s_{F^{A_n}}(z) = -z^{-1} + z^{-1}\frac{1}{n}tr\{A_n(A_n - zI_n)^{-1}\},$$

$$s_{F^{\forall_n}}(z) = -z^{-1} + z^{-1}\frac{1}{N}tr\{\forall_n(\forall_n - zI_N)^{-1}\}.$$

Using relation (4.1.2), we then get

$$tr\{A_n(A_n - zI_n)^{-1}\} = tr\{\forall_n(\forall_n - zI_N)^{-1}\}. \tag{4.2.3}$$

As will be seen in later sections, Theorem 4.2.1 and its consequences provide us special help in obtaining estimates for quantities involving the resolvent matrices $(\forall_n - zI)^{-1}$ and $(A_n - zI)^{-1}$.

4.3 Preliminary Results

In the present section, we prove some preliminary results concerning the empirical spectral distribution functions $F^{\forall_n}$ and F^{A_n} and their Stieltjes transforms. The purpose of these derivations is just to collect some basic facts needed for proving Theorem 1.2.1. Let us denote throughout the section $S_n = (1/N)X_n^*X_n$.

4.3.1 Preliminary Results: Part I

Lemma 4.3.1. *For any $M = M_1M_2M_3$ with $M_1 > 0$, $M_2 > 0$, $M_3 > 0$,*

$$\begin{aligned} & F^{\forall_n}\{(-\infty, -M) \cup (M, \infty)\} \qquad (4.3.1) \\ \leq\ & 2F^{T_{2n}}\{(M_1, \infty)\} + 2\frac{n}{N}F^{S_n}\{(M_2, \infty)\} + \frac{n}{N}F^{\sqrt{T_{1n}^2}}\{(M_3, \infty)\}. \end{aligned}$$

Proof. By Lemma 2.1.8, we have

$$
\begin{aligned}
& F^{\forall_n}\{(-\infty,-M)\cup(M,\infty)\} \\
= \; & F^{\frac{1}{N}T_{2n}^{1/2}X_nT_{1n}X_n^*T_{2n}^{1/2}}\{(-\infty,-M)\cup(M,\infty)\} \\
= \; & F^{\sqrt{(\frac{1}{N}T_{2n}^{1/2}X_nT_{1n}X_n^*T_{2n}^{1/2})^2}}\{(M,\infty)\} \\
\leq \; & 2F^{T_{2n}^{1/2}}\{(\sqrt{M_1},\infty)\} + F^{\sqrt{(\frac{1}{N}X_nT_{1n}X_n^*)^2}}\{(M_2M_3,\infty)\} \\
= \; & 2F^{T_{2n}}\{(M_1,\infty)\} + F^{\frac{1}{N^2}X_nT_{1n}X_n^*X_nT_{1n}X_n^*}\{(M_2^2M_3^2,\infty)\} \\
= \; & 2F^{T_{1n}}\{(M_1,\infty)\} + \frac{n}{N}F^{T_{1n}S_nT_{1n}S_n}\{(M_2^2M_3^2,\infty)\} \\
= \; & 2F^{T_{2n}}\{(M_1,\infty)\} + \frac{n}{N}F^{\sqrt{(S_n^{1/2}T_{1n}S_n^{1/2})^2}}\{(M_2M_3,\infty)\} \\
\leq \; & 2F^{T_{2n}}\{(M_1,\infty)\} + 2\frac{n}{N}F^{S_n}\{(M_2,\infty)\} + \frac{n}{N}F^{\sqrt{T_{1n}^2}}\{(M_3,\infty)\}.
\end{aligned}
$$

□

We can deduce two consequences from Lemma 4.3.1.

Lemma 4.3.2. *With probability* 1, $\{F^{\forall_n}\}$ *and* $\{F^{A_n}\}$ *are tight sequences.*

Proof. By Theorem 2.2.2, under assumption (i) of Theorem 1.2.1, F^{S_n} almost surely converges weakly to Marcĕnko-Pastur's law $F_{M-P}^{c,1}$ so that $\{F^{S_n}\}$ must be a tight sequence with probability one. Again, assumption (iii) in Theorem 1.2.1 implies that $F^{T_{1n}}$ and $F^{T_{2n}}$ are tight sequences. Hence from (4.3.1), with probability one $\{F^{\forall_n}\}$ is tight sequence. Note that

$$F^{A_n}\{(-\infty,-M)\cup(M,\infty)\} = \frac{N}{n}F^{\forall_n}\{(-\infty,-M)\cup(M,\infty)\}. \qquad (4.3.2)$$

Thus with Probability one, $\{F^{A_n}\}$ is also tight. $\square$

Another consequence of Lemma 4.3.1 is as follows.

Theorem 4.3.1. *If $H_1 \equiv 1_{[0,\infty)}$ or $H_2 \equiv 1_{[0,\infty)}$, then with probability 1, $F^{\forall_n}$ and F^{A_n} converge weakly to $1_{[0,\infty)}$.*

Proof. By (4.3.2), $F^{\forall_n}$ and F^{A_n} converge weakly to $1_{[0,\infty)}$ simultaneously. We need only show the theorem holds for $F^{\forall_n}$.

For any $\varepsilon > 0$ and $M > 0$, choose $M_1 = M_2 = \sqrt{M}$ and $M_3 = \varepsilon/M$ in (4.3.1). We get

$$\begin{aligned} & F^{\forall_n}\{(-\infty, -\varepsilon) \cup (\varepsilon, \infty)\} \\ \leq\ & 2F^{T_{2n}}\{(\sqrt{M}, \infty)\} + 2\frac{n}{N} F^{S_n}\{(\sqrt{M}, \infty)\} + \frac{n}{N} F^{\sqrt{T_{1n}^2}}\{(\varepsilon/M, \infty)\}. \end{aligned}$$

If $H_1 \equiv 1_{[0,\infty)}$, then with probability 1 for all $\varepsilon > 0$ and $M > 0$,

$$\limsup_{n\to\infty} F^{\sqrt{T_{1n}^2}}\{(\frac{\varepsilon}{M}, \infty)\} = 0.$$

Let $\sqrt{M} \to \infty$ only through continuity points of H_2. It follows, by assumption (iii) of Theorem 1.2.1 and the fact F^{S_n} converges weakly to $F^{c,1}_{M-P}$, with probability one,

$$\begin{aligned} & \limsup_{n\to\infty} F^{\forall_n}\{(-\infty, -\varepsilon) \cup (\varepsilon, \infty)\} \\ \leq\ & \lim_{M\to\infty} \left(2H_2\{(\sqrt{M}, \infty)\} + 2cF^{c,1}_{M-P}\{(\sqrt{M}, \infty)\}\right) \\ =\ & 0. \end{aligned}$$

Since ε is arbitrarily chosen, this implies if $H_1 \equiv 1_{[0,\infty)}$, then $F^{\forall_n}$ converges weakly to $1_{[0,\infty)}$ with probability one.

On the other hand, by choosing $M_2 = M_3 = \sqrt{M}$ and $M_1 = \varepsilon/M$, we get

$$\begin{aligned} & F^{\forall_n}\{(-\infty,-\varepsilon)\cup(\varepsilon,\infty)\} \\ \leq \;& 2F^{T_{2n}}\{(\varepsilon/M,\infty)\} + 2\frac{n}{N}F^{S_n}\{(\sqrt{M},\infty)\} + \frac{n}{N}F^{\sqrt{T_{1n}^2}}\{(\sqrt{M},\infty)\}, \end{aligned}$$

from which it follows if $H_2 \equiv 1_{[0,\infty)}$, then $F^{\forall_n}$ converges weakly to $1_{[0,\infty)}$ with probability one. This completes the proof. $\square$

In view of Theorem 4.3.1, throughout the remainder of the present chapter, we assume $H_1 \neq 1_{[0,\infty)}$ and $H_2 \neq 1_{[0,\infty)}$. We now present a proof of the unique solubility of the system of equations (1.2.2).

Lemma 4.3.3. *When $H_1 \neq 1_{[0,\infty)}$ and $H_2 \neq 1_{[0,\infty)}$, for each $z \in \mathbb{C}^+$, there is at most one vector $(s(z), p(z), q(z))$ satisfying the system of equations (1.2.2) with $Im(zp(z)) > 0$, $Imq(z) > 0$.*

Proof. From the three equations, it is easy to deduce

$$\begin{aligned} & -zp(z)q(z) \\ = \;& c\int \frac{q(z)x}{1+q(z)x}dH_1(x) = \int \frac{yp(z)}{1+p(z)y}dH_2(y). \end{aligned}$$

Since $Im(zp(z)) > 0$ and $Imq(z) > 0$ imply that $p(z) \neq 0$ and $q(z) \neq 0$, we get

$$zp(z) = -c\int \frac{x}{1+q(z)x}dH_1(x), \tag{4.3.3}$$

$$zq(z) = -\int \frac{y}{1+p(z)y}dH_2(y). \tag{4.3.4}$$

Thus we only need to show that if $(p(z), q(z))$ and $(\tilde{p}(z), \tilde{q}(z))$ both satisfy these two equations with $Im(zp(z)) > 0$, $Imq(z) > 0$, $Im(z\tilde{p}(z)) > 0$, $Im\tilde{q}(z) > 0$, then $p(z) = \tilde{p}(z)$ and $q(z) = \tilde{q}(z)$. Write $z = z_1 + iz_2$, $p(z) = p_1(z) + ip_2(z)$, $q(z) = q_1(z) + iq_2(z)$. Then (4.3.3) and (4.3.4) give us

$$z_1p_1(z) - z_2p_2(z) = -c\int \frac{x(1+q_1(z)x)}{|1+q(z)x|^2}dH_1(x), \tag{4.3.5}$$

$$z_1p_2(z) + z_2p_1(z) = c\int \frac{x^2q_2(z)}{|1+q(z)x|^2}dH_1(x), \tag{4.3.6}$$

$$z_1q_1(z) - z_2q_2(z) = -\int \frac{y(1+p_1(z)y)}{|1+p(z)y|^2}dH_2(y), \tag{4.3.7}$$

$$z_1q_2(z) + z_2q_1(z) = \int \frac{y^2p_2(z)}{|1+p(z)y|^2}dH_2(y). \tag{4.3.8}$$

From (4.3.7) and (4.3.8),

$$(z_1^2+z_2^2)q_2(z) = \int \frac{yz_2 + y^2(z_1p_2(z)+z_2p_1(z))}{|1+p(z)y|^2}dH_2(y).$$

Substitute this relation into (4.3.6). It follows

$$\begin{aligned}
&(z_1^2+z_2^2)(z_1p_2(z)+z_2p_1(z)) \\
= \;& c\int \frac{x^2}{|1+q(z)x|^2}dH_1(x)(z_1^2+z_2^2)q_2(z) \\
= \;& cz_2\int \frac{x^2}{|1+q(z)x|^2}dH_1(x)\int \frac{y}{|1+p(z)y|^2}dH_2(y) \\
& +c\int \frac{x^2}{|1+q(z)x|^2}dH_1(x)\int \frac{y^2}{|1+p(z)y|^2}dH_2(y)(z_1p_2(z)+z_2p_1(z)).
\end{aligned}$$

Note that $Im(zp(z)) > 0$ implies $z_1p_2(z) + z_2p_1(z) > 0$. By the facts that $z_2 > 0$, $H_1 \neq 1_{[0,\infty)}$ and $H_2 \neq 1_{[0,\infty)}$, the first term appearing on the right-hand side of the last equality in the above relation is positive. It follows

$$c\int \frac{x^2}{|1+q(z)x|^2}dH_1(x)\int \frac{y^2}{|1+p(z)y|^2}dH_2(y) < z_1^2 + z_2^2 = |z|^2.$$

This inequality remains true when $p(z)$ and $q(z)$ are replaced by $\tilde{p}(z)$ and $\tilde{q}(z)$ respectively.

We can compute that

$$z(p(z)-\tilde{p}(z)) = c\int \frac{x^2}{(1+q(z)x)(1+\tilde{q}(z)x)}dH_1(x)(q(z)-\tilde{q}(z)),$$

and so

$$\begin{aligned}
& z^2(q(z)-\tilde{q}(z)) \\
= & \int \frac{y^2}{(1+p(z)y)(1+\tilde{p}(z)y)}dH_2(y)z(p(z)-\tilde{p}(z)) \\
= & c\int \frac{x^2}{(1+q(z)x)(1+\tilde{q}(z)x)}dH_1(x) \\
& \int \frac{y^2}{(1+p(z)y)(1+\tilde{p}(z)y)}dH_2(y)(q(z)-\tilde{q}(z)).
\end{aligned}$$

However,

$$\begin{aligned}
& |c\int \frac{x^2}{(1+q(z)x)(1+\tilde{q}(z)x)}dH_1(x)\int \frac{y^2}{(1+p(z)y)(1+\tilde{p}(z)y)}dH_2(y)| \\
\le & \left(c\int \frac{x^2}{|1+q(z)x|^2}dH_1(x)\int \frac{y^2}{|1+p(z)y|^2}dH_2(y)\right)^{1/2} \\
& \left(c\int \frac{x^2}{|1+\tilde{q}(z)x|^2}dH_1(x)\int \frac{y^2}{|1+\tilde{p}(z)y|^2}dH_2(y)\right)^{1/2} \\
< & |z|^2.
\end{aligned}$$

Therefore, we must have $q(z)-\tilde{q}(z)=0$ and then $p(z)-\tilde{p}(z)=0$. This completes the proof. $\square$

4.3.2 Preliminary Results: Part II

Let us define two functions:

$$g_{1n}(z) = (1/N)tr\{(A_n - zI)^{-1}T_{1n}\}, \tag{4.3.9}$$

and

$$g_{2n}(z) = (1/N)tr\{(\forall_n - zI)^{-1}T_{2n}\}. \tag{4.3.10}$$

Their appearance is not accidental. Remember that in Theorem 1.2.1, we assert there are two functions $g_1(z)$ and $g_2(z)$ associated with the limit $s_{\underline{F}}(z)$ of the Stieltjes transform $s_{F^{\forall_n}}(z)$. In fact, $g_1(z)$ and $g_2(z)$ are just the almost sure limits of $g_{1n}(z)$ and $g_{2n}(z)$ respectively. It is during discussion of the limiting behavior of the two relations (4.1.8) and (4.1.12) that we shall first observe the appearance of these two functions. In this part, for later use, we shall derive some elementary properties concerning the Stieltjes transforms $s_{F^{\forall_n}}(z)$, $s_{F^{A_n}}(z)$ and the functions $g_{1n}(z)$, $g_{2n}(z)$.

In the following, let us denote by $\lambda_i(A)$ the i-th largest eigenvalue of any matrix A having real eigenvalues, i.e. $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \lambda_n(A)$. Also, we shall use the fact that for any $r \times r$ Hermitian matrix A, letting its first $s \times s$ principle sub-matrix be A_s, then $\lambda_i(A_s) \geq \lambda_{r-s+i}(A)$ and so $tr(A_s) \geq \sum_{i=r-s+1}^{r} \lambda_i(A)$.

Lemma 4.3.4. *Under the assumption that $H_2 \neq 1_{[0,\infty)}$, for any $z \in \mathbb{C}^+$, with probability* 1

$$\delta_{2z} \equiv \liminf_{n \to \infty} Im g_{2n}(z) > 0. \tag{4.3.11}$$

Proof. By Theorem 4.2.1,

$$g_{2n}(z) = \frac{1}{N} tr\{T_{2n}^{1/2} U \begin{pmatrix} G_{11}(z) & O \\ O & O \end{pmatrix} U^* T_{2n}^{1/2}\}, \tag{4.3.12}$$

where U is $N \times N$ unitary such that $U^* T_{2n} U = \Delta = diag(\lambda_1(T_{2n}), \cdots, \lambda_N(T_{2n}))$, and $G_{11}^{-1}(z) = (C_r - zI_r)^{-1}$, with $C_r = \frac{1}{N}\tilde{\Delta}^{1/2} U_1^* X_n T_{1n} X_n^* U_1 \tilde{\Delta}^{1/2}$, $r = rank(T_{2n})$, U_1 consisting of the first r columns of U and $\tilde{\Delta}$ being the diagonal matrix of the r positive eigenvalues of T_{2n}.

From (4.3.12), it follows $g_{2n}(z) = \frac{1}{N} tr\{\tilde{\Delta}^{1/2} G_{11}^{-1}(z) \tilde{\Delta}^{1/2}\}$. Denote the eigenvalues of C_r by μ_1, μ_2, $\cdots$, μ_r. Write $z = u + iv$. Define

$$\Phi_r = diag\left(\frac{\mu_1 - u}{(\mu_1 - u)^2 + v^2}, \cdots, \frac{\mu_r - u}{(\mu_r - u)^2 + v^2}\right),$$
$$\Psi_r = diag\left(\frac{v}{(\mu_1 - u)^2 + v^2}, \cdots, \frac{v}{(\mu_r - u)^2 + v^2}\right).$$

Then there exists a $r \times r$ unitary matrix Q_r such that

$$G_{11}^{-1}(z) = (C_r - zI_r)^{-1} = Q_r \Phi_r Q_r^* + i Q_r \Psi_r Q_r^*.$$

Hence we get

$$Im g_{2n}(z) = \frac{1}{N} tr\{\tilde{\Delta}^{1/2} Q_r \Psi_r Q_r^* \tilde{\Delta}^{1/2}\}.$$

For any constant $M > 0$, let

$$q_M = \sum_{i=1}^{r} I_{(|\mu_i| > M)},$$
$$c_M(z) = \frac{v}{2(M^2 + u^2) + v^2},$$
$$E_r = diag(I_{(|\mu_1| \le M)}, \cdots, I_{(|\mu_r| \le M)}).$$

We have $\Psi_r \geq c_M(z)E_r$. By the preceding claimed fact, we get

$$\begin{aligned}
Img_{2n}(z) &\geq c_M(z)\frac{1}{N}tr\{E_rQ_r^*\tilde{\Delta}Q_r\} \\
&= c_M(z)\frac{1}{N}tr\{E_r^2Q_r^*\tilde{\Delta}Q_r\} \\
&= c_M(z)\frac{1}{N}tr\{E_rQ_r^*\tilde{\Delta}Q_rE_r\} \\
&= c_M(z)\frac{1}{N}tr\{P_r\begin{pmatrix} I_{r-q_M} & O \\ O & O \end{pmatrix}P_r^*Q_r^*\tilde{\Delta}Q_rP_r\begin{pmatrix} I_{r-q_M} & O \\ O & O \end{pmatrix}P_r^*\} \\
&= c_M(z)\frac{1}{N}tr\{\begin{pmatrix} I_{r-q_M} & O \\ O & O \end{pmatrix}P_r^*Q_r^*\tilde{\Delta}Q_rP_r\begin{pmatrix} I_{r-q_M} & O \\ O & O \end{pmatrix}\} \\
&\geq c_M(z)\frac{1}{N}\sum_{i=q_M+1}^{r}\lambda_i(P_r^*Q_r^*\tilde{\Delta}Q_rP_r) \\
&= c_M(z)\frac{1}{N}\sum_{i=q_M+1}^{r}\lambda_i(\tilde{\Delta}) \\
&> \varepsilon c_M(z)\frac{1}{N}\sum_{i=q_M+1}^{r}I_{(\lambda_i(\tilde{\Delta})>\varepsilon)} \\
&= \varepsilon c_M(z)\frac{1}{N}\left(\sum_{i=1}^{r}I_{(\lambda_i(\tilde{\Delta})>\varepsilon)} - q_M\right) \\
&= \varepsilon c_M(z)\frac{1}{N}\left(\sum_{i=1}^{N}I_{(\lambda_i(T_{2n})>\varepsilon)} - q_M\right) \\
&= \varepsilon c_M(z)\left(F^{T_{2n}}\{(\varepsilon,\infty)\} - \frac{q_M}{N}\right).
\end{aligned}$$

It is easy to see

$$U^*\frac{1}{N}T_{2n}^{1/2}X_nT_{1n}X_n^*T_{2n}^{1/2}U = \begin{pmatrix} C_r & O \\ O & O \end{pmatrix}. \tag{4.3.13}$$

Then

$$\begin{aligned}
\frac{q_M}{N} &= \frac{1}{N}\sum_{i=1}^{N} 1_{(|\lambda_i(\frac{1}{N}T_{2n}^{1/2}X_nT_{1n}X_n^*T_{2n}^{1/2})|>M)} \\
&= \frac{1}{N}\sum_{i=1}^{N} 1_{(|\lambda_i(\forall_n)|>M)} \\
&= F^{\forall_n}\{(-\infty,-M)\cup(M,\infty)\}.
\end{aligned}$$

It follows

$$Img_{2n}(z) \geq \varepsilon c_M(z)\left(F^{T_{2n}}\{(\varepsilon,\infty)\} - F^{\forall_n}\{(-\infty,-M)\cup(M,\infty)\}\right). \quad (4.3.14)$$

The assumption $H_2 \neq 1_{[0,\infty)}$ implies that with probability one,

$$\liminf_{\varepsilon\downarrow 0}\liminf_{n\to\infty} F^{T_{2n}}\{(\varepsilon,\infty)\} > 0.$$

While, by the tightness of $\{F^{\forall_n}\}$,

$$\limsup_{M\to\infty}\limsup_{n\to\infty} F^{\forall_n}\{(-\infty,-M)\cup(M,\infty)\} = 0.$$

Let Ω be the subspace with probability one such that for every $\omega\in\Omega$, both of the above two relations hold. Then for each $\omega\in\Omega$ fixed, we first choose $\varepsilon>0$ such that

$$\liminf_{n\to\infty} F^{T_{2n}}\{(\varepsilon,\infty)\} = \varepsilon' > 0.$$

Then for this ε' we choose M such that

$$\limsup_{n\to\infty} F^{\forall_n}\{(-\infty,-M)\cup(M,\infty)\} < \varepsilon'/2.$$

Therefore for the fixed ω, we have $\delta_{2z} > \varepsilon c_M(z)\varepsilon'/2 > 0$. The proof is complete. $\square$

Lemma 4.3.5. *Under the assumption that $H_1 \neq 1_{[0,\infty)}$ and $H_2 \neq 1_{[0,\infty)}$, for any $z \in \mathbb{C}^+$, with probability* 1

$$\delta_{1z} \equiv \liminf_{n\to\infty} Im(zg_{1n}(z)) > 0. \tag{4.3.15}$$

Proof. To use Theorem 4.2.1, denote in this case

$$B_n \equiv (1/N)X_n^* T_{2n} X_n = U\Delta U^*,$$

$$G_{11}^{-1}(z) \equiv (\tilde{\Delta}^{1/2} U_1^* T_{1n} U_1 \tilde{\Delta}^{1/2} - zI_r)^{-1},$$

where U is $n \times n$ unitary, $\Delta = diag(\lambda_1(B_n), \cdots, \lambda_n(B_n))$, $r = rank(B_n)$ and $\tilde{\Delta} = diag(\lambda_1(B_n), \cdots, \lambda_r(B_n))$. Then by the resolvent identity,

$$\begin{aligned}
zg_{1n}(z) &= z\frac{1}{N}tr\{(A_n - zI)^{-1}T_{1n}\} \\
&= -\frac{1}{N}trT_{1n} + \frac{1}{N}tr\{A_n(A_n - zI)^{-1}T_{1n}\} \\
&= -\frac{1}{N}trT_{1n} + \frac{1}{N}tr\{T_{1n}B_n^{1/2}U \begin{pmatrix} G_{11}^{-1}(z) & O \\ O & O \end{pmatrix} U^* B_n^{1/2} T_{1n}\} \\
&= -\frac{1}{N}trT_{1n} + \frac{1}{N}tr\{T_{1n}U_1\tilde{\Delta}^{1/2}G_{11}^{-1}(z)\tilde{\Delta}^{1/2}U_1^*T_{1n}\}.
\end{aligned}$$

Write $C_r = \tilde{\Delta}^{1/2}U_1^*T_{1n}U_1\tilde{\Delta}^{1/2}$. Similarly define Φ_r and Ψ_r as in Lemma 4.3.4. Then there exists a $r \times r$ unitary matrix Q_r such that $G_{11}^{-1}(z) = (C_r - zI_r)^{-1} = Q_r\Phi_r Q_r^* + iQ_r\Psi_r Q_r^*$. Hence we get

$$Im(zg_{1n}(z)) = \frac{1}{N}tr\{T_{1n}U_1\tilde{\Delta}^{1/2}Q_r\Psi_r Q_r^*\tilde{\Delta}^{1/2}U_1^*T_{1n}\}.$$

Again define $c_M(z)$, q_M and E_r as in the proof of Lemma 4.3.4, but of course with the meaning interpreted within the present context. We then get

$$\begin{aligned}
Im(zg_{1n}(z)) &\geq c_M(z)\frac{1}{N}\sum_{i=q_M+1}^{r}\lambda_i(\tilde{\Delta}^{1/2}U_1^*T_{1n}^2U_1\tilde{\Delta}^{1/2}) \\
&> \varepsilon c_M(z)\frac{1}{N}\left(\sum_{i=1}^{n} I_{\left(\lambda_i(\Delta^{1/2}U^*T_{1n}^2U\Delta^{1/2})>\varepsilon\right)} - q_M\right) \\
&= \varepsilon c_M(z)\frac{1}{N}\left(\sum_{i=1}^{n} I_{\left(\lambda_i(T_{1n}^2B_n)>\varepsilon\right)} - q_M\right) \\
&= \varepsilon c_M(z)\frac{n}{N}\left(F^{\frac{1}{N}T_{1n}X_n^*T_{2n}X_nT_{1n}}\{(\varepsilon,\infty)\} - \frac{q_M}{n}\right).
\end{aligned}$$

By the definition of q_M, we obtain

$$\begin{aligned}
\frac{q_M}{n} &= \frac{1}{n}\sum_{i=1}^{r} I_{(|\lambda_i(\tilde{\Delta}^{1/2}U_1^*T_{1n}U_1\tilde{\Delta}^{1/2})|>M)} \\
&= \frac{1}{n}\sum_{i=1}^{n} I_{(|\lambda_i(\Delta^{1/2}U^*T_{1n}U\Delta^{1/2})|>M)} \\
&= \frac{1}{n}\sum_{i=1}^{n} I_{(|\lambda_i(B_n^{1/2}T_{1n}B_n^{1/2})|>M)} \\
&= \frac{1}{n}\sum_{i=1}^{n} I_{(|\lambda_i(A_n)|>M)} \\
&= F^{A_n}\{(-\infty,-M)\cup(M,\infty)\}.
\end{aligned}$$

It follows

$$\begin{aligned}
Im(zg_{1n}(z)) \geq\ & \varepsilon c_M(z)\frac{n}{N}\Big(F^{\frac{1}{N}T_{1n}X_n^*T_{2n}X_nT_{1n}}\{(\varepsilon,\infty)\} \\
& -F^{A_n}\{(-\infty,-M)\cup(M,\infty)\}\Big). \qquad (4.3.16)
\end{aligned}$$

It can be shown under the assumption that $H_1 \neq 1_{[0,\infty)}$ and $H_2 \neq 1_{[0,\infty)}$,

$$\liminf_{\varepsilon\downarrow 0}\liminf_{n\to\infty} F^{\frac{1}{N}T_{1n}X_n^*T_{2n}X_nT_{1n}}\{(\varepsilon,\infty)\} > 0.$$

By following the same argument as used in proving Lemma 4.3.4, we then finish the proof of Lemma 4.3.5. $\square$

Lemma 4.3.6. *For each $n \geq 1$ and each $\omega \in \Omega$, $s_{F^{A_n}}(z)$, $s_{F^{\forall_n}}(z)$, $g_{1n}(z)$ and $g_{2n}(z)$ are analytic functions on $\mathbb{C}^+$. Moreover, for any given $z \in \mathbb{C}^+$ with $Imz = v$, $|s_{F^{A_n}}(z)| \leq 1/v$, $|s_{F^{\forall_n}}(z)| \leq 1/v$, $|g_{2n}(z)| \leq \tau/v$, and $|g_{1n}(z)| \leq \frac{\tau}{v}\frac{n}{N} + \frac{\tau^3}{v^2}\frac{1}{N^2}tr(X_n^* X_n)$, where τ is any constant bigger than the value of* $\max(\|T_{1n}\|, \|T_{2n}\|)$.

Proof. Write $B_n = (1/N)X_n^* T_{2n} X_n$ and $H_n = (1/N)X_n T_{1n} X_n^*$. From Corollary 4.2.2,

$$s_{F^{A_n}}(z) = \frac{1}{n}tr(B_n^{1/2} T_{1n} B_n^{1/2} - zI_n)^{-1},$$
$$s_{F^{\forall_n}}(z) = \frac{1}{N}tr(T_{2n}^{1/2} H_n T_{2n}^{1/2} - zI_N)^{-1}.$$

It follows immediately $s_{F^{A_n}}(z)$ and $s_{F^{\forall_n}}(z)$ are analytic on $\mathbb{C}^+$ with $|s_{F^{A_n}}(z)| \leq 1/v$, $|s_{F^{\forall_n}}(z)| \leq 1/v$.

By Theorem 4.2.1,

$$|g_{2n}(z)| = |\frac{1}{N}tr\{T_{2n}^{1/2} G(z) T_{2n}^{1/2}\}| \leq \frac{\tau}{v}.$$

By Theorem 4.2.1, the resolvent identity (2.1.5) and (2) of Lemma 4.2.1,

$$\begin{aligned} |g_{1n}(z)| &= | - z^{-1}\frac{1}{N}trT_{1n} + z^{-1}\frac{1}{N}tr\{T_{1n}B_n^{1/2}G(z)B_n^{1/2}T_{1n}\}| \\ &\leq \frac{\tau}{v}\frac{n}{N} + \frac{\tau^3}{v^2}\frac{1}{N^2}tr(X_n^* X_n). \end{aligned}$$

In the above and in the sequel, whenever we use $G(z)$, its meaning should be interpreted according to the context where it is used. Thus below we need only prove $g_{1n}(z)$ and $g_{2n}(z)$ are analytic on $\mathbb{C}^+$.

For any z, $z_0 \in \mathbb{C}^+$ with $v = Imz$, $v_0 = Imz_0$, by Theorem 4.2.1 and Lemma

4.2.1,

$$
\begin{aligned}
& |\frac{1}{N} tr\{(\forall_n - zI)^{-1}(\forall_n - z_0 I)^{-2} T_{2n}\}| \\
= & |\frac{1}{N} tr\{(\forall_n - zI)^{-1} T_{2n}^{1/2} [G(z_0)]^2 T_{2n}^{1/2}\}| \\
= & |\frac{1}{N} tr\{T_{2n}^{1/2} G(z) [G(z_0)]^2 T_{2n}^{1/2}\}| \\
\leq & \frac{\tau}{v v_0^2}.
\end{aligned}
$$

It follows from the resolvent identity, as $z \to z_0$,

$$
\begin{aligned}
\frac{g_{2n}(z) - g_{2n}(z_0)}{z - z_0} &= \frac{1}{N} tr\{(\forall_n - zI)^{-1}(\forall_n - z_0 I)^{-1} T_{2n}\} \\
&= \frac{1}{N} tr\{(\forall_n - z_0 I)^{-2} T_{2n}\} \\
&\quad +(z - z_0)\frac{1}{N} tr\{(\forall_n - zI)^{-1}(\forall_n - z_0 I)^{-2} T_{2n}\} \\
&\to \frac{1}{N} tr\{(\forall_n - z_0 I)^{-2} T_{2n}\}.
\end{aligned}
$$

Write $B_n = (1/N) X_n^* T_{2n} X_n$. From Theorem 4.2.1, we have

$$
\begin{aligned}
& |\frac{1}{N} tr\{A_n (A_n - zI)^{-1}(A_n - z_0 I)^{-2} T_{1n}\}| \\
= & |\frac{1}{N} tr\{T_{1n} B_n^{1/2} G(z) [G(z_0)]^2 B_n^{1/2} T_{1n}\}| \\
\leq & \frac{n\tau^3}{v v_0^2} \frac{1}{N^2} tr(X_n^* X_n).
\end{aligned}
$$

Hence by the resolvent identity as $z \to z_0$,

$$
\begin{aligned}
& (z - z_0)\frac{1}{N} tr\{(A_n - zI)^{-1}(A_n - z_0 I)^{-2} T_{1n}\} \\
= & -z^{-1}(z - z_0)\frac{1}{N} tr\{(A_n - z_0 I)^{-2} T_{1n}\} \\
& +z^{-1}(z - z_0)\frac{1}{N} tr\{A_n (A_n - zI)^{-1}(A_n - z_0 I)^{-2} T_{1n}\} \\
\to & \; 0.
\end{aligned}
$$

It follows

$$\begin{aligned}\frac{g_{1n}(z)-g_{1n}(z_0)}{z-z_0} &= \frac{1}{N}tr\{(A_n-zI_n)^{-1}(A_n-z_0I_n)^{-1}T_{1n}\}\\ &\to \frac{1}{N}tr\{(A_n-z_0I)^{-2}T_{1n}\}.\end{aligned}$$

Thus $g_{1n}(z)$ and $g_{2n}(z)$ are analytic on $\mathbb{C}^+$. $\square$

Lemma 4.3.7. *For any positive integer l,*

$$\frac{1}{N}tr\{T_{1n}^l A_n(A_n-zI)^{-1}\} = \frac{1}{N}tr\{\forall_n^{(l+1)}(\forall_n-zI)^{-1}\}, \tag{4.3.17}$$

$$\frac{1}{N}tr\{(\forall_n-zI)^{-1}\forall_n T_{2n}^l\} = \frac{1}{N}tr\{A_n^{(l+1)}(A_n-zI)^{-1}\}, \tag{4.3.18}$$

for all $z\in\mathbb{C}^+$.

Proof. Given any $n\geq 1$ and $\omega\in\Omega$, the four functions involved in (4.3.17) and (4.3.18) are analytic on $\mathbb{C}^+$. In fact, let $\delta=\|(1/N)X_n^*X_n\|$ and τ be any positive constant bigger than the value of $\max(\|T_{1n}\|,\|T_{2n}\|)$. Consider $z\to z_0$, where z, $z_0\in\mathbb{C}^+$ with $v=Imz$ and $v_0=Imz_0$.

By Theorem 4.2.1,

$$\begin{aligned}&|\frac{1}{N}tr\{T_{1n}^l A_n(A_n-zI)^{-1}(A_n-z_0I)^{-2}\}|\\ =\ &|\frac{1}{N}tr\{T_{1n}^{l+1}B_n^{1/2}G(z)[G(z_0)]^2B_n^{1/2}\}|\\ \leq\ &\frac{n}{N}\frac{\delta\tau^{l+2}}{vv_0^2}.\end{aligned}$$

It follows

$$
\begin{aligned}
& \frac{\frac{1}{N} tr\{T_{1n}^l A_n[(A_n - zI)^{-1} - (A_n - z_0 I)^{-1}]\}}{z - z_0} \\
= \quad & \frac{1}{N} tr\{T_{1n}^l A_n (A_n - zI)^{-1} (A_n - z_0)^{-1}\} \\
= \quad & \frac{1}{N} tr\{T_{1n}^l A_n (A_n - z_0 I)^{-2}\} \\
& +(z - z_0) \frac{1}{N} tr\{T_{1n}^l A_n (A_n - zI)^{-1} (A_n - z_0 I)^{-2}\} \\
\to \quad & \frac{1}{N} tr\{T_{1n}^l A_n (A_n - z_0 I)^{-2}\}.
\end{aligned}
$$

Thus $(1/N) tr\{T_{1n}^l A_n (A_n - zI)^{-1}\}$ is analytic on $\mathbb{C}^+$. Also,

$$
\begin{aligned}
& |\frac{1}{N} tr\{\forall_n^{(l+1)} (\forall_n - zI)^{-1} (\forall_n - z_0 I)^{-2}\}| \\
= \quad & |\frac{1}{N^2} tr\{X_n T_{1n}^{l+1} X_n^* (\forall_n - zI)^{-1} (\forall_n - z_0)^{-2} T_{2n}| \\
= \quad & |\frac{1}{N^2} tr\{X_n T_{1n}^{l+1} X_n^* T_{2n}^{1/2} G(z) [G(z_0)]^2 T_{2n}^{1/2}\}| \\
\leq \quad & \frac{\delta \tau^{l+2}}{v v_0^2}
\end{aligned}
$$

implies

$$
\begin{aligned}
& \frac{\frac{1}{N} tr\{\forall_n^{(l+1)} [(\forall_n - zI)^{-1} - (\forall_n - z_0 I)^{-1}]\}}{z - z_0} \\
= \quad & \frac{1}{N} tr\{\forall_n^{(l+1)} (\forall_n - zI)^{-1} (\forall_n - z_0 I)^{-1}\} \\
= \quad & \frac{1}{N} tr\{\forall_n^{(l+1)} (\forall_n - z_0)^{-2}\} \\
& +(z - z_0) \frac{1}{N} tr\{\forall_n^{(l+1)} (\forall_n - zI)^{-1} (\forall_n - z_0 I)^{-2}\} \\
\to \quad & \frac{1}{N} tr\{\forall_n^{(l+1)} (\forall_n - z_0 I)^{-2}\}.
\end{aligned}
$$

Hence $(1/N)tr\{\forall_n^{(l+1)}(\forall_n - zI)^{-1}\}$ is analytic on $\mathbb{C}^+$. Similarly, by

$$\begin{aligned}
& |\frac{1}{N}tr\{(\forall_n - zI)^{-1}(\forall_n - z_0)^{-2}\forall_n T_{2n}^l\}| \\
= & |\frac{1}{N^2}tr\{T_{2n}^{1/2}G(z)[G(z_0)]^2 T_{2n}^{1/2}X_n T_{1n} X_n^* T_{2n}^l\}| \\
\leq & \frac{\delta\tau^{l+2}}{vv_0^2},
\end{aligned}$$

we get $(1/N)tr\{(\forall_n - zI)^{-1}\forall_n T_{2n}^l\}$ is analytic on $\mathbb{C}^+$, and by

$$\begin{aligned}
& |\frac{1}{N}tr\{A_n^{(l+1)}(A_n - zI)^{-1}(A_n - z_0 I)^{-2}\}| \\
= & |-z^{-1}\frac{1}{N}tr\{A_n^{(l+1)}(A_n - z_0 I)^{-2}\} \\
& +z^{-1}\frac{1}{N}tr\{A_n^{(l+1)}A_n(A_n - zI)^{-1}(A_n - z_0 I)^{-2}\}|,
\end{aligned}$$

in which

$$\begin{aligned}
& |\frac{1}{N}tr\{A_n^{(l+1)}A_n(A_n - zI)^{-1}(A_n - z_0 I)^{-2}\}| \\
= & |\frac{1}{N}tr\{A_n^{(l+1)}T_{1n}B_n^{1/2}G(z)[G(z_0)]^2 B_n^{1/2}\}| \\
\leq & \frac{n}{N}\frac{\delta^2\tau^{l+4}}{vv_0^2},
\end{aligned}$$

we get $(1/N)tr\{A_n^{(l+1)}(A_n - zI)^{-1}\}$ is analytic on $\mathbb{C}^+$.

By using similar arguments, from Theorem 4.2.1, we can compute that

$$\begin{aligned}
& |\frac{1}{N}tr\{T_{1n}^l A_n^{k+2}(A_n - zI)^{-1}\}| \leq \frac{n\tau^l}{Nv}[\tau^2\delta]^{k+2}, \\
& |\frac{1}{N}tr\{\forall_n^{(l+1)}\forall_n^{k+1}(\forall_n - zI)^{-1}\}| \leq \frac{\tau^l}{v}[\tau^2\delta]^{k+2}, \\
& |\frac{1}{N}tr\{A_n^{(l+1)}A_n^{k+1}(A_n - zI)^{-1}\}| \leq \frac{n\tau^l}{Nv}[\tau^2\delta]^{k+2}, \\
& |\frac{1}{N}tr\{(\forall_n - zI)^{-1}\forall_n^{k+2}T_{2n}^l\}| \leq \frac{\tau^l}{v}[\tau^2\delta]^{k+2}.
\end{aligned}$$

Now we introduce the following formula which is a consequence of the resolvent identity

$$(A - zI)^{-1} = -\sum_{j=0}^{k} z^{-j-1}A^j + z^{-k-1}A^{k+1}(A - zI)^{-1},$$

where the second term on the right-hand side can also be $z^{-k-1}(A - zI)^{-1}A^{k+1}$.

Using this formula yields the following equalities:

$$\begin{aligned}
(1) \qquad & \frac{1}{N}tr\{T_{1n}^l A_n(A_n - zI)^{-1}\} \\
= \quad & -\sum_{j=1}^{k+1} z^{-j}\frac{1}{N}tr\{T_{1n}^l A_n^j\} + z^{-k-1}\frac{1}{N}tr\{T_{1n}^l A_n^{k+2}(A_n - zI)^{-1}\}, \\
(2) \qquad & \frac{1}{N}tr\{\forall_n^{(l+1)}(\forall_n - zI)^{-1}\} \\
= \quad & -\sum_{j=1}^{k+1} z^{-j}\frac{1}{N}tr\{\forall_n^{(l+1)}\forall_n^{j-1}\} + z^{-k-1}\frac{1}{N}tr\{\forall_n^{(l+1)}\forall_n^{k+1}(\forall_n - zI)^{-1}\}, \\
(3) \qquad & \frac{1}{N}tr\{(\forall_n - zI)^{-1}\forall_n T_{2n}^l\} \\
= \quad & -\sum_{j=1}^{k+1} z^{-j}\frac{1}{N}tr\{\forall_n^j T_{2n}^l\} + z^{-k-1}\frac{1}{N}tr\{(\forall_n - zI)^{-1}\forall_n^{k+2}T_{2n}^l\}, \\
(4) \qquad & \frac{1}{N}tr\{A_n^{(l+1)}(A_n - zI)^{-1}\} \\
= \quad & -\sum_{j=1}^{k+1} z^{-j}\frac{1}{N}tr\{A_n^{(l+1)}A_n^{j-1}\} + z^{-k-1}\frac{1}{N}tr\{A_n^{(l+1)}A_n^{k+1}(A_n - zI)^{-1}\}.
\end{aligned}$$

It is easy to check for each integer j,

$$tr(T_{1n}^l A_n^j) = tr(\forall_n^{(l+1)}\forall_n^{j-1}), \quad tr(\forall_n^j T_{2n}^l) = tr(A_n^{(l+1)}A_n^{j-1}).$$

It follows from (1) – (4) that

$$\begin{aligned}
& |\frac{1}{N}tr\{T_{1n}^l A_n(A_n - zI)^{-1}\} - \frac{1}{N}tr\{\forall_n^{(l+1)}(\forall_n - zI)^{-1}\}| \\
= \quad & |z^{-k-1}||\frac{1}{N}tr\{T_{1n}^l A_n^{k+2}(A_n - zI)^{-1}\} - \frac{1}{N}tr\{\forall_n^{(l+1)}\forall_n^{k+1}(\forall_n - zI)^{-1}\}| \\
\leq \quad & (1 + \frac{n}{N})\tau^l[\frac{\tau^2\delta}{v}]^{k+2},
\end{aligned}$$

and

$$\begin{aligned}
& |\frac{1}{N}tr\{(\forall_n - zI)^{-1}\forall_n T_{2n}^l\} - \frac{1}{N}tr\{A_n^{(l+1)}(A_n - zI)^{-1}\}| \\
= & |z^{-k-1}||\frac{1}{N}tr\{(\forall_n - zI)^{-1}\forall_n^{k+2}T_{2n}^l\} - \frac{1}{N}tr\{A_n^{(l+1)}A_n^{k+1}(A_n - zI)^{-1}\}| \\
\leq & (1+\frac{n}{N})\tau^l[\frac{\tau^2\delta}{v}]^{k+2},
\end{aligned}$$

which hold for every integer $k \geq 1$. It therefore follows when $z \in \mathbb{C}^+$ is such that $Imz = v > \tau^2\delta$, (4.3.17) and (4.3.18) hold. In view of the previously shown result that the four functions involved in the two identities are all analytic on $\mathbb{C}^+$, we conclude the two identities hold for all $z \in \mathbb{C}^+$. □

The special case of $l = 1$ of Lemma 4.3.7 gives us the following result.

Corollary 4.3.1. *For all $z \in \mathbb{C}^+$,*

$$zg_{1n}(z) = -\frac{1}{N}tr(T_{1n}) + \frac{1}{N}tr\{\forall_n^{(2)}(\forall_n - zI)^{-1}\}, \tag{4.3.19}$$

$$zg_{2n}(z) = -\frac{1}{N}tr(T_{2n}) + \frac{1}{N}tr\{A_n^{(2)}(A_n - zI)^{-1}\}. \tag{4.3.20}$$

4.4 Truncation and Centralization Treatment

The main target of the present section is to show the proof of Theorem 1.2.1 can be done with the conditions included in the following assumption added. Note that we have automatically taken the assumption that $H_1 \neq 1_{[0,\infty)}$ and $H_2 \neq 1_{[0,\infty)}$.

Assumption 4.4.1.

(i) $\|T_{1n}\|$ *and* $\|T_{2n}\|$ *are uniformly bounded for* n*, where* $\|\cdot\|$ *denotes the spectral norm of a matrix.*

(ii) $Ex_{ij} = 0$*,* $E|x_{ij}|^2 \le 1$*,* $|x_{ij}| \le \delta_n\sqrt{n}$*, with* $\delta_n \to 0$*,*

$$\frac{1}{\delta_n^2 nN}\sum_{ij}(1 - E|x_{ij}|^2) \to 0,$$

as $n \to \infty$*.*

(iii) T_{1n} *and* T_{2n} *are non-random.*

To verify the first two conditions in Assumption 4.4.1, suppose Theorem 1.2.1 is true for those matrices which satisfy both assumptions $(i)-(v)$ in Theorem 1.2.1 and conditions (i), (ii) in Assumption 4.4.1.

For any constant $\tau > 0$, replace in the spectral decompositions of T_{1n} and T_{2n}, those eigenvalues whose absolute values are bigger than τ with 0, and denote the resulting matrices by T_{1n}^τ and T_{2n}^τ respectively. Then for every $\tau \in \mathcal{T}$, where $\mathcal{T}$ is defined by

$$\begin{aligned} \mathcal{T} \equiv \{\tau > 0 : \quad & \tau \text{ is a continuity point of both } H_1 \text{ and } H_2 \\ & \text{and } -\tau \text{ is a continuity point of } H_1\}, \end{aligned} \tag{4.4.1}$$

it can be verified with probability one, $F^{T_{1n}^\tau}$, $F^{T_{2n}^\tau}$ and $F^{(T_{1n}^\tau)^2}$ all converge weakly. Denote their limits by H_1^τ, H_2^τ and $\tilde{H}_1^\tau$ respectively. Note that we have assumed $H_1 \neq 1_{[0,\infty)}$ and $H_2 \neq 1_{[0,\infty)}$. It follows that for all $\tau \in \mathcal{T}$ large, none of H_1^τ, H_2^τ and $\tilde{H}_1^\tau$ is equal to $1_{[0,\infty)}$.

Further let $\tilde{x}_{ij} = \hat{x}_{ij} - E\hat{x}_{ij}$ with $\hat{x}_{ij} = x_{ij}I_{(|x_{ij}|\leq \delta_n\sqrt{n})}$, where $\delta_n \downarrow 0$ is chosen such that condition (1.2.1) in Theorem 1.2.1 still holds with δ replaced with δ_n. Let $\hat{X}_n = [\hat{x}_{ij}]$ and $\tilde{X}_n = [\tilde{x}_{ij}]$ with (i,j)th entry $\hat{x}_{ij}$ and $\tilde{x}_{ij}$ respectively. Then it can be verified that $E\tilde{x}_{ij} = 0$, $|\tilde{x}_{ij}| \leq 2\delta_n\sqrt{n}$, $E|\tilde{x}_{ij}|^2 \leq 1$ and

$$\begin{aligned}
& \frac{1}{\delta_n^2 nN}\sum_{ij}(1 - E|\tilde{x}_{ij}|^2) \\
\leq \quad & \frac{2}{\delta_n^2 nN}\sum_{ij} E|x_{ij}|^2 I_{(|x_{ij}|>\delta_n\sqrt{n})} \\
\rightarrow \quad & 0
\end{aligned}$$

Also, $\frac{1}{N^2}tr(E\hat{X}_n E\hat{X}_n^*)^2 \rightarrow 0$ and $\frac{1}{N^3}tr(\tilde{X}_n^*\tilde{X}_n)^2 \rightarrow c + c^2$, almost surely.

Define $\forall_n^\tau = \frac{1}{N}T_{2n}^\tau X_n T_{1n}^\tau X_n^*$, $\hat{\forall}_n^\tau = \frac{1}{N}T_{2n}^\tau \hat{X}_n T_{1n}^\tau \hat{X}_n^*$ and $\tilde{\forall}_n^\tau = \frac{1}{N}T_{2n}^\tau \tilde{X}_n T_{1n}^\tau \tilde{X}_n^*$. Then $\tilde{\forall}_n^\tau$ satisfies assumptions $(i)-(v)$ in Theorem 1.2.1 and conditions (i), (ii) in Assumption 4.4.1. By hypothesis, Theorem 1.2.1 is true for $\tilde{\forall}_n^\tau$.

By the rank inequality in Lemma 2.1.1 and Bernstein's inequality in Lemma 2.1.3,

$$\begin{aligned}
& \|F^{\forall_n^\tau} - F^{\hat{\forall}_n^\tau}\| \leq \frac{2}{N}rank(X_n - \hat{X}_n) \\
\leq \quad & \frac{2}{N}\sum_{ij} I_{(|x_{ij}|>\delta_n\sqrt{n})} \rightarrow 0, \text{ almost surely.}
\end{aligned} \tag{4.4.2}$$

By the difference inequality in Lemma 2.1.2,

$$\begin{aligned}
& L^3(F^{\hat{\forall}_n^\tau}, F^{\tilde{\forall}_n^\tau}) \\
\leq \quad & \frac{\tau^2}{N^3}tr(\hat{X}_n T_{1n}^\tau \hat{X}_n^* - \tilde{X}_n T_{1n}^\tau \tilde{X}_n^*)^2 \\
\leq \quad & \frac{2\tau^4}{N^3}\{4\left(tr(E\hat{X}_n^* E\hat{X}_n)^2\right)^{1/2}\left(tr(\tilde{X}_n^*\tilde{X}_n)^2\right)^{1/2} + tr(E\hat{X}_n^* E\hat{X}_n)^2\} \\
\rightarrow \quad & 0, \text{ almost surely.}
\end{aligned} \tag{4.4.3}$$

Combining (4.4.2) and (4.4.3) gives with probability one,

$$L(F^{\forall_n^\tau}, F^{\tilde{\forall}_n^\tau}) \to 0. \tag{4.4.4}$$

This implies with probability one, $F^{\forall_n^\tau}$ and $F^{\tilde{\forall}_n^\tau}$ converges weakly to the same limit. In another word, Theorem 1.2.1 must also be true for the matrices $\forall_n^\tau$.

Denote by $\underline{F}^\tau$ the limiting spectral distribution of $\forall_n^\tau$. Then by Theorem 1.2.1,

$$\begin{cases} \underline{s}_\tau(z) &= -z^{-1}(1-c) - z^{-1}c\int \frac{1}{1+g_2^\tau(z)x} dH_1^\tau(x), \\ \underline{s}_\tau(z) &= -z^{-1}\int \frac{1}{1+g_1^\tau(z)y} dH_2^\tau(y), \\ \underline{s}_\tau(z) &= -z^{-1} - g_1^\tau(z) g_2^\tau(z), \end{cases} \tag{4.4.5}$$

where $\underline{s}_\tau(z)$ denotes the Stieltjes transform of $\underline{F}^\tau$, $g_1^\tau(z)$ and $g_2^\tau(z)$ are the two functions associated with $\underline{s}_\tau(z)$.

Let $\tau \to \infty$ only through points in the set $\mathcal{T}$. Then we have

$$\liminf_{\tau\to\infty} Im g_2^\tau(z) > 0, \quad \liminf_{\tau\to\infty} Im(z g_1^\tau(z)) > 0. \tag{4.4.6}$$

For the integrality, the proof of this fact will be presented later in the present section. This implies, by the third equation in (4.4.5) and the fact $|\underline{s}_\tau(z)| \le \frac{1}{v}$,

$$\limsup_{\tau\to\infty} |g_1^\tau(z)| < \infty \quad \text{and} \quad \limsup_{\tau\to\infty} |g_2^\tau(z)| < \infty. \tag{4.4.7}$$

Thus for a given $z \in \mathbb{C}^+$, $\{(\underline{s}_\tau(z), g_1^\tau(z), g_2^\tau(z))\}$ is bounded for all large $\tau \in \mathcal{T}$.

Consider any subsequence $\{(\underline{s}_{\tau_m}(z), g_1^{\tau_m}(z), g_2^{\tau_m}(z))\} \to (\underline{s}(z), g_1(z), g_2(z))$. Then (4.4.6) implies that $Im(z g_1(z)) > 0$ and $Im g_2(z) > 0$. Let $y \equiv \frac{\limsup_{\tau\to\infty} |g_2^\tau(z)|}{\liminf_{\tau\to\infty} Im g_2^\tau(z)}$. Then

$$\limsup_{\tau\to\infty} |\frac{1}{1+g_2^\tau(z)x}| \le y < \infty.$$

By the dominated convergence theorem, it follows

$$\lim_{\tau_m\to\infty}\int\frac{1}{1+g_2^{\tau_m}(z)x}dH_1(x)=\int\frac{1}{1+g_2(z)x}dH_1(x). \tag{4.4.8}$$

It can be verified

$$\begin{aligned}H_1^\tau(x) &= 1_{(\tau,\infty)}(x)+(H_1(x)+1-H_1(\tau))1_{[0,\tau]}(x)\\ &\quad +(H_1(x)-H_1(-\tau))1_{[-\tau,0)}(x),\end{aligned}$$

with $H_1^\tau\{0\}=H_1\{0\}+H_1\{(-\infty,-\tau)\cup(\tau,\infty)\}$, so that

$$\begin{aligned}&|\int\frac{1}{1+g_2^{\tau_m}(z)x}(dH_1^{\tau_m}(x)-dH_1(x))|\\ =\ &|(H_1^{\tau_m}\{0\}-H_1\{0\})+\int_{[-\tau_m,0)\cup(0,\tau_m]}\frac{1}{1+g_2^{\tau_m}(z)x}(dH_1^{\tau_m}(x)-dH_1(x))\\ &-\int_{(-\infty,-\tau_m)\cup(\tau_m,\infty)}\frac{1}{1+g_2^{\tau_m}(z)x}dH_1(x)|\\ =\ &|H_1\{(-\infty,-\tau_m)\cup(\tau_m,\infty)\}-\int_{(-\infty,-\tau_m)\cup(\tau_m,\infty)}\frac{1}{1+g_2^{\tau_m}(z)x}dH_1(x)|\\ =\ &|\int_{(-\infty,-\tau_m)\cup(\tau_m,\infty)}\left(1-\frac{1}{1+g_2^{\tau_m}(z)x}\right)dH_1(x)|\\ \leq\ &(1+y)H_1\{(-\infty,-\tau_m)\cup(\tau_m,\infty)\}\\ \to\ &0.\end{aligned} \tag{4.4.9}$$

Combining (4.4.8) and (4.4.9) gives at once

$$\lim_{\tau_m\to\infty}\int\frac{1}{1+g_2^{\tau_m}(z)x}dH_1^{\tau_m}(x)=\int\frac{1}{1+g_2(z)x}dH_1(x). \tag{4.4.10}$$

Now we prove, by the same type of argument,

$$\lim_{\tau_m\to\infty}\int\frac{1}{z+zg_1^{\tau_m}(z)y}dH_2^{\tau_m}(y)=\int\frac{1}{z+zg_1(z)y}dH_2(y). \tag{4.4.11}$$

In this case, we have, since $Im(zg_1^{\tau_m}(z))\geq 0$, $|\frac{1}{z+zg_1^{\tau_m}(z)y}|\leq\frac{1}{v}$, for $y\in[0,\infty)$ (recall that H_2^τ is supported on $[0,\infty)$). Then by the dominated convergence theorem,

$$\lim_{\tau_m\to\infty}\int\frac{1}{z+zg_1^{\tau_m}(z)y}dH_2(y)=\int\frac{1}{z+zg_1(z)y}dH_2(y). \tag{4.4.12}$$

Again, it can be shown

$$H_2^\tau(y) = 1_{(\tau,\infty)}(y) + (H_2(y) + 1 - H_2(\tau))1_{[0,\tau]}(y),$$

with $H_2^\tau\{0\} - H_2\{0\} = H_2\{(\tau,\infty)\}$, so that

$$\begin{aligned} & |\int \frac{1}{z + zg_1^{\tau_m}(z)y}(dH_2^{\tau_m}(y) - dH_2(y))| \\ = & |\int_{(\tau_m,\infty)} \left(\frac{1}{z} - \frac{1}{z + zg_1^{\tau_m}(z)y}\right) dH_2(y)| \\ \leq & \frac{2}{v} H_2\{(\tau_m,\infty)\} \\ \to & 0. \end{aligned} \qquad (4.4.13)$$

Combining (4.4.12) and (4.4.13) gives (4.4.11) immediately.

From the three equations of (4.4.5), (4.4.10) and (4.4.11), we see $(\underline{s}(z), g_1(z), g_2(z))$ satisfies (1.2.2). Since it is the limit of an arbitrarily chosen convergent subsequence and is a point in the set (1.2.3), by Lemma 4.3.3, we get

$$\lim_{\tau\to\infty} (\underline{s}_\tau(z), g_1^\tau(z), g_2^\tau(z)) = (\underline{s}(z), g_1(z), g_2(z)). \qquad (4.4.14)$$

By the rank inequality,

$$\begin{aligned} \|F^{\forall_n} - F^{\forall_n^\tau}\| & \leq \frac{1}{N}(rank(T_{1n} - T_{1n}^\tau) + 2 \times rank(T_{2n} - T_{2n}^\tau)) \\ & = \frac{n}{N} F^{T_{1n}}\{(-\infty,-\tau) \cup (\tau,\infty)\} + 2F^{T_{2n}}\{(\tau,\infty)\}, \end{aligned}$$

and so almost surely $\limsup_{\tau\to\infty} \limsup_{n\to\infty} \|F^{\forall_n} - F^{\forall_n^\tau}\| = 0$. Let

$$\begin{aligned} \theta_M & \equiv 2H_2\{(M^{\frac{1}{3}},\infty)\} + 2cF_{M-P}\{(M^{\frac{1}{3}},\infty)\} \\ & +cH_1\{(-\infty,-M^{\frac{1}{3}}) \cup (M^{\frac{1}{3}},\infty)\}. \end{aligned}$$

Then as $M \to \infty$, $\theta_M \to 0$. From Lemma 4.3.1 (taking $M_1 = M_2 = M_3 = M^{\frac{1}{3}}$), one gets almost surely

$$\limsup_{n\to\infty} F^{\forall_n}\{(-\infty, -M) \cup (M, \infty)\} \le \theta_M,$$

$$\limsup_{\tau\to\infty} \limsup_{n\to\infty} F^{\forall_n^\tau}\{(-\infty, -M) \cup (M, \infty)\} \le \theta_M.$$

It follows then, with the aid of integration by parts,

$$\begin{aligned}
& \limsup_{\tau\to\infty} \limsup_{n\to\infty} |s_{F^{\forall_n}}(z) - s_{F^{\forall_n^\tau}}(z)| \\
= & \limsup_{\tau\to\infty} \limsup_{n\to\infty} |\int_{[-M,M]} \frac{1}{x-z} d(F^{\forall_n}(x) - F^{\forall_n^\tau}(x)) \\
& + \int_{(-\infty,-M)\cup(M,\infty)} \frac{1}{x-z} d(F^{\forall_n}(x) - F^{\forall_n^\tau}(x))| \\
\le & \limsup_{\tau\to\infty} \limsup_{n\to\infty} |\frac{1}{M-z}(F^{\forall_n}(M) - F^{\forall_n^\tau}(M)) \\
& - \frac{1}{-M-z}(F^{\forall_n}(-M) - F^{\forall_n^\tau}(-M)) \\
& + \int_{[-M,M]} \frac{1}{(x-z)^2}(F^{\forall_n}(x) - F^{\forall_n^\tau}(x))dx| + \frac{2}{v}\theta_M \\
\le & \left(\frac{2}{v} + \frac{2M}{v^2}\right) \limsup_{\tau\to\infty} \limsup_{n\to\infty} \|F^{\forall_n} - F^{\forall_n^\tau}\| + \frac{2}{v}\theta_M \\
= & \frac{2}{v}\theta_M, \quad \text{for all } M > 0.
\end{aligned}$$

Hence we get $\limsup_{\tau\to\infty} \limsup_{n\to\infty} |s_{F^{\forall_n}}(z) - s_{F^{\forall_n^\tau}}(z)| = 0$.

Finally, we obtain with probability one,

$$\begin{aligned}
& \limsup_{n\to\infty} |s_{F^{\forall_n}}(z) - \underline{s}(z)| \\
\le & \limsup_{\tau\to\infty} \limsup_{n\to\infty} |s_{F^{\forall_n}}(z) - s_{F^{\forall_n^\tau}}(z)| + \limsup_{\tau\to\infty} \limsup_{n\to\infty} |s_{F^{\forall_n^\tau}}(z) - \underline{s}_\tau(z)| \\
& + \limsup_{\tau\to\infty} |\underline{s}_\tau(z) - \underline{s}(z)| \\
= & 0.
\end{aligned}$$

By Theorem 2.3.9, since we have shown $(\underline{s}(z), g_1(z), g_2(z))$ satisfies (1.2.2), it then follows Theorem 1.2.1 holds for the matrix $\forall_n$. Thus to show Theorem 1.2.1, without loss of generality, we may assume conditions (i), (ii) of Assumption 4.4.1 hold.

Let us now consider condition (iii) in Assumption 4.4.1. Suppose Theorem 1.2.1 is true for all those matrices which satisfy, besides assumptions $(i)-(v)$ in Theorem 1.2.1, also condition $(i)-(iii)$ in Assumption 4.4.1. Consider the matrix $\forall_n$ which satisfy assumptions $(i)-(v)$ in the theorem and conditions $(i)-(ii)$ in Assumption 4.4.1.

By assumption (iii) of Theorem 1.2.1, there exists a subspace $\Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$ such that for each $\omega \in \Omega_0$, $F^{T_{1n}(\omega)} \to H_1$ and $F^{T_{2n}(\omega)} \to H_2$. Define

$$\forall_n^\omega = \frac{1}{N} T_{2n}(\omega) X_n T_{1n}(\omega) X_n^*, \quad \text{for each } \omega \in \Omega,$$

where for an arbitrarily given ω, $T_{1n}(\omega)$ and $T_{2n}(\omega)$ are the observation of the matrices T_{1n} and T_{2n} at the basic element ω of the probability space Ω. By hypothesis, for every $\omega \in \Omega_0$, Theorem 1.2.1 holds for $\forall_n^\omega$. Then with probability one $F^{\forall_n^\omega}$ converges weakly to a non-random limit whose Stieltjes transform satisfies (1.2.2). Since H_1 and H_2 are the limits of $F^{T_{1n}(\omega)}$ and $F^{T_{2n}(\omega)}$ for every $\omega \in \Omega_0$, by Lemma 4.3.3 and the inversion formula of Theorem 2.3.1, for all $\omega \in \Omega_0$, $F^{\forall_n^\omega}$ converge to the same limit, which we now denote by $\underline{F}$. It follows that

$$Es_{F^{\forall_n^\omega}}(z) \to s_{\underline{F}}(z), \quad \text{for any } \omega \in \Omega_0, \tag{4.4.15}$$

and from Lemma 4.5.19 of the present chapter

$$E|s_{F^{\forall_n^\omega}}(z) - Es_{F^{\forall_n^\omega}}(z)|^4 \le Kn^{-2+\delta},$$

where $\delta > 0$ can be arbitrarily small, K is a constant depending only on δ, τ and $v = Im(z)$.

Note that since X_n is independent of T_{1n}, T_{2n}, we have for any $\omega \in \Omega$

$$E(s_{F^{\forall_n}}(z)|T_{1n} = T_{1n}(\omega), T_{2n} = T_{2n}(\omega)) = Es_{F^{\forall_n^\omega}}(z) \tag{4.4.16}$$

and

$$\begin{aligned}
& E|s_{F^{\forall_n}}(z) - E(s_{F^{\forall_n}}(z)|T_{1n} = T_{1n}(\omega), T_{2n} = T_{2n}(\omega))|^4 \\
= \; & E\left\{E\left(|s_{F^{\forall_n}}(z) - E(s_{F^{\forall_n}}(z)|T_{1n} = T_{1n}(\omega), T_{2n} = T_{2n}(\omega))|^4\right.\right. \\
& \left.\left.|T_{1n} = T_{1n}(\omega), T_{2n} = T_{2n}(\omega)\right)\right\} \\
= \; & E\left(E|s_{F^{\forall_n^\omega}}(z) - Es_{F^{\forall_n^\omega}}(z)|^4\right) \\
= \; & \int_{\Omega_0} E|s_{F^{\forall_n^\omega}}(z) - Es_{F^{\forall_n^\omega}}(z)|^4 dP(\omega) \\
\leq \; & Kn^{-2+\delta}.
\end{aligned} \tag{4.4.17}$$

By (4.4.15), (4.4.16), and $P(\Omega_0) = 1$, $E(s_{F^{\forall_n}}(z)|T_{1n} = T_{1n}(\omega), T_{2n} = T_{2n}(\omega)) \to s_{\underline{F}}(z)$ almost surely. From (4.4.17), $s_{F^{\forall_n}}(z) - E(s_{F^{\forall_n}}(z)|T_{1n} = T_{1n}(\omega), T_{2n} = T_{2n}(\omega)) \to 0$ almost surely. Thus, $s_{F^{\forall_n}}(z) \to s_{\underline{F}}(z)$ almost surely. That is, Theorem 1.2.1 must hold for $\forall_n$. Thus, to prove Theorem 1.2.1, without loss of generality, we may further assume condition (iii) of Assumption 4.4.1 holds. This completes our verification of the sufficiency of Assumption 4.4.1.

At the end of the present section, we prove the following result, which has played role in proving (4.4.6) and Lemma 4.3.5. Corresponding to $\tilde{\forall}_n^\tau$, let us define

$\tilde{A}_n^\tau = (1/N)T_{1n}^\tau \tilde{X}_n^* T_{2n}^\tau \tilde{X}_n$. Let

$$\tilde{g}_{1n}^\tau(z) \equiv (1/N)tr\{(\tilde{A}_n^\tau - zI)^{-1}T_{1n}^\tau\}, \quad \tilde{g}_{2n}^\tau(z) \equiv \frac{1}{N}tr\{(\tilde{\forall}_n^\tau - zI)^{-1}T_{2n}^\tau\}.$$

Proposition 4.3.1. *Under the assumption that* $H_1 \neq 1_{[0,\infty)}$ *and* $H_2 \neq 1_{[0,\infty)}$,

(1) $\lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty} F^{D_n}\{(\varepsilon,\infty)\} > 0$, *where* $D_n = (1/N)T_{1n}X_n^*T_{2n}X_nT_{1n}$.

(2) *Let* $\tau \to \infty$ *through points in the set* $\mathcal{T}$. *Then for any* $z \in \mathbb{C}^+$, *with probability* 1,

$$\delta_{2z} \equiv \liminf_{\tau\to\infty}\liminf_{n\to\infty} Im\tilde{g}_{2n}^\tau(z) > 0. \tag{4.4.18}$$

$$\delta_{1z} \equiv \liminf_{\tau\to\infty}\liminf_{n\to\infty} Im(z\tilde{g}_{1n}^\tau(z)) > 0. \tag{4.4.19}$$

Proof. Let $\tilde{D}_n = (1/N)T_{1n}\tilde{X}_n^*T_{2n}\tilde{X}_nT_{1n}$. Then by the proof of (4.4.4), we may in parallel obtain with probability one $L(F^{D_n}, F^{\tilde{D}_n}) \to 0$. This implies with probability one for any $\varepsilon' < \varepsilon$,

$$\liminf_{n\to\infty} F^{D_n}\{(\varepsilon',\infty)\} \geq \liminf_{n\to\infty} F^{\tilde{D}_n}\{(\varepsilon,\infty)\},$$

which implies with probability one

$$\lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty} F^{D_n}\{(\varepsilon,\infty)\} \geq \lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty} F^{\tilde{D}_n}\{(\varepsilon,\infty)\}.$$

By the symmetry of F^{D_n} and $F^{\tilde{D}_n}$, we then get with probability one,

$$\lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty} F^{D_n}\{(\varepsilon,\infty)\} = \lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty} F^{\tilde{D}_n}\{(\varepsilon,\infty)\}.$$

Moreover, if we define $\tilde{D}_n^\tau = (1/N)T_{1n}^\tau\tilde{X}_n^*T_{2n}^\tau\tilde{X}_nT_{1n}^\tau$, then

$$\liminf_{n\to\infty} F^{\tilde{D}_n}\{(\varepsilon,\infty)\} \geq \liminf_{n\to\infty} F^{\tilde{D}_n^\tau}\{(\varepsilon,\infty)\}.$$

Since we have asserted for all large $\tau \in \mathcal{T}$, none of the limiting spectral distributions of $F^{T_{1n}^{\tau}}$ and $F^{T_{2n}^{\tau}}$ is equal to $1_{[0,\infty)}$. We therefore claim that in the proof of (1), without loss of generality, we may assume there is a constant $\tau > 0$ bigger than the maximum of $\|T_{1n}\|$ and $\|T_{2n}\|$. It suffices, with this assumption, to show $\lim_{\varepsilon\downarrow 0}\liminf_{n\to\infty} F^{\tilde{D}_n}\{(\varepsilon,\infty)\} > 0$. Suppose not. Then there must exists some subsequence $\{n_m\}$ on which $\{F^{\tilde{D}_{n_m}}\}$ converges weakly to $1_{[0,\infty)}$.

Now let us define $h_n(z) = 1 + z\frac{1}{n}tr(\tilde{D}_n - zI)^{-1}$. Then we must have $h_{n_m}(z) \to 0$ for every $z \in \mathbb{C}^+$. By using the resolvent identity,

$$(iv)h_n(iv) = -\frac{1}{n}tr\tilde{D}_n + \frac{1}{n}tr\{\tilde{D}_n^2(\tilde{D}_n - ivI)^{-1}\},$$

in which almost surely $(1/n)tr\tilde{D}_n \to \int x^2 dH_1(x)\int y dH_2(y)$ and by Lemma 4.2.1(2)

$$|\frac{1}{n}tr\{\tilde{D}_n^2(\tilde{D}_n - ivI)^{-1}\}| \le \frac{\tau^6}{vnN^2}tr(\tilde{X}_n^*\tilde{X}_n)^2 \to \tau^6(1+c)/v.$$

It follows for all large v, by the assumption $H_1 \neq 1_{[0,\infty)}$ and $H_2 \neq 1_{[0,\infty)}$,

$$\begin{aligned} -v\liminf_{n\to\infty} Imh_n(iv) &= \limsup_{n\to\infty} Re(ivh_n(iv)) \\ &\le -\int x^2 dH_1(x)\int y dH_2(y) + \tau^6(1+c)/v < 0. \end{aligned}$$

Hence we get for all large v, $\liminf_{n\to\infty} Imh_n(iv) > 0$. This contradicts with the previously claimed fact $h_{n_m}(z) \to 0$ for every $z \in \mathbb{C}^+$. The proof of (1) is complete.

Let us proceed into the proof of (2). To show (4.4.18), we use (4.3.14). Then we have

$$Im\tilde{g}_{2n}^{\tau}(z) \ge \varepsilon c_M(z)(F^{T_{2n}^{\tau}}\{(\varepsilon,\infty)\} - F^{\tilde{\forall}_n^{\tau}}\{(-\infty,-M)\cup(M,\infty)\}).$$

From Lemma 4.3.1, choosing $M_1 = M_2 = M_3 = M^{1/3}$, we have

$$
\begin{aligned}
& F^{\tilde{\forall}^\tau_n}\{(-\infty, -M) \cup (M, \infty)\} \\
\leq \; & 2F^{T^\tau_{2n}}\{(M^{1/3}, \infty)\} + 2\frac{n}{N} F^{\tilde{S}_n}\{(M^{1/3}, \infty)\} + \frac{n}{N} F^{\sqrt{(T^\tau_{1n})^2}}\{(M^{1/3}, \infty)\} \\
\leq \; & 2F^{T_{2n}}\{(M^{1/3}, \infty)\} + 2\frac{n}{N} F^{\tilde{S}_n}\{(M^{1/3}, \infty)\} + \frac{n}{N} F^{\sqrt{(T_{1n})^2}}\{(M^{1/3}, \infty)\} \\
\equiv \; & \epsilon_n(M).
\end{aligned}
$$

Hence with probability 1,

$$
\begin{aligned}
& \lim_{M\to\infty} \limsup_{\tau\to\infty} \limsup_{n\to\infty} F^{\tilde{\forall}^\tau_n}\{(-\infty, -M) \cup (M, \infty)\} \\
\leq \; & \lim_{M\to\infty} \limsup_{n\to\infty} \epsilon_n(M) \\
\to \; & 0.
\end{aligned}
$$

Letting $\varepsilon \downarrow 0$ only through continuity points of H_2, noting that $H_2 \neq 1_{[0,\infty)}$, we have with probability 1,

$$
\begin{aligned}
\lim_{\varepsilon\downarrow 0} \liminf_{\tau\to\infty} \liminf_{n\to\infty} F^{T^\tau_{2n}}\{(\varepsilon, \infty)\} &= \lim_{\varepsilon\downarrow 0} \liminf_{\tau\to\infty} H^\tau_2\{(\varepsilon, \infty)\} \\
&= \lim_{\varepsilon\downarrow 0} H_2\{(\varepsilon, \infty)\} > 0.
\end{aligned}
$$

Let Ω be the subspace with probability 1 such that for every $\omega \in \Omega$, both of the above two relations hold. Then for each $\omega \in \Omega$ fixed, we firstly choose $\varepsilon > 0$ such that

$$\liminf_{\tau\to\infty} \liminf_{n\to\infty} F^{T^\tau_{2n}}\{(\varepsilon, \infty)\} = \varepsilon' > 0.$$

Then we choose M such that

$$\limsup_{\tau\to\infty} \limsup_{n\to\infty} F^{\forall^\tau_n}\{(-\infty, -M) \cup (M, \infty)\} < \frac{\varepsilon'}{2}.$$

It follows then

$$\begin{aligned}\delta_{2z} &\geq \varepsilon c_M(z)\Big(\liminf_{\tau\to\infty}\liminf_{n\to\infty} F^{T_{2n}^{\tau}}\{(\varepsilon,\infty)\} \\ &\quad - \limsup_{n\to\infty}\limsup_{n\to\infty} F^{\tilde{\forall}_n}\{(-\infty,-M)\cup(M,\infty)\}\Big) \\ &> \varepsilon c_M(z)\frac{\varepsilon'}{2}.\end{aligned}$$

To show (4.4.19), we use (4.3.16). Then we get

$$\begin{aligned}& Im(z\tilde{g}_{1n}^{\tau}(z)) \\ \geq\ & \varepsilon c_M(z)\frac{n}{N}\left(F^{\frac{1}{N}T_{1n}^{\tau}\tilde{X}_n^{*}T_{2n}^{\tau}\tilde{X}_nT_{1n}^{\tau}}\{(\varepsilon,\infty)\} - F^{\tilde{A}_n^{\tau}}\{(-\infty,-M)\cup(M,\infty)\}\right)\end{aligned}$$

From the proof of (1) in the present proposition,

$$\lim_{\varepsilon\downarrow 0}\liminf_{\tau\to\infty}\liminf_{n\to\infty} F^{\frac{1}{N}T_{1n}^{\tau}\tilde{X}_n^{*}T_{2n}^{\tau}\tilde{X}_nT_{1n}^{\tau}}\{(\varepsilon,\infty)\} > 0.$$

Thus by following the same argument, (4.4.19) is proved. □

Finally, we need to point out that (4.4.6) is a consequence of (4.4.18) and (4.4.19) in this proposition. However, to see this, we need to use one result in the last section of the present chapter, *i.e.* Corollary 4.6.2. All results of the last section are to be derived under the assumptions in both Theorem 1.2.1 and Assumption 4.4.1, all of which are satisfied by the matrices $\tilde{\forall}_n^{\tau}$. Hence by Corollary 4.6.2, almost surely, $\tilde{g}_{1n}^{\tau}(z) \to g_1^{\tau}(z)$ and $\tilde{g}_{2n}^{\tau}(z) \to g_2^{\tau}(z)$. Therefore, (4.4.18) and (4.4.19) yields (4.4.6).

Due to the work in the present section, in the next section, assuming Assumption 4.4.1 to be true, we shall focus on constructing some useful bounds for the quantities involved in relations (4.1.8) and (4.1.12).

4.5 Construction of Bounds for Quantities Involved in the Main Relations

At the first place, we clarify all derivations in the present section will be done under, besides assumptions $(i)-(v)$ in Theorem 1.2.1, also conditions $(i)-(iii)$ in Assumption 4.4.1.

As we have indicated earlier, the asymptotic behaviors of relations (4.1.8) and (4.1.12) will lead us to the answer of our question. In this section, we shall then deal with constructing bounds appropriate for estimating those quantities involved in these relations as well as in (4.1.10), (4.1.11) and (4.1.14), (4.1.15).

Let us begin by showing two basic results. The estimates they state are conceivable. As for their proof, a solution can be obtained from the proof of Theorem 2.10 in Bai(1999). However, since in that work a stronger conclusion is pursued, the proof needs a rather complicated combinatoric argument. Our result, focusing on providing suitable bounds, can be established by using only the induction method and some basic properties of matrices.

Lemma 4.5.8. *For any integer $\ell \geq 1$, if $\|T_n\|$ is bounded by τ, then there exists a constant K depending only on τ and ℓ such that*

$$|Etr(X_n T_n X_n^*)^\ell| \leq K n^{1+\ell}. \tag{4.5.1}$$

Proof. By using the fact $|Etr(A)| \leq \sqrt{n}\{Etr(AA^*)\}^{1/2}$ valid for $n \times n$ matrix

A, we have for any ℓ,

$$|Etr(X_nT_nX_n^*)^\ell| \le \sqrt{n}\{Etr(X_nT_nX_n^*)^{2\ell}\}^{1/2},$$

which implies we need only show that the asserted result holds for every even ℓ, say $\ell = 2v$. We first show for any positive integers a and v,

$$A(a,v) \equiv Etr([(X_nX_n^*)T_n]^v(X_n^*X_n)^a[T_n(X_n^*X_n)]^v) \le Kn^{1+a+2v}. \qquad (4.5.2)$$

When $v = 1$, by using the fact $|Etr(AB)| \le \{Etr(AA^*)\}^{1/2}\{Etr(BB^*)\}^{1/2}$,

$$\begin{aligned} A(a,1) &= Etr((X_nX_n^*)^aT_n(X_n^*X_n)^2T_n) \\ &\le \tau^2\{Etr(X_n^*X_n)^{2a}\}^{1/2}\{Etr(X_n^*X_n)^4\}^{1/2} \\ &\le Kn^{a+3}, \end{aligned}$$

and hence (4.5.2) holds for all a when $v = 1$. By the way of induction, when $v \ge 2$, the same type of argument yields

$$\begin{aligned} A(a,v) &= Etr((X_n^*X_n)^a[T_n(X_n^*X_n)]^{v-1}T_n(X_n^*X_n)^2[T_n(X_n^*X_n)]^{v-1}T_n) \\ &\le \tau^2\{A(2a,v-1)\}^{1/2}\{A(4,v-1)\}^{1/2} \\ &\le Kn^{1+a+2v}. \end{aligned}$$

Thus (4.5.2) is proved. Note that when $l = 2v$,

$$\begin{aligned} & Etr(X_nT_nX_n^*)^{2v} = Etr(T_nX_n^*X_n)^{2v} \\ \le\ & Etr(T_n[(X_n^*X_n)T_n]^{v-1}(X_n^*X_n)T_n(X_n^*X_n)[T_n(X_n^*X_n)]^{v-1}) \\ \le\ & \tau^2A(2,v-1) \le Kn^{1+2v}. \end{aligned}$$

Thus (4.5.1) is shown for the case when ℓ is even. $\square$

A consequence of (4.5.1) is that for any positive integer l, there exists a constant K depending only on τ, l, k such that

$$Etr(T_{1n}^k X_n^* X_n T_{1n}^k)^l \leq K n^{1+l}. \tag{4.5.3}$$

To see this, simply notice that the left-hand term is $Etr(X_n T_{1n}^{2k} X_n^*)^l$.

Lemma 4.5.9. *For any positive integer l, there exists a constant K depending only on τ, l, k such that*

$$Etr(T_{2n}^k (X_n T_{1n} X_n^*)^2 T_{2n}^k)^l \leq K n^{1+2l}. \tag{4.5.4}$$

Proof. By the same type of argument as used in proving (4.5.1), we only need to show (4.5.4) for the case of $l = 2v$. We first prove for all positive integers a and v,

$$\begin{aligned} B(a, v) &\equiv Etr((X_n T_{1n} X_n^*)^a [T_{2n}^{2k} (X_n T_{1n} X_n^*)^2]^v [(X_n T_{1n} X_n^*)^2 T_{2n}^{2k}]^v) \\ &\leq K n^{1+a+4v}. \end{aligned} \tag{4.5.5}$$

We again use the method of induction. When $v = 1$, from (4.5.1), we have

$$\begin{aligned} B(a, 1) &\leq \tau^{4k} \{Etr(X_n T_{1n} X_n^*)^{2a}\}^{1/2} \{Etr(X_n T_{1n} X_n^*)^8\}^{1/2} \\ &\leq K n^{5+a} = K n^{1+a+4v}. \end{aligned}$$

Thus (4.5.5) holds for all a when $v = 1$. When $v \geq 2$, by using (1) of Lemma 4.2.1, we have

$$\begin{aligned} B(a,v) &= Etr((X_nT_{1n}X_n^*)^a[T_{2n}^{2k}(X_nT_{1n}X_n^*)^2]^{v-1}T_{2n}^{2k} \\ &\quad (X_nT_{1n}X_n^*)^4[T_{2n}^{2k}(X_nT_{1n}X_n^*)^2]^{v-1}T_{2n}^{2k}) \\ &\leq \tau^{4k}\{B(2a,v-1)\}^{1/2}\{B(8,v-1)\}^{1/2} \\ &\leq Kn^{1+a+4v}. \end{aligned}$$

Thus (4.5.5) holds for all positive integers a and v.

Note that for $l = 2v$,

$$Etr(T_{2n}^k(X_nT_{1n}X_n^*)^2T_{2n}^k)^{2v} \leq \tau^{4k}B(4,v-1) \leq Kn^{1+4v} = Kn^{1+2l}.$$

This completes the proof. □

The variables concerned in the next lemma will make their appearance frequently in our latter arguments. For more convenient use, we regard the regular rule about them.

Lemma 4.5.10. *For each pair (i,j) and each positive integer k, let*

$$\alpha(i,j,k) = f_j'T_{1n}^kX_{ij}^*X_{ij}T_{1n}^kf_j, \qquad \alpha_{jj}(k) = f_j'T_{1n}^kX_n^*X_nT_{1n}^kf_j.$$

Moreover, let

$$\beta(i,j,k) = e_i'T_{2n}^kX_{ij}T_{1n}X_{ij}^*X_{ij}T_{1n}X_{ij}^*T_{2n}^ke_i,$$

$$\tilde{\beta}(i,j,k) = e_i'T_{2n}^kX_{ij}T_{1n}X_n^*X_nT_{1n}X_{ij}^*T_{2n}^ke_i,$$

$$\hat{\beta}(i,j,k) = e_i'T_{2n}^kX_nT_{1n}X_{ij}^*X_{ij}T_{1n}X_n^*T_{2n}^ke_i,$$

$$\beta_{ii}(k) = e_i'T_{2n}^kX_nT_{1n}X_n^*X_nT_{1n}X_n^*T_{2n}^ke_i.$$

Then there exists constant K depending only on τ and k but not on i and j such that

(1) $\alpha_{jj}(k) \leq K(|x_{ij}|^2 + \alpha(i,j,k))$.

(2) $\alpha(i,j,k) \leq K(|x_{ij}|^2 + \alpha_{jj}(k))$.

(3) $\tilde{\beta}(i,j,k) \leq K(|x_{ij}|^2\alpha(i,j,1) + \beta(i,j,k))$,

(4) $\hat{\beta}(i,j,k) \leq K(|x_{ij}|^2\alpha(i,j,1) + \beta(i,j,k))$,

(5) $\beta_{ii}(k) \leq K(|x_{ij}|^4 + |x_{ij}|^2\alpha(i,j,1) + \beta(i,j,k))$,

(6) $\beta(i,j,k) \leq K(|x_{ij}|^2\alpha(i,j,1) + \tilde{\beta}(i,j,k))$,

(7) $\tilde{\beta}(i,j,k) \leq K(|x_{ij}|^2\alpha_{jj}(1) + \beta_{ii}(k))$,

(8) $\beta(i,j,k) \leq K(|x_{ij}|^4 + |x_{ij}|^2\alpha_{jj}(1) + \beta_{ii}(k))$.

Proof. By using the expressions $X_n = X_{ij} + x_{ij}e_if_j'$ and $X_n^* = X_{ij}^* + \bar{x}_{ij}f_je_i'$, the relations $|xy| \leq (x^2+y^2)$ $(x,\ y \in \mathbb{R})$, $|a^*b| \leq (a^*a)^{1/2}(b^*b)^{1/2}$ (a, b complex vectors),

$$
\begin{aligned}
& |\alpha_{jj}(k) - \alpha(i,j,k)| \\
\leq\ & K(2|\zeta_{jj}^{(k)}||x_{ij}||e_i'X_{ij}T_{1n}^kf_j| + |\zeta_{jj}^{(k)}|^2|x_{ij}|^2) \\
\leq\ & K(|x_{ij}||e_i'X_{ij}T_{1n}^kf_j| + |x_{ij}|^2) \\
\leq\ & K(|x_{ij}|^2 + |e_i'X_{ij}T_{1n}^kf_j|^2) \\
\leq\ & K(|x_{ij}|^2 + f_j'T_{1n}^kX_{ij}^*X_{ij}T_{1n}^kf_j) \\
=\ & K(|x_{ij}|^2 + \alpha(i,j,k)),
\end{aligned}
$$

which implies (1). Similarly, using the expressions $X_{ij} = X_n - x_{ij}e_i f_j'$ and $X_{ij}^* = X_n^* - \bar{x}_{ij} f_j e_i'$ gives (2).

The same type of arguments can be used to establish (3) − (8). For instance, we have

$$\begin{aligned}
& |\hat{\beta}(i,j,k) - \beta(i,j,k)| \\
\leq\ & K(2|\xi_{ii}^{(k)}||x_{ij}||f_j' T_{1n} X_{ij}^* X_{ij} T_{1n} X_{ij}^* T_{2n}^k e_i| + |\xi_{ii}^{(k)}|^2 |x_{ij}|^2 f_j' T_{1n} X_{ij}^* X_{ij} T_{1n} f_j) \\
\leq\ & K(|x_{ij}||f_j' T_{1n} X_{ij}^* X_{ij} T_{1n} X_{ij}^* T_{2n}^k e_i| + |x_{ij}|^2 f_j' T_{1n} X_{ij}^* X_{ij} T_{1n} f_j) \\
\leq\ & K(|x_{ij}|^2 f_j' T_{1n} X_{ij}^* X_{ij} T_{1n} f_j + e_i' T_{2n}^k X_{ij} T_{1n} X_{ij}^* X_{ij} T_{1n} X_{ij}^* T_{2n}^k e_i) \\
=\ & K(|x_{ij}|^2 \alpha(i,j,1) + \beta(i,j,k)),
\end{aligned}$$

which implies (4) and

$$\begin{aligned}
& |\tilde{\beta}(i,j,k) - \beta(i,j,k)| \\
\leq\ & K(|x_{ij}||e_i' T_{2n}^k X_{ij} T_{1n} X_{ij}^* e_i||f_j' T_{1n} X_{ij}^* T_{2n}^k e_i| \\
& +|x_{ij}|^2 |e_i' T_{2n}^k X_{ij} T_{1n} f_j||f_j' T_{1n} X_{ij}^* T_{2n}^k e_i|) \\
\leq\ & K(|x_{ij}|^2 |f_j' T_{1n} X_{ij}^* T_{2n}^k e_i|^2 + |e_i' T_{2n}^k X_{ij} T_{1n} X_{ij}^* e_i|^2) \\
\leq\ & K(|x_{ij}|^2 \alpha(i,j,1) + \beta(i,j,k)),
\end{aligned}$$

which implies (3). Further, we have

$$\begin{aligned}
&|\beta_{ii}(k) - \tilde{\beta}(i,j,k)| \\
\leq\ & K(|x_{ij}||f_j'T_{1n}X_n^*X_nT_{1n}X_{ij}^*T_{2n}^ke_i| + |x_{ij}|^2f_j'T_{1n}X_n^*X_nT_{1n}f_j) \\
\leq\ & K(|x_{ij}|^2f_j'T_{1n}X_n^*X_nT_{1n}f_j + e_i'T_{2n}^kX_{ij}T_{1n}X_n^*X_nT_{1n}X_{ij}^*T_{2n}^ke_i) \\
=\ & K(|x_{ij}|^2\alpha_{jj}(1) + \tilde{\beta}(i,j,k)) \\
\leq\ & K(|x_{ij}|^4 + |x_{ij}|^2\alpha(i,j,1) + \tilde{\beta}(i,j,k)),
\end{aligned}$$

where in the last step we have used (1). This relation together with (3) imply (5).

With $X_{ij} = X_n - x_{ij}e_if_j'$ and $X_{ij}^* = X_n^* - \bar{x}_{ij}f_je_i'$, we have

$$\begin{aligned}
&|\beta(i,j,k) - \tilde{\beta}(i,j,k)| \\
\leq\ & K(|x_{ij}||e_i'T_{2n}^kX_{ij}T_{1n}X_n^*e_i||f_j'T_{1n}X_{ij}^*T_{2n}^ke_i| \\
&+|x_{ij}|^2|e_i'T_{2n}^kX_{ij}T_{1n}f_j|^2) \\
\leq\ & K(|x_{ij}|^2f_j'T_{1n}X_{ij}^*X_{ij}T_{1n}f_j + e_i'T_{2n}^kX_{ij}T_{1n}X_n^*X_nT_{1n}X_{ij}^*T_{2n}^ke_i) \\
=\ & K(|x_{ij}|^2\alpha(i,j,1) + \tilde{\beta}(i,j,k)) \\
\leq\ & K(|x_{ij}|^4 + |x_{ij}|^2\alpha_{jj}(1) + \tilde{\beta}(i,j,k))
\end{aligned}$$

and

$$\begin{aligned}
&|\tilde{\beta}(i,j,k) - \beta_{ii}(k)| \\
\leq\ & K(|x_{ij}||f_j'T_{1n}X_n^*X_nT_{1n}X_n^*T_{2n}^ke_i| + |x_{ij}|^2f_j'T_{1n}X_n^*X_nT_{1n}f_j) \\
\leq\ & K(|x_{ij}|^2\alpha_{jj}(1) + \beta_{ii}(k)),
\end{aligned}$$

that is, (6) and (7) hold. Combining (6) and (7) gives (8). From the derivations, we can see that K depends only on τ and k but not on i and j. $\square$

Lemma 4.5.11. *For each positive integer l, there exists a constant K depending on l, τ, k only, such that*

$$(1) E\sum_j [\alpha_{jj}(k)]^l \leq Kn^{1+l}. \quad (2) E\sum_i [\beta_{ii}(k)]^l \leq Kn^{1+2l}.$$
$$(3) E\sum_{ij} [\alpha(i,j,k)]^l \leq Kn^{2+l}. \ (4) E\sum_{ij} [\beta(i,j,k)]^l \leq Kn^{2+2l}.$$

Proof. Note that for any Hermitian $A = [a_{ij}]$,

$$\sum_i a_{ii}^{2l} \leq \sum_i [\lambda_i(A)]^{2l} = tr(A^{2l}),$$

since $f(x) = x^{2l}$ is convex. Therefore, from (4.5.3),

$$\begin{aligned} & E\sum_j [\alpha_{jj}(k)]^l \leq \sqrt{n}\{E\sum_j [\alpha_{jj}(k)]^{2l}\}^{1/2} \\ \leq\ & \sqrt{n}\{Etr(T_{1n}^k X_n^* X_n T_{1n}^k)^{2l}\}^{1/2} \\ \leq\ & Kn^{1+l}. \end{aligned}$$

Similarly, from (4.5.4),

$$\begin{aligned} & E\sum_i [\beta_{ii}(k)]^l \leq \sqrt{N}\{E\sum_i [\beta_{ii}(k)]^{2l}\}^{1/2} \\ \leq\ & \sqrt{N}\{Etr(T_{2n}^k (X_n T_{1n} X_n^*)^2 T_{2n}^k)^{2l}\}^{1/2} \\ \leq\ & Kn^{1+2l}. \end{aligned}$$

Thus far we have proved (1) and (2). Further, by (2) and (8) in Lemma 4.5.10, we have

$$\begin{aligned}
& E\sum_{ij}[\alpha(i,j,k)]^l \\
\leq\ & KE\sum_{ij}(|x_{ij}|^2+\alpha_{jj}(k))^l \\
\leq\ & KE\sum_{ij}(|x_{ij}|^{2l}+[\alpha_{jj}(k)]^l) \\
\leq\ & Kn^{2+l},
\end{aligned}$$

and

$$\begin{aligned}
& E\sum_{ij}[\beta(i,j,k)]^l \\
\leq\ & KE\sum_{ij}(|x_{ij}|^4+|x_{ij}|^2\alpha_{jj}(1)+\beta_{ii}(k))^l \\
\leq\ & KE\sum_{ij}(|x_{ij}|^{4l}+|x_{ij}|^{2l}[\alpha_{jj}(1)]^l+[\beta_{ii}(k)]^l) \\
\leq\ & Kn^{2+2l}.
\end{aligned}$$

Thus (3) and (4) are also proved. $\square$

In the following, we come to the main purpose of the present section, that is, to define three types of bounds for use in later arguments. In the remainder of the present section, we shall take k to be a fixed positive integer. Also, in the sequel, we shall frequently use the fact that for any numbers $a_l \geq 0$ $(1 \leq l \leq L)$,

$$\left(\sum_{l=1}^{L} a_l\right)^{\frac{p}{2}} \leq L^{\frac{p}{2}-1}\sum_{l=1}^{L} a_l^{\frac{p}{2}}. \tag{4.5.6}$$

The first type bound is as follows.

Lemma 4.5.12. *For each pair (i,j), define*

$$\begin{aligned}\mathbf{B}(i,j) &= \sqrt{n}+\sum_{p=1}^{2} n^{-\frac{p-1}{2}}\left([\alpha(i,j,1)]^{\frac{p}{2}}+[\alpha(i,j,k)]^{\frac{p}{2}}\right)\\ &\quad+\sum_{p=1}^{2} n^{-p+\frac{1}{2}}\left([\beta(i,j,1)]^{\frac{p}{2}}+[\beta(i,j,k)]^{\frac{p}{2}}\right). \qquad (4.5.7)\end{aligned}$$

Then for any positive integer l, there exists a constant K depending only on τ, k and l such that

$$E\sum_{ij}[\mathbf{B}(i,j)]^{\frac{l}{2}} \leq Kn^{2+\frac{l}{4}}. \qquad (4.5.8)$$

Proof. By Hölder's inequality, we only need to prove

$$E\sum_{ij}[\mathbf{B}(i,j)]^{l} \leq Kn^{2+\frac{l}{2}}.$$

However, as an easy consequence of Lemma 4.5.11, we have

$$E\sum_{ij}\left(n^{-\frac{p-1}{2}}[\alpha(i,j,k)]^{\frac{p}{2}}\right)^{l} \leq Kn^{2+\frac{l}{2}},$$

and

$$E\sum_{ij}\left(n^{-p+\frac{1}{2}}[\beta(i,j,k)]^{\frac{p}{2}}\right)^{l} \leq Kn^{2+\frac{l}{2}}.$$

They give, by the fact in (4.5.6), immediately the result. □

We now use the bound defined in (4.5.7) to estimate the quantities appearing the those relations of our interest.

Lemma 4.5.13. *For the quantities* $\tilde{a}_{ji}^{(k)}$, $a_{ji}^{(k)}$, $\hat{a}_{ij}^{(1)}$, $\tilde{p}_{jj}$, p_{jj} *and* $d_{ii}^{(k)}$ *appearing the relations (4.1.8), (4.1.10) and (4.1.11), there exists constant* K *depending on* τ, v, k *only such that*

$$
\begin{aligned}
&|\tilde{a}_{ji}^{(k)}| \leq K\mathbf{B}(i,j),\ \ |a_{ji}^{(k)}| \leq K\mathbf{B}(i,j),\ \ |\hat{a}_{ij}^{(1)}| \leq K\mathbf{B}(i,j),\\
&|\tilde{p}_{jj}| \leq K\mathbf{B}(i,j)/\sqrt{n},\ \ |p_{jj}| \leq K\mathbf{B}(i,j)/\sqrt{n},\\
&|d_{ii}^{(k)}| \leq K\sqrt{n}\mathbf{B}(i,j).
\end{aligned}
$$

Proof. By using Lemma 4.2.1(6), Corollary 4.2.1(4) and Lemma 4.5.10,

$$
\begin{aligned}
|\tilde{a}_{ji}^{(k)}| \ &\leq \ \frac{1}{v}|f_j'T_{1n}X_{ij}^*T_{2n}^k e_i| + \frac{1}{Nv^2}(f_j'T_{1n}X_n^*T_{2n}X_nT_{1n}f_j)^{1/2}\\
&\quad\ (e_i'T_{2n}^kX_{ij}T_{1n}X_n^*T_{2n}X_nT_{1n}X_{ij}^*T_{2n}^k e_i)^{1/2}\\
&\leq \ K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{n}[\alpha_{jj}(1)]^{1/2}[\tilde{\beta}(i,j,k)]^{1/2}\}\\
&\leq \ K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}\alpha_{jj}(1) + \frac{1}{n\sqrt{n}}\tilde{\beta}(i,j,k)\}\\
&\leq \ K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}(|x_{ij}|^2 + \alpha(i,j,1))\\
&\quad\ +\frac{1}{n\sqrt{n}}(|x_{ij}|^2\alpha(i,j,1) + \beta(i,j,k))\}\\
&\leq \ K\{\delta_n\sqrt{n} + [\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}\alpha(i,j,1)\\
&\quad\ +\frac{\delta_n^2}{\sqrt{n}}\alpha(i,j,1) + \frac{1}{n\sqrt{n}}\beta(i,j,k)\}\\
&\leq \ K\{\sqrt{n} + [\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}\alpha(i,j,1) + \frac{1}{n\sqrt{n}}\beta(i,j,k)\}\\
&\leq \ K\mathbf{B}(i,j).
\end{aligned}
$$

In the above, we also frequently use the basic fact, $|xy| \leq (|x|^2 + |y|^2)/2$, valid for any numbers x and y. The constant K appearing all the way thus takes different value in its different appearance. However, as claimed its values only depend on τ, v and k.

The same type of arguments can be developed for all the other quantities in the lemma. In details, for $a_{ji}^{(k)}$, we have

$$
\begin{aligned}
|a_{ji}^{(k)}| &\leq \frac{1}{v}|f_j' T_{1n} X_{ij}^* T_{2n}^k e_i| + \frac{1}{Nv^2}(f_j' T_{1n} X_{ij}^* T_{2n} X_{ij} T_{1n} f_j)^{1/2} \\
&\quad (e_i' T_{2n}^k X_{ij} T_{1n} X_{ij}^* T_{2n} X_{ij} T_{1n} X_{ij}^* T_{2n}^k e_i)^{1/2} \\
&\leq \frac{\tau^k}{v}[\alpha(i,j,1)]^{1/2} + \frac{\tau}{Nv^2}[\alpha(i,j,1)]^{1/2}[\beta(i,j,k)]^{1/2} \\
&\leq K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{n}[\alpha(i,j,1)]^{1/2}[\beta(i,j,k)]^{1/2}\} \\
&\leq K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}\alpha(i,j,1) + \frac{1}{n\sqrt{n}}\beta(i,j,k)\} \\
&\leq K\mathbf{B}(i,j).
\end{aligned}
$$

For $\hat{a}_{ij}^{(1)}$,

$$
\begin{aligned}
|\hat{a}_{ij}^{(1)}| &\leq \frac{1}{v}|e_i' T_{2n} X_{ij} T_{1n} f_j| + \frac{1}{Nv^2}(f_j' T_{1n} X_{ij}^* T_{2n} X_{ij} T_{1n} f_j)^{1/2} \\
&\quad (e_i' T_{2n} X_{ij} T_{1n} X_{ij}^* T_{2n} X_{ij} T_{1n} X_{ij}^* T_{2n} e_i)^{1/2} \\
&\leq K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{N}[\alpha(i,j,1)]^{1/2}[\beta(i,j,k)]^{1/2}\} \\
&\leq K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}\alpha(i,j,1) + \frac{1}{n\sqrt{n}}\beta(i,j,1)\} \\
&\leq K\mathbf{B}(i,j).
\end{aligned}
$$

For $\tilde{p}_{jj}$,

$$\begin{aligned}
|\tilde{p}_{jj}| &\leq \frac{1}{v}|f_j' T_{1n} f_j| + \frac{1}{Nv^2} f_j' T_{1n} X_n^* T_{2n} X_n T_{1n} f_j \\
&= K(1 + \frac{1}{n}\alpha_{jj}(1)) \\
&\leq K(1 + \frac{1}{n}(|x_{ij}|^2 + \alpha(i,j,1))) \\
&\leq K(1 + \frac{1}{n}\alpha(i,j,1)) \\
&\leq K\mathbf{B}(i,j)/\sqrt{n}.
\end{aligned}$$

For p_{jj},

$$\begin{aligned}
|p_{jj}| &\leq \frac{1}{v}|f_j' T_{1n} f_j| + \frac{1}{Nv^2} f_j' T_{1n} X_{ij}^* T_{2n} X_{ij} T_{1n} f_j \\
&\leq K(1 + \frac{1}{n}\alpha(i,j,1)) \\
&\leq K\mathbf{B}(i,j)/\sqrt{n}.
\end{aligned}$$

For $d_{ii}^{(k)}$,

$$\begin{aligned}
|d_{ii}^{(k)}| &\leq \frac{1}{v}|e_i' T_{2n} X_{ij} T_{1n} X_{ij}^* T_{2n}^k e_i| + \frac{1}{Nv^2}(e_i' T_{2n} X_{ij} T_{1n} X_{ij}^* T_{2n} X_{ij} T_{1n} X_{ij}^* T_{2n} e_i)^{1/2} \\
&\quad (e_i' T_{2n}^k X_{ij} T_{1n} X_{ij}^* T_{2n} X_{ij} T_{1n} X_{ij}^* T_{2n}^k e_i)^{1/2} \\
&\leq K\{[\beta(i,j,1)]^{1/2} + \frac{1}{n}\beta(i,j,1) + \frac{1}{n}\beta(i,j,k)\} \\
&\leq K\sqrt{n}\mathbf{B}(i,j).
\end{aligned}$$

This completes the proof. □

Lemma 4.5.14. *For the quantities $\tilde{\sigma}_{ij}^{(k)}$, $\sigma_{ij}^{(k)}$, $\hat{\sigma}_{ji}^{(1)}$, $\tilde{q}_{ii}$, q_{ii} and $b_{jj}^{(k)}$ appearing the relations (4.1.12), (4.1.14) and (4.1.15), there exists constant K depending on*

τ, v, k only such that

$$|\tilde{\sigma}_{ij}^{(k)}| \le K\mathbf{B}(i,j),\ \ |\sigma_{ij}^{(k)}| \le K\mathbf{B}(i,j),\ \ |\hat{\sigma}_{ji}^{(1)}| \le K\mathbf{B}(i,j),$$

$$|\tilde{q}_{ii}| \le K \le K\mathbf{B}(i,j)/\sqrt{n},\ \ |q_{ii}| \le K \le K\mathbf{B}(i,j)/\sqrt{n},$$

$$|b_{jj}^{(k)}| \le K\sqrt{n}\mathbf{B}(i,j).$$

Proof. The proof is similar to that of the previous lemma. By using Lemma 4.2.1(6), Corollary 4.2.1(2) and Lemma 4.5.10, the absolute values of $\tilde{\sigma}_{ij}^{(k)}$ and $\sigma_{ij}^{(k)}$ are both bounded by $(\tau/v)[\alpha(i,j,k)]^{1/2}$ and hence by $K\mathbf{B}(i,j)$. The absolute values of $\hat{\sigma}_{ji}^{(1)}$ is bounded by $(\tau/v)[\alpha(i,j,1)]^{1/2}$ and hence by $K\mathbf{B}(i,j)$. The absolute values of $\tilde{q}_{ii}$ and q_{ii} are bounded by τ/v and hence by $K\mathbf{B}(i,j)/\sqrt{n}$. Finally, for $b_{jj}^{(k)}$,

$$\begin{aligned}
|b_{jj}^{(k)}| &\le \frac{1}{v}(f_j'T_{1n}X_{ij}^*T_{2n}X_{ij}T_{1n}f_j)^{1/2}(f_j'T_{1n}^kX_{ij}^*T_{2n}X_{ij}T_{1n}^kf_j)^{1/2} \\
&\le K(\alpha(i,j,1)+\alpha(i,j,k)) \\
&\le K\sqrt{n}\mathbf{B}(i,j).
\end{aligned}$$

The proof is complete. □

As a consequence of the above two lemmas, we have

Lemma 4.5.15.

$$\begin{aligned}
&(1) \qquad |\tilde{a}_{ji}^{(k)} - a_{ji}^{(k)}| \le K\frac{1}{n}|x_{ij}|[\mathbf{B}(\mathbf{i},\mathbf{j})]^2,\\
&(2) \qquad |\tilde{\sigma}_{ij}^{(k)} - \sigma_{ij}^{(k)}| \le K\frac{|x_{ij}|}{\sqrt{n}}\mathbf{B}(i,j) \le K\frac{|x_{ij}|}{n}[\mathbf{B}(i,j)]^2,\\
&(3) \qquad |\tilde{p}_{jj} - p_{jj}| \le K\frac{|x_{ij}|}{n\sqrt{n}}[\mathbf{B}(i,j)]^2,\\
&(4) \qquad |\tilde{q}_{ii} - q_{ii}| \le K\frac{|x_{ij}|}{n}\mathbf{B}(i,j) \le K\frac{|x_{ij}|}{n\sqrt{n}}[\mathbf{B}(i,j)]^2.
\end{aligned}$$

The estimates in the next lemmas are very helpful in our calculations.

Lemma 4.5.16. *There exists a constant K depending only on τ, v, k such that*

$$\begin{aligned}
&(1) \quad E\sum_{ji}|\underline{a}_{ji}^{(k)}|^2 \le Kn^2, \ \textit{where } \underline{a}_{ji}^{(k)} \equiv f_j'(A_n - zI)^{-1}T_{1n}X_n^*T_{2n}^k e_i;\\
&(2) \quad E\sum_{ji}|\tilde{a}_{ji}^{(k)}|^2 \le Kn^2, \ \textit{where } \tilde{a}_{ji}^{(k)} \equiv f_j'(A_n - zI)^{-1}T_{1n}X_{ij}^*T_{2n}^k e_i;\\
&(3) \quad E\sum_{ji}|a_{ji}^{(k)}|^2 \le Kn^2, \ \textit{where } a_{ji}^{(k)} \equiv f_j'(A_{ij} - zI)^{-1}T_{1n}X_{ij}^*T_{2n}^k e_i.
\end{aligned}$$

Proof. We first prove $Etr\{X_nT_{1n}(A_n^* - \bar{z}I)^{-1}(A_n - zI)^{-1}T_{1n}X_n^*\} \le Kn^2$. Let $B_n = (1/N)X_n^*T_{2n}X_n$. By Theorem 4.2.1 and Lemmas 4.2.1 and 4.5.8, we have

$$\begin{aligned}
& |Etr\{X_nT_{1n}A_n(A_n - zI)^{-1}T_{1n}X_n^*\}|\\
= \ & |Etr\{X_nT_{1n}^2B_n^{1/2}G(z)B_n^{1/2}T_{1n}X_n^*\}|\\
\le \ & \frac{1}{v}\{Etr(X_nT_{1n}^2B_nT_{1n}^2X_n^*)\}^{1/2}\{Etr(X_nT_{1n}B_nT_{1n}X_n^*)\}^{1/2}\\
\le \ & \frac{\tau}{v}\{\frac{1}{N}Etr(X_nT_{1n}^2X_n^*)^2\}^{1/2}\{\frac{1}{N}Etr(X_nT_{1n}X_n^*)^2\}^{1/2}\\
\le \ & Kn^2
\end{aligned}$$

and

$$
\begin{aligned}
& Etr\{X_nT_{1n}(A_n^*-\bar{z}I)^{-1}A_n^*A_n(A_n-zI)^{-1}T_{1n}X_n^*\} \\
= \;& Etr\{X_nT_{1n}B_n^{1/2}G(z)^*B_n^{1/2}T_{1n}^2B_n^{1/2}G(z)B_n^{1/2}T_{1n}X_n^*\} \\
\le \;& \tau^2 Etr\{B_n^{1/2}T_{1n}X_n^*X_nT_{1n}B_n^{1/2}G(z)^*B_nG(z)\} \\
\le \;& \frac{\tau^2}{v^2}\{Etr(B_n^{1/2}T_{1n}X_n^*X_nT_{1n}B_n^{1/2})^2\}^{1/2}\{Etr(B_n^2)\}^{1/2} \\
= \;& \frac{\tau^2}{v^2}\{Etr(T_{1n}X_n^*X_nT_{1n}B_n)^2\}^{1/2}\{\frac{1}{N^2}Etr(X_n^*T_{2n}X_n)^2\}^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
= \;& \frac{\tau^2}{v^2}\{\frac{1}{N^2}Etr((X_nT_{1n}X_n^*)^2T_{2n})^2\}^{1/2}\{\frac{1}{N^2}Etr((X_nX_n^*)T_{2n})^2\}^{1/2} \\
\le \;& \frac{\tau^4}{v^2}\{\frac{1}{N^2}Etr(X_nT_{1n}X_n^*)^4\}^{1/2}\{\frac{1}{N^2}Etr(X_n^*X_n)^2\}^{1/2} \\
\le \;& Kn^2.
\end{aligned}
$$

Note that from the resolvent identity (2.1.5),

$$
\begin{aligned}
& |z|^2tr\{X_nT_{1n}(A_n^*-\bar{z}I)^{-1}(A_n-zI)^{-1}T_{1n}X_n^*\} \\
= \;& tr\{X_nT_{1n}(I-(A_n^*-\bar{z}I)^{-1}A_n^*)(I-A_n(A_n-zI)^{-1})T_{1n}X_n^*\} \\
= \;& tr(T_{1n}X_n^*X_nT_{1n})-tr\{X_nT_{1n}A_n(A_n-zI)^{-1}T_{1n}X_n^*\} \\
& -tr\{X_nT_{1n}(A_n^*-zI)^{-1}A_n^*T_{1n}X_n^*\} \\
& +tr\{X_nT_{1n}(A_n^*-zI)^{-1}A_n^*A_n(A_n-zI)^{-1}T_{1n}X_n^*\}.
\end{aligned}
$$

The asserted result is proved.

Now we proceed with our proof of the three inequalities in the lemma. For (1), we have

$$
\begin{aligned}
E\sum_{ji}|\underline{a}_{ji}^{(k)}|^2 &= Etr\{T_{2n}^k X_n T_{1n}(A_n^* - \bar{z}I)^{-1}(A_n - zI)^{-1}T_{1n}X_n^* T_{2n}^k\} \\
&\leq \tau^{2k} Etr\{X_n T_{1n}(A_n^* - \bar{z}I)^{-1}(A_n - zI)^{-1}T_{1n}X_n^*\} \\
&\leq Kn^2.
\end{aligned}
$$

For (2), we have

$$
\begin{aligned}
& E\sum_{ji}|\tilde{a}_{ji}^{(k)} - \underline{a}_{ji}^{(k)}|^2 \\
= \;& E\sum_{ji}|x_{ij}|^2(\xi_{ii}^{(k)})^2|f'(A_n - zI)^{-1}T_{1n}f_j|^2 \\
\leq \;& KE\sum_{ji}|x_{ij}|^2(1+\frac{1}{n}\alpha(i,j,1))^2 \\
\leq \;& K(n^2 + \frac{1}{n^2}E\sum_{ij}[\alpha(i,j,1)]^2) \\
\leq \;& Kn^2,
\end{aligned}
$$

and hence (2) follows. To show (3), we use Lemma 4.5.15(1). It follows

$$
\begin{aligned}
E\sum_{ji}|\tilde{a}_{ji}^{(k)} - a_{ji}^{(k)}|^2 &\leq KE\sum_{ji}\frac{1}{N^2}|x_{ij}|^2[\mathbf{B}(i,j)]^4 \\
&\leq KE\sum_{ji}\frac{1}{N^2}[\mathbf{B}(i,j)]^4 \\
&\leq Kn^2.
\end{aligned}
$$

Thus (3) is proved. □

Lemma 4.5.17. *There exists a constant K depending only on τ, v, k such* *that*

$$(1)\quad E\sum_{ji}|\underline{\sigma}_{ij}^{(k)}|^2 \le Kn^2,\ \textit{where}\ \underline{\sigma}_{ij}^{(k)} \equiv e_i'(\forall_n - zI)^{-1}T_{2n}X_nT_{1n}^kf_j;$$

$$(2)\quad E\sum_{ji}|\tilde{\sigma}_{ij}^{(k)}|^2 \le Kn^2,\ \textit{where}\ \tilde{\sigma}_{ij}^{(k)} \equiv e_i'(\forall_n - zI)^{-1}T_{2n}X_{ij}T_{1n}^kf_j;$$

$$(3)\quad E\sum_{ji}|\sigma_{ij}^{(k)}|^2 \le Kn^2,\ \textit{where}\ \sigma_{ij}^{(k)} \equiv e_i'(\forall_{ij} - zI)^{-1}T_{2n}X_{ij}T_{1n}^kf_j.$$

Proof. By Theorem 4.2.1, taking $B_n = T_{2n}$, we have

$$\begin{aligned}
E\sum_{ji}|\underline{\sigma}_{ij}^{(k)}|^2 &= Etr(T_{1n}^kX_n^*T_{2n}(\forall_n^* - \bar{z}I)^{-1}(\forall_n - zI)^{-1}T_{2n}X_nT_{1n}^k)\\
&= Etr(T_{1n}^kX_n^*T_{2n}^{1/2}G(z)^*T_{2n}G(z)T_{2n}^{1/2}X_nT_{1n}^k)\\
&\le \frac{\tau^2}{v^2}Etr(T_{1n}^kX_n^*X_nT_{1n}^k)\\
&\le \frac{\tau^{2+2k}}{v^2}Etr(X_n^*X_n)\\
&\le Kn^2.
\end{aligned}$$

Since

$$|\underline{\sigma}_{ij}^{(k)} - \tilde{\sigma}_{ij}^{(k)}| = |x_{ij}\zeta_{jj}^{(k)}\tilde{q}_{ii}| \le K|x_{ij}|,$$

$E\sum_{ij}|\underline{\sigma}_{ij}^{(k)} - \tilde{\sigma}_{ij}^{(k)}|^2 \le Kn^2$ and hence (2) follows. From Lemma 4.5.15(2), we obtain (3).□

We now define the second type of bounds.

Lemma 4.5.18. *Let*

$$\mathbf{B}_1(i,j) = \sqrt{n} + \sum_{p=1}^{6}\left(\frac{1}{\sqrt{n}}\right)^{p-1}\|X_{ij}^*X_{ij}\|^{\frac{p}{2}}.$$

Then for each $\delta > 0$ and each positive integer l, there exists constant K depending on l and δ such that

$$E\sum_{ij}[\mathbf{B}_1(i,j)]^{\frac{l}{2}} \leq Kn^{2+\frac{l}{4}+\delta}. \tag{4.5.9}$$

Proof. Without loss of generality, we only need show that

$$E\sum_{ij}[\mathbf{B}_1(i,j)]^{l} \leq Kn^{2+\frac{l}{2}+\delta}. \tag{4.5.10}$$

From the fact in (4.5.6), it suffices to show for each $\delta > 0$ and each positive integer l, there exists constant K depending on l and δ such that

$$E\sum_{ij}\|X_{ij}^*X_{ij}\|^{\frac{l}{2}} \leq Kn^{2+\frac{l}{2}+\delta}. \tag{4.5.11}$$

We first observe that for any positive integer m, there is a constant K depending only on l and m such that

$$\begin{aligned} E\|X_n^*X_n\|^{\frac{l}{2}} &= E\{\lambda_{max}((X_n^*X_n)^{ml})\}^{\frac{1}{2m}} \\ &\leq \{E\lambda_{max}((X_n^*X_n)^{ml})\}^{\frac{1}{2m}} \leq \{Etr(X_n^*X_n)^{ml}\}^{\frac{1}{2m}} \\ &\leq Kn^{\frac{l}{2}+\frac{1}{2m}}. \end{aligned}$$

This implies for each $\delta > 0$ and each positive integer l, there exists constant K depending on l and δ such that $E\|X_n^*X_n\|^{\frac{l}{2}} \leq Kn^{\frac{l}{2}+\delta}$.

From the fact that $\|X_{ij}^*X_{ij}\| \leq \|X_n^*X_n\| + 2|x_{ij}|\|X_n^*X_n\|^{1/2} + |x_{ij}|^2$, it then follows

$$
\begin{aligned}
& E\sum_{ij}\|X_{ij}^*X_{ij}\|^{\frac{l}{2}} \\
\leq\ & K\left(E\sum_{ij}\|X_n^*X_n\|^{\frac{l}{2}}+(\delta_n\sqrt{n})^{\frac{l}{2}}E\sum_{ij}\|X_n^*X_n\|^{\frac{l}{4}}+(\delta_n\sqrt{n})^{l-2}n^2\right) \\
\leq\ & Kn^{2+\frac{l}{2}+\delta}.
\end{aligned}
$$

Thus we get (4.5.10). □

As an illustration of the use of this second bound, we use Burkholder's inequality in Lemma 2.1.4 to derive the following result.

Lemma 4.5.19. *For each positive integer k and each $z \in \mathbb{C}^+$, let $\Phi_n^{(k)}(z)$ and $\Psi_n^{(k)}(z)$ be defined as in section 4.1 (4.1.5) and let $g_{1n}(z)$, $g_{2n}(z)$ be defined as in Section 4.3 (4.3.9) and (4.3.10). Further denote*

$$\phi_n^{(k)}(z) = E\Phi_n^{(k)}(z), \quad \psi_n^{(k)}(z) = E\Psi_n^{(k)}(z).$$

Then with probability 1, as $n \to \infty$,

$$\Phi_n^{(k)}(z) - \phi_n^{(k)}(z) \to 0, \quad \Psi_n^{(k)}(z) - \psi_n^{(k)}(z) \to 0, \tag{4.5.12}$$

$$g_{1n}(z) - Eg_{1n}(z) \to 0, \quad g_{2n}(z) - Eg_{2n}(z) \to 0. \tag{4.5.13}$$

Proof. By Corollary 4.3.1, $\Phi_n^{(2)}(z) - \phi_n^{(2)}(z) = z(g_{2n}(z) - Eg_{2n}(z))$, $\Psi_n^{(2)}(z) - \psi_n^{(2)}(z) = z(g_{1n}(z) - Eg_{1n}(z))$. This means we only need to show (4.5.12).

For each pair (i, j), let $m = (i-1)n + j$ and

$$\mathcal{F}_m = \sigma(\{\cup_{a=1}^{i-1}\sigma(x_{a1}, x_{a2}, \cdots, x_{an})\} \cup \sigma(x_{i1}, \cdots, x_{ij})),$$

and let $E_m(\cdot)$ denote the conditional expectation given the σ-field $\mathcal{F}_m$. Further let $E_0(\cdot)$ denote the mathematical expectation of $(\cdot)$. Note that each pair (i, j) corresponds to a unique m. Write

$$\begin{aligned} y_m^{(1)} &\equiv \frac{1}{N} tr\{[(A_{ij} - zI)^{-1} - (A_n - zI)^{-1}] T_{1n}^{k-1}\}, \\ y_m^{(2)} &\equiv \frac{1}{N} tr\{[(\forall_{ij} - zI)^{-1} - (\forall_n - zI)^{-1}] T_{2n}^{k-1}\}. \end{aligned}$$

We first prove $|y_m^{(l)}| \le \frac{K}{n^3} |x_{ij}| [\mathbf{B}_1(i,j)]^3$, for $l = 1, 2$.

By the resolvent identity, the assumption $|x_{ij}| \le \delta_n \sqrt{n}$ and Theorem 4.2.1, we have

$$\begin{aligned} \|(A_n - zI)^{-1}\| &\le \frac{1}{v} + \frac{\tau^2}{v^2} \|\frac{1}{N} X_n^* X_n\| \\ &\le K(1 + \frac{1}{N}\|X_{ij}^* X_{ij}\| + \frac{1}{N}|x_{ij}| \|X_{ij}^* X_{ij}\|^{\frac{1}{2}} + \frac{1}{N}|x_{ij}|^2) \\ &\le K \frac{1}{\sqrt{n}} \mathbf{B}_1(i,j), \\ \|(A_{ij} - zI)^{-1}\| &\le K(1 + \frac{1}{N}\|X_{ij}^* X_{ij}\|) \le K \frac{1}{\sqrt{n}} \mathbf{B}_1(i,j), \end{aligned}$$

and

$$\|(\forall_n - zI)^{-1} T_{2n}\| \le \frac{\tau}{v}, \quad \|(\forall_{ij} - zI)^{-1} T_{2n}\| \le \frac{\tau}{v}.$$

It follows

$$\begin{aligned} |y_m^{(1)}| &= |\frac{1}{N^2} x_{ij} f_j' (A_n - zI)^{-1} T_{1n}^{k-1} (A_{ij} - zI)^{-1} T_{1n} X_{ij}^* T_{2n} e_i \\ &\quad + \frac{1}{N^2} \bar{x}_{ij} e_i' T_{2n} X_{ij} (A_n - zI)^{-1} T_{1n}^{k-1} (A_{ij} - zI)^{-1} T_{1n} f_j \\ &\quad + \frac{1}{N^2} |x_{ij}|^2 \xi_{ii}^{(1)} f_j' (A_n - zI)^{-1} T_{1n}^{k-1} (A_{ij} - zI)^{-1} T_{1n} f_j| \\ &\le K \left(\frac{1}{n^3} |x_{ij}| [\mathbf{B}_1(i,j)]^2 \|X_{ij}^* X_{ij}\|^{1/2} + \frac{1}{n^3} |x_{ij}|^2 [\mathbf{B}_1(i,j)]^2 \right) \\ &\le K \frac{1}{n^3} |x_{ij}| [\mathbf{B}_1(i,j)]^3, \end{aligned} \tag{4.5.14}$$

and

$$\begin{aligned}
|y_m^{(2)}| &= |\frac{1}{N^2} x_{ij} f_j' T_{1n} X_{ij}^* (\forall_n - zI)^{-1} T_{2n}^{k-1} (\forall_{ij} - zI)^{-1} T_{2n} e_i \\
&\quad + \frac{1}{N^2} \bar{x}_{ij} e_i' (\forall_n - zI)^{-1} T_{2n}^{k-1} (\forall_{ij} - zI)^{-1} T_{2n} X_{ij} T_{1n} f_j \\
&\quad + \frac{1}{N^2} |x_{ij}|^2 \zeta_{jj}^{(1)} e_i' (\forall_n - zI)^{-1} T_{2n}^{k-1} (\forall_{ij} - zI)^{-1} T_{2n} e_i| \qquad (4.5.15) \\
&\leq K \left(\frac{1}{n^2} |x_{ij}| \mathbf{B}_1(i,j) + \frac{1}{n^2} |x_{ij}|^2 \right) \\
&\leq K \frac{1}{n^3} |x_{ij}| [\mathbf{B}_1(i,j)]^3 .
\end{aligned}$$

Note that by Lemma 4.3.7,

$$\begin{aligned}
&\Phi_n^{(k)}(z) - \phi_n^{(k)}(z) \\
= \ & z \left(\frac{1}{N} tr\{(\forall_n - zI)^{-1} T_{2n}^{k-1}\} - E \frac{1}{N} tr\{(\forall_n - zI)^{-1} T_{2n}^{k-1}\} \right) \\
= \ & z \sum_{m=1}^{nN} (E_m - E_{m-1}) \frac{1}{N} tr\{(\forall_n - zI)^{-1} T_{2n}^{k-1}\} \\
= \ & z \sum_{m=1}^{nN} (E_m - E_{m-1}) \frac{1}{N} tr\{[(\forall_n - zI)^{-1} - (\forall_{ij} - zI)^{-1}] T_{2n}^{k-1}\} \\
= \ & -z \sum_{m=1}^{nN} (E_m - E_{m-1}) y_m^{(2)},
\end{aligned}$$

and

$$\begin{aligned}
&\Psi_n^{(k)}(z) - \psi_n^{(k)}(z) \\
= \ & z \left(\frac{1}{N} tr\{(A_n - zI)^{-1} T_{1n}^{k-1}\} - E \frac{1}{N} tr\{(A_n - zI)^{-1} T_{1n}^{k-1}\} \right) \\
= \ & -z \sum_{m=1}^{nN} (E_m - E_{m-1}) y_m^{(1)}.
\end{aligned}$$

By Burkholder's inequality, using the facts, $E_{m-1}|(E_m - E_{m-1})Z|^2 \leq 4E_{m-1}|Z|^2$,

$E|(E_m - E_{m-1})Z|^p \le 2^p E|Z|^p$, we get for any $p \ge 2$,

$$
\begin{aligned}
& E|\Psi_n^{(k)}(z) - \psi_n^{(k)}(z)|^p \\
= \;& |z|^p E|\sum_{m=1}^{nN}(E_m - E_{m-1})y_m^{(1)}|^p \\
\le \;& K\left(E\left(\sum_{m=1}^{nN} E_{m-1}|y_m^{(1)}|^2\right)^{\frac{p}{2}} + \sum_{m=1}^{nN} E|y_m^{(1)}|^p\right).
\end{aligned}
$$

Replacing the $y_m^{(1)}$ on the right-hand side of this inequality with $y_m^{(2)}$ will give an analogous result for $E|\Phi_n^{(k)}(z) - \phi_n^{(k)}(z)|^p$.

By means of (4.5.9), it can be computed that

$$
\begin{aligned}
& E\left(\sum_{m=1}^{Nn} E_{m-1}\left(\frac{1}{n^3}|x_{ij}|[\mathbf{B}_1(i,j)]^3\right)^2\right)^{p/2} \\
= \;& E\left(\sum_{m=1}^{Nn}\frac{1}{n^6}E_{m-1}[\mathbf{B}_1(i,j)]^6\right)^{p/2} \\
\le \;& n^{-3p}(Nn)^{p/2-1}E\sum_{m=1}^{Nn}(E_{m-1}[\mathbf{B}_1(i,j)]^6)^{p/2} \\
\le \;& Kn^{-2p-2}E\sum_{m=1}^{Nn}[\mathbf{B}_1(i,j)]^{3p} \\
\le \;& Kn^{-p/2+\delta},
\end{aligned}
$$

and

$$
\begin{aligned}
& E\sum_{m=1}^{Nn}\left(\frac{1}{n^3}|x_{ij}|[\mathbf{B}_1(i,j)]^3\right)^p \\
= \;& n^{-3p}\sum_{m=1}^{Nn} E|x_{ij}|^p E[\mathbf{B}_1(i,j)]^{3p} \\
\le \;& K\delta_n^{p-2}n^{-\frac{5}{2}p-1}E\sum_{m=1}^{Nn}[\mathbf{B}_1(i,j)]^{3p} \\
\le \;& Kn^{-p+1+\delta}.
\end{aligned}
$$

Hence we get for any $p \geq 2$,

$$E|\Phi_n^{(k)}(z) - \phi_n^{(k)}(z)|^p \leq Kn^{-p/2+\delta},$$

$$E|\Psi_n^{(k)}(z) - \psi_n^{(k)}(z)|^p \leq Kn^{-p/2+\delta}.$$

By Borel-Cantelli's lemma, (4.5.12) follows. □

We now define the third type of bound. However, this needs us define first the matrix $\Gamma_n(z) = -zI - zEg_{2n}(z)T_{1n}$. By Lemmas 4.3.4 and 4.3.6, it can be proven

$$\limsup_{n\to\infty} \|\Gamma_n^{-1}(z)\| \leq \frac{\tau}{v^2 E\delta_{2z}}, \tag{4.5.16}$$

where $\delta_{2z} \equiv \liminf_{n\to\infty} Img_{2n}(z)$. Note that $E\delta_{2z} > 0$, since by Lemma 4.3.4, δ_{2z} is positive almost surely. To see (4.5.16), use the fact $|g_{2n}(z)| \leq \frac{\tau}{v}$ and the following derivation:

$$\begin{aligned} \|\Gamma_n^{-1}(z)\| &= \frac{1}{|zEg_{2n}(z)|}\|(T_{1n} + \frac{1}{Eg_{2n}(z)})^{-1}\| \\ &\leq \frac{1}{|zEg_{2n}(z)|}\frac{1}{|Im\left(\frac{1}{Eg_{2n}(z)}\right)|} = \frac{|Eg_{2n}(z)|}{|zImEg_{2n}(z)|} \\ &\leq \frac{\tau}{v^2 ImEg_{2n}(z)} \end{aligned}$$

and so by Fatou's Lemma,

$$\begin{aligned} \limsup_{n\to\infty} \|\Gamma_n^{-1}(z)\| &\leq \frac{\tau}{v^2 \liminf_{n\to\infty} ImEg_{2n}(z)} \\ &\leq \frac{\tau}{v^2 E \liminf_{n\to\infty} Img_{2n}(z)} \\ &= \frac{\tau}{v^2 E\delta_{2z}}. \end{aligned}$$

This result guarantees us to define the third type of bound as follows.

Lemma 4.5.20. *Let*

$$\mathbf{B}_2(i,j) = \mathbf{B}(i,j) + \gamma^{1/2}(i,j) + \frac{1}{\sqrt{n}}\gamma(i,j),$$

where

$$\gamma(i,j) \equiv f_j' T_{1n}^{k+1}\Gamma_n^{-1}(z)^* X_{ij}^* X_{ij}\Gamma_n^{-1}(z) T_{1n}^{k+1} f_j.$$

Then there exists constant K depending only on k, τ, v such that

$$E\sum_{ij}[\mathbf{B}_2(i,j)]^l \le Kn^{2+\frac{l}{2}}. \tag{4.5.17}$$

Proof. Let

$$\gamma_{jj} \equiv f_j' T_{1n}^{k+1}\Gamma_n^{-1}(z)^* X_n^* X_n \Gamma_n^{-1}(z) T_{1n}^{k+1} f_j.$$

Then by taking T_n to be $\Gamma_n^{-1}(z)$ in Lemma 4.5.8 and using the same argument as we prove Lemma 4.5.11(1), it follows for every integer $l \ge 1$,

$$E\sum_j [\gamma_{jj}]^l \le Kn^{1+l}.$$

Also by following the same procedure as for Lemma 4.5.10(2), we have

$$\gamma(i,j) \le K(|x_{ij}|^2 + \gamma_{jj}),$$

so that $E\sum_j [\gamma(i,j)]^l \le Kn^{1+l}$. Now from (4.5.8), we straightforwardly get (4.5.17).

$\square$

Up to this point, the aim of the present section has been achieved. At the end, we complement some results which are of the same type as those in Lemma 4.5.13

and 4.5.14. We shall consider some quantities which are closely related with those therein. Let us introduce then

$$\underline{a}_{ji}^{(k)} \equiv f_j'(A_n - zI)^{-1}T_{1n}X_n^*T_{2n}^k e_i,\ \tilde{\hat{a}}_{ij}^{(1)} \equiv e_i'T_{2n}X_{ij}(A_n - zI)^{-1}T_{1n}f_j,$$

$$\tilde{d}_{ii}^{(k)} \equiv e_i'T_{2n}X_{ij}(A_n - zI)^{-1}T_{1n}X_{ij}^*T_{2n}^k e_i,$$

$$\underline{\sigma}_{ij}^{(k)} \equiv e_i'(\forall_n - zI)^{-1}T_{2n}X_nT_{1n}^k f_j,\ \tilde{\hat{\sigma}}_{ji}^{(1)} \equiv f_j'T_{1n}X_{ij}^*(\forall_n - zI)^{-1}T_{2n}e_i,$$

$$\tilde{b}_{jj}^{(k)} \equiv f_j'T_{1n}X_{ij}^*(\forall_n - zI)^{-1}T_{2n}X_{ij}T_{1n}^k f_j.$$

Lemma 4.5.21. *There exists constant K depending on τ, v, k only such that*

$$|\underline{a}_{ji}^{(k)}| \le K\mathbf{B}(i,j),\ |\tilde{\hat{a}}_{ij}^{(1)}| \le K\mathbf{B}(i,j),\ |\tilde{d}_{ii}^{(k)}| \le K\sqrt{n}\mathbf{B}(i,j),$$

$$|\underline{\sigma}_{ij}^{(k)}| \le K\mathbf{B}(i,j),\ |\tilde{\hat{\sigma}}_{ji}^{(1)}| \le K\mathbf{B}(i,j),\ |\tilde{b}_{jj}^{(k)}| \le K\sqrt{n}\mathbf{B}(i,j).$$

Proof. By using Lemma 4.2.1(6), Corollary 4.2.1(4) and Lemma 4.5.10, for $\underline{a}_{ji}^{(k)}$,

$$\begin{aligned}
|\underline{a}_{ji}^{(k)}| &\le \frac{1}{v}|f_j'T_{1n}X_n^*T_{2n}^k e_i| + \frac{1}{Nv^2}(f_j'T_{1n}X_n^*T_{2n}X_nT_{1n}f_j)^{1/2} \\
&\quad (e_i'T_{2n}^kX_nT_{1n}X_n^*T_{2n}X_nT_{1n}X_n^*T_{2n}^k e_i)^{1/2} \\
&\le K\{[\alpha_{jj}(1)]^{1/2} + \frac{1}{n}[\alpha_{jj}(1)]^{1/2}[\beta_{ii}(k)]^{1/2}\} \\
&\le K\{[\alpha_{jj}(1)]^{1/2} + \frac{1}{\sqrt{n}}\alpha_{jj}(1) + \frac{1}{n\sqrt{n}}\beta_{ii}(k)\} \\
&\le K\{(|x_{ij}|^2 + \alpha(i,j,1))^{1/2} + \frac{1}{\sqrt{n}}(|x_{ij}|^2 + \alpha(i,j,1)) \\
&\quad + \frac{1}{n\sqrt{n}}(|x_{ij}|^4 + |x_{ij}|^2\alpha(i,j,1) + \beta(i,j,k))\} \\
&\le K\{\sqrt{n} + [\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}\alpha(i,j,1) + \frac{1}{n\sqrt{n}}\beta(i,j,k)\} \\
&\le K\mathbf{B}(i,j).
\end{aligned}$$

For $\tilde{\hat{a}}_{ij}^{(1)}$,

$$
\begin{aligned}
|\tilde{\hat{a}}_{ij}^{(1)}| &\leq \frac{1}{v}|e_i' T_{2n} X_{ij} T_{1n} f_j| + \frac{1}{Nv^2}(f_j' T_{1n} X_n^* T_{2n} X_n T_{1n} f_j)^{1/2} \\
&\quad (e_i' T_{2n} X_{ij} T_{1n} X_n^* T_{2n} X_n T_{1n} X_{ij}^* T_{2n} e_i)^{1/2} \\
&\leq K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{n}[\alpha_{jj}(1)]^{1/2}[\tilde{\beta}(i,j,1)]^{1/2}\} \\
&\leq K\{[\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}(|x_{ij}|^2 + \alpha(i,j,1)) \\
&\quad + \frac{1}{n\sqrt{n}}(|x_{ij}|^2 \alpha(i,j,1) + \beta(i,j,1))\} \\
&\leq K\{\sqrt{n} + [\alpha(i,j,1)]^{1/2} + \frac{1}{\sqrt{n}}\alpha(i,j,1) + \frac{1}{n\sqrt{n}}\beta(i,j,1)\} \leq K\mathbf{B}(i,j).
\end{aligned}
$$

For $\tilde{d}_{ii}^{(k)}$,

$$
\begin{aligned}
|\tilde{d}_{ii}^{(k)}| &\leq \frac{1}{v}|e_i' T_{2n} X_{ij} T_{1n} X_{ij}^* T_{2n}^k e_i| + \frac{1}{Nv^2}(e_i' T_{2n} X_{ij} T_{1n} X_n^* T_{2n} X_n T_{1n} X_{ij}^* T_{2n} e_i)^{1/2} \\
&\quad (e_i' T_{2n}^k X_{ij} T_{1n} X_n^* T_{2n} X_n T_{1n} X_{ij}^* T_{2n}^k e_i)^{1/2} \\
&\leq K\{[\beta(i,j,1)]^{1/2} + \frac{1}{n}\tilde{\beta}(i,j,1) + \frac{1}{n}\tilde{\beta}(i,j,k)\} \\
&\leq K\{[\beta(i,j,1)]^{1/2} + \frac{1}{n}(|x_{ij}|^2 \alpha(i,j,1) + \beta(i,j,1)) \\
&\quad + \frac{1}{n}(|x_{ij}|^2 \alpha(i,j,1) + \beta(i,j,k))\} \\
&\leq \{\alpha(i,j,1) + [\beta(i,j,1)]^{1/2} + \frac{1}{n}\beta(i,j,1) + \frac{1}{n}\beta(i,j,k)\} \\
&\leq K\sqrt{n}\mathbf{B}(i,j).
\end{aligned}
$$

By using Lemma 4.2.1(6), Corollary 4.2.1(1) and Lemma 4.5.10, for $\underline{\sigma}_{ij}^{(k)}$,

$$
\begin{aligned}
|\underline{\sigma}_{ij}^{(k)}| &\leq \frac{1}{v}[\xi_{ii}^{(1)}]^{1/2}(f_j' T_{1n}^k X_n^* T_{2n} X_n T_{1n}^k f_j)^{1/2} \leq K[\alpha_{jj}(k)]^{1/2} \\
&\leq K(|x_{ij}|^2 + \alpha(i,j,k))^{1/2} \leq K(\sqrt{n} + [\alpha(i,j,k)]^{1/2}) \\
&\leq K\mathbf{B}(i,j).
\end{aligned}
$$

For $\tilde{\hat{\sigma}}_{ji}^{(1)}$,

$$
\begin{aligned}
|\tilde{\hat{\sigma}}_{ji}^{(1)}|) &\leq \frac{1}{v}[\xi_{ii}^{(1)}]^{1/2}(f_j'T_{1n}X_{ij}^*T_{2n}X_{ij}T_{1n}f_j)^{1/2} \\
&\leq K[\alpha(i,j,1)]^{1/2} \leq K\mathbf{B}(i,j).
\end{aligned}
$$

For $\tilde{b}_{jj}^{(k)}$,

$$
\begin{aligned}
|\tilde{b}_{jj}^{(k)}| &\leq \frac{1}{v}(f_j'T_{1n}X_{ij}^*T_{2n}X_{ij}T_{1n}f_j)^{1/2}(f_j'T_{1n}^kX_{ij}^*T_{2n}X_{ij}T_{1n}^kf_j)^{1/2} \\
&\leq K(\alpha(i,j,1)+\alpha(i,j,k)) \leq K\sqrt{n}\mathbf{B}(i,j).
\end{aligned}
$$

This completes the proof. □

Previously we have asserted from Theorem 2.3.9 and Lemma 4.3.2 that Theorem 1.2.1 will follow if it is shown $(s_{F^{\forall_n}}(z), g_{1n}(z), g_{2n}(z))$ converges almost surely to some non-random limit satisfying the system of equations (1.2.2). By Lemma 4.5.19 in the present section, however, since $s_{F^{\forall_n}}(z) = -z^{-1} - z^{-1}\Psi_n^{(1)}(z)$, it suffices to show $(Es_{F^{\forall_n}}(z), Eg_{1n}(z), Eg_{2n}(z))$ converges to some limit satisfying the system of equations (1.2.2). This will then be the main task of the next section.

4.6 Proof of Theorem 1.2.1.

The results in the previous sections have furnished us all necessary preliminary material needed for proving Theorem 1.2.1. The main task of the present section is then to finish the proof of the theorem. However, we have noticed that by Lemma 2.3.9 as well as Lemmas 4.3.2 and 4.5.19, it suffices to establish that $(Es_{F^{\forall_n}}(z), Eg_{1n}(z), Eg_{2n}(z))$ converges to some limit satisfying the system of equations (1.2.2). This is exactly the place where the procedure we proposed in Section

4.1 concerning manipulating the the Stieltjes transform method for the class of large general sample covariance matrices $(1/N)T_{2n}^{1/2}X_nT_{1n}X_n^*T_{2n}^{1/2}$ comes into play.

In Section 4.1, we emphasized the importance of the asymptotic behavior of those relations regarding the quantities, *i.e.* $\Phi_n^{(k)}(z)$ and $\Psi_n^{(k)}(z)$, which are closely related with the Stieltjes transforms of $F^{\forall_n}$ and F^{A_n}. These relations, namely, (4.1.8) equipped with (4.1.10), (4.1.11) and (4.1.12) equipped with (4.1.14), (4.1.15), will then be the primary main concern in the present section.

All derivations of the present section will again be done under all assumptions in Theorem 1.2.1 and conditions in Assumption 4.4.1.

4.6.1 Asymptotic Behavior of the Main Relations

Using the bound defined in Lemma 4.5.12, also with the aid of the results summarized in Lemmas 4.5.13-4.5.17, the asymptotic behavior of the above mentioned relations is first attained in the following theorem.

Theorem 4.6.1. *Given any positive integer k and any $z \in \mathbb{C}^+$, let $\varepsilon_n^{(k)}(z)$ and $\epsilon_n^{(k)}(z)$ be the residual terms in respectively,*

$$\phi_n^{(k)}(z) = \left(\frac{1}{N}tr(T_{2n}^k) - \phi_n^{(k+1)}(z)\right) Eg_{1n}(z) + \varepsilon_n^{(k)}(z), \tag{4.6.1}$$

and

$$\psi_n^{(k)}(z) = \left(\frac{1}{N}tr(T_{1n}^k) - \psi_n^{(k+1)}(z)\right) Eg_{2n}(z) + \epsilon_n^{(k)}(z). \tag{4.6.2}$$

Then

$$(1) \lim_{n\to\infty} \varepsilon_n^{(k)}(z) = 0. \quad (2) \lim_{n\to\infty} \epsilon_n^{(k)}(z) = 0. \tag{4.6.3}$$

Proof. By the proof of Lemma 4.5.19 and Hölder's inequality, we have

$$\lim_{n\to\infty} \left(E(\Phi_n^{(k+1)}(z) g_{1n}(z)) - \phi_n^{(k+1)}(z) E g_{1n}(z) \right) = 0, \tag{4.6.4}$$

$$\lim_{n\to\infty} \left(E(\Psi_n^{(k+1)}(z) g_{2n}(z)) - \psi_n^{(k+1)}(z) E g_{2n}(z) \right) = 0. \tag{4.6.5}$$

They imply that we can prove (4.6.3) through proving

$$(1)' \lim_{n\to\infty} \tilde{\varepsilon}_n^{(k)}(z) = 0 \quad \mathit{and} \quad (2)' \lim_{n\to\infty} \tilde{\epsilon}_n^{(k)}(z) = 0. \tag{4.6.6}$$

Here $\tilde{\varepsilon}_n^{(k)}(z)$ and $\tilde{\epsilon}_n^{(k)}(z)$ are the residual terms in

$$\phi_n^{(k)}(z) = \frac{1}{N} tr(T_{2n}^k) E g_{1n}(z) - E\left(g_{1n}(z) \Phi_n^{(k+1)}(z) \right) + \tilde{\varepsilon}_n^{(k)}(z),$$

and

$$\psi_n^{(k)}(z) = \frac{1}{N} tr(T_{1n}^k) E g_{2n}(z) - E\left(g_{2n}(z) \Psi_n^{(k+1)}(z) \right) + \tilde{\epsilon}_n^{(k)}(z).$$

We now develop the proof of (4.6.6).

Proof of (4.6.6)$(1)'$**.** In this proof, we shall base our derivations on relation (4.1.8) equipped with (4.1.10), (4.1.11). From (4.1.8), we see we only need prove

$$\lim_{n\to\infty} E\left(\frac{1}{N^2} \sum_{ij} x_{ij} \tilde{a}_{ji}^{(k)} + g_{1n}(z) \Phi_n^{(k+1)}(z) \right) = 0, \tag{4.6.7}$$

$$\lim_{n\to\infty} E\left(\frac{1}{N^2} \sum_{ij} |x_{ij}|^2 \xi_{ii}^{(k)} \tilde{p}_{jj} - \frac{1}{N} tr(T_{2n}^k) g_{1n}(z) \right) = 0. \tag{4.6.8}$$

However, by (4.1.11), (4.6.7) is a consequence of the following facts:

$$E\left(\frac{1}{N^2}\sum_{ij} x_{ij}a_{ji}^{(k)}\right) = 0, \tag{4.6.9}$$

$$E\left(\frac{1}{N^3}\sum_{ij} x_{ij}^2\tilde{a}_{ji}^{(1)}a_{ji}^{(k)}\right) = o(\frac{1}{\sqrt{n}}), \tag{4.6.10}$$

$$E\left(\frac{1}{N^3}\sum_{ij} x_{ij}^2\bar{x}_{ij}\xi_{ii}^{(1)}\tilde{p}_{jj}a_{ji}^{(k)}\right) = o(\frac{1}{\sqrt{n}}), \tag{4.6.11}$$

$$E\left(\frac{1}{N^3}\sum_{ij} |x_{ij}|^2\tilde{p}_{jj}d_{ii}^{(k)}\right) = E\left(g_{1n}(z)\Phi_n^{(k+1)}(z)\right) + o(1). \tag{4.6.12}$$

Now let us show these asymptotic relations successively. Since x_{ij} is independent of $a_{ji}^{(k)}$, (4.6.9) follows immediately. By Lemmas 4.5.13 and 4.5.16 and Hölder's inequality, we have

$$\begin{aligned}
& |E\left(\frac{1}{N^3}\sum_{ij} x_{ij}^2\tilde{a}_{ji}^{(1)}a_{ji}^{(k)}\right)| \\
\leq\ & \frac{1}{N^3}\left(E\sum_{ij}|x_{ij}|^4|a_{ji}^{(k)}|^2\right)^{1/2}\left(E\sum_{ij}|\tilde{a}_{ji}^{(1)}|^2\right)^{1/2} \\
\leq\ & K\frac{1}{n^2}\left(\delta_n^2 nE\sum_{ij}|a_{ji}^{(k)}|^2\right)^{1/2} \\
\leq\ & K\frac{\delta_n}{\sqrt{n}} \\
=\ & o(\frac{1}{\sqrt{n}}),
\end{aligned}$$

and

$$
\begin{aligned}
& |E\left(\frac{1}{N^3}\sum_{ij} x_{ij}^2 \bar{x}_{ij}\xi_{ii}^{(1)}\tilde{p}_{jj}a_{ji}^{(k)}\right)| \\
\leq\ & K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|^3\mathbf{B}(i,j)|a_{ji}^{(k)}| \\
\leq\ & K\frac{\delta_n}{n^3}\left(E\sum_{ij}[\mathbf{B}(i,j)]^2\right)^{1/2}\left(E\sum_{ij}|a_{ji}^{(k)}|^2\right)^{1/2} \\
\leq\ & K\frac{\delta_n}{\sqrt{n}} \\
=\ & o(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

Thus (4.6.10) and (4.6.11) are also proved.

To show (4.6.12), we define $\underline{d}_{ii}^{(k)} = e_i' T_{2n} X_n (A_n - zI)^{-1} T_{1n} X_n^* T_{2n}^k e_i$. It is easy to see

$$
E\frac{1}{N^3}\sum_{ij}\tilde{p}_{jj}\underline{d}_{ii}^{(k)} = E\left(g_{1n}(z)\Phi_n^{(k+1)}(z)\right). \tag{4.6.13}
$$

Now we estimate, recalling $\tilde{d}_{ii}^{(k)} \equiv e_i' T_{2n} X_{ij} (A_n - zI)^{-1} T_{1n} X_{ij}^* T_{2n}^k e_i$, the difference

$$
\underline{d}_{ii}^{(k)} - d_{ii}^{(k)} \equiv (\underline{d}_{ii}^{(k)} - \tilde{d}_{ii}^{(k)}) - (d_{ii}^{(k)} - \tilde{d}_{ii}^{(k)}).
$$

By their definition and the resolvent identity with (4.1.9), we have

$$
\underline{d}_{ii}^{(k)} - \tilde{d}_{ii}^{(k)} = x_{ij}\xi_{ii}^{(1)}\underline{a}_{ji}^{(k)} + \bar{x}_{ij}\xi_{ii}^{(k)}\tilde{\hat{a}}_{ij}^{(1)},
$$

and

$$\begin{aligned}
& d_{ii}^{(k)} - \tilde{d}_{ii}^{(k)} \\
= \;& \frac{1}{N} x_{ij} e_i' T_{2n} X_{ij} (A_{ij} - zI)^{-1} T_{1n} X_{ij}^* T_{2n} e_i f_j' (A_n - zI)^{-1} T_{1n} X_{ij}^* T_{2n}^k e_i \\
& + \frac{1}{N} \bar{x}_{ij} e_i' T_{2n} X_{ij} (A_{ij} - zI)^{-1} T_{1n} f_j e_i' T_{2n} X_{ij} (A_n - zI)^{-1} T_{1n} X_{ij}^* T_{2n}^k e_i \\
& + \frac{1}{N} |x_{ij}|^2 \xi_{ii}^{(1)} e_i' T_{2n} X_{ij} (A_{ij} - zI)^{-1} T_{1n} f_j f_j' (A_n - zI)^{-1} T_{1n} X_{ij}^* T_{2n}^k e_i \\
= \;& \frac{1}{N} x_{ij} d_{ii}^{(1)} \tilde{a}_{ji}^{(k)} + \frac{1}{N} \bar{x}_{ij} \hat{a}_{ij}^{(1)} \tilde{d}_{ii}^{(k)} + \frac{1}{N} |x_{ij}|^2 \xi_{ii}^{(1)} \hat{a}_{ij}^{(1)} \tilde{a}_{ji}^{(k)}.
\end{aligned}$$

Using the estimates given in Lemmas 4.5.13 and 4.5.20, it follows then

$$\begin{aligned}
& |\underline{d}_{ii}^{(k)} - \tilde{d}_{ii}^{(k)}| \le K |x_{ij}| \mathbf{B}(i,j), \\
& |d_{ii}^{(k)} - \tilde{d}_{ii}^{(k)}| \\
\le \;& K \left(\frac{1}{\sqrt{n}} |x_{ij}| [\mathbf{B}(i,j)]^2 + \frac{1}{\sqrt{n}} |x_{ij}| [\mathbf{B}(i,j)]^2 + \frac{1}{n} |x_{ij}|^2 [\mathbf{B}(i,j)]^2 \right) \\
\le \;& K \frac{1}{\sqrt{n}} |x_{ij}| [\mathbf{B}(i,j)]^2,
\end{aligned}$$

and hence

$$|\underline{d}_{ii}^{(k)} - d_{ii}^{(k)}| \le K \frac{1}{\sqrt{n}} |x_{ij}| [\mathbf{B}(i,j)]^2. \tag{4.6.14}$$

We then write

$$\begin{aligned}
& E \frac{1}{N^3} \sum_{ij} |x_{ij}|^2 \tilde{p}_{jj} d_{ii}^{(k)} \\
= \;& E \frac{1}{N^3} \sum_{ij} \tilde{p}_{jj} \underline{d}_{ii}^{(k)} - E \frac{1}{N^3} \sum_{ij} \tilde{p}_{jj} (\underline{d}_{ii}^{(k)} - d_{ii}^{(k)}) \\
& - E \frac{1}{N^3} \sum_{ij} (\tilde{p}_{jj} - p_{jj}) d_{ii}^{(k)} - \frac{1}{N^3} \sum_{ij} (1 - E|x_{ij}|^2) E(p_{jj} d_{ii}^{(k)}), \\
& + E \frac{1}{N^3} \sum_{ij} |x_{ij}|^2 (\tilde{p}_{jj} - p_{jj}) d_{ii}^{(k)}.
\end{aligned} \tag{4.6.15}$$

By the estimate given in Lemma 4.5.13 for $d_{ii}^{(k)}$ and the estimate given in Lemma 4.5.15(3) for the difference $|\tilde{p}_{jj} - p_{jj}|$, we have

$$\begin{aligned} & |E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2(\tilde{p}_{jj} - p_{jj})d_{ii}^{(k)}| \\ \leq\ & K\frac{1}{n^4}E\sum_{ij}|x_{ij}|^3[\mathbf{B}(i,j)]^3 \\ \leq\ & K\delta_n \\ =\ & o(1), \end{aligned} \tag{4.6.16}$$

and

$$\begin{aligned} & |E\frac{1}{N^3}\sum_{ij}(\tilde{p}_{jj} - p_{jj})d_{ii}^{(k)}| \\ \leq\ & K\frac{1}{n^4}E\sum_{ij}|x_{ij}|[\mathbf{B}(i,j)]^3 \\ \leq\ & K/\sqrt{n}. \end{aligned} \tag{4.6.17}$$

Similarly by (4.6.14), we have

$$|E\frac{1}{N^3}\sum_{ij}\tilde{p}_{jj}(\underline{d}_{ii}^{(k)} - d_{ii}^{(k)})| \leq K\frac{1}{n^4}E\sum_{ij}|x_{ij}|[\mathbf{B}(i,j)]^3 \leq K\frac{1}{\sqrt{n}}. \tag{4.6.18}$$

By Hölder's inequality and the estimates given in Lemma 4.5.13,

$$\begin{aligned} & |\frac{1}{N^3}\sum_{ij}(1 - E|x_{ij}|^2)E(p_{jj}d_{ii}^{(k)})| \\ \leq\ & \frac{1}{N^3}\sum_{ij}(1 - E|x_{ij}|^2)E[\mathbf{B}(i,j)]^2 \\ \leq\ & K\left(\frac{1}{nN}\sum_{ij}(1 - E|x_{ij}|^2)^2\right)^{1/2}\left(\frac{1}{n^4}E\sum_{ij}[\mathbf{B}(i,j)]^4\right)^{1/2} \\ \leq\ & K\left(\frac{1}{nN}\sum_{ij}(1 - E|x_{ij}|^2)\right)^{1/2} \\ \to\ & 0. \end{aligned} \tag{4.6.19}$$

In view of (4.6.15), combining the results in (4.6.13), (4.6.16)-(4.6.19) shows (4.6.12). Thus the proof of (4.6.7) is complete. To finish the proof of (4.6.6)(1)$'$, we still need to prove (4.6.8). We next show (4.6.8).

By using the same type of arguments as we used above, we have

$$\begin{aligned}
& E|\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\xi_{ii}^{(k)}(\tilde{p}_{jj}-p_{jj})| \le K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|^3[\mathbf{B}(i,j)]^2 \le K\delta_n \to 0,\\
& E|\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(k)}(p_{jj}-\tilde{p}_{jj})| \le K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|[\mathbf{B}(i,j)]^2 \le K\frac{1}{\sqrt{n}},\\
& |E\frac{1}{N^2}\sum_{ij}(1-|x_{ij}|^2)\xi_{ii}^{(k)}p_{jj}|\\
\le\ & K\frac{1}{n^2\sqrt{n}}\sum_{ij}(1-E|x_{ij}|^2)E\mathbf{B}(i,j)\\
\le\ & K\frac{1}{n\sqrt{n}}\left(\sum_{ij}(1-E|x_{ij}|^2)^2\right)^{1/2}\left(E\sum_{ij}[\mathbf{B}(i,j)]^2\right)^{1/2}\\
\le\ & K\left(\sum_{ij}(1-E|x_{ij}|^2)\right)^{1/2}\\
\to\ & 0,
\end{aligned}$$

and

$$E\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(k)}\tilde{p}_{jj} = E\left(\frac{1}{N}tr(T_{2n}^k)g_{1n}(z)\right).$$

Therefore we get

$$\begin{aligned}
& E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\xi_{ii}^{(k)}\tilde{p}_{jj}\\
=\ & E\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(k)}\tilde{p}_{jj} + E\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(k)}(p_{jj}-\tilde{p}_{jj})\\
& -E\frac{1}{N^2}\sum_{ij}(1-|x_{ij}|^2)\xi_{ii}^{(k)}p_{jj} + E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\xi_{ii}^{(k)}(\tilde{p}_{jj}-p_{jj})\\
=\ & E\left(\frac{1}{N}tr(T_{2n}^k)g_{1n}(z)\right) + o(1).
\end{aligned}$$

Thus (4.6.8) is also proved. The proof of (1)$'$ is complete.

Proof of (4.6.6)(2)$'$**.** Similarly, we shall derive the result from relation (4.1.12) equipped with (4.1.14) and (4.1.15).

From (4.1.12) we need only prove

$$\lim_{n\to\infty} E\left(\frac{1}{N^2}\sum_{ij}\bar{x}_{ij}\tilde{\sigma}_{ij}^{(k)} + g_{2n}(z)\Psi_n^{(k+1)}(z)\right) = 0, \tag{4.6.20}$$

$$\lim_{n\to\infty} E\left(\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\zeta_{jj}^{(k)}\tilde{q}_{ii} - \frac{1}{N}tr(T_{1n}^k)g_{2n}(z)\right) = 0. \tag{4.6.21}$$

By (4.1.15), we prove (4.6.20) by proving

$$E\left(\frac{1}{N^2}\sum_{ij}\bar{x}_{ij}\sigma_{ij}^{(k)}\right) = 0, \tag{4.6.22}$$

$$E\left(\frac{1}{N^3}\sum_{ij}\bar{x}_{ij}^2\tilde{\sigma}_{ij}^{(1)}\sigma_{ij}^{(k)}\right) = o(\frac{1}{\sqrt{n}}), \tag{4.6.23}$$

$$E\left(\frac{1}{N^3}\sum_{ij}x_{ij}\bar{x}_{ij}^2\zeta_{jj}^{(1)}\tilde{q}_{ii}\sigma_{ij}^{(k)}\right) = o(\frac{1}{\sqrt{n}}), \tag{4.6.24}$$

$$E\left(\frac{1}{N^3}\sum_{ij}|x_{ij}|^2\tilde{q}_{ii}b_{jj}^{(k)}\right) = E\left(g_{2n}(z)\Psi_n^{(k+1)}(z)\right) + o(1). \tag{4.6.25}$$

The result in (4.6.22) is again a consequence of the fact that x_{ij} is independent of $\sigma_{ij}^{(k)}$. Also, from Hölder's inequality and Lemmas 4.5.14 and 4.5.17, we have

$$\begin{aligned}
&|E\frac{1}{N^3}\sum_{ij}\bar{x}_{ij}^2\tilde{\sigma}_{ij}^{(1)}\sigma_{ij}^{(k)}| \\
\le\ & \frac{1}{N^3}\left(E\sum_{ij}|x_{ij}|^4|\sigma_{ij}^{(k)}|^2\right)^{1/2}\left(E\sum_{ij}|\tilde{\sigma}_{ij}^{(1)}|^2\right)^{1/2} \\
=\ & o(\frac{1}{\sqrt{n}}),
\end{aligned}$$

and

$$\begin{aligned}
& |E\frac{1}{N^3}\sum_{ij} x_{ij}\bar{x}_{ij}^2\zeta_{jj}^{(1)}\tilde{q}_{ii}\sigma_{ij}^{(k)}| \\
\leq & \; K\frac{\delta_n}{n\sqrt{n}}\left(E\sum_{ij}|\sigma_{ij}^{(k)}|^2\right)^{1/2} \\
= & \; o(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Thus (4.6.23) and (4.6.24) are proved.

To show (4.6.25), we define $\underline{b}_{jj}^{(k)} = f_j' T_{1n} X_n^* (\forall_n - zI)^{-1} T_{2n} X_n T_{1n}^k f_j$. Then

$$\frac{1}{N^3}\sum_{ij}\tilde{q}_{ii}\underline{b}_{jj}^{(k)} = g_{2n}(z)\Psi_n^{(k+1)}(z). \tag{4.6.26}$$

Recall that $\tilde{b}_{jj}^{(k)} = f_j' T_{1n} X_{ij}^* (\forall_n - zI)^{-1} T_{2n} X_{ij} T_{1n}^k f_j$. It can be calculated

$$\begin{aligned}
\underline{b}_{jj}^{(k)} - \tilde{b}_{jj}^{(k)} &= \bar{x}_{ij}\zeta_{jj}^{(1)}\underline{\sigma}_{ij}^{(k)} + x_{ij}\zeta_{jj}^{(k)}\tilde{\hat{\sigma}}_{ji}^{(1)}, \\
b_{jj}^{(k)} - \tilde{b}_{jj}^{(k)} &= \frac{1}{N}\bar{x}_{ij}b_{jj}^{(1)}\tilde{\sigma}_{ij}^{(k)} + \frac{1}{N}x_{ij}\hat{\sigma}_{ji}^{(1)}\tilde{b}_{jj}^{(k)} + \frac{1}{N}|x_{ij}|^2\zeta_{ii}^{(1)}\hat{\sigma}_{ji}^{(1)}\tilde{\sigma}_{ij}^{(k)}.
\end{aligned}$$

Using the estimates given in Lemma 4.3.14, we get

$$\begin{aligned}
&|\underline{b}_{jj}^{(k)} - \tilde{b}_{jj}^{(k)}| \leq K|x_{ij}|\mathbf{B}(i,j) \leq K\frac{1}{\sqrt{n}}|x_{ij}|[\mathbf{B}(i,j)]^2, \\
&|b_{jj}^{(k)} - \tilde{b}_{jj}^{(k)}| \leq K\frac{1}{\sqrt{n}}|x_{ij}|[\mathbf{B}(i,j)]^2,
\end{aligned}$$

and so

$$|\underline{b}_{jj}^{(k)} - b_{jj}^{(k)}| \leq K\frac{1}{\sqrt{n}}|x_{ij}|[\mathbf{B}(i,j)]^2. \tag{4.6.27}$$

We then write

$$
\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2\tilde{q}_{ii}b_{jj}^{(k)} \\
= \quad & E\frac{1}{N^3}\sum_{ij}\tilde{q}_{ii}\underline{b}_{jj}^{(k)} + E\frac{1}{N^3}\sum_{ij}\tilde{q}_{ii}(b_{jj}^{(k)} - \underline{b}_{jj}^{(k)}) \\
& -E\frac{1}{N^3}\sum_{ij}(\tilde{q}_{ii} - q_{ii})b_{jj}^{(k)} - \frac{1}{N^3}\sum_{ij}(1 - E|x_{ij}|^2)E(q_{ii}b_{jj}^{(k)}) \\
& +E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2(\tilde{q}_{ii} - q_{ii})b_{jj}^{(k)}. \qquad (4.6.28)
\end{aligned}
$$

By the estimate given in Lemma 4.5.14 for $b_{jj}^{(k)}$ and the estimate given in Lemma 4.5.15(4) for the difference $\tilde{q}_{ii} - q_{ii}$, we have

$$
\begin{aligned}
& |E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2(\tilde{q}_{ii} - q_{ii})b_{jj}^{(k)}| \\
\leq \quad & K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|^3[\mathbf{B}(i,j)]^2 \\
\leq \quad & K\delta_n \to 0, \qquad (4.6.29)
\end{aligned}
$$

and

$$
\begin{aligned}
& |E\frac{1}{N^3}\sum_{ij}(\tilde{q}_{ii} - q_{ii})b_{jj}^{(k)}| \\
\leq \quad & K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|[\mathbf{B}(i,j)]^2 \\
\leq \quad & K/\sqrt{n}. \qquad (4.6.30)
\end{aligned}
$$

Similarly, from (4.6.27), we get

$$
\begin{aligned}
& \frac{1}{N^3}|E\sum_{ij}\tilde{q}_{ii}(b_{jj}^{(k)} - \tilde{b}_{jj}^{(k)})| \\
\leq \quad & K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|[\mathbf{B}(i,j)]^2 \\
\leq \quad & K\frac{1}{\sqrt{n}}. \qquad (4.6.31)
\end{aligned}
$$

By Hölder's inequality and Lemma 4.5.14, we further have

$$\begin{aligned}
& |\frac{1}{N^3}\sum_{ij}(1-|x_{ij}|^2)E(q_{ii}b_{jj}^{(k)})| \\
\le\ & K\left(\frac{1}{nN}\sum_{ij}(1-E|x_{ij}|^2)^2\right)^{1/2}\left(\frac{1}{n^3}E\sum_{ij}[\mathbf{B}(i,j)]^2\right)^{1/2} \\
\le\ & K\left(\frac{1}{n^2}\sum_{ij}(1-E|x_{ij}|)\right)^{1/2} \\
\to\ & 0. \qquad (4.6.32)
\end{aligned}$$

In view of (4.6.28), the results in (4.6.26), (4.6.29)-(4.6.32) give (4.6.25). Hence the proof of (4.6.20) is complete.

Now we show (4.6.21). By using the same type of arguments, we have

$$\begin{aligned}
&|E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\zeta_{jj}^{(k)}(\tilde{q}_{ii}-q_{ii})| \le K\frac{1}{N^3}E\sum_{ij}|x_{ij}|^3\mathbf{B}(i,j)\le K\delta_n\to 0,\\
&|E\frac{1}{N^2}\sum_{ij}(1-|x_{ij}|^2)\zeta_{jj}^{(k)}q_{ii}| \le K\frac{1}{N^2}\sum_{ij}(1-E|x_{ij}|^2)\to 0,\\
&|E\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(k)}(q_{ii}-\tilde{q}_{ii})| \le K\frac{1}{N^3}E\sum_{ij}|x_{ij}|\mathbf{B}(i,j)\le K\frac{1}{\sqrt{n}},
\end{aligned}$$

and

$$\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(k)}\tilde{q}_{ii} = \left(\frac{1}{N}tr(T_{1n}^k)g_{2n}(z)\right).$$

Therefore, we get

$$\begin{aligned}
& E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\zeta_{jj}^{(k)}\tilde{q}_{ii} \\
=\ & E\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(k)}\tilde{q}_{ii} + E\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(k)}(q_{ii}-\tilde{q}_{ii}) - E\frac{1}{N^2}\sum_{ij}(1-E|x_{ij}|^2)\zeta_{jj}^{(k)}q_{ii} \\
& +E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\zeta_{jj}^{(k)}(\tilde{q}_{ii}-q_{ii}) \\
=\ & E\left(\frac{1}{N}tr(T_{1n}^k)g_{2n}(z)\right)+o(1).
\end{aligned}$$

Thus (4.6.21) is proved. The proof of (4.6.6)(2)′ is also complete. This completes the proof of Theorem 4.6.1.□

We now take a look at the special case of $k = 1$ of Theorem 4.6.1. In this case, we note that by Corollary 4.3.1,

$$-zEg_{1n}(z) = \frac{1}{N}tr(T_{1n}) - \psi_n^{(2)}(z),$$
$$-zEg_{2n}(z) = \frac{1}{N}tr(T_{2n}) - \phi_n^{(2)}(z).$$

On the other hand, we have by definition

$$\psi_n^{(1)}(z) = 1 + zEs_{F^{\forall_n}}(z). \tag{4.6.33}$$

$$\phi_n^{(1)}(z) = \frac{n}{N} + z\frac{n}{N}Es_{F^{A_n}}(z), \tag{4.6.34}$$

Thus in case of $k = 1$, further with Lemma 4.5.19, Theorem 4.6.1 gives us the following result.

Corollary 4.6.1. *For any $z \in \mathbb{C}^+$,*

$$\lim_{n\to\infty}(Es_{F^{A_n}}(z) + z^{-1} + \frac{N}{n}Eg_{1n}(z)Eg_{2n}(z)) = 0,$$

$$\lim_{n\to\infty}(Es_{F^{\forall_n}}(z) + z^{-1} + Eg_{1n}(z)Eg_{2n}(z)) = 0,$$

and almost surely,

$$\lim_{n\to\infty}(s_{F^{A_n}}(z) + z^{-1} + \frac{N}{n}g_{1n}(z)g_{2n}(z)) = 0,$$

$$\lim_{n\to\infty}(s_{F^{\forall_n}}(z) + z^{-1} + g_{1n}(z)g_{2n}(z)) = 0.$$

4.6.2 Understanding the Asymptotic Results Established

Recall that the proof of Theorem 1.2.1 lies in proving $(Es_{F^{\mathbf{W}_n}}(z), Eg_{1n}(z), Eg_{2n}(z))$ converges to some limit satisfying the system of equations (1.2.2). However, since by Lemma 4.3.6, the whole sequence is tight, and by Lemma 4.3.3, the system of equations has no more than one solution, this in turn lies in proving the result for any convergent subsequence. Therefore, we consider $(Es_{F^{\mathbf{W}_{n_m}}}(z), Eg_{1n_m}(z), Eg_{2n_m}(z))$ with limit $(\underline{s}(z), g_1(z), g_2(z))$. Our present concern is to explore the asymptotic results in Theorem 4.6.1 to get some understanding of the limit $(\underline{s}(z), g_1(z), g_2(z))$.

Theorem 4.6.1 indeed established two recursive relations for us. Note that Lemmas 4.3.4 and 4.3.5 imply $g_1(z) \neq 0$ and $g_2(z) \neq 0$. This means, in (4.6.1) and (4.6.2), if $\phi_{n_m}^{(k)}(z)$ ($\psi_{n_m}^{(k)}(z)$) converges then $\phi_{n_m}^{(k+1)}(z)$ ($\psi_{n_m}^{(k+1)}(z)$) must also converge. However, by (4.2.3) in Section 4.2, we have

$$\phi_{n_m}^{(1)}(z) = \psi_{n_m}^{(1)}(z) = 1 + zEs_{F^{\mathbf{W}_n}}(z) \to 1 + z\underline{s}(z),$$

It therefore follows for each k, $\phi_{n_m}^{(k)}(z)$ and $\psi_{n_m}^{(k)}(z)$ converge with their limits, say denoted by $\phi_k(z)$ and $\psi_k(z)$, satisfying

$$\phi_1(z) = \psi_1(z) = 1 + z\underline{s}(z), \tag{4.6.35}$$

$$\psi_{k+1}(z) = c\int x^k dH_1(x) - g_2^{-1}(z)\psi_k(z). \tag{4.6.36}$$

$$\phi_{k+1}(z) = \int y^k dH_2(y) - g_1^{-1}(z)\phi_k(z). \tag{4.6.37}$$

The two recursive relations in (4.6.36) and (4.6.37) with initial values given by (4.6.35) are the results provided by Theorem 4.6.1, or more pertinent, by the procedure of applying the Stieltjes transform method in Section 4.1. We shall in the following get from them new directions to the final answer of the problem, that is, to finding the equations uniquely determining the limit $(\underline{s}(z), g_1(z), g_2(z))$.

first, also obviously, we get from Corollary 4.6.1,

$$\underline{s}(z) = -z^{-1} - g_1(z)g_2(z). \tag{4.6.38}$$

Then, by the two iterative relations, let us roughly deduce that

$$\begin{aligned}
\psi_k(z) &= c\int x^k dH_1(x)g_2(z) - \psi_{k+1}(z)g_2(z) \\
&= c\int x^k dH_1(x)g_2(z) - c\int x^{k+1}dH_1(x)(g_2(z))^2 + \psi_{k+2}(z)(g_2(z))^2 \\
&= \sum_{m=1}^{\infty}(-1)^{m-1}(g_2(z))^m c\int x^{k+m-1}dH_1(x) \\
&= c\int x^k \sum_{m=1}^{\infty}(-1)^{m-1}(xg_2(z))^{m-1}g_2(z)dH_1(x) \\
&= c\int \frac{x^k g_2(z)}{1+xg_2(z)}dH_1(x),
\end{aligned} \tag{4.6.39}$$

and similarly,

$$\begin{aligned}
\phi_k(z) &= \sum_{m=1}^{\infty}(-1)^{m-1}(g_1(z))^m \int y^{k+m-1}dH_2(y) \\
&= \int \frac{y^k g_1(z)}{1+yg_1(z)}dH_2(y).
\end{aligned} \tag{4.6.40}$$

Letting $k=1$ in the above two equalities gives

$$\psi_1(z) = c\int \frac{xg_2(z)}{1+xg_2(z)}dH_1(x) = \phi_1(z) = \int \frac{yg_1(z)}{1+yg_1(z)}dH_2(y). \tag{4.6.41}$$

From (4.6.35), it therefore follows

$$\begin{aligned}
\underline{s}(z) &= -z^{-1} + z^{-1}\psi_1(z) \\
&= -z^{-1}(1-c) - z^{-1}c\int \frac{1}{1+g_2(z)x}dH_1(x), \qquad (4.6.42) \\
&= -z^{-1} + z^{-1}\phi_1(z) \\
&= -z^{-1}\int \frac{1}{1+g_1(z)y}dH_2(y). \qquad (4.6.43)
\end{aligned}$$

It can be seen (4.6.38), (4.6.42) and (4.6.43) show that $(\underline{s}(z), g_1(z), g_2(z))$ satisfies the system of equations in (1.2.2). Thus we arrives at our target claimed. However, the above argument is not rigorous. To make the result rigorously hold, we need to prove (4.6.39) and (4.6.40) to be true in the strict sense. Thus we are now led to this new direction to complete our proof of Theorem 1.2.1.

4.6.3 Proof of Theorem 1.2.1

As discussed above, we shall in the following prove (4.6.39) and (4.6.40). For that purpose, write $R_n(z) = -zI - zEg_{1n}(z)T_{2n}$ and $\Gamma_n(z) = -z - zEg_{2n}(z)T_{1n}$. Then it is conceivable the proof of (4.6.39) and (4.6.40) can be obtained by showing

$$\phi_n^{(k)}(z) - \frac{1}{N}tr\{-zEg_{1n}(z)T_{2n}^k R_n^{-1}(z)\} \to 0,$$

and

$$\psi_n^{(k)}(z) - \frac{1}{N}tr\{-zEg_{2n}(z)T_{1n}^k \Gamma_n^{-1}(z)\} \to 0.$$

Let us now give some new notations for better understanding of the problem. Define for each nonnegative integer k,

$$g_{1n}^{(k)}(z) = \frac{1}{N}tr\{(A_n - zI)^{-1}T_{1n}^k\},$$
$$g_{2n}^{(k)}(z) = \frac{1}{N}tr\{(\forall_n - zI)^{-1}T_{2n}^k\}.$$

Then for $k = 1$, $g_{1n}^{(k)}(z)$ and $g_{2n}^{(k)}(z)$ will coincide with $g_{1n}(z)$ and $g_{2n}(z)$ respectively.

By Lemma 4.3.7, using the trivial fact $A(A-zI)^{-1} = I + z(A-zI)^{-1}$,

$$g_{1n}^{(k)}(z) = -z^{-1}\frac{1}{N}tr(T_{1n}^k) + z^{-1}\Psi_n^{(k+1)}(z), \tag{4.6.44}$$

$$g_{2n}^{(k)}(z) = -z^{-1}\frac{1}{N}tr(T_{2n}^k) + z^{-1}\Phi_n^{(k+1)}(z). \tag{4.6.45}$$

Simultaneously, we have

$$-zEg_{1n}(z)T_{2n}R_n^{-1}(z) = I + zR_n^{-1}(z),$$

$$-zEg_{2n}(z)T_{1n}\Gamma_n^{-1}(z) = I + z\Gamma_n^{-1}(z),$$

which imply

$$\frac{1}{N}tr\{-zEg_{1n}(z)T_{2n}^kR_n^{-1}(z)\} = \frac{1}{N}tr(T_{2n}^{k-1}) + z\frac{1}{N}tr\{T_{2n}^{k-1}R_n^{-1}(z)\}, \tag{4.6.46}$$

$$\frac{1}{N}tr\{-zEg_{2n}(z)T_{1n}^k\Gamma_n^{-1}(z)\} = \frac{1}{N}tr(T_{1n}^{k-1}) + z\frac{1}{N}tr\{T_{1n}^{k-1}\Gamma_n^{-1}(z)\}. \tag{4.6.47}$$

In view of (4.6.44)-(4.6.47), we get

$$\begin{aligned} &\phi_n^{(k)}(z) - \frac{1}{N}tr\{-zEg_{1n}(z)T_{2n}^kR_n^{-1}(z)\} \\ = \; & z\left(Eg_{2n}^{(k-1)}(z) - \frac{1}{N}tr(T_{2n}^{k-1}R_n^{-1}(z))\right), \end{aligned} \tag{4.6.48}$$

and

$$\begin{aligned} &\psi_n^{(k)}(z) - \frac{1}{N}tr\{-zEg_{2n}(z)T_{1n}^k\Gamma_n^{-1}(z)\} \\ = \; & z\left(Eg_{1n}^{(k-1)}(z) - \frac{1}{N}tr(\Gamma_n^{-1}(z)T_{1n}^{k-1})\right). \end{aligned} \tag{4.6.49}$$

The advantage of introducing $g_{1n}^{(k)}(z)$ and $g_{2n}^{(k)}(z)$ is that we can apply the resolvent identity to the right-hand terms appearing in the above two equalities. Note that $R_n(z)T_{2n} = T_{2n}R_n(z)$ implies $R_n^{-1}(z)T_{2n}^k = T_{2n}^kR_n^{-1}(z)$ for all k (and similarly,

$\Gamma_n^{-1}(z)T_{1n}^k = T_{1n}^k\Gamma_n^{-1}(z))$. Applying the resolvent identity, we can get

$$\begin{aligned} & \frac{1}{N}tr(R_n^{-1}(z)T_{2n}^k) - Eg_{2n}^{(k)}(z) \\ = \;& E\frac{1}{N}tr\{(\forall_n - zI)^{-1}\forall_n R_n^{-1}(z)T_{2n}^k\} \\ & + zEg_{1n}(z)E\frac{1}{N}tr\{(\forall_n - zI)^{-1}R_n^{-1}(z)T_{2n}^{k+1}\}, \end{aligned} \tag{4.6.50}$$

and

$$\begin{aligned} & \frac{1}{N}tr(\Gamma_n^{-1}(z)T_{1n}^k) - Eg_{1n}^{(k)}(z) \\ = \;& E\frac{1}{N}tr\{A_n(A_n - zI)^{-1}\Gamma_n^{-1}(z)T_{1n}^k\} \\ & + zEg_{2n}(z)E\frac{1}{N}tr\{(A_n - zI)^{-1}\Gamma_n^{-1}(z)T_{1n}^{k+1}\}. \end{aligned} \tag{4.6.51}$$

Hence, at this stage, we can see (4.6.39) and (4.6.40) is a consequence of the following theorem.

Theorem 4.6.2. *For each nonnegative integer k, let*

$$\begin{aligned} \theta_n^{(k)}(z) &= E\frac{1}{N}tr\{(\forall_n - zI)^{-1}\forall_n R_n^{-1}(z)T_{2n}^k\} \\ & \quad + zEg_{1n}(z)E\frac{1}{N}tr\{(\forall_n - zI)^{-1}R_n^{-1}(z)T_{2n}^{k+1}\}, \end{aligned}$$

and

$$\begin{aligned} \vartheta_n^{(k)}(z) &= E\frac{1}{N}tr\{A_n(A_n - zI)^{-1}\Gamma_n^{-1}(z)T_{1n}^k\} \\ & \quad + zEg_{2n}(z)E\frac{1}{N}tr\{(A_n - zI)^{-1}\Gamma_n^{-1}(z)T_{1n}^{k+1}\}. \end{aligned}$$

Then

$$(1)\ \lim_{n\to\infty}\theta_n^{(k)}(z) = 0. \qquad (2)\ \lim_{n\to\infty}\vartheta_n^{(k)}(z) = 0. \tag{4.6.52}$$

Proof. We first note that from the proof of Lemmas 4.3.4 and 4.3.5, one can see that $Im(zg_{1n}(z) \geq 0$ and $Img_{2n}(z) \geq 0$ always hold. It follows that $\|R_n^{-1}(z)\| \leq \frac{1}{v}$ and, as we proved in (4.5.16) before Lemma 4.5.20, $\limsup_n \|\Gamma_n^{-1}(z)\| \leq \frac{\tau}{v^2 E\delta_{2z}}$, where $\delta_{2z} \equiv \liminf_{n\to\infty} Img_{2n}(z)$. So $\|R_n^{-1}(z)\|$ and $\|\Gamma_n^{-1}(z)\|$ are both uniformly bounded for all n large.

Let us now start with the proof of (4.6.52)(1). Write

$$\forall_n = \frac{1}{N}\sum_{ij} \bar{x}_{ij} T_{2n} X_{ij} T_{1n} f_j e_i' + \frac{1}{N}\sum_{ij} |x_{ij}|^2 \zeta_{jj}^{(1)} T_{2n} e_i e_i'.$$

Then

$$\begin{aligned} & \frac{1}{N} tr\{(\forall_n - zI)^{-1} \forall_n R_n^{-1}(z) T_{2n}^k\} \\ = \;& \frac{1}{N^2}\sum_{ij} \bar{x}_{ij} e_i' R_n^{-1}(z) T_{2n}^k (\forall_n - zI)^{-1} T_{2n} X_{ij} T_{1n} f_j \\ & + \frac{1}{N^2}\sum_{ij} \zeta_{jj}^{(1)} |x_{ij}|^2 e_i' R_n^{-1}(z) T_{2n}^k (\forall_n - zI)^{-1} T_{2n} e_i. \end{aligned} \tag{4.6.53}$$

By Corollary 4.3.1, we first show

$$\begin{aligned} & E\frac{1}{N^2}\sum_{ij} \bar{x}_{ij} e_i' R_n^{-1}(z) T_{2n}^k (\forall_n - zI)^{-1} T_{2n} X_{ij} T_{1n} f_j \\ = \;& -\psi_n^{(2)}(z) E\frac{1}{N} tr\{(\forall_n - zI)^{-1} R_n^{-1}(z) T_{2n}^{k+1}\} + o(1). \end{aligned} \tag{4.6.54}$$

By using the resolvent identity together with the equality of $\forall_n - \forall_{ij}$ given in

(4.1.13), we get

$$\begin{aligned}
& E\frac{1}{N^2}\sum_{ij}\bar{x}_{ij}e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}X_{ij}T_{1n}f_j \\
= \quad & E\frac{1}{N^2}\sum_{ij}\bar{x}_{ij}e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij} - zI)^{-1}T_{2n}X_{ij}T_{1n}f_j \\
& -E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2(e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i)\times b_{jj}^{(1)} \\
& -E\frac{1}{N^3}\sum_{ij}\bar{x}_{ij}^2(e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}X_{ij}T_{1n}f_j)\times \sigma_{ij}^{(1)} \\
& -E\frac{1}{N^3}\sum_{ij}\zeta_{jj}^{(1)}x_{ij}\bar{x}_{ij}^2(e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i)\times \sigma_{ij}^{(1)}, \quad (4.6.55)
\end{aligned}$$

where the notations $b_{jj}^{(1)}$ and $\sigma_{ij}^{(1)}$ are as defined in Section 4.1.

It is easy to see

$$E\frac{1}{N^2}\sum_{ij}\bar{x}_{ij}e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij} - zI)^{-1}T_{2n}X_{ij}T_{1n}f_j = 0. \quad (4.6.56)$$

By using Theorem 4.2.1 and the fact $\|\Gamma_n^{-1}(z)\|$ is bounded, we have

$$|e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}X_{ij}T_{1n}f_j| \leq K\mathbf{B}(i,j), \quad (4.6.57)$$

and

$$|e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i| \leq K. \quad (4.6.58)$$

It follows that, using Lemma 4.5.17 and Holder's inequality,

$$\begin{aligned}
& |E\frac{1}{N^3}\sum_{ij}\zeta_{jj}^{(1)}x_{ij}\bar{x}_{ij}^2(e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i)\times \sigma_{ij}^{(1)}| \\
\leq \quad & \frac{1}{N^3}E\sum_{ij}|x_{ij}|^3|\sigma_{ij}^{(1)}| \leq K\frac{\delta_n}{n^2\sqrt{n}}E\sum_{ij}|\sigma_{ij}^{(1)}| \\
\leq \quad & K\frac{\delta_n}{\sqrt{n}},
\end{aligned}$$

and

$$
\begin{aligned}
& |E\frac{1}{N^3}\sum_{ij}\bar{x}_{ij}^2(e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}X_{ij}T_{1n}f_j) \times \sigma_{ij}^{(1)}| \\
\leq\ & K\frac{1}{N^3}\sum_{ij}|x_{ij}|^2[\mathbf{B}(i,j)]|\sigma_{ij}^{(1)}| \\
\leq\ & K\frac{1}{N^3}\left(E\sum_{ij}[\mathbf{B}(i,j)]^2\right)^{1/2}\left(E\sum_{ij}|\sigma_{ij}^{(1)}|^2\right)^{1/2} \\
\leq\ & K\frac{1}{\sqrt{n}}.
\end{aligned}
$$

Hence from (4.6.55), to show (4.6.54), it suffices to show

$$
\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2 e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i b_{jj}^{(1)} \\
=\ & \psi_n^{(2)}(z)E\frac{1}{N}tr\{(\forall_n - zI)^{-1}R_n^{-1}(z)T_{2n}^{k+1}\} + o(1). \qquad (4.6.59)
\end{aligned}
$$

Then we write

$$
\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2 e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i b_{jj}^{(1)} \\
=\ & E\frac{1}{N^3}\sum_{ij}e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i \underline{b}_{jj}^{(1)} \\
& +E\frac{1}{N^3}\sum_{ij}e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i(b_{jj}^{(1)} - \underline{b}_{jj}^{(1)}) \\
& +E\frac{1}{N^3}\sum_{ij}e_i'R_n^{-1}(z)T_{2n}^k\{(\forall_{ij} - zI)^{-1} - (\forall_n - zI)^{-1}\}T_{2n}e_i b_{jj}^{(1)} \\
& -E\frac{1}{N^3}\sum_{ij}(1 - E|x_{ij}|^2)E(e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij} - zI)^{-1}T_{2n}e_i b_{jj}^{(1)}) \\
& -E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2 e_i'R_n^{-1}(z)T_{2n}^k\{(\forall_{ij} - zI)^{-1} - (\forall_n - zI)^{-1}\}T_{2n}e_i b_{jj}^{(1)}. \qquad (4.6.60)
\end{aligned}
$$

Note that by the proof Lemma 4.5.19 as well as the fact $R_n^{-1}(z)T_{2n} = T_{2n}R_n^{-1}(z)$,

$$
\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}e_i'R_n^{-1}(z)T_{2n}^k(\forall_n - zI)^{-1}T_{2n}e_i \underline{b}_{jj}^{(1)} \\
=\ & E\left(\Psi_n^{(2)}(z)\frac{1}{N}tr\{(\forall_n - zI)^{-1}T_{2n}R_n^{-1}(z)T_{2n}^k\}\right) \\
=\ & \psi_n^{(2)}(z)E\frac{1}{N}tr\{(\forall_n - zI)^{-1}R_n^{-1}(z)T_{2n}^{k+1}\} + o(1). \qquad (4.6.61)
\end{aligned}
$$

By using the estimate given in (4.6.58) and the estimate of $b_{jj}^{(1)}$ in Lemma 4.5.14, we have

$$
\begin{aligned}
& |\frac{1}{N^3}\sum_{ij}(1-E|x_{ij}|^2)E(e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij}-zI)^{-1}T_{2n}e_ib_{jj}^{(1)})| \\
\leq\ & K\frac{1}{n^2\sqrt{n}}\sum_{ij}(1-E|x_{ij}|^2)E\mathbf{B}(i,j) \\
\leq\ & K\left(\frac{1}{nN}\sum_{ij}(1-E|x_{ij}|^2)^2\right)^{1/2}\left(\frac{1}{n^3}E\sum_{ij}[\mathbf{B}(i,j)]^2\right)^{1/2} \\
\leq\ & K\left(\frac{1}{nN}\sum_{ij}(1-E|x_{ij}|)^2\right)^{1/2} \\
\to\ & 0. \qquad (4.6.62)
\end{aligned}
$$

By using the estimate given in (4.6.58) and the estimate on $|b_{jj}^{(1)}-\underline{b}_{jj}^{(1)}|$ in (4.6.27), we get

$$
\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}|e_i'R_n^{-1}(z)T_{2n}^k(\forall_n-zI)^{-1}T_{2n}e_i||b_{jj}^{(1)}-\underline{b}_{jj}^{(1)}| \\
\leq\ & K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|[\mathbf{B}(i,j)]^2 \\
\leq\ & K/\sqrt{n}. \qquad (4.6.63)
\end{aligned}
$$

Further, there is constant K depending only on τ, v and k such that

$$
\begin{aligned}
& |e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij}-zI)^{-1}(\forall_n-\forall_{ij})(\forall_n-zI)^{-1}T_{2n}e_i| \\
\leq\ & \frac{1}{N}|x_{ij}||e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij}-zI)^{-1}T_{2n}e_if_j'T_{1n}X_{ij}^*(\forall_n-zI)^{-1}T_{2n}e_i| \\
& +\frac{1}{N}|x_{ij}||e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij}-zI)^{-1}T_{2n}X_{ij}T_{1n}f_je_i'(\forall_n-zI)^{-1}T_{2n}e_i| \\
& +\frac{\tau}{N}|x_{ij}|^2|e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij}-zI)^{-1}T_{2n}e_ie_i'(\forall_n-zI)^{-1}T_{2n}e_i| \\
\leq\ & K\frac{1}{N}|x_{ij}|\mathbf{B}(i,j). \qquad (4.6.64)
\end{aligned}
$$

Together with the estimate of $|b_{jj}^{(1)}| \leq K\sqrt{n}\mathbf{B}(i,j)$ given in Lemma 4.5.14, this implies

$$\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2|e_i'R_n^{-1}(z)T_{2n}^k\{(\forall_{ij}-zI)^{-1}-(\forall_n-zI)^{-1}\}T_{2n}e_i||b_{jj}^{(1)}| \\
\leq\ & K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|^3[\mathbf{B}(i,j)]^2 \\
\leq\ & K\delta_n, \qquad (4.6.65)
\end{aligned}$$

and

$$\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}|e_i'R_n^{-1}(z)T_{2n}^k\{(\forall_{ij}-zI)^{-1}-(\forall_n-zI)^{-1}\}T_{2n}e_i||b_{jj}^{(1)}| \\
\leq\ & K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|[\mathbf{B}(i,j)]^2 \\
\leq\ & K/\sqrt{n}. \qquad (4.6.66)
\end{aligned}$$

Combining (4.6.61)-(4.6.63), (4.6.65)-(4.6.66) gives us, in view of (4.6.60), immediately (4.6.59). Therefore, the proof of (4.6.54) is complete.

In the following, to finish the proof of (4.6.52)(1), we prove

$$\begin{aligned}
& E\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(1)}|x_{ij}|^2e_i'R_n^{-1}(z)T_{2n}^k(\forall_n-zI)^{-1}T_{2n}e_i \\
=\ & \frac{1}{N}tr(T_{1n})E\frac{1}{N}tr\{(\forall_n-zI)^{-1}R_n^{-1}(z)T_{2n}^{k+1}\}+o(1). \qquad (4.6.67)
\end{aligned}$$

By using (4.6.64), we have

$$\begin{aligned}
& |E\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(1)}|x_{ij}|^2e_i'R_n^{-1}(z)T_{2n}^k\{(\forall_{ij}-zI)^{-1}-(\forall_n-zI)^{-1}\}T_{2n}e_i| \\
\leq\ & K\frac{1}{n^3}\sum_{ij}|x_{ij}|^3\mathbf{B}(i,j) \\
\leq\ & K\delta_n,
\end{aligned}$$

and

$$
\begin{aligned}
& |E\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(1)}e_i'R_n^{-1}(z)T_{2n}^k\{(\forall_{ij}-zI)^{-1}-(\forall_n-zI)^{-1}\}T_{2n}e_i| \\
\le\ & K\frac{1}{n^3}\sum_{ij}|x_{ij}|\mathbf{B}(i,j) \\
\le\ & K/\sqrt{n}.
\end{aligned}
$$

Note that the estimate in (4.6.58) remains true with the matrix $(\forall_n - zI)^{-1}$ replaced by $(\forall_{ij} - zI)^{-1}$. It follows then

$$
\begin{aligned}
& |\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(1)}(1-E|x_{ij}|^2)E(e_i'R_n^{-1}(z)T_{2n}^k(\forall_{ij}-zI)^{-1}T_{2n}e_i)| \\
\le\ & K\frac{1}{nN}\sum_{ij}(1-E|x_{ij}|^2) \\
\to\ & 0.
\end{aligned}
$$

With these results, (4.6.67) follows from the following fact

$$
\begin{aligned}
& E\frac{1}{N^2}\sum_{ij}\zeta_{jj}^{(1)}e_i'R_n^{-1}(z)T_{2n}^k(\forall_n-zI)^{-1}T_{2n}e_i \\
=\ & \frac{1}{N}tr(T_{1n})E\frac{1}{N}tr\{(\forall_n-zI)^{-1}R_n^{-1}(z)T_{2n}^{k+1}\}.
\end{aligned}
$$

Note that from Corollary 4.3.1,

$$
-zg_{1n}(z) = \frac{1}{N}tr(T_{1n}) - \Psi_n^{(2)}(z).
$$

Therefore, (4.6.54) and (4.6.67) together complete the proof of (4.6.52)(1).

We now proceed into the proof of (4.6.52)(2). Write

$$
A_n = \frac{1}{N}\sum_{ij}\bar{x}_{ij}T_{1n}f_je_i'T_{2n}X_{ij} + \frac{1}{N}\sum_{ij}|x_{ij}|^2\xi_{ii}^{(1)}T_{1n}f_jf_j'.
$$

For notational convenience, write

$$
\Lambda_n^{(k)}(z) = (A_n - zI)^{-1}\Gamma_n^{-1}(z)T_{1n}^k \quad\text{and}\quad \Lambda_{ij}^{(k)}(z) = (A_{ij} - zI)^{-1}\Gamma_n^{-1}(z)T_{1n}^k.
$$

Then

$$\begin{aligned}
& \frac{1}{N} tr\{A_n(A_n - zI)^{-1}\Gamma_n^{-1}(z)T_{1n}^k\} \\
= \; & \frac{1}{N^2}\sum_{ij} \bar{x}_{ij} e_i' T_{2n} X_{ij} \Lambda_n^{(k+1)}(z) f_j + \frac{1}{N^2}\sum_{ij} |x_{ij}|^2 \xi_{ii}^{(1)} f_j' \Lambda_n^{(k+1)}(z) f_j.
\end{aligned}$$

To develop the following derivations, we need to use the bound defined in Lemma 4.5.20 to prove some preliminary results which are in parallel with those we proved in Lemmas 4.5.13, 4.5.15 and 4.5.16. Since their proofs demand nothing new, we shall not prove them in details. These results are outlined below.

Proposition 4.6.1. *There is a constant K depending only on τ, v and k such that*

$$\begin{aligned}
&(a) \qquad E\sum_{ij} |\hat{a}_{ij}^{(k)}|^2 \le Kn^2, \\
&(b) \qquad |e_i' T_{2n} X_{ij} \Lambda_n^{(k+1)}(z) f_j| \le K\mathbf{B}_2(i,j). \\
&(c) \qquad |e_i' T_{2n} X_{ij} \Lambda_{ij}^{(k+1)}(z) f_j| \le K\mathbf{B}_2(i,j). \\
&(d) \qquad |f_j' \Lambda_n^{(k+1)}(z) f_j| \le K\frac{1}{\sqrt{n}}\mathbf{B}_2(i,j). \\
&(e) \qquad |f_j' \Lambda_{ij}^{(k+1)}(z) f_j| \le K\frac{1}{\sqrt{n}}\mathbf{B}_2(i,j). \\
&(f) \qquad |f_j' \{\Lambda_{ij}^{(k+1)}(z) - \Lambda_n^{(k+1)}(z)\} f_j| \le K\frac{|x_{ij}|}{n\sqrt{n}}[\mathbf{B}_2(i,j)]^2. \\
&(g) \qquad |e_i' T_{2n} X_{ij} \{\Lambda_{ij}^{(k+1)}(z) - \Lambda_n^{(k+1)}(z)\} f_j| \le K\frac{|x_{ij}|}{n}[\mathbf{B}_2(i,j)]^2.
\end{aligned}$$

By Corollary 4.3.1, $-zEg_{2n}(z) = \frac{1}{N}tr(T_{2n}) - \phi_n^{(2)}(z)$. Thus (4.6.52)(2) is a consequence of the following two facts:

$$E\frac{1}{N^2}\sum_{ij}\bar{x}_{ij}e_i'T_{2n}X_{ij}\Lambda_n^{(k+1)}(z)f_j \to -\phi_n^{(2)}(z)E\frac{1}{N}tr\Lambda_n^{(k+1)}(z), \quad (4.6.68)$$

$$E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\xi_{ii}^{(1)}f_j'\Lambda_n^{(k+1)}(z)f_j \to \frac{1}{N}tr(T_{2n})E\frac{1}{N}tr\Lambda_n^{(k+1)}(z). \quad (4.6.69)$$

We first show (4.6.68). Using the resolvent identity together with the equality of $A_n - A_{ij}$ in (4.1.9) gives

$$\begin{aligned}
& E\frac{1}{N^2}\sum_{ij}\bar{x}_{ij}e_i'T_{2n}X_{ij}\Lambda_n^{(k+1)}(z)f_j \\
= \; & E\frac{1}{N^2}\sum_{ij}\bar{x}_{ij}(e_i'T_{2n}X_{ij}\Lambda_{ij}^{(k+1)}(z)f_j) - E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2 d_{ii}^{(1)}(f_j'\Lambda_n^{(k+1)}(z)f_j) \\
& -E\frac{1}{N^3}\sum_{ij}\bar{x}_{ij}^2\hat{a}_{ij}^{(1)}(e_i'T_{2n}X_{ij}\Lambda_n^{(k+1)}(z)f_j) \\
& -E\frac{1}{N^3}\sum_{ij}x_{ij}\bar{x}_{ij}^2\xi_{ii}^{(1)}\hat{a}_{ij}^{(1)}(f_j'\Lambda_n^{(k+1)}(z)f_j).
\end{aligned}$$

Again, the first term on the right-hand side is equal to 0. Recall that $\hat{a}_{ij}^{(1)} = e_i'T_{2n}X_{ij}(A_{ij} - zI)^{-1}T_{1n}f_j$ is independent of x_{ij}. Thus from Hölder's inequality and Proposition 4.6.1(a), (b) and (d), it follows

$$\begin{aligned}
& |E\frac{1}{N^3}\sum_{ij}\bar{x}_{ij}^2\hat{a}_{ij}^{(1)}(e_i'T_{2n}X_{ij}\Lambda_n^{(k+1)}(z)f_j)| \\
\le \; & K\frac{1}{n^3}\sum_{ij}|x_{ij}|^2|\hat{a}_{ij}^{(1)}|\mathbf{B}_2(i,j) \\
\le \; & K\frac{1}{n^3}\left(E\sum_{ij}[\mathbf{B}_2(i,j)]^2\right)^{1/2}\left(E\sum_{ij}|\hat{a}_{ij}^{(1)}|^2\right)^{1/2} \\
\le \; & K/\sqrt{n},
\end{aligned}$$

and

$$\begin{aligned}
& |E\frac{1}{N^3}\sum_{ij}x_{ij}\bar{x}_{ij}^2\xi_{ii}^{(1)}\hat{a}_{ij}^{(1)}(f_j'\Lambda_n^{(k+1)}(z)f_j)| \\
\le \; & K\frac{1}{n^3}E\sum_{ij}|\hat{a}_{ij}^{(1)}|\mathbf{B}_2(i,j) \\
\le \; & K/\sqrt{n}.
\end{aligned}$$

Hence (4.6.68) is a consequence of

$$E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2 d_{ii}^{(1)}(f_j'\Lambda_n^{(k+1)}(z)f_j) \to \phi_n^{(2)}(z)E\frac{1}{N}tr\Lambda_n^{(k+1)}(z). \tag{4.6.70}$$

We write

$$\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2 d_{ii}^{(1)}(f_j'\Lambda_n^{(k+1)}(z)f_j) \\
= \; & E\frac{1}{N^3}\sum_{ij}\underline{d}_{ii}^{(1)}(f_j'\Lambda_n^{(k+1)}(z)f_j) + E\frac{1}{N^3}\sum_{ij}(d_{ii}^{(1)} - \underline{d}_{ii}^{(1)})(f_j'\Lambda_n^{(k+1)}(z)f_j) \\
& +E\frac{1}{N^3}\sum_{ij} d_{ii}^{(1)}(f_j'\Lambda_{ij}^{(k+1)}(z)f_j - f_j'\Lambda_n^{(k+1)}(z)f_j) \\
& -\frac{1}{N^3}\sum_{ij}(1-E|x_{ij}|^2)E(d_{ii}^{(1)}f_j'\Lambda_{ij}^{(k+1)}(z)f_j) \\
& -E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2 d_{ii}^{(1)}(f_j'\Lambda_{ij}^{(k+1)}(z)f_j - f_j'\Lambda_n^{(k+1)}(z)f_j).
\end{aligned} \tag{4.6.71}$$

By definition and the proof of Lemma 4.5.19, we have

$$\begin{aligned}
& E\frac{1}{N^3}\sum_{ij}\underline{d}_{ii}^{(1)}(f_j'\Lambda_n^{(k+1)}(z)f_j) \\
= \; & E\left(\Phi_n^{(2)}(z)\frac{1}{N}tr\Lambda_n^{(k+1)}(z)\right) \\
= \; & \phi_n^{(2)}(z)E\frac{1}{N}tr\Lambda_n^{(k+1)}(z) + o(1).
\end{aligned} \tag{4.6.72}$$

By Hölder's inequality, using Proposition 4.6.1(e) as well as the estimate in Lemma 4.5.13 that $|d_{ii}^{(1)}| \le K\sqrt{n}\mathbf{B}(i,j) \le K\sqrt{n}\mathbf{B}_2(i,j)$,

$$\begin{aligned}
& |\frac{1}{N^3}\sum_{ij}(1-E|x_{ij}|^2)E(d_{ii}^{(1)}f_j'\Lambda_{ij}^{(k+1)}(z)f_j)| \\
\le \; & K\left(\frac{1}{nN}\sum_{ij}(1-E|x_{ij}|^2)^2\right)^{1/2}\left(\frac{1}{n^4}E\sum_{ij}[\mathbf{B}_2(i,j)]^4\right)^{1/2} \\
\le \; & K\left(\frac{1}{nN}\sum_{ij}(1-E|x_{ij}|^2)\right)^{1/2} \\
\to \; & 0.
\end{aligned} \tag{4.6.73}$$

By Proposition 4.6.1(f),

$$\begin{aligned} & |E\frac{1}{N^3}\sum_{ij}|x_{ij}|^2 d_{ii}^{(1)}(f_j'\Lambda_{ij}^{(k+1)}(z)f_j - f_j'\Lambda_n^{(k+1)}(z)f_j)| \\ \leq \;& K\frac{1}{n^4}E\sum_{ij}|x_{ij}|^3[\mathbf{B}_2(i,j)]^3 \\ \leq \;& K\delta_n, \end{aligned} \tag{4.6.74}$$

and

$$\begin{aligned} & |E\frac{1}{N^3}\sum_{ij} d_{ii}^{(1)}(f_j'\Lambda_{ij}^{(k+1)}(z)f_j - f_j'\Lambda_n^{(k+1)}(z)f_j)| \\ \leq \;& K\frac{1}{n^4}E\sum_{ij}|x_{ij}|[\mathbf{B}_2(i,j)]^3 \\ \leq \;& K/\sqrt{n}. \end{aligned} \tag{4.6.75}$$

By Proposition 4.6.1(d) as well as the estimate on $|d_{ii}^{(1)} - \underline{d}_{ii}^{(1)}|$ in (4.6.14),

$$\begin{aligned} & |E\frac{1}{N^3}\sum_{ij}(d_{ii}^{(1)} - \underline{d}_{ii}^{(1)})(f_j'\Lambda_n^{(k+1)}(z)f_j)| \\ \leq \;& K\frac{1}{n^4}E\sum_{ij}|x_{ij}|[\mathbf{B}_2(i,j)]^3 \\ \leq \;& K/\sqrt{n}. \end{aligned} \tag{4.6.76}$$

Combining results (4.6.72)-(4.6.76) gives (4.6.70). Hence we finished the proof of (4.6.68). We next show (4.6.69).

Similarly, we write

$$\begin{aligned}
& E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\xi_{ii}^{(1)}f_j'\Lambda_n^{(k+1)}(z)f_j \\
= \;& E\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(1)}f_j'\Lambda_n^{(k+1)}(z)f_j \\
& +E\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(1)}(f_j'\Lambda_{ij}^{(k+1)}(z)f_j - f_j'\Lambda_n^{(k+1)}(z)f_j) \\
& -\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(1)}(1-E|x_{ij}|^2)E(f_j'\Lambda_{ij}^{(k+1)}(z)f_j) \\
& -E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\xi_{ii}^{(1)}(f_j'\Lambda_{ij}^{(k+1)}(z)f_j - f_j'\Lambda_n^{(k+1)}(z)f_j). \qquad (4.6.77)
\end{aligned}$$

By definition,

$$E\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(1)}f_j'\Lambda_n^{(k+1)}(z)f_j = \frac{1}{N}tr(T_{2n})E\frac{1}{N}tr\Lambda_n^{(k+1)}(z). \qquad (4.6.78)$$

By Hölder's inequality and Proposition 4.6.1(e),

$$\begin{aligned}
& |\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(1)}(1-E|x_{ij}|^2)E(f_j'\Lambda_{ij}^{(k+1)}(z)f_j)| \\
\leq \;& K\left(\frac{1}{nN}\sum_{ij}(1-E|x_{ij}|^2)^2\right)^{1/2}\left(\frac{1}{n^3}E\sum_{ij}[\mathbf{B}_2(i,j)]^2\right)^{1/2} \\
\leq \;& K\left(\frac{1}{nN}\sum_{ij}(1-E|x_{ij}|^2)\right)^{1/2} \\
\to \;& 0. \qquad (4.6.79)
\end{aligned}$$

By Proposition 4.6.1(f), we get

$$\begin{aligned}
& |E\frac{1}{N^2}\sum_{ij}|x_{ij}|^2\xi_{ii}^{(1)}(f_j'\Lambda_{ij}^{(k+1)}(z)f_j - f_j'\Lambda_n^{(k+1)}(z)f_j)| \\
\leq \;& K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|^3[\mathbf{B}_2(i,j)]^2 \\
\leq \;& K\delta_n, \qquad (4.6.80)
\end{aligned}$$

and

$$\begin{aligned}
& |E\frac{1}{N^2}\sum_{ij}\xi_{ii}^{(1)}(f_j'\Lambda_{ij}^{(k+1)}(z)f_j - f_j'\Lambda_n^{(k+1)}(z)f_j)| \\
\leq\ & K\frac{1}{n^3\sqrt{n}}E\sum_{ij}|x_{ij}|[\mathbf{B}_2(i,j)]^2 \\
\leq\ & K/\sqrt{n}. \qquad (4.6.81)
\end{aligned}$$

In view of (4.6.77), the results in (4.6.78)-(4.6.81) imply (4.6.69). Thus (4.6.52)(2) is also proved. Up to this point, the proof of Theorem 4.6.2 is complete. □

As intended, the results in Theorem 4.6.2 fulfilled our tasks we proposed at the beginning of this section. This can be seen from the following two important consequences derived from it.

Corollary 4.6.2. *For each $z \in \mathbb{C}^+$, almost surely, $(s_{F^{\forall_n}}(z), g_{1n}(z), g_{2n}(z))$ converges to some non-random limit $(\underline{s}(z), g_1(z), g_2(z))$ satisfying the system of equations (1.2.2) in the sense that it is the unique solution to (1.2.2) in the set (1.2.3).*

Proof. By Lemmas 4.5.19, 4.3.3-4.3.5, to show $(Es_{F^{\forall_n}}(z), Eg_{1n}(z), Eg_{2n}(z))$ converges to some limit satisfying the system (1.2.2) is already sufficient. By Lemma 4.3.6, consider any convergent sequence $(Es_{F^{\forall_{n_m}}}(z), Eg_{1n_m}(z), Eg_{2n_m}(z))$, whose limit is denoted by $(\underline{s}(z), g_1(z), g_2(z))$. From Corollary 4.6.1, we have

$$\underline{s}(z) = -z^{-1} - g_1(z)g_2(z). \qquad (4.6.82)$$

By (4.6.48) and (4.6.49), we have

$$\phi_n^{(k)}(z) + zEg_{1n}(z)\frac{1}{N}tr(R_n^{-1}(z)T_{2n}^k) = -z\theta_n^{(k-1)}(z) \to 0, \qquad (4.6.83)$$

$$\psi_n^{(k)}(z) + zEg_{2n}(z)\frac{1}{N}tr(\Gamma_n^{-1}(z)T_{1n}^k) = -z\vartheta_n^{(k-1)}(z) \to 0. \qquad (4.6.84)$$

Let

$$f_{1n}(y) = \frac{y^k}{-z - zEg_{1n}(z)y}, \quad f_1(y) = \frac{y^k}{-z - zEg_1(z)y},$$
$$f_{2n}(x) = \frac{x^k}{-z - zEg_{2n}(z)x}, \quad f_2(x) = \frac{x^k}{-z - zEg_2(z)x}.$$

Due to $Im(zg_{1n}(z)) \geq 0$ and $Im(zg_1(z)) > 0$, it is easy to see for $y \in [0, \tau]$, $|f_{1n}(y)| \leq \frac{\tau^k}{v}$, $|f_1(y)| \leq \frac{\tau^k}{v}$ and $f_1(y)$ is bounded continuous function. At the same time, using the same type of argument as we used to prove (4.5.16) before Lemma 4.5.20 in Section 4.5, we can prove for $x \in [-\tau, \tau]$, $\limsup_n |f_{2n}(x)| \leq \frac{\tau^{k+1}}{v^2 E\delta_{2z}}$, $|f_2(x)| \leq \frac{\tau^{k+1}}{v^2 E\delta_{2z}}$ and $f_2(x)$ is bounded continuous function. Further, it can be verified that

$$|f_{1n}(y) - f_1(y)| \leq \frac{\tau^{k+1}}{v^2}|z||Eg_{1n}(z) - g_1(z)|,$$

and

$$|f_{2n}(x) - f_2(x)| \leq \frac{\tau^{k+3}}{v^3(E\delta_{2z})}\frac{|Eg_{2n}(z) - Eg_2(z)|}{EImg_{2n}(z)},$$

so that as $n_m \to \infty$

$$\sup_{y\in[0,\tau]} |f_{1n_m}(y) - f_1(y)| \to 0, \qquad (4.6.85)$$

$$\sup_{x\in[-\tau,\tau]} |f_{2n_m}(x) - f_2(x)| \to 0. \qquad (4.6.86)$$

It follows, by (4.6.85) and assumption (iii) in Theorem 1.2.1, that

$$\begin{aligned}
& |\frac{1}{N} tr(T_{2n_m}^k R_{n_m}^{-1}(z)) - \int f_1(y) dH_2(y)| \\
= \; & |\int f_{1n_m}(y) dF^{T_{2n_m}}(y) - \int f_1(y) dH_2(y)| \\
\leq \; & |\int (f_{1n_m}(y) - f_1(y)) dF^{T_{2n_m}}(y)| + |\int f_1(y) dF^{T_{2n_m}}(y) - \int f_1(y) dH_2(y)| \\
\to \; & 0,
\end{aligned}$$

and hence

$$\frac{1}{N} tr(T_{2n_m}^k R_{n_m}^{-1}(z)) \to \int f_1(y) dH_2(y).$$

Similarly, (4.6.86) together with assumption (iii) give us

$$\frac{1}{N} tr(T_{1n_m}^k \Gamma_{n_m}^{-1}(z)) \to c \int f_2(x) dH_1(x).$$

Therefore, from (4.6.83) and (4.6.84), we get

$$\phi_{n_m}^{(k)}(z) \to \int \frac{g_1(z) y^k}{1 + g_1(z) y} dH_2(y), \tag{4.6.87}$$

$$\psi_{n_m}^{(k)}(z) \to c \int \frac{g_2(z) x^k}{1 + g_2(z) x} dH_1(x). \tag{4.6.88}$$

In case of $k = 1$, we get (4.6.41) and hence (4.6.42) and (4.6.43). Therefore, the limit $(\underline{s}(z), g_1(z), g_2(z))$ satisfies the system of equations (1.2.2). The corollary follows. $\square$

Corollary 4.6.3. *Given any positive integer k, for every $z \in \mathbb{C}^+$, almost* *surely,*

$$\Phi_n^{(k)}(z) \to \int \frac{g_1(z)y^k}{1+g_1(z)y} dH_2(y), \tag{4.6.89}$$

$$\Psi_n^{(k)}(z) \to c\int \frac{g_2(z)x^k}{1+g_2(z)x} dH_1(x), \tag{4.6.90}$$

$$g_{1n}^{(k)}(z) \to c\int \frac{x^k}{-z-zg_2(z)x} dH_1(x), \tag{4.6.91}$$

$$g_{2n}^{(k)}(z) \to \int \frac{y^k}{-z-zg_1(z)y} dH_2(y). \tag{4.6.92}$$

Proof. In view of the relationship between $\Phi_n^{(k)}(z)$, $\Psi_n^{(k)}(z)$ and $g_{1n}^{(k)}(z)$, $g_{2n}^{(k)}(z)$ in (4.6.44) and (4.6.45), we need only show (4.6.89) and (4.6.90).

Now that $Eg_{1n}(z) \to g_1(z)$ and $Eg_{2n}(z) \to g_2(z)$, we have

$$\frac{1}{N} tr(T_{2n}^k R_n^{-1}(z)) \to \int f_1(y) dH_2(y),$$

and

$$\frac{1}{N} tr(T_{1n}^k \Gamma_n^{-1}(z)) \to c\int f_2(x) dH_1(x).$$

By (4.6.83) and (4.6.84), we then get (4.6.89) and (4.6.90). By Lemma 4.5.19, the corollary follows. □

Chapter 5

Sparse Random Matrices

The main concern of the present chapter is to prove the semicircle law for the class of large sparse random matrices which can be expressed as the Hadamard products A_p of a normalized sample covariance matrix B_m and a sparsing matrix D_m. The definitions of these matrices were formulated in Definitions 1.3.1-1.3.3. Our object in the present chapter is then to use the moment method to prove Theorem 1.3.1.

The organization of the chapter is as follows. In the first section, as a preliminary preparation, we the truncation and centralization technique to simply the underlying matrices. In the second section, we prove the semicircle law for the matrices after simplification. We have introduced in Section 1.3 the main idea underlying our assumptions on the matrices, but due to consideration of the organization of contents, we did not explain in details there their possible effects. So in the last section, we give a detailed discussion on the implications of our conditions.

5.1 Truncation and Centralization Treatment

The aim of applying the truncation and centralization technique is again to possess more convenient conditions for later arguments. This procedure regarding the matrices of the present concern will involve two steps:

Step 1. Replace the diagonal elements of the Hadamard products A_p by zeros;

Step 2. Truncate and centralize the elements of the matrix $X_{m,n}$.

5.1.1 Removal of the Diagonal Elements of A_p.

For any $\varepsilon > 0$, denote by $\widehat{A}_p$ the matrix obtained from A_p by replacing the diagonal elements of A_p whose absolute values are greater than ε with 0 and denote by $\widetilde{A}_p$ the matrix obtained from A_p by replacing all diagonal elements of A_p with 0.

Proposition 5.1.1. *Under the assumptions of Theorem 1.3.1,*

$$L^3(F^{\widehat{A}_p}, F^{\widetilde{A}_p}) \leq \varepsilon^2,$$

and almost surely,

$$\|F^{\widehat{A}_p} - F^{A_p}\| \to 0.$$

Proof. The first conclusion is a trivial consequence of the difference inequality in Lemma 2.1.2. As for the second conclusion, we use the rank inequality in Lemma 2.1.1 which gives

$$\|F^{\widehat{A}_p} - F^{A_p}\| \quad \leq \quad \frac{1}{m}\sum_{i=1}^{m} I_{(|\frac{1}{\sqrt{np}}\sum_{k=1}^{n}(|x_{ik}|^2-\sigma^2)d_{ii}|>\varepsilon)}.$$

By condition (X3) in Definition 1.3.2, we have

$$\sum_{i=1}^{m} P(|\frac{1}{\sqrt{np}}\sum_{k=1}^{n}(|x_{ik}|^2-\sigma^2)d_{ii}|>\varepsilon)=o(m).$$

Using Bernstein's inequality in Lemma 2.1.3, it then follows for any constant $\eta > 0$ there exists some constant $b > 0$ such that

$$\begin{aligned} P(\|F^{\widehat{A}_p}-F^{A_p}\|\geq\eta) &\leq P(\sum_{i=1}^{m} I_{(|\frac{1}{\sqrt{np}}\sum_{k=1}^{n}(|x_{ik}|^2-\sigma^2)d_{ii}|>\varepsilon)}\geq\eta m) \\ &\leq 2e^{-bm}. \end{aligned}$$

By Borel-Cantelli's lemma in Lemma 2.1.5, we therefore get the second conclusion in the proposition holds. This complete the proof. □

Combining the two conclusions in Proposition 5.1.1 gives us with probability one $L(F^{A_p}, F^{\widetilde{A}_p}) \to 0$, which implies without loss of generality, we may assume the diagonal elements of A_p are zero. Hence, in the sequel, we conveniently assume $d_{ii} = 0$ for all $i = 1, \cdots, m$.

5.1.2 Truncation and Centralization of the Entries of $X_{m,n}$.

Select a sequence of numbers $\eta_n \downarrow 0$ such that conditions $(X2.1)$, $(X2.2)$ still hold with η replaced with η_n. Then correspondingly, we get

$$\frac{1}{mn\eta_n^2}\sum_{ij} E|x_{ij}|^2 I(|x_{ij}|>\eta_n\sqrt[4]{np})\to 0.$$
$$\sum_{u=1}^{\infty}\frac{1}{mn\eta_n^2}\sum_{ij} E|x_{ij}|^2 I[|x_{ij}|>\eta_n\sqrt[4]{np}]<\infty,$$

where still u takes $[p]$, m or n.

Let $\tilde{x}_{ij} = x_{ij} I_{(|x_{ij}| \le \eta_n \sqrt[4]{np})} - E x_{ij} I_{(|x_{ij}| \le \eta_n \sqrt[4]{np})}$ and $\hat{x}_{ij} = x_{ij} - \tilde{x}_{ij}$. Define $\widetilde{B}_m$ with (i,j)th entry $\widetilde{B}[i,j] = \frac{1}{\sqrt{np}} \sum_{k=1}^n \tilde{x}_{ik} \bar{\tilde{x}}_{jk}$ $(i \ne j)$, and denote by $\widetilde{A}_p$ the Hadamard product of $\widetilde{B}_m$ with D_m.

Proposition 5.1.2. *Under condition* $(X2.1)$ *in Definition 1.3.2 and the other assumptions of Theorem 1.3.1, in probability*

$$L(F^{\widetilde{A}_p},\ F^{A_p}) \to 0.$$

If condition $(X2.1)$ *is strengthened to* $(X2.2)$*, then almost surely*

$$L(F^{\widetilde{A}_p},\ F^{A_p}) \to 0, \text{ as } \ u \to \infty,$$

where $u = [p],\ m,$ *or* n *in accordance with the choice of* u *in condition* $(X2.2)$*.*

Proof. By the Difference inequality,

$$\begin{aligned} L^3(F^{\widetilde{A}_p},\ F^{A_p}) &\le \frac{1}{m} \mathrm{tr}[(B_m - \widetilde{B}_m) \circ D_m]^2 \\ &= \frac{1}{mnp} \sum_{i \ne j} \left| \sum_{k=1}^n (x_{ik} \bar{x}_{jk} - \tilde{x}_{ik} \bar{\tilde{x}}_{jk}) d_{ij} \right|^2. \end{aligned}$$

But we have

$$\begin{aligned} & E\left(\frac{1}{mnp} \sum_{i \ne j} \left| \sum_{k=1}^n (x_{ik} \bar{x}_{jk} - \tilde{x}_{ik} \bar{\tilde{x}}_{jk}) d_{ij} \right|^2 \right) \\ = \ & \frac{1}{mnp} \sum_{i \ne j} \sum_{k=1}^n E|x_{ik} \bar{x}_{jk} - \tilde{x}_{ik} \bar{\tilde{x}}_{jk})|^2 E|d_{ij}^2| \\ \le \ & \frac{8\sigma^2}{mnp} \sum_{j=1}^m \sum_{k=1}^n E|\hat{x}_{jk}|^2 \sum_{i=1}^m p_{ij} \\ \le \ & \frac{16\sigma^2}{mn} \sum_{j=1}^m \sum_{k=1}^n E|x_{jk}|^2 I[|x_{jk}| > \eta_n \sqrt[4]{np}], \end{aligned}$$

where in the last step we have used condition $(D2)$ in Definition 1.3.1, which can be easily shown remaining true after the removal of the diagonal elements of D_m.

Thus we can see that if condition $(X2.1)$ is assumed, then the right-hand side of the above inequality converges to 0 and hence the first conclusion of the present proposition follows.

However, if condition $(X2.1)$ is strengthened to $(X2.2)$, then we have

$$\sum_{u=1}^{\infty} \frac{16\sigma^2}{mn} \sum_{ij} E|x_{ij}|^2 I(|x_{ij}| > \eta_n \sqrt[4]{np}) < \infty,$$

thus it follows almost surely

$$L^3(F^{\widetilde{A}_p},\ F^{A_p}) \to 0,$$

as $u \to \infty$, where u takes $[p]$, m or n in accordance with the choice of u in condition $(X2.2)$. The proof of the proposition is complete. □

Proposition 5.1.3. *Under the assumptions of Theorem 1.3.1, letting* $\sigma_{ij}^2 \equiv E|\tilde{x}_{ij}|^2$*, then for any* i*,* j*,* $\sigma_{ij}^2 \le \sigma^2$ *and* $\frac{1}{mn}\sum_{ij} \sigma_{ij}^2 \to \sigma^2$*.*

Proof. The first conclusion is trivial. The second conclusion is a consequence of condition $(X2.1)$ and the following fact:

$$0 \le \sigma^2 - \frac{1}{mn}\sum_{ij} \sigma_{ij}^2 \le \frac{2}{mn}\sum_{ij} E|x_{ij}|^2 I(|x_{ij}| > \eta_n \sqrt[4]{np}) \to 0.$$

Based on the above three propositions, in proving Theorem 1.3.1, we may assume further the conditions included in the following assumption holds.

Assumption 5.1.1. *(i)* $d_{ii} = 0$.

(ii) $Ex_{ij} = 0$, $|x_{ij}| \le \eta_n \sqrt[4]{np}$, $E|x_{ij}|^2 \le \sigma^2$ *and* $\sigma^2 - \frac{1}{mn}\sum_{ij} E|x_{ij}|^2 \to 0$.

Thus in the following section, we shall apply the moment method under both the assumptions in Theorem 1.3.1 and the conditions in Assumption 1.3.1.

5.2 Proof of Theorem 1.3.1 by Moment Method

In this section, under both those assumptions of Theorem 1.3.1 and those of Assumption 5.1.1, we shall deduce that, with probability one, the empirical spectral distributions of A_p, F^{A_p}, converge weakly to the semicircle law F_{sc,σ^2}. Note that the statement in Theorem 1.3.1 that the convergence is in the sense of in probability (almost surely) if condition $(X2.1)$ ($(X2.2)$) is assumed has been proven during the procedure of simplifying our matrices.

We use the moment method introduced in Section 2.2. Denote by M_k and m_k respectively the kth moments of F^{A_p} and $F_{sc,\sigma^2}(x)$. It is easy to calculate

$$m_k = \begin{cases} \frac{\sigma^{4s}(2s)!}{s!(s+1)!}, & \text{if } k = 2s, \\ 0, & \text{if } k = 2s+1. \end{cases}$$

Since $m_{2k} \le \sigma^{4k}2^{2k}$, thus $\{m_k\}_{k=1}^{\infty}$ satisfies Carleman's condition. Based on the moment convergence theorem in Lemma 2.2.1, to show Theorem 1.3.1, we need only show that for each k, almost surely $M_k \to m_k$. However, by Borel-Cantelli's lemma in Lemma 2.1.5, this is obviously a consequence of the following two results:

$$\text{(I)} \qquad E(M_k) = m_k + o(1),$$

$$\text{(II)} \qquad E|M_k - EM_k|^4 = O(\frac{1}{m^2}).$$

5.2.1 Graphs and Their Isomorphic Classes

For any indices $1 \le i_1, \cdots, i_k \le m$, $1 \le j_1, \cdots, j_k \le n$, write

$$\mathbf{i} = (i_1, \cdots, i_k), \quad \mathbf{j} = (j_1, \cdots, j_k).$$

Taking the convention that $i_{k+1} \equiv i_1$, define index set

$$\begin{aligned} \mathcal{I} &= \{(\mathbf{i}, \mathbf{j}) : 1 \le i_v \le m, 1 \le j_v \le n, \\ &\qquad \text{with } i_v \ne i_{v+1} \text{ , for each } 1 \le v \le k\}. \end{aligned}$$

Using the above notations, we can express

$$M_k = \frac{1}{mn^{k/2}p^{k/2}} \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{I}} d_{(\mathbf{i},\mathbf{j})} X_{(\mathbf{i},\mathbf{j})},$$

where

$$d_{(\mathbf{i},\mathbf{j})} = d_{i_1 i_2} \cdots d_{i_k i_1},$$

$$X_{(\mathbf{i},\mathbf{j})} = x_{i_1 j_1} \overline{x}_{i_2 j_1} x_{i_2 j_2} \overline{x}_{i_3 j_2} \cdots \overline{x}_{i_k j_{k-1}} x_{i_k j_k} \overline{x}_{i_1 j_k}.$$

For each pair $(\mathbf{i}, \mathbf{j}) = ((i_1, \cdots, i_k), (j_1, \cdots, j_k)) \in \mathcal{I}$, construct a graph $G(\mathbf{i}, \mathbf{j})$ by plotting the i_v's and j_v's on two parallel straight lines respectively, and then drawing k down edges (i_v, j_v) from i_v to j_v, k up edges (j_v, i_{v+1}) from j_v to i_{v+1}, and another k horizontal edges (i_v, i_{v+1}) from i_v to i_{v+1}. A down edge (i_v, j_v)

corresponds to the variable $x_{i_v j_v}$, an up edge (j_v, i_{v+1}) corresponds to the variable $\overline{x}_{i_{v+1} j_v}$, and a horizontal edge (i_v, i_{v+1}) corresponds to the variable $d_{i_v i_{v+1}}$.

A so defined graph thus corresponds to the product of the variables corresponding to the edges making up this graph. We shall call the subgraph of horizontal edges and their vertices of $G(\mathbf{i}, \mathbf{j})$ the roof of $G(\mathbf{i}, \mathbf{j})$ and denote it as $\overline{G}(\mathbf{i}, \mathbf{j})$ and call the subgraph of vertical edges and their vertices of $G(\mathbf{i}, \mathbf{j})$ the base of $G(\mathbf{i}, \mathbf{j})$ and denote it as $\underline{G}(\mathbf{i}, \mathbf{j})$. By noting that the roof of $G(\mathbf{i}, \mathbf{j})$ depends on $\mathbf{i}$ only, we may simplify the notation of roofs as $\overline{G}(\mathbf{i})$. For any two pairs $(\mathbf{i}, \mathbf{j}) = ((i_1, \cdots, i_k), (j_1, \cdots, j_k))$ and $(\mathbf{i}', \mathbf{j}') = ((i'_1, \cdots, i'_k), (j'_1, \cdots, j'_k))$, the two graphs $G(\mathbf{i}, \mathbf{j})$ and $G(\mathbf{i}', \mathbf{j}')$ are said to be isomorphic if for any $1 \le a_1, a_2, b_1, b_2 \le k$, $i_{a_1} = i_{a_2}$, $j_{b_1} = j_{b_2}$ if and only if $i'_{a_1} = i'_{a_2}$, $j'_{b_1} = j'_{b_2}$. All graphs are classified into isomorphic classes. An isomorphic class is denoted by $\mathcal{G}$. Similarly, two roofs $\overline{G}(\mathbf{i})$ and $\overline{G}(\mathbf{i}')$ are said to be isomorphic if for any $1 \le a_1, a_2 \le k$, $i_{a_1} = i_{a_2}$ if and only if $i'_{a_1} = i'_{a_2}$. An isomorphic roof class is denoted by $\overline{\mathcal{G}}$. For a given $\mathbf{i}$, two graphs $G(\mathbf{i}, \mathbf{j})$ and $G(\mathbf{i}, \mathbf{j}')$ are said to be isomorphic given $\mathbf{i}$ if for any $1 \le b_1, b_2 \le k$, $j_{b_1} = j_{b_2}$ if and only if $j'_{b_1} = j'_{b_2}$. An isomorphic class given $\mathbf{i}$ is denoted by $\underline{\mathcal{G}}(\mathbf{i})$.

These definitions enable us to write further

$$\begin{aligned} M_k &= \frac{1}{mn^{k/2}p^{k/2}} \sum_{\mathbf{i},\mathbf{j}} d_{\overline{G}(\mathbf{i})} X_{\underline{G}(\mathbf{i},\mathbf{j})} \\ &= \frac{1}{mn^{k/2}p^{k/2}} \sum_{\mathcal{G}} \sum_{G(\mathbf{i},\mathbf{j}) \in \mathcal{G}} d_{\overline{G}(\mathbf{i})} X_{\underline{G}(\mathbf{i},\mathbf{j})}. \end{aligned} \tag{5.2.1}$$

It follows

$$E(M_k) = \frac{1}{mn^{k/2}p^{k/2}} \sum_{\mathcal{G}} \sum_{G(\mathbf{i},\mathbf{j}) \in \mathcal{G}} E d_{\overline{G}(\mathbf{i})} E X_{\underline{G}(\mathbf{i},\mathbf{j})}. \tag{5.2.2}$$

Note that when $G(\mathbf{i},\mathbf{j})$ contains a single vertical edge, $EX_{\underline{G}(\mathbf{i},\mathbf{j})} = 0$. Thus we may assume in the following each graph appearing in the summation of EM_k does not contain single vertical edges. Also note that from the definition of the set $\mathcal{I}$, each graph also does not contain any loops of horizontal edges.

Let us denote by l, r, s respectively the numbers of non-coincident vertical edges, non-coincident i_v vertices and non-coincident j_v vertices contained in a graph. It is obvious these numbers must be the same for isomorphic graphs. Further, notice that for each isomorphic class $\mathcal{G}$, there is a unique graph $G(\mathbf{i},\mathbf{j})$ satisfying

$$i_1 = 1, i_{v+1} \leq \max\{i_1, \cdots, i_v\} + 1,$$

$$j_1 = 1, j_{v+1} \leq \max\{j_1, \cdots, j_v\} + 1,$$

which will be called the canonical (or representative) graph of $\mathcal{G}$. We can further define a number q as follows.

Suppose that $\mathcal{G}$ is an isomorphic class having the index r, the number of non-coincident i-vertices in its canonical graph, $G(\mathbf{i},\mathbf{j})$. Then since $\overline{G}(\mathbf{i})$, the roof of $G(\mathbf{i},\mathbf{j})$, is a connected graph, we can select a tree which contains all the r vertices of $\overline{G}(\mathbf{i})$ and exactly $r-1$ edges. Excluding the $r-1$ edges in $\overline{G}_1(\mathbf{i})$, we have $k-(r-1)$ edges left in $\overline{G}(\mathbf{i})$. Then the remaining $k-(r-1)$ edges together with all of the vertices of $\overline{G}(\mathbf{i})$ form another subgraph of $\overline{G}(\mathbf{i})$, which will be denoted by $\overline{G}_2(\mathbf{i})$. The remainder subgraph $\overline{G}_2(\mathbf{i})$ may not be connected. Then with the understanding that each isolated vertex is considered as a connected block, the number q is defined to be the number of connected blocks contained in $\overline{G}_2(\mathbf{i})$.

From the definition of isomorphic class, it is easy to see we may use q for all the other graphs in this isomorphic class.

5.2.2 Preliminary Results

Denote by $\mathcal{G}(l, r, s, q)$ the collection of isomorphic classes with the indices l, r, s and q. Then we have the following two propositions, which will be useful for estimating the terms involved in EM_k.

Proposition 5.2.1. *Under the conditions in Theorem 1.3.1 and Assumption 5.1.1, for any $\mathcal{G} \in \mathcal{G}(l, r, s, q)$, there is a constant K such that*

$$\sum_{\overline{G}(\mathbf{i}) \in \overline{\mathcal{G}}} E|d_{\overline{G}(\mathbf{i})}| \leq K m^{1+\delta(q-1)} p^{r-1-\delta(q-1)}, \tag{5.2.3}$$

and

$$\sum_{\substack{\overline{G}(\mathbf{i}) \in \overline{\mathcal{G}} \\ I fixed}} E|d_{\overline{G}(\mathbf{i})}| \leq K m^{\delta(q-1)} p^{r-1-\delta(q-1)}, \tag{5.2.4}$$

where I is an arbitrarily selected non-coincident i-vertex of $\overline{G}(\mathbf{i})$ while K does not depend on I.

Proof. It is straightforward to see (5.2.3) is a consequence of (5.2.4). We thus need only prove (5.2.4). Since $\overline{G}(\mathbf{i})$ is connected, there must exist $q-1$ edges in $\overline{G}_1(\mathbf{i})$ which make the q blocks of $\overline{G}_2(\mathbf{i})$ connected. From the definition of $\overline{G}_1(\mathbf{i})$, these $q-1$ edges are single in $\overline{G}(\mathbf{i})$ and, together with their vertices, cannot form any cycles. We shall call them bridge edges. Suppose the $(q-1)$ bridge edges are

$$(i_{b_1}, i_{b_1+1}), (i_{b_2}, i_{b_2+1}), \cdots, (i_{b_{q-1}}, i_{b_{q-1}+1});$$

the other $r-q$ edges in $\overline{G}_1(\mathbf{i})$ are

$$(i_{a_1}, i_{a_1+1}), (i_{a_2}, i_{a_2+1}), \cdots, (i_{a_{r-q}}, i_{a_{r-q}+1});$$

the $k-r+1$ edges in $\overline{G}_2(\mathbf{i})$ are

$$(i_{c_1}, i_{c_1+1}), (i_{c_2}, i_{c_2+1}), \cdots, (i_{c_{k-r+1}}, i_{c_{k-r+1}+1}).$$

(Note that q may be equal to 1 so that there are no bridge edges at all, but it is easy to see the proof that follows is still valid). Then by Hölder's inequality, we have

$$\begin{aligned}
& \sum_{\substack{\overline{G}(\mathbf{i})\in\overline{\mathcal{G}} \\ I fixed}} E|d_{\overline{G}(\mathbf{i})}| \\
= \;& E \sum_{\substack{\overline{G}(\mathbf{i})\in\overline{\mathcal{G}} \\ I fixed}} \prod_{u=1}^{q-1} |d_{i_{b_u} i_{b_u+1}}| \prod_{v=1}^{r-q} |d_{i_{a_v} i_{a_v+1}}| \prod_{w=1}^{k-r+1} |d_{i_{c_w} i_{c_w+1}}| \\
\leq \;& \left(E \sum_{\substack{\overline{G}(\mathbf{i})\in\overline{\mathcal{G}} \\ I fixed}} \prod_{u=1}^{q-1} |d_{i_{b_u} i_{b_u+1}}| \prod_{v=1}^{r-q} |d_{i_{a_v} i_{a_v+1}}|^2 \right)^{1/2} \\
& \left(E \sum_{\substack{\overline{G}(\mathbf{i})\in\overline{\mathcal{G}} \\ I fixed}} \prod_{u=1}^{q-1} |d_{i_{b_u} i_{b_u+1}}| \prod_{w=1}^{k-r+1} |d_{i_{c_w} i_{c_w+1}}|^2 \right)^{1/2} \\
\leq \;& K m^{\delta(q-1)} p^{r-1-\delta(q-1)}.
\end{aligned}$$

To get this inequality, we have used the following two results. Namely, by assumptions $(D2)$, $(D3.1)$ and $(D3.2)$,

$$\begin{aligned}
& E \sum_{\substack{\overline{G}(\mathbf{i})\in\overline{\mathcal{G}} \\ I fixed}} \prod_{u=1}^{q-1} |d_{i_{b_u} i_{b_u+1}}| \prod_{v=1}^{r-q} |d_{i_{a_v} i_{a_v+1}}|^2 \\
= \;& \sum_{\substack{\overline{G}(\mathbf{i})\in\overline{\mathcal{G}} \\ I fixed}} \prod_{u=1}^{q-1} E|d_{i_{b_u} i_{b_u+1}}| \prod_{v=1}^{r-q} E|d_{i_{a_v} i_{a_v+1}}|^2 \\
\leq \;& K(m^{\delta} p^{1-\delta})^{q-1} p^{r-q} \\
= \;& K m^{\delta(q-1)} p^{r-1-\delta(q-1)}
\end{aligned}$$

and further by the fact that the p_{ij}'s are uniformly bounded,

$$\begin{aligned}
& E \sum_{\substack{\overline{G}(\mathbf{i}) \in \overline{\mathcal{G}} \\ I fixed}} \prod_{u=1}^{q-1} |d_{i_{b_u} i_{b_u+1}}| \prod_{w=1}^{k-r+1} |d_{i_{c_w} i_{c_w+1}}|^2 \\
= & \sum_{\substack{\overline{G}(\mathbf{i}) \in \overline{\mathcal{G}} \\ I fixed}} \prod_{u=1}^{q-1} E|d_{i_{b_u} i_{b_u+1}}| E \prod_{w=1}^{k-r+1} |d_{i_{c_w} i_{c_w+1}}|^2 \\
\leq & \; K m^{\delta(q-1)} p^{r-1-\delta(q-1)}.
\end{aligned}$$

This completes the proof of the proposition. □

Note that since each graph does not contain single vertical edge, we have $l \leq k$. Since each graph does not contain any loops of horizontal edges, we further have $l \geq 2s$. And as a basic property of connected graph, we have $r + s \leq l + 1$. To estimate the integer q, we need the following proposition.

Proposition 5.2.2. *Let l, s, q be defined as above. Then*

$$l - 2s \geq q - 1. \tag{5.2.5}$$

Proof. Since $\overline{G}(\mathbf{i})$ contains no loops, it follows that each j-vertex is connected with at least two non-coincident vertical edges and hence that $l - 2s \geq 0$, which implies (5.2.5) for the case of $q = 1$. Now assume $q > 1$. Then we have $q - 1$ bridge edges. If (i_v, i_{v+1}) is a bridge edge, then we call the vertex j_v its supporting vertex, while the edges (i_v, j_v), (j_v, i_{v+1}) its supporting edges.

Denote the s non-coincident j-vertices by $J_1, J_2, \cdots, J_s$. For each $1 \leq a \leq s$, denote by l_a the number of non-coincident vertical edges connected with J_a. Then obviously $l = l_1 + l_2 + \cdots + l_s$. Note that each non-coincident j-vertex is composed of at least two j-vertices coincident with each other. For each $1 \leq a \leq s$, denote by t_a the number of bridge edges supported by J_a. Here, of course, t_a is the total number of bridge edges whose supporting j-vertex is from those coincident j-vertices constituting J_a. Then $q - 1 = t_1 + t_2 + \cdots + t_s$. To prove $l - 2s \geq q - 1$, it is sufficient to prove for each $1 \leq a \leq s$, $l_a \geq t_a + 2$.

If $t_a = 0$, then $l_a \geq t_a + 2$ follows simply from the previously stated fact that each j-vertex is connected with at least two non-coincident vertical edges. Now assume $t_a \geq 1$. In view of the property that bridge edges together with their vertices do not form any cycles and may be disconnected among themselves, we shall consider two cases, when the t_a bridge edges (together with their vertices) form exactly one tree and when they form more than one trees disjoint with each other. For the first case, since each supporting edge connected with J_a must take one vertex of the tree, there are exactly $t_a + 1$ non-coincident supporting edges connected with J_a. The same reasoning shows that for the second case, there are at least $t_a + 2$ non-coincident supporting edges connected with J_a and hence $l_a \geq t_a + 2$.

To complete the proof of the proposition, we need only proceed with the proof of the first case. In this case, the tree formed by the t_a bridge edges possesses $(t_a + 1)$ vertices. Arbitrarily select two vertices of the tree. Then these two vertices are joined up by one path composed of only bridge edges from the tree. We assert that there cannot be any edge of $\overline{G}(\mathbf{i})$, which does not belong to the tree, taking the two vertices as its two end points. To see this, by the way of contradiction, we suppose one such edge exists. Then this edge and the prescribed path form a cycle and consequently we see that this edge must not belong to $\overline{G}_1(\mathbf{i})$ and that the cycle belong to the graph consisting of $\overline{G}_2(\mathbf{i})$ and all bridge edges. Denote this later mentioned graph by $\overline{G}_3(\mathbf{i})$. It follows that removing any edge of the path from $\overline{G}_3(\mathbf{i})$ does not cause the graph disconnected and so any edge arbitrarily selected from the bridge edges forming the path is not cutting in $\overline{G}_3(\mathbf{i})$. However, by the definition of bridge edges, $\overline{G}_3(\mathbf{i})$ is a connected graph and each of the $(q-1)$ bridge edges should be cutting in the graph. Thus we reach a contradiction and so we conclude our assertion is true.

Now arbitrarily select one vertex of degree one in the tree. Then the supporting edge connecting this vertex and J_a must be single among supporting edges and so must be coincident with one non-supporting vertical edge. Note that this non-supporting vertical edge may be a down edge and also may be an up edge. We first consider the case when this vertical edge is a down edge, say (i_v, j_v). Then i_v is coincident with the vertex we selected which has degree one in the tree and j_v is one of the coincident j-vertices constituting J_a. Note that since (i_v, j_v) is non-supporting, (i_v, i_{v+1}) is not a bridge edge. Note that $i_v \neq i_{v+1}$. By the preceding argument, i_{v+1} cannot be coincident with any of the other t_a vertices of the tree either. This implies the up vertical edge (j_v, i_{v+1}) cannot be coincident with any of the $(t_a + 1)$ non-coincident supporting edges connected with J_a. Therefore if follows $l_a \geq t_a + 2$. For the other case when the vertical edge is an up edge, say (j_v, i_{v+1}), similar argument can be used to conclude that the anterior down edge (i_v, j_v) cannot be coincident with any of the $(t_a + 1)$ non-coincident supporting edges connected with J_a. The proof of the proposition is complete. $\square$

5.2.3 Convergence of the Expectation: Proof of (I)

From Proposition 5.2.1, it follows for each $\mathcal{G} \in \mathcal{G}(l, r, s, q)$,

$$\begin{aligned}
& \frac{1}{mn^{k/2}p^{k/2}} \Big| \sum_{G(\mathbf{i},\mathbf{j}) \in \mathcal{G}} Ed_{\overline{G}(\mathbf{i})} EX_{\underline{G}(\mathbf{i},\mathbf{j})} \Big| \\
\leq \quad & \frac{1}{mn^{k/2}p^{k/2}} \sum_{\overline{G}(\mathbf{i}) \in \overline{\mathcal{G}}} |Ed_{\overline{G}(\mathbf{i})}| \sum_{\underline{G}(\mathbf{i},\mathbf{j}) \in \underline{\mathcal{G}}(\mathbf{i})} |EX_{\underline{G}(\mathbf{i},\mathbf{j})}| \\
\leq \quad & \frac{1}{mn^{k/2}p^{k/2}} K m^{1+\delta(q-1)} p^{r-1-\delta(q-1)} n^s (\eta_n \sqrt[4]{np})^{2k-2l} \\
= \quad & K \eta_n^{2(k-l)} m^{\delta(q-1)} n^{s-l/2} p^{r-l/2-1-\delta(q-1)} \\
= \quad & K \eta_n^{2(k-l)} (m/n)^{\delta(q-1)} (p/n)^{\frac{l}{2}-s-\delta(q-1)} p^{r+s-l-1}. \qquad (5.2.6)
\end{aligned}$$

Based on above relation, we separate the terms involved in EM_k into three parts, *i.e.* let S_1 be the sum of terms with $l < k$, S_2 be the sum of terms with $l = k$, but either $r + s < l + 1$ or $l > 2s$, and S_3 be the sum of terms with $l = k$, $r + s = l + 1$, $l = 2s$, and hence from Proposition 5.2.2, $q = 1$. Noting the relations between the numbers l, r, s analyzed previously, we have

$$EM_k = S_1 + S_2 + S_3.$$

We first prove $S_1 \to 0$ and $S_2 \to 0$. Note that under assumption (iv) of THEOREM 1.3.1, $(m/n)^{\delta(q-1)}$ is bounded, while from Proposition 5.2.2, $l/2 - s - \delta(q-1)$ is always nonnegative. Thus we get

$$|S_1| = o((m/n)^{\delta(q-1)} (p/n)^{\frac{l}{2}-s-\delta(q-1)} p^{r+s-l-1}) = o(1).$$

Moreover, when $\delta(q-1) = 0$, by the fact either $k > 2s$ or $r + s < k + 1$, we have

$$|S_2| = O((p/n)^{\frac{k}{2}-s} p^{r+s-k-1}) = o(1),$$

When $\delta \in (0, 1/2)$ and $q > 1$, by the fact $k/2 > s + \delta(q-1)$, we have

$$|S_2| = O((p/n)^{\frac{k}{2}-s-\delta(q-1)}) = o(1).$$

When $\delta = 1/2$ and $q > 1$, since $m/n \to 0$, we still have

$$|S_2| = O((m/n)^{\frac{1}{2}(q-1)}) = o(1).$$

Note that when k is odd, no terms involved in the summation of EM_k belong to S_3 and hence we must have

$$EM_k \to 0.$$

In the following we only need to evaluate S_3 for the case k is even, by definition of S_3, $k = 2s$.

We first note that $r + s = k + 1$ implies that there cannot be cycles of non-coincident vertical edges in the base of the graph. Also note that $l = k$ implies that each non-coincident vertical edge must consist of exactly two vertical edges. It follows that each down (up) edge must coincide with one and only one up (down) edge because the coincidence of a down edge (up) with another down (up) edge would imply that the non-coincident edges of the base contain a cycle. Therefore, if we denote the non-coincident vertical edges by $\{(u_1, v_1), \cdots, (u_k, v_k)\}$, then

$$EX_{\underline{G}(\mathbf{i},\mathbf{j})} = \prod_{j=1}^{k} \sigma^2_{u_j v_j}$$

and hence for each isomorphic class $\mathcal{G} \in \mathcal{G}(2s, s+1, s, 1)$ (all isomorphic classes involved in S_3 belong to $\mathcal{G}(2s, s+1, s, 1)$), we have

$$\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} Ed_{\overline{G}(\mathbf{i})} EX_{\underline{G}(\mathbf{i},\mathbf{j})} = \sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} Ed_{\overline{G}(\mathbf{i})} \prod_{j=1}^{k} \sigma^2_{u_j v_j}.$$

Now we show that

$$\frac{1}{m(np)^s}\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} Ed_{\overline{G}(\mathbf{i})}\prod_{j=1}^{k}\sigma^2_{u_jv_j} = \frac{1}{m(np)^s}\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} Ed_{\overline{G}(\mathbf{i})}\sigma^{2k} + o(1). \tag{5.2.7}$$

By (ii) of Assumption 5.1.1 and (5.2.4) of Proposition 5.2.1 for the case $q = 1$, we have

$$\begin{aligned}
0 &\le \frac{1}{m(np)^s}\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} |Ed_{\overline{G}(\mathbf{i})}|[\sigma^{2k} - \prod_{j=1}^{k}\sigma^2_{u_jv_j}] \\
&\le \frac{1}{m(np)^s}\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} |Ed_{\overline{G}(\mathbf{i})}|\sum_{\ell=1}^{k}[\sigma^{2(k-\ell)}(\sigma^2 - \sigma^2_{u_\ell v_\ell})\prod_{j=1}^{\ell-1}\sigma^2_{u_jv_j}] \\
&\le \frac{1}{m(np)^s}\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} |Ed_{\overline{G}(\mathbf{i})}|\sum_{\ell=1}^{k}[\sigma^{2(k-1)}(\sigma^2 - \sigma^2_{u_\ell v_\ell})] \\
&\le \frac{\sigma^{2(k-1)}}{mnp^s}\sum_{\overline{G}(\mathbf{i})\in\overline{\mathcal{G}}} |Ed_{\overline{G}(\mathbf{i})}|\sum_{\ell=1}^{k}\sum_{v_\ell}(\sigma^2 - \sigma^2_{u_\ell v_\ell}) \\
&\le \sum_{\ell=1}^{k}\frac{K\sigma^{2(k-1)}}{mn}\sum_{v_\ell}\sum_{u_\ell}(\sigma^2 - \sigma^2_{u_\ell v_\ell}) \to 0,
\end{aligned}$$

from which (5.2.7) follows.

For a graph corresponding to a term in S_3, we claim that each horizontal edge (v_1, v_2) must coincide with a horizontal edge (v_2, v_1). In fact, suppose $(i_\ell, i_{\ell+1})$ is the first appearance of (v_1, v_2), *i.e.* $i_\ell = v_1,\ i_{\ell+1} = v_2$ and v_2 is not in $\{i_1, \cdots, i_\ell\}$. We claim that j_ℓ is not in $\{j_1, \cdots, j_{\ell-1}\}$. Otherwise, assuming j_ℓ is coincident with j_a with $a < \ell$, then $i_a,\ i_{a+1}$ and $i_{\ell+1}$ are three non-coincident i-vertices so that (i_a, j_a), (j_a, i_{a+1}) and $(j_\ell, i_{\ell+1})$ are three non-coincident vertical edges. It follows that there are at least three non-coincident vertical edges connected with the non-coincident j-vertex which contains j_a, j_ℓ. This obviously violates the assumption $k = 2s$. As a consequence of the assertion, both of the two vertical edges (i_ℓ, j_ℓ)

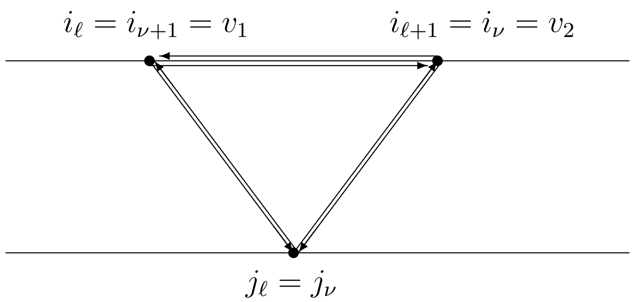

Figure 3.1 (v_1, v_2) **coincides with** (v_2, v_1)

and $(j_\ell, i_{\ell+1})$ are single up to the vertex $i_{\ell+1}$. In the future development of the graph, there must be one down edge (i_ν, j_ν) coincident with the single up-edge $(j_\ell, i_{\ell+1})$, that is, $i_\nu = i_{\ell+1} = v_2$ and $j_\nu = j_\ell$. Then the next up edge $(j_\nu, i_{\nu+1})$ must coincide with (i_ℓ, j_ℓ) since, otherwise, the vertex $j_\nu = j_\ell$ will be connected with at least 3 non-coincident vertical edges. Thus $i_{\nu+1} = i_\ell = v_1$ and so the horizontal edge $(i_\ell, i_{\ell+1}) = (v_1, v_2)$ coincides with the horizontal edge $(i_\nu, i_{\nu+1}) = (v_2, v_1)$ (see Figure 3.1). In view that the total number of non-coincident i-vertices contained in $\overline{G}(\mathbf{i})$ is $r = s+1$, we conclude that the non-coincident horizontal edges of $\overline{G}(\mathbf{i})$ form a tree of s edges, each edge consisting of exactly two horizontal edges of converse directions.

It follows

$$Ed_{\overline{G}(\mathbf{i})} = \prod_{\ell=1}^{s} p_{a_\ell, b_\ell},$$

where (a_ℓ, b_ℓ), $1 \le \ell \le s$, denote the edges of the tree of non-coincident horizontal edges. By (5.2.7) and condition $(D2)$, we have

$$\begin{aligned} & \frac{1}{m(np)^s} \sum_{G(\mathbf{i},\mathbf{j}) \in \mathcal{G}} Ed_{\overline{G}(\mathbf{i})} EX_{\underline{G}(\mathbf{i},\mathbf{j})} \\ = & \frac{\sigma^{2k}}{mp^s} \sum_{\overline{G}(\mathbf{i}) \in \overline{\mathcal{G}}} \prod_{\ell=1}^{s} p_{a_\ell, b_\ell} + o(1) \\ = & \sigma^{2k} + o(1). \end{aligned}$$

Therefore, to evaluate EM_k, what remains is to count the number of isomorphic classes in $\mathcal{G}(2s, s+1, s, 1)$. Note that for the graphs defined earlier, one only needs to arrange the vertical edges since the positions of the horizontal edges will then be automatically determined by the positions of the i-vertices. When one draws the graph edge by edge, starting from i_1, an edge is called an innovation if it is the first appearance of a non-coincident edge and called a Type 3 edge otherwise. As we have shown in the previous paragraph, for a graph corresponding to a term in S_3, a down innovation must be followed by an up innovation and a down edge of Type 3 must be followed by an up edge of Type 3. Thus, we only need to arrange the s down innovations and the s down edges of Type 3. Define $a_\ell = 1$ if the ℓ-th down edge is an innovation and $= -1$ otherwise. Before any Type 3 edge, there must be a single innovation. That is, for every $\ell \le k$, we should have

$$a_1 + \cdots + a_\ell \ge 0.$$

Thus, the number of isomorphic classes in $\mathcal{G}(2s, s+1, s, 1)$ is the number of sequences of s ones and s minus ones subject to the nonnegative partial sum requirement. By the reflection theorem, it is easy to show the number of such sequences is

$$\binom{2s}{s} - \binom{2s}{s-1} = \frac{(2s)!}{s!(s+1)!}.$$

Conclusion (I) is proved.

5.2.4 Estimation of the Fourth Moment: Proof of (II)

Since M_k is real, we may write

$$\begin{aligned} & E|M_k - EM_k|^4 = E(M_k - EM_k)^4 \\ = \quad & \frac{1}{m^4 n^{2k} p^{2k}} \sum_{\substack{\mathbf{i}_\ell, \mathbf{j}_\ell \\ 1 \le \ell \le 4}} E(\prod_{\ell=1}^{4} [d_{\overline{G}(\mathbf{i}_\ell)} X_{\underline{G}(\mathbf{i}_\ell, \mathbf{j}_\ell)} - E d_{\overline{G}(\mathbf{i}_\ell)} E X_{\underline{G}(\mathbf{i}_\ell, \mathbf{j}_\ell)}], \end{aligned}$$

where $G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ is the graph defined by $(\mathbf{i}_\ell, \mathbf{j}_\ell)$ in the way given in the proof of part (I).

If $G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ has no edges coincident with edges of the other three, then the corresponding term in the summation is zero by independence. Furthermore, the term is also zero if $\bigcup_{\ell=1}^{4} G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ contains a single vertical edge. Hence we only need to consider the following two cases:

(1) The four graphs are connected together through edges.

(2) $\bigcup_{\ell=1}^{4} G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ consists of two separated pieces, each of which is composed of two graphs connected together. Split $E|M_k - EM_k|^4 = S_I + S_{II}$ according to the two cases. Denote the collection of graphs in case (1) by $\mathcal{C}_1$ and similarly denote the collection of graphs in case (2) by $\mathcal{C}_2$.

In case (1), the graph $G = \bigcup_{\ell=1}^{4} G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ has a connected roof. Let r, s, and l be, respectively, the numbers of non-coincident i-vertices, non-coincident j-vertices and non-coincident vertical edges contained in G. And similarly, we can define the number q. Then, similar to the estimation of EM_k, under the conditions in Theorem 1.3.1 and Assumption 5.1.1, we have

$$\begin{aligned} |S_I| &\leq \frac{16}{m^4 n^{2k} p^{2k}} \sum_{G \in \mathcal{C}_1} E(\prod_{\ell=1}^{4} |d_{\overline{G}(\mathbf{i}_\ell)}||X_{\underline{G}(\mathbf{i}_\ell,\mathbf{j}_\ell)}|) \\ &\leq \frac{K}{m^4 n^{2k} p^{2k}} \sum_{r,s,l,q} m^{\delta(q-1)} p^{r-1-\delta(q-1)} (\eta_n \sqrt[4]{np})^{(8k-2l)} n^s \\ &\leq Km^{-3} \sum_{r,s,l,q} \eta_n^{8k-2l} (m/n)^{\delta(q-1)} (p/n)^{\frac{1}{2}-s-\delta(q-1)} p^{r+s-l-1} = O(m^{-3}), \end{aligned}$$

where we have used Proposition 5.2.1 and the facts $l \leq 4k$, $r + s \leq l + 1$ and $l - 2s \geq q - 1$.

To estimate S_{II}, one only need to note that for each piece of the roof subgraph, one factor m is obtained. Then, totally, we get one more m and hence

$$S_{II} = O(m^{-2}).$$

Combining the above, (II) is proved. Consequently, the proof of Theorem 1.3.1 is complete.

Using the same approach as in proving (II), one can easily show that

$$E|M_k - EM_k|^{2\mu} = O(m^{-\mu}),$$

for any fixed integer μ. This result will be useful when the a.s. convergence is considered for $n \to \infty$.

5.3 Discussion

In this chapter, with the aid of the moment method, we have accomplished proving the empirical spectral distributions of a class of large sparse random matrices converge to the semicircle law. The result generalized to a large extent known works on the spectral analysis of large sparse random matrices. The main accomplishment is a new formulation of the sparsing matrix, whose non-zero-one and non-homogeneous nature provides a new and more relevant understanding of the sparseness relating to a certain class of random matrices. As a consequence, the result can be useful in more circumstances where sparse random matrices play their effects.

The conclusion of semicircle law for our matrices is consistent with findings appearing in other works. Indeed, the phenomenon can be traced back to the basic behavior of the normalized sample covariance matrices. So let us introduce the formally referred definition of the normalized sample covariance matrices (Bai and Yin (1988a)).

Definition 5.3.1. *Let* $B_m = (1/\sqrt{mn})(X_{m,n}X_{m,n}^* - n\sigma^2 I_m)$, *where* $X_{m,n}$ *is* $m \times n$ *consisting of independent and identically distributed complex random variables with* $Ex_{11} = 0$, $E|x_{11}|^2 = \sigma^2$ *and* $E|x_{11}|^4 < \infty$. *Then* B_m *is said to be a normalized sample covariance matrix.*

For such defined normalized sample covariance matrix, Bai and Yin (1988a) showed as $m \to \infty$ and $m/n \to 0$, with probability one F^{B_m} converges weakly to the semicircle law. Noticing that when $E|x_{11}|^4 < \infty$ and m is fixed, the matrix $(1/\sqrt{n})(X_{m,n}X_{m,n}^* - n\sigma^2 I_m)$ tends to a Gaussian matrix, with Wigner's pioneering result on the semicircle law for Gaussian matrix, the result is thus conceivable.

In our context, the definition of normalized sample covariance matrix has been adjusted to suit the sparse nature of the Hadamard products. More specifically, the place of m in the square root of the denominator of B_m of Definition 5.1.1 has been taken place by the sparseness level p of Definition 1.3.2. Thus when the sparsing matrix D_m is taken to be the special case of $d_{ij} = 1$ for all i,j, then $p = m$ and the condition $m/n \to 0$ in Bai and Yin (1988) coincides with the condition $p/n \to 0$ we have assumed. So our result covers Bai and Yin's result and is more general than theirs.

Furthermore, since Bernoulli trials satisfy our conditions on the sparsing matrix, those sparse matrices for which the sparsing factors are chosen to be Bernoulli trials are included in the class of sparse matrices of our concern. A typical example is the result in Kohrunzhy and Rodgers (1997). We can check easily that if their conditions on the matrices are satisfied, then our conditions are all satisfied. Specifically, if $P(d_{ii} \neq 0) = 0)$ then condition (X3) is automatically true. If there is a positive and increasing function $\varphi(x)$ defined on $\mathbb{R}^+$ such that

$$Q_n \equiv \frac{1}{mn} \sum_{ij} E|x_{ij}^2|\varphi(|x_{ij}|)I[|x_{ij}| > \eta\sqrt[4]{np}] \to 0, \tag{5.3.1}$$

then condition (X2.1) holds. Letting $\varphi(x) = x^{4(2\nu-1)}$ with $\frac{1}{2} \leq \nu < 1$, (5.3.1) reduces to condition (2.4) of Kohrunzhy and Rodgers (1997), if we change their notation as $p_{ij} = P(d_{ij} = 1) = p/m$ with $p = n^{2\nu-1}$ and $m/n \to c \in (0, \infty)$. Thus Theorem 1.3.1 covers Kohrunzhy and Rodgers (1997) as a special case for all ν's in the interval $[1/2, 1)$. Furthermore, it can be seen if $Q_n \to 0$ with a suitable rate such that

$$\sum_{u=1}^{\infty} Q_n/\varphi(\eta\sqrt[4]{np}) < \infty, \tag{5.3.2}$$

then condition $(X2.2)$ is satisfied. Indeed, under their assumptions, $\varphi(x) = x^{4(2\nu-1)}$, with $(1+\sqrt{5})/4 < \nu < 1$, makes (5.3.2) hold. Then for these ν's Theorem 1.3.1 states the a.s. convergence holds and hence is stronger than the conclusion of i.p. convergence proven by Kohrunzhy and Rodgers (1997).

Concerning our conditions and results, we have the following several remarks which are expected to give some useful interpretations.

Remark 1. Condition $(D3.2)$ implies that p_{ij} are uniformly bounded. In fact,

$$p_{ij} = E|d_{ij}|^2 \leq (E|d_{ij}|^4)^{1/2} \leq C_4.$$

Combining this fact with condition $(D2)$, we see $p \leq Km$, for some constant $K > 0$.

In view of the relation between p and m, we notice that if condition $(D3.1)$ holds for some $\delta_0 \in [0, 1/2]$, then it must hold for every $\delta \geq \delta_0$, $\delta \in [0, 1/2]$. Therefore, we clarify here when we say that condition $(D3.1)$ holds for some $\delta \in [0, 1/2]$ we are referring to δ as the smallest value in $[0, 1/2]$ for which condition $(D3.1)$ holds.

With this understanding of the parameter δ, we can see in the case of $\delta = 1/2$, condition $(D3.1)$ is a direct consequence of Hölder's inequality and condition (D2), *i.e.* no additional assumption is imposed on the first moments of the sparsing factors.

Remark 2. A new contribution of Theorem 1.3.1 is to allow the sparsing factors d_{ij}'s to be very non-homogenous. Consider the following example. Let $D_m = [d_{ij}]$ be symmetric. Let $m = kL$ with L fixed, and let for all $(\ell - 1)k < i, j \le \ell k$ with $\ell \le L$,

$$P(d_{ij} = 1) = p/k = 1 - P(d_{ij} = 0),$$

and for all other indices i, j, $d_{ij} \equiv 0$. Then conditions $(D1)$, $(D2)$, $(D3.1)$ and $(D3.2)$ are true whenever $p \le k$.

Remark 3. In the case of $\delta = 0$, condition $(D3.1)$ seems not to allow the d_{ij}'s to take large values. In fact, it is not the case. For example, consider

$$d_{ij} = c^{-1}|z_{ij}|I(|z_{ij}| > c)$$

where z_{ij} are i.i.d. $N(0,1)$ subject to the condition $d_{ij} = d_{ji}$ and c is a positive constant uniquely solving the equation $Ez_{ij}^2 I(|z_{ij}| > c) = c^2 p/m$. Then obviously d_{ij} can take very large values, and D_m is symmetric with

$$\sum_{i=1}^{m} p_{ij} = c^{-2} \sum_{i=1}^{m} E z_{ij}^2 I(|z_{ij}| > c) = p,$$

i.e. conditions $(D1)$ and $(D2)$ are satisfied. It is trivial to see $c^{-1}E|z_{ij}|I_{(|z_{ij}|>c)} < c^{-2}E|z_{ij}|^2 I_{(|z_{ij}|>c)}$, that is, $E|d_{ij}| < E|d_{ij}|^2$. Thus condition $(C3.1)$ holds for $\delta = 0$.

Now we show condition $(D3.2)$ holds if $p/m \to 0$. In fact, we can see that if

$p/m \to 0$, then $c \to \infty$ and consequently for any positive integer k,

$$Ez_{ij}^k I(|z_{ij}| > c) \simeq 2c^{k-1}\varphi(c),$$

where the notation "$\simeq$" is used to represent the relation that the two quantities on its two sides have a ratio which tends to 1 as $c \to \infty$, while $\varphi(\cdot)$ is the density function of standard normal variables. It follows

$$E|d_{ij}|^2 = c^{-2}E|z_{ij}|^2 I_{(|z_{ij}|>c)} \simeq 2c^{-1}\varphi(c),$$

and for integer $k > 2$

$$E|d_{ij}|^k = c^{-k}E|z_{ij}|^k I(|z_{ij}| > c) \simeq 2c^{-1}\phi(c).$$

Thus condition $(D3.2)$ holds.

Remark 4. However, if condition $(D3.1)$ is assumed for $\delta = 0$, then it does happen that the d_{ij}'s are not allowed to take small values with large probabilities. For example,

$$d_{ij} = \sqrt{p/m}, \text{ with probability 1.}$$

Then obviously $(D3.1)$ can only hold for $\delta = 1/2$.

Remark 5. Condition $(D3.2)$ assuming that higher moments of d_{ij} are not larger than a multiple of their second moments is not seriously restrictive because the d_{ij}'s are usually small random variables. Moreover, this condition still allows the first moments of d_{ij}'s to be much larger than their second moments, *e.g.*

$d_{ij} = c\sqrt{p/m}|z_{ij}|$ where z_{ij} are i.i.d. random variables and c makes $Ed_{ij}^2 = p/m$.

Remark 6. Note that p may not be an integer and it may increase very slowly as n increases. Thus, the limit for $p \to \infty$ may not be true for a.s. convergence. So, we consider the limit when the integer part of p tends to infinity. However, if we consider the convergence in probability, Theorem 1.3.1 is true for $p \to \infty$.

Remark 7. From the proof given in the previous sections, one can see that the almost sure convergence is true for $m \to \infty$ in all places except the part for the truncation on the entries of $X_{m,n}$ which was guaranteed by condition $(X2.2)$. Thus, if condition $(X2.2)$ holds for $u = m$, then the almost sure convergence is true in the sense of $m \to \infty$. Sometimes, it may be of interest to consider the almost sure convergence in the sense of $n \to \infty$. Examining the proof, one can find that to guarantee the almost sure convergence for $n \to \infty$, the truncation on the entries of D_m and the removal of diagonal elements require $m/\log n \to \infty$; the truncation on the entries of $X_{m,n}$ require condition $(X2.2)$ to be true for $u = n$. As we have claimed at the end of the previous section, one may modify the conclusion of (II) as

$$E|M_k - EM_k|^{2\mu} = O(m^{-\mu})$$

for any fixed integer μ. Thus, if $m \geq n^{\varepsilon}$ for some positive constant ε, then the almost sure convergence for the ESD after the truncation and centralization is true for $n \to \infty$. Therefore, the conclusion of Theorem 1.3.1 can be strengthened to the almost sure convergence as $n \to \infty$ under the additional assumptions that, for some small positive constant ε, $m \geq n^{\varepsilon}$ and condition $(X2.2)$ holds for $u = n$.

Remark 8. Conditions $(D2)$ and $(D3.2)$ imply that $p \leq Km$, that is, the order of p cannot be larger than m. In the theorem, it is assumed that $p/n \to 0$, that is, p also has a lower order than n. This is essential. However, the relation between m and n can be arbitrary if condition $(D3.1)$ holds for $\delta = 0$.

It should be useful to remind that the statement "the relation between m and n can be arbitrary" only says that there are examples with $m/n \to \infty$ as well as examples with $m/n \to 0$, for which the result in Theorem 1.3.1 is applicable equally well. For example, if the d_{ij}'s (subject to the condition $d_{ij} = d_{ji}$) are the Bernoulli trials defined by $P(d_{ij} = 1) = p/m = 1 - P(d_{ij} = 0)$ for any i, j, then $E|d_{ij}| = E|d_{ij}|^2 = E|d_{ij}|^k$ for any $k > 2$. This implies conditions $(D1)$, $(D2)$, $(D3.1)$(with $\delta = 0$) and $(D3.2)$ always hold. But no matter $m/n \to 0$, $m/n \leq K < \infty$, or $m/n \to \infty$, Theorem 1.3.1 always holds provided that $p \leq m$ abd $p/n \to 0$.

It may be of interest whether our conditions put on the relation between m and n for the case of $\delta \neq 0$ are arguable. Nevertheless, based on the following two examples, we take the point that they are necessary for the validity of the semicircle law for the general class of matrices we considered. We firstly present an example to show that when condition $(D3.1)$ is assumed for $\delta = 1/2$, to ensure the convergence of the semicircle law of F^{A_p}, it is necessary to require $m/n \to 0$.

Example 5.3.1. Let $D_m = [d_{ij}]$ be consisting of $d_{ii} = 0$ and $d_{ij} = \sqrt{p/m}$ for $i \neq j$. Let $X_{m,n} = [x_{ij}]$ be consisting of independent, identically distributed standard normal random variables. Now assume $m/n \to c > 0$ and $p/n \to 0$. Then conditions (D2), (D3.1) and (D3.2) hold. Specifically, $1/2$ is the smallest parameter in $[0, 1/2]$ such that condition $(D3.1)$ is satisfied by A_p.

Consider the k-th moment of F^{A_p}. Using the definitions we gave in proving Proposition 3.1, for any isomorphic class $\mathcal{G}$ whose canonical graph possesses r non-coincident i-vertices and s non-coincident j-vertices and does not contain loops of horizontal edges, we have

$$\begin{aligned} S_{\mathcal{G}} &= \frac{1}{mn^{k/2}p^{k/2}} \sum_{G(\mathbf{i},\mathbf{j}) \in \mathcal{G}} Ed_{\overline{G}(\mathbf{i})} EX_{\underline{G}(\mathbf{i},\mathbf{j})} \\ &= K_{\mathcal{G}} m^{-1} n^{-k/2} p^{-k/2} (p/m)^{k/2} m^r n^s + o(1) \\ &= K_{\mathcal{G}} m^{r+s-k-1} (n/m)^{s-k/2} + o(1), \end{aligned}$$

where $K_{\mathcal{G}} = EX_{\underline{G}(\mathbf{i},\mathbf{j})}$. It is easy to see that

$$S_{\mathcal{G}} \to \begin{cases} 0, & \text{if } r+s < k+1, \\ K_{\mathcal{G}} c^{k/2-s}, & \text{if } r+s = k+1. \end{cases}$$

Note that when $r+s = k+1$, since $r+s \leq l+1$ where l is the number of non-coincident vertical edges contained in the canonical graph of $\mathcal{G}$, it follows $l \geq k$. If $l > k$, then there must exist single vertical edge and hence $K_{\mathcal{G}} = 0$. Otherwise, $l = k$, then every non-coincident vertical edge is composed of exactly 2 vertical edges of opposite directions and hence $K_{\mathcal{G}} = 1$. Therefore, noticing the restriction that, since there are no loops of horizontal edges, every non-coincident j-vertex must be connected with at least 2 non-coincident vertical edges, we get

$$EM_k \to m_k = \sum_{s=1}^{[\frac{k}{2}]} c^{k/2-s}\mu_s,$$

where μ_s is the number of isomorphic classes whose canonical graphs satisfy the following condition: Each canonical graph contains exactly s non-coincident j-vertices and $(k+1-s)$ non-coincident i-vertices; Each canonical graph contains exactly k non-coincident vertical edges each of which consists of two edges of opposite directions; Each canonical graph possesses the property that, supposing one person starts a walk along its edges, then whenever a down-edge leads to a new non-coincident j-vertex the next up-edge must lead to a new i-vertex.

To estimate the limit m_k, let us observe further

$$\begin{aligned} & \left(\frac{m}{n}\right)^{k/2} \times EM_k \\ = \;& m^{-1}n^{-k} \sum_{i_1,\cdots,i_k}^{res} \sum_{j_1,\cdots,j_k} E(x_{i_1j_1}x_{i_2j_1}x_{i_2j_2}x_{i_3,j_2}\cdots x_{i_kj_k}x_{i_1j_k}), \qquad (5.3.3) \\ \le \;& Em^{-1}tr\left(\frac{1}{n}X_{m,n}X_{m,n}^*\right)^k \end{aligned}$$

where the summation $\sum_{i_1,\cdots,i_k}^{res}$ is taken over all possible values of $i_1, \cdots, i_k$ satisfying the restriction that $i_1 \neq i_2$, $i_2 \neq i_3$, $\cdots$, $i_k \neq i_1$. Thus it follows, by Theorem 2.5 of Bai(1999), $c^{k/2}m_k$ is bounded by the kth moment of the Marcĕnko-Pastur law with ratio index c and scale index 1. Thus $\{m_k\}_{k=1}^{\infty}$ satisfies the Carleman condition.[1]

[1]Note that there is a one-to-one correspondence between $\mathcal{G}$ and its base and that the base of $\mathcal{G}$ must be a canonical graph defined in deriving the Marcĕnko-Pastur law. Thus, we indeed have

$$\mu_s \le \frac{1}{k+1-s}\binom{k}{s}\binom{k-1}{s-1}.$$

Using (5.3.3), one can easily show that

$$E(M_k - EM_k)^{2\mu} = O(m^{-2\mu}).$$

Therefore, with probability one, F^{A_p} converges to a non-random limiting distribution, say F. It is easy to verify that when $k = 3$, we have $s = 1$ and $r = 3$ so that $i_1 \neq i_2 \neq i_3$ and $j_1 = j_2 = j_3$, *i.e.* there is exactly one contributing isomorphic class $\mathcal{G}$. Thus,

$$m_3 \quad = \quad \sqrt{c}.$$

Since the third moment of F is not 0, F is not the semicircle law. That is, we have shown with probability one, F^{A_p} converges weakly but the limiting spectral distribution is not the semicircle law. $\square$

For the case $\delta \in (0, 1/2)$, we present the following example to show the condition m/n is bounded is also necessary for the convergence to the semicircle law.

Example 5.3.2. Let $D_m = [d_{ij}]$ be defined the same as in Example 5.3.1. We assume the same conditions $m/n \to c > 0$ and $p/n \to 0$. Now we define $\tilde{D}_h = D_m \otimes I_h$ and $\tilde{B}_h = \frac{1}{\sqrt{np}} \left(X_{mh,n} X^*_{mh,n} - \sigma^2 n I_{mh} \right)$, where "$\otimes$" denotes the Kronecker product of matrices, $h = [m^\eta]$ with $\eta > 0$ and $X_{mh,n}$ is $mh \times n$ consisting of independent and identically distributed standard normal random variables.

Let $\tilde{A}_p = \tilde{B}_h \circ \tilde{D}_h$. Then $\tilde{A}_p = \mathrm{diag}[A_{1,m}, \cdots, A_{h,m}]$ where

$$A_{i,m} = B_{ii} \circ D_m, \; i = 1, \cdots, h$$

and B_{ii} is the i-th $m \times m$ major sub-matrix of $\tilde{B}_h$.

Note that $A_{1,m}, \cdots, A_{h,m}$ are independent with the same distribution as A_p defined in Example 5.3.1. Denote by $\tilde{M}_k$, $M_{i,k}$ and M_k, respectively, the kth moment of $\tilde{A}_p$, the kth moment of $A_{i,m}$ and the kth moment of A_p. Then it follows $EM_{i,k} = EM_k$ and $E(M_{i,k} - EM_{i,k})^{2\mu} = E(M_k - EM_k)^{2\mu}$. Since $F^{\tilde{A}_p} = \frac{1}{h}\sum_{i=1}^{h} F^{A_{i,m}}$ so that $\tilde{M}_k = \frac{1}{h}\sum_{i=1}^{h} M_{i,k}$, we get $E\tilde{M}_k = EM_k$ and $E(\tilde{M}_k - E\tilde{M}_k)^{2\mu} \leq E(M_k - EM_k)^{2\mu}$. By the results we proved in Example 5.3.1, it follows with probability one, $F^{\tilde{A}_p}$ converges weakly but the limiting spectral distribution is not the semicircle law.

Let us now check the validity of the assumptions of Theorem 1.3.1 for $\tilde{A}_p$. Conditions $(D1)$, $(D2)$ and $(D3,2)$ hold for $\tilde{A}_p$ automatically by definition. We now show that for any $\delta \in (0, 1/2)$ by choosing $\eta > 0$ such that $2\delta(1+\eta) = 1$, condition $(D3,1)$ is satisfied by $\tilde{A}_p$ for the given δ. To see this, note that the dimension of $\tilde{A}_p$ is mh and we have

$$\sum_i Ed_{ij} \leq \sqrt{mp} \leq \frac{\sqrt{m}}{(mh)^\delta}(mh)^\delta p^{1-\delta} \leq C_1 (mh)^\delta p^{1-\delta}.$$

By requiring $p = O(\log m)$, we further can see for any $\delta_0 < \delta$,

$$(\sum_i Ed_{ij}) / \left((mh)^{\delta_0} p^{1-\delta_0}\right) \geq \frac{1}{2} m^{\frac{1}{2}(1-\delta_0/\delta)} p^{\delta_0 - \frac{1}{2}} \to \infty,$$

which confirms δ is the smallest parameter in $(0, 1/2)$ such that condition $(D3.1)$ is satisfied by $\tilde{A}_p$. Noticing that $mh/n \to \infty$, we see $\tilde{A}_p$ satisfies all assumptions of Theorem 1.3.1 except only the condition that in case of $\delta \in (0, 1/2)$ the ratio between the vector dimension and the sample size should be bounded. We achieved our target. $\square$

The class of sparse random matrices presently studied are Hadamard products of the sparsing matrix with sample covariance matrices. Future development may concern the Hadamard products with other interesting random matrices. A relevant case in point is given by the sample covariance matrices with some correlation structure assumed among the entries of the matrix $X_{m,n}$. Investigations can also be given to alternative definition of the sparsing matrix, which has been formulated in the present thesis in terms of the moments of its entries. Especially, when there are important physical factors affecting the systems modelled by the matrices, the definition of the sparsing matrix needs to take into account of their effects. Examples on this concern are still not available, whereas the consideration is of interest in practical field. For instance, in wireless communications, the Hadamard product of a sample covariance matrix and an appropriately defined sparsing matrix can be used as the channel matrix for a channel which strategically allocate different powers to different users.

Bibliography

[1] Achlioptas, D. Random matrices in data analysis. *http://www.cs.ucsc.edu/ optas/papers/matrices.pdf*

[2] Altland, A. and Zirnbauer, M. R. (1996). Random matrix theory of a chaotic Andreev quantum dot. *Physics Review Letters* **76**, 3420–3423.

[3] Arnold, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *Journal of Mathematical Analysis and Applications* **20**, 262–268.

[4] Arnold, L. (1971). On Wigner's semicircle law for the eigenvalues of random matrices. *Probability Theory and Related Fields* **19**, 191–198.

[5] Bai, Z. D. (1993a). Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices. *Annals of Probability* **21**, 625–648.

[6] Bai, Z. D. (1993b). Convergence rate of expected spectral distributions of large random matrices. Part II. Sample Covariance Matrices. *Annals of Probability* **21**, 649–672.

[7] Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica* **9**, 611-677.

[8] Bai Z. D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *Annals of Probability* **26**, 316-345.

[9] Bai, Z.D. and Silverstein, J. W. (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. *Annals of Probability* **27**, 1536-1555.

[10] Bai, Z.D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Annals of Probability* **32**, 553-605.

[11] Bai, Z.D., Silverstein, J. W. and Yin, Y. Q. (1998). A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis* **26**, 166–168.

[12] Bai, Z. D. and Yao, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11**, 1059–1092.

[13] Bai, Z. D. and Yin, Y. Q. (1986). Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang. *Probability Theory and Related Fields* **73**, 555–569.

[14] Bai, Z. D. and Yin, Y. Q. (1988a). Convergence to the semicircle law. *Annals of Probability* **16,** 863-875.

[15] Bai, Z. D. and Yin, Y. Q. (1988b). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of Wigner matrix. *Annals of Probability* **16**, 1729-1741.

[16] Bai, Z. D., and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of large dimensional covariance matrix. *Annals of Probability* **21**, 1275-1294.

[17] Barash, D. (2004). Spectral decomposition for the search and analysis of RNA secondary structure. *Journal of Computational Biology* **11**, 1169–1174.

[18] Barry, R. P. and Pace, R. K. (1999). Monte Carlo estimates of the log determinant of large sparse matrices. *Linear Algebra and its Applications* **289**, 41–54.

[19] Bekakos, M. P. and Bartzi, A. A. (1999). Sparse matrix and network minimization schemes. *Computational Methods and Neural Networks* , 61–94.

[20] Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons, New York.

[21] Boley, D. and Goehring, T. (2000). LQ-Schur projection on large sparse matrix equations. *Numerical Linear Algebra with Applications* **7**, 491–503.

[22] Botta, E. F. F. and Wubs, F. W. (1999). Matrix renumbering ILU: an effective algebraic multilevel ILU preconditioner for sparse matrices. *SIAM Journal on Matrix Analysis and Applications* **20**, 1007-1026.

[23] Bronk, B. V. (1964). Accuracy of the semicircle approximation for the density of eigenvalues of random matrices. *Journal of Mathematical Physics* **5**, 215–220.

[24] Bryc, W., Dembo, A. and Jiang, T. F. (2006). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Annals of Probability* **34**, 1–38.

[25] Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Annals of Probability* **1**, 19–42.

[26] Deift, P., Kriecherbauer, T., McLauglin, K. T. -R., Venakides, S. and Zhou, X. (1999). Uniform asymptotics for polynomials orthogonal with respect to varying exponential weight and applications to universality questions in random matrix theory. *Communications on Pure and Applied Mathematics* **52**, 1335–1425.

[27] Dozier, R. B. and Silverstein, J. W. (2004). On the empirical distribution of large dimensional information-plus-noise type matrices. Preprint.

[28] Dyson, F. J. (1962). Statistical theory of the energy levels of complex systems, I II & III. *Journal of Mathematical Physics* **3**, 140–175.

[29] Fisher, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Annals of Eugenics* **9,** 238–249.

[30] Gao, H. and Smith, P. J. (2000). A determinant representation for the distribution of quadratic forms in complex normal vectors. *Journal of Multivariate Analysis* **73**, 155–165.

[31] Gaudin, M. (1961). Sur la loi limite de I'espacement des valeurs propres d'une matrice aléatoire. *Nuclear Physics* **25**, 447–458.

[32] Geman, S. (1980). A limit theorem for the norm of random matrices. *Annals of Probability* **8**, 252-261.

[33] Geman, S. (1986). The spectral radius of large random matrices. *Annals of Probability* **14**, 1318–1328.

[34] Geronimo, J. S. and Hill, T. P. (2003). Necessary and sufficient condition that the limit of Stieltjes Transforms is a Stieltjes Transform. *Journal of Approximation Theory* **121**, 54–60.

[35] Ginibre, J. (1965). Statistical ensembles of complex, quaternion, and real matrices. *Journal of Mathematical Physics* **6**, 440–449.

[36] Girko, V. L., Kirsch, W. and Kutzelnigg, A. (1994). A necessary and sufficient condition for the semicircle law. *Random Operators and Stochastic Equations* **2**, 195-202.

[37] Girko, V. L. (1990). *Theory of random determinants.* Translated from the Russian. Mathematics and its Applications (Soviet Series). 45. Kluwer Academic Publishers Group, Dordrecht.

[38] Girshick, M. A. (1939). On the sampling theory of roots of determinantal equations. *The Annals of Mathematical Statistics* **10,** 203–224.

[39] Grenander, U. (1963). *Probabilities on Algebraic Structures.* John Wiley, New York-London.

[40] Grenander, U. and Silverstein, J.W. (1977). Spectral analysis of networks with random topologies. *SIAM Journal on Applied Mathematics* **32**, 499–519.

[41] Gurevich, I. I. and Pevsner, M. I. (1957). Repulsion of nuclear levels. *Nuclear Physics* **2**, 575–581.

[42] Hoyle, D. C. and Rattray, M. (2004). Principal-component-analysis eigenvalue spectra from data with symmetry-breaking structure. *Physical Review E* **69**, 026124.

[43] Hsu, P. L. (1939). On the distribution of roots of certain determinantal equations. *Annals of Eugenics* **9**, 250–258.

[44] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics* **29**, 295–327.

[45] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis* **12**, 1–38.

[46] Katz, N. M. and Sanak, P. (1998). *Random matrices, Frobenius eigenvalues, and Monodromy.* American Mathematical Society Colloquium Publications, 45.

[47] Khatri, C. G. (1964). Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations. *The Annals of Mathematical Statistics* **35**, 1807–1810.

[48] Khatri, C. G. (1969). Non-central distributions of i-th largest characteristic roots of three matrices concerning complex multivariate normal populations. *The Annals of Institute of Statistical Mathematics* **21,** 23–32.

[49] Khorunzhy, A. and Rodgers, G. J. (1997). Eigenvalue distribution of large dilute random matrices. *Journal of Mathematical Physics* **38,** 3300-3320.

[50] Khorunzhy, A. and Rodgers, G. J. (1998). On the Wigner law in dilute random matrices. *Reports on Mathematical Physics* **42,** 297-319.

[51] Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J. P. (1999). Random matrix theory and financial correlations. *http://www.science-finance.fr/papers/dublin.pdf.*

[52] Marchěnko, V.A. and Pastur, L.A. (1967). Distribution for some sets of random matrices. *Math. USSR-Sb.* **1**, 457-483.

[53] Marti, R., Laguna, M., Glover, F. and Campos, V. (2001). Reducing the bandwidth of a sparse matrix with tabu search. *European Journal of Operational Research* **135**, 450-459.

[54] McKenzie, B. J. and Timothy, B. (2001). Compression of sparse matrices by blocked Rice coding. *IEEE Transactions on Information Theory* **47**, 1223–1230.

[55] Mehta, M. L. (1960). On the statistical properties of the level-spacings in nuclear spectra. *Nuclear Physics* **18**, 395–419.

[56] Mehta, M. L. and Gaudin, M. (1960). On the density of eigenvalues of a random matrix. *Nuclear Physics* **18**, 420–427.

[57] Mehta, M. L. (1991). *Random Matrices.* Academic Press, New York.

[58] Monvel, A. B. and Khorunzhy, A. (1999). On the norm and eigenvalue distribution of large random matrices. *Annals of Probability* **27**, 913–944.

[59] M*ü*ller, R. and Verd*ú*, S. (2001). Design and analysis of low-complexity interference mitigation on vector channels. *IEEE Journal on Selected Areas in Communications* **19**, 1429–1441.

[60] Naulin, Jean-Marc (2002). A contribution of sparse matrices tools to matrix population model analysis. *Mathematical Biosciences* **177/178**, 25-38.

[61] NeuB, N. (2002). A new sparse-matrix storage method for adaptively solving large systems of reaction-diffusion-transport equations. *Computing* **68**, 19–36.

[62] Oppermann, R. (1990). Anderson localization problems in gapless conducting phases. *Physica* **A 167**, 301–312.

[63] Pafka, S. and Kondor, I. (2004). Estimated correlation matrices and portfolio optimization. *http://arxiv.org/abs/cond mat/0305475.*

[64] Plerou, V., Copikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T. and Stanley, H. E. (2001). A random matrix approach to cross-correlations in financial data. *http://arxiv:cond-mat/0108023.*

[65] Porter, C.E. and Thomas, R.G. (1956). Fluctuations of nuclear reaction widths. *Physical Review* **104**, 483–491.

[66] Potters, M., Bouchaud, J. P. and Laloux, L. (2005). Financial applications of random matrix theory. *Acta physica polonica B* 36, 2767–2784.

[67] Rosenzweig, N. and Porter, C. E. (1960). "Repulsion of energy levels" in complex atomic spectra. *Physical Review* **120**, 1698–1714.

[68] Roy, S. N. (1939). p-statistics or some generalizations in the analysis of variance appropriate to multivariate problems. *Sankhya* **4**, 381–396.

[69] Silverstein, J. W. (1979). On the randomness of eigenvectors generated from networks with random topologies. *SIAM Journal on Applied Mathematics* **37**, 235–245.

[70] Silverstein, J. W. (1981). Describing the behavior of eigenvectors of random matrices using sequences of measures on orthogonal groups. *SIAM Journal on Mathematical Analysis* **12**, 174–181.

[71] Silverstein, J. W. (1984). Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. *Journal of Multivariate Analysis* **15**, 295–324.

[72] Silverstein, J. W. (1989). On the eigenvectors of large dimensional sample covariance matrices. *Journal of Multivariate Analysis* **30**, 1–16.

[73] Silverstein, J. W. (1990). Weak convergence of random functions defined by the eigenvectorsof sample covariance matrices. *Annals of Probability* **18**, 1174–1194.

[74] Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis* **54**, 175–192.

[75] Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis* **54**, 331–339.

[76] Silverstein, J. W. and Choi, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis* **54**, 295–309.

[77] Soshnikov, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Communications in Mathematical Physics* **207**, 697–733.

[78] Soshnikov, A. (2002). A note on the universality of the distribution of the largest eigenvalue in certain classes of sample covariance matrices. *Journal of Statistical Physics* **108**, 1033–1056.

[79] Stariolo, D. A., Curado, E. M. F. and Tamarit, F. A. (1996). Distributions of eigenvalues of ensembles of asymmetrically diluted Hopfield matrices. *Journal of Physics A* **29**, 4733-4739.

[80] Tracy, C. A. and Widom, H. (1993). Level-spacing distribution and the Airy kernel. *Physics Letters. B* **305**, 115–118.

[81] Tracy, C. A. and Widom, H. (1994). Level-spacing distribution and the Airy kernel. *Communications in Mathematical Physics* **159**, 151–174.

[82] Tracy, C. A. and Widom, H. (1996). On orthognal and symplectic matrix ensembles. *Communications in Mathematical Physics* **177**, 727–754.

[83] Tracy, C. A. and Widom, H. (2002). Distribution Functions for largest eigenvalues and their applications. *Proceedings of the ICM,* **1**, 587-596.

[84] Tulino, A. M. and Verdú, S. (2005) *Random matrix theory and wireless communications.* Foundation and Trends in communications and information theory. Now Publishers.

[85] Vassilevski, Panayot, S. (2002). Sparse matrix element topology with application to AMG(e) and preconditioning. *Numerical Linear Algebra with Applications* **9**, 429–444.

[86] Verbaarschot, J. J. M. and Wetting, T. (2000). Random matrix theory and chiral symmetry in QCD. *http://arXiv:hep-ph/0003017 v2.*

[87] Verbaarschot, J. J. M. (2006). The suppersymmetric method in random matrix theory and applications to QCD. *http://arXiv:hep-th/0410211 v1.*

[88] Vernizzi, G. and Orland, H. (2005). Large-N random matrices for RNA folding. *Acta Physica Polonica. B* **36**, 2821–2827.

[89] Voiculescu, D. (2000). Lectures on free probability theory. *Lecture Notes in Mathematics* **1738**, 283-349.

[90] Wachter, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Annals of Probability* **6**, 1–18.

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

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

[93] Wigner, E. P. (1958). On the distributions of the roots of certain symmetric matrices. *Annals of Mathematics* **67**, 325–327.

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

[95] Wishart, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A**, 32–52.

[96] Yin, Y. Q. (1986). LSD's for a class of random matrices. *Journal of Multivariate Analysis* **20**, 50–68.

[97] Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1983). Limiting behavior of the eigenvalues of a multivariate F matrix. *Journal of Multivariate Analysis* **13**, 508–516.

[98] Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields* **78**, 509–521.

[99] Yin, Y. Q. and Krishnaiah, P. R. (1983). A limit theorem for the eigenvalues of product of two random matrices. *Journal of Multivariate Analysis* **13**, 489–507.

[100] Zirnbauer, M. R. (1996). Riemann symmetric superspaces and their origin in random matrix theory. *Journal of Mathematical Physics* **37**, 4986–5018.

[101] http://www.maths.ex.ac.uk/ mwatkins/zeta/random.htm

Random matrices and the Riemann zeta function.